Digital vocal editing, plug-ins and the fetishism of voice and technology in 21st-century Western commercial popular music production

Angharad Thomas

The Queen's College

Faculty of Music

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Trinity Term 2022

Abstract

This thesis describes and interrogates the increasingly significant role of vocal editing in the production of Western commercial popular music in the 21st-century, drawing on two years of multi-sited ethnographic research with 'producer-songwriters' working in this genre in the UK. In it, I argue that producer-songwriters' digital vocal editing practices fetishize both singing voices and the digital technologies which are used to edit them. These two different fetishisms which exist in the practice—of the voice, and of technology—are synergistic and strengthen one another. Furthermore, this co-productive fetishism is reflected in the specific ways in which the music spotlights the lead vocal part. Additionally, key companies within the Western music technology industry (including Waves Audio, Universal Audio and FabFilter) both exploit and drive these fetishisms through the design and marketing of plug-ins (software programs) for vocal editing.

The thesis is divided into five main chapters. Following a review of relevant literature in ethnomusicology and popular music studies in the introduction, Chapter 1 explains the theoretical frame of fetishism. Chapter 2 illustrates how the fetishism of the voice is perpetuated among producer-songwriters by their aesthetic discourse which privileges ideas of the voice occupying an 'up-front and central' position within the metaphorical space of the popular music 'mix'—an aesthetic aim which I term 'vocal foregrounding'. The subsequent chapters of the thesis focus on a set of creative practices employing different technologies which are central to vocal editing and which are deployed serially across "vocal chains"—tuning & comping (Chapter 3), equalization (EQ) (Chapter 4) and compression (Chapter 5). Each contributes to the commercial pop vocal sound, and each is associated with a different kind of fetishism of technology which interacts with and reinforces the fetishism of the voice. These chapters also address the main plug-ins and marketing techniques used by the plug-in industry to perpetuate such double fetishism.

This thesis contributes to existing musicological literature in two ways. First, I approach the practice of vocal editing and production from an aesthetic perspective, explaining how the sound of commercial popular vocals is created through multiple, cumulative digital processes which are influenced by external cultural and economic factors, rather than focusing on individual technologies

in the process and their associated histories (e.g. *AutoTune*). Secondly, I build on existing ethnomusicological approaches to music and technology by not only engaging fully with the practices and discourses of the digital production of commercial popular music through my fieldwork, but also by supplementing this ethnographic focus with a variety of Western pedagogical sources about sound processing, production and mixing, contextualizing my informants' practices within a transnational culture of music production.

Contents

Abstract ... 1
List of Figures ... 6
List of Audio Examples ... 9
Acknowledgements ... 10
Introduction ... 12
Context and definitions: vocal editing in a networked songwriting and production industry ... 15
Genre: 21st-century 'commercial' popular music? ... 24
Methodology ... 32
Methodological literature review: understanding Western popular music production through technology and aesthetic discourses ... 40
Chapter summary ... 47
Chapter One: Contextualizing Voice and Technology Fetishism in Vocal Editing ... 50
1.1. Defining fetishism ... 53
1.2. Contextualizing voice fetishism in vocal editing ... 62
Contextualizing erotic voice fetishism in Western popular music culture ... 62
Voice fetishism in commercial pop vocal editing: erotic voice fetishism vs commodity fetishism of vocals ... 71
1.3. Contextualizing technology fetishism in vocal editing ... 75
Two forms of plug-in commodity fetishism: analogue emulations and digital novelties ... 81
Plug-in commodity fetishism and the online cultures of home music production ... 88
Plug-in commodity fetishism and gender ... 92
1.4. Voice and technology fetishism in other cultures ... 94
Conclusion ... 96
Chapter Two: Up-front and Central, Up-close and Personal: Voice Fetishism and the Aesthetic Discourse of Vocal Editing ... 100
2.1. Foregrounding the voice in the aesthetic discourse of vocal editing ... 104
Vocal foregrounding in 20th-century commercial popular music mixing ... 104
Vocal foregrounding in the aesthetic discourse of 21st-century commercial pop vocal editing ... 111
"Presence" ... 118
2.2. Vocal "tone" and "dynamics": voices are special and unpredictable ... 125
Vocal "Tone" ... 126
Vocal "Dynamics" ... 131
Conclusion ... 137
Chapter Three: Comping, Tuning and the Visual Objectification of the Voice ... 141
3.1. Contextualizing vocal comping, tuning and related techniques ... 144
Comping ... 144

Tuning 148
Timing 157
Manual gain adjustment ("clip-gain") 159
Audio restoration tools 161
3.2. Visual objectifications of the voice in comping, tuning and related techniques in commercial pop vocal editing 163
Visual vocal editing and its effect on interpretations of the commercial pop vocal style 163
The visual objectification of vocal "emotion" 173
Visual objectifications of the voice and the gendered dialogues between producer-songwriter and vocalist 178
Conclusion 186
Chapter Four: "You Want it to Sound Real, Like a Person": Erotic Voice Fetishism, Para-graphic EQ Plug-ins and Realism in Commercial Pop Vocal Equalization (EQ) 189
4.1. Contextualizing vocal EQ 192
A brief history of EQ in audio engineering and music production 192
Digital EQ plug-ins: para-graphic and rotary interfaces 196
EQ in Western popular music mixing: 'carving' and universal models of auditory perception 198
4.2. EQ in commercial pop vocal editing 203
Para-graphic EQ and assumptions about vocal realism 203
The cultural specificity of realist discourse in commercial pop vocal EQ 218
4.3. Vocal realism, surgical metaphors and gender 220
"Subtractive" vocal EQing: vocal realism and surgical metaphors 220
"Presence", "harshness", and gendered interpretations of vocal realism 225
4.4. "Intelligent" EQ plug-ins and vocal realism: Fabfilter's *Pro-Q 3*, oeksound's *Soothe* and SoundTheory's *Gullfoss* 229
FabFilter's *Pro-Q 3* 231
oeksound's *Soothe2* 234
SoundTheory's *Gullfoss* 236
Conclusion 239
Chapter Five: The Commodity Fetishism of Analogue Emulations in Commercial Pop Vocal Compression 243
5.1. Dynamic range compression: definitions, history and context 245
Vocal compression in commercial pop vocal editing: contributing to vocal foregrounding by disciplining vocal "dynamics" 252
5.2. The commodity fetishism of *1176* and *LA-2A* emulations 254
The *1176* and *LA-2A* as vocal compressors: commodity fetishism and voice fetishism combined 258
Waves' advertising partnership with Chris Lord-Alge and the masculine tone of *1176* and *LA-2A* commodity fetishism 261

5.3. Iterative compression processes in commercial pop vocal editing 265
Serial compression 266
Parallel compression 269
Frequency-sensitive compression, digital novelty commodity fetishism and the interactions between compression and EQ 274
Conclusion 283
Conclusion 285
Voice and technology fetishism combined 285
The commercial pop vocal aesthetic 293
Gender and voice and technology fetishism 295
The plug-in industry drives voice and technology fetishism 297
Reference List 304
Appendix 1: Fieldwork Sources 321

List of Figures

Figure 1: 'Musical Networks' by Leyshon, additional dotted circle added (2000: 61). 16
Figure 2: A producer-songwriter editing vocals within a Digital Audio Workstation (DAW), Apple's *Logic Pro X*. (Image from my informant, Chris). 18
Figure 3: Simplified diagram of vital production → consumption stages in the Western commercial popular music industry, by the author. 21
Figure 4: A selection of plug-ins used by one of my informants for vocal editing. (Image from my informant, Chris). 24
Figure 5: A producer-songwriter's vocal chain in the DAW *Logic Pro X*, with reference to Figures 2 and 4. 30
Figure 6: Tabular summary of the intersections between theoretical and ethnographic uses of the concept of fetishism in the thesis. 53
Figure 7: Screenshot from the homepage of www.vocalkitchen.com, accessed 17 Feb 2022. 74
Figure 8: Introducing the central plug-ins of the thesis: *AutoTune*, *Melodyne*, FabFilter *Pro-Q 3*, *1176*, *LA-2A*. 77
Figure 9: *Nectar 3* interface. 80
Figure 10: iZotope *VocalSynth 2* interface. 81
Figure 11: iZotope's Vinyl before and after 2020. 86
Figure 12: Screenshot from a tutorial video by Alex Rome (Alex Rome 2021: 3'05"). 90
Figure 13: Two screenshots from the advertisement video for plug-in suite *CLA Epic* (Waves Audio 2020b). 92
Figure 14: Screenshot from an advert for *AutoTune unlimited* (Auto-Tune / Antares Audio Technologies 2022). 92
Figure 15: Sound box transcription for Brenda Holloway's 'You've Made Me So Very Happy.' (from Dockwray & Moore 2010: 190). 105
Figure 16: Hodgson's diagram of the three 'planes' which exist within a mix (2010: 156). 107
Figure 17: Hodgson's screenshot of a virtual mixing console, with pan pots noted (2010: 163). 107
Figure 18: Dibben's extension of Hall's 'rings' of interpersonal distance which form the foundations of proxemic theory. (From Dibben 2013: 114). 108
Figure 19: Visual representation of an 'out front' lead vocal in 'middle-of-the-road music' by Gibson (1990: 1:46:57). 110
Figure 20: Screenshots from 'The Art of Mixing' by David Gibson (TVHomeStudio 2012). 134
Figure 21: Comping a vocal in Pro Tools 8 using the playlists function. (from Thornton 2009). 146
Figure 22: Complex comp in a track by Jason (Top track is lead vocal). (2020J) 148
Figure 23: The auto-mode interface for *AutoTune Pro*, with scale and retune speed noted (Image from Antares Audio Technologies 2018: 11). 152
Figure 24: An example of a user manually tuning a voice using *AutoTune Pro*'s graphical mode (Image from Antares Audio Technologies 2018: 13). 152
Figure 25: An example like the above but including the 'make notes' function. 153
Figure 26: *Melodyne 5* interface. 154
Figure 27: Using the pitch modulation tool in *Melodyne*. (Celemony Software GmbH 2018: 92) ... 154
Figure 28: Using the note separation tool in *Melodyne*. This tool allows the user to split up notes and then drag the newly separated notes into new positions. (Celemony Software GmbH 2018: 108). ... 155
Figure 29: An example of how *Melodyne*'s promotional materials encourage a more holistic way of working with vocal sound, rather than just focusing on pitch. (Celemony 2020) 155
Figure 30: Complex manual timing of a very short vocal phrase against the DAWs tempo grid by Jason (2020J) 158

Figure 31: Using *VocAlign* to time-align one vocal track with another. The yellow trace over the guide vocal shows the proposed alignment (from Synchro Arts 2013: 12). 159
Figure 32: An example of a vocal phrase before and after Jason drew a clip-gain line to regulate the voice's loudness. (2020O). 160
Figure 33: An informant using the 'spectral repair' function in iZotope *RX* to remove "harshness" – before and after. (anonymized) 162
Figure 34: Musical theatre "straightness" labelled in red. 166
Figure 35: Slides labelled with red arrows. 166
Figure 36: Before and after the clip gaining process. 168
Figure 37: Chris's detailed approach to fading in the chorus of Audio 2. 170
Figure 38: 'GANG' vocals (2020H). 173
Figure 39: Robin's *Melodyne* interface, with his arrows pointing to his emphasis of "emotion". 175
Figure 40: Comp for phrase 'Don't you ever cut and run', before and after Jason's edit (2020M). . 177
Figure 41: iZotope *RX-7* audio restoration plug-ins. Screenshots from fieldwork (anonymized) 182
Figure 42: The anonymous producer-songwriter's *Melodyne* edit window, before and after "straightening" out the vocal syllable in question. 185
Figure 43: Glen Campbell's graphical representations of low-pass and high-pass wave-filters (Campbell 1922: 16), juxtaposed with a figure explaining 'low-cut' (=high pass) and 'high-cut' (=low pass) filters in EQ (Sound on Sound 2009). 194
Figure 44: Diagram of parametric equalizer parameters, taken from Bohn & Rane (1982/87). 195
Figure 45: An example of an EQ curve created using the in-built para-graphic EQ in the *Logic Pro X* DAW. 197
Figure 46: The 'rotary' interface of a Waves *SSL G-Equalizer*, modelled on the SSL *G Series EQ292* hardware EQ. 197
Figure 47: Instrument frequency chart from automated music mastering company LANDR (Seydel 2016). 200
Figure 48: A selection of first-round para-graphic EQ curves from producer-songwriters working in commercial pop. 201
Figure 49: An example of a 'separation in the mix' infographic used frequently by informal music production pedagogy website masteringthemix.com. (Mastering the Mix 2018) 203
Figure 50: Chris' vocal chain (inserts only) for Audio Track 2 (2020H). 205
Figure 51: 'Vocal EQ Cheat Sheet' from ProducerHive (Fitzjohn 2020). 206
Figure 52: Oliver's EQ curve, with the boost for "air" at 6130Hz circled. 215
Figure 53: Typical 'Parlak' EQ settings, taken from Bates (2010: 99).. 219
Figure 54: Chris and Jason experiment with surgical EQ (2020F, 2020N). 222
Figure 55: A producer makes a small cut specifically to address a "harsh" frequency in a female vocal, in a pedagogical video by SoundOracle.net (SoundOracle 2018). 228
Figure 56: Jason uses Fabfilter *Pro-Q 3* to create a dynamic threshold at certain frequencies associated with 'harshness' and 'air' (green and blue respectively). 233
Figure 57: *Soothe2* interface (oeksound 2021). 235
Figure 58: *Gullfoss* interface (from White, 2018). 237
Figure 59: Diagram of the process of peak limiting, taken from Case (2007) (Ch. 6, Figure 6.3). 246
Figure 60: 'The effect of different ratios on the level envelope of a waveform' (from Izhaki 2017: Ch.17, Figure 17.8). 250
Figure 61: 'The effect of different attack and release settings on a waveform. (a) The original levels before compression. (b–e) The resultant levels after compression.' (From Izhaki 2017: Ch. 17, Figure 17. 11). 250
Figure 62: 'Uplift compression' represented diagrammatically by engineer Hugh Robjohns in a Sound on Sound article (Robjohns 2013). 251
Figure 63: Hardware interface (a) and plug-in interfaces by Universal Audio (b) and Waves (c) of the *1176* compressor. 257

Figure 64: Hardware interface (a) and plug-in interfaces by Universal Audio (b) and Waves (c) of the *LA-2A* compressor. 258
Figure 65: *CLA Vocals* plug-in interface. 263
Figure 66: Oliver's vocal chain (inserts only) for Audio 3, with plug-in names and functions in Oliver's own words (2020Q). 267
Figure 67: Fabfilter *Pro DS* interface. 276
Figure 68: *Pro-MB* interface. 277
Figure 69: The Orban 5*36A Dynamic Sibilance Controller* (from SteamPowered Radio.com, 2020). 278
Figure 70: The vocal chain used by Austin Hull (Make Pop Music 2020), with each plug-in labelled in more detail. 280
Figure 71: Settings for the two rounds of de-essing applied by Austin. 281
Figure 72: *RESO* plug-in interface. Note the similarity to the interface of Soothe2 (shown in Figure 57, p. 236). (Mastering The Mix 2021). 300
Figure 73: Comments on an Instagram post by @masteringthemix, November 2021. 300

List of Audio Examples

Audio 1: A commercial pop track with vocals edited by key informant Jason, later re-edited by another vocal producer (2019-20).

Audio 2: A commercial pop track with vocals edited by my key informant Chris (2020).

Audio 3: A commercial pop track with vocals edited by my key informant Oliver (2020).

Audio 4: Excerpt of a commercial pop track with vocals edited by my informant Robin (2019).

Audio 5: An excerpt of Oliver's track, "raw" lead vocals only.

Audio 6: An excerpt of Oliver's track, edited vocals only.

Audio 7: An excerpt of Oliver's track, full mix.

Acknowledgements

First and foremost, I am indebted to my principal supervisor Professor Georgina Born. Her steadfast guidance and intellectual stimulation have substantially shaped the content of this thesis, and her moral support was indispensable during what was at times a long and arduous process. I also want to thank my secondary supervisor Dr Thomas Hodgson for joining the team at the last minute, and for providing feedback on my drafts as well as invaluable pastoral advice. Closely following, I am sincerely grateful to my field informants, especially my three key informants, all of whom must unfortunately remain nameless for ethical purposes. The account I give of vocal editing in this thesis would not have been nearly as rich without the wealth of data they generously provided.

My thanks go to the Arts and Humanities Research Council for supporting the project as well as providing me with further financial assistance during prolonged periods of uncertainty in my 3rd and 4th years. The Faculty of Music at Oxford University additionally provided useful financial assistance during my fieldwork. I am also much obliged to Professor Gascia Ouzounian and Professor Eric Clarke for their comments and advice at earlier stages of the DPhil. Several other academic mentors and collaborators also contributed valuable recommendations and discussions throughout this project, including Daniel Hulme, Dr Asaf Peres, Dr Toby Young, Professor Simon Zagorski-Thomas, Dr Jacob Kingsbury Downs, Dr Christabel Stirling, Dr Ed Spencer, Dr Adam Harper, Dr Daniel Walden and Dr Christopher Haworth.

My decision to embark on this dissertation was also influenced by important mentors from my time spent studying musicology at previous institutions. During my time at Cambridge University in particular, Dr Myles Eastwood encouraged my early interest in popular music studies, and Francis Knights provided vital support during my undergraduate degree and the transition to my master's. I would also like to thank Professor Jason Stanyek, whose supervision during my master's degree also significantly informed the topic choice and approach taken in this DPhil.

I have benefitted from a great sense of community throughout my time at Oxford. My DPhil colleagues in the music department were a source of great solidarity, especially during the Covid-19

pandemic. I also appreciated my choral scholarship from The Choir of the Queen's College and Owen Rees, which provided regular respite from the DPhil in my 1st and 2nd years.

I am especially thankful to my family for their unfailing support and care throughout the four years of this project and before. Their willingness to put me up occasionally and facilitate my erratic travel plans during fieldwork was particularly helpful. My close friends Eleanor Rashid and Isis Pinner also provided great advice, company and humour throughout this undertaking in spite of the long periods of separation instigated by the global pandemic. Lastly, I am eternally grateful to my partner Laurence, whose dependable companionship, conversation and proof-reading skills helped make this thesis a reality.

Introduction

The digital editing of singing voices is an increasingly dominant feature of commercial popular music production processes in the West. The voices heard on Western commercial popular music tracks in the late 2010s and early 2020s are painstakingly manufactured using digital technologies, and the degree of such editing is intensifying. A variety of music production experts—both during and after the act of recording—use a whole host of digital processes including tuning, comping, equalization and compression to refine the voice's sound. Both reflecting and driving this practice over the past two decades, the commercial popular music songwriting and production industries are increasingly coalescing around voices. In this developing economy, the copyright identity and potential economic value of a track are heavily influenced by the character of its vocal part, and it is increasingly common for music production professionals to advertise themselves as vocal editing specialists. In short, vocal editing has become a critical component of the labour process of popular music production.

This dissertation describes and interrogates the increasingly significant role of vocal editing in the landscape of Western commercial popular music production. I focus on three main research questions: 1. What are the technological processes that define the unique sound of the commercial pop singing voice, and how does the envisaging of this aesthetic ideal among producer-songwriters drive the practice further? 2. What conceptions of both the voice and technology underpin the escalating importance of vocal editing in the labour process of commercial pop production? And 3. How do the digital music technology industries both respond to and fuel developments in vocal editing practices?

To answer such questions, I conducted two years of multi-sited ethnographic research with music production professionals working in commercial popular music in the UK, culminating in a focus on three key informants. These informants spanned both producer and songwriter roles in their work, hence I use the term 'producer-songwriter' throughout to label their primary occupation. The study is limited to the UK context for reasons of timing and scope, however I concentrate on trends in the practice which seem to be more transatlantic or broadly Western and commercialized. More generally, the digital manipulation of the voice is now ubiquitous across most Western popular music

genres. However, whereas other contemporary genres are defined by the clearly audible, obvious sound of digital vocal manipulation (e.g. trap), in commercial pop, the total results of routine vocal editing procedures are intentionally understated and often difficult to detect in released tracks. My fieldwork focused on this subtle, routine work, as it felt necessary to try to account for these lesser-known features of what is currently a pervasive practice. My key informants were all embedded in this production culture where the voice inhabited a privileged position, and vocal editing occupied a significant portion of their creative labour.

I was alerted to the significance of the voice and vocal editing in the working lives of these individuals very early on in this study, as my attempts to find fieldwork contacts by advertising the research as a musicological study of vocal editing (including production, mixing and processing) frequently generated enthusiastic responses, such as this excerpt from an anonymous producer and university educator:

> I lecture in music production at [university] and I have been using Melodyne [tuning software] for years. My favourite part of engineering has always been vocal processing. I also have a few highflyer degree vocal processing *addicts* who might be useful to chat with. (Email exchange with engineer Brian,[1] italics added)

I introduce my fieldwork with this message, as it gives a sense of how the obsessive nature of vocal editing practices was a defining feature of the culture. This informant was quick to highlight how his students considered vocal editing their specialism—even a fixation—also implying that special digital tools (in this instance, *Melodyne*) were instrumental in such a practice. More firmly, and directly addressing the second of my three research questions, my overall argument is that producer-songwriters' digital vocal editing practices *fetishize* both the voice and the digital technologies which are used to edit it. Furthermore, these two different fetishisms which exist in the culture of vocal editing—of the voice, and of technology—are complementary and strengthen one another. I also show how specific companies within the Western music technology industry which develop *plug-ins*—purchasable software programs which add separate, new functions to a Digital Audio Workstation

[1] All my informants are referred to using pseudonyms.

(DAW)[2]—stimulate the increasing complexity of vocal editing practices by exploiting these two different kinds of fetishism in the design and marketing of their new products. I use fetishism as the main conceptual framework in this thesis because, throughout my ethnography, there were underlying similarities in the ways that both the recorded voice and music production technologies were regarded by producer-songwriters, which seemed to motivate the huge aesthetic importance they placed on vocal editing across their music production activities. As I will illustrate, my informants viewed the voice as radically different and special in comparison to other elements of popular music tracks, and they were constantly searching for and obsessing over the newest vocal editing plug-ins, each of which seemed to offer them some innovative yet elusive way to refine the voice's unique sound.

In response to the first research question, I illustrate how the sound of the commercial pop singing voice is, in its totality, produced through a host of cumulative digital processes. I focus on tuning, comping, EQ and compression as they were the four most central techniques in my informants' work. I also claim that these combined processes are oriented towards an overall aesthetic result which I term 'vocal foregrounding'. To use my informants' terms, the voice is intended to occupy an "up-front and central" position within the metaphorical space of the popular music "mix".[3] This aesthetic is explained and contextualized deeply in Chapter 2, but its character and the discourse which surrounds it are mentioned throughout the dissertation. Combined with the former themes, the thesis also implies that this vocal aesthetic is reinforced by the combined fetishism of the voice and technology.

A subsidiary theme of this dissertation is the gendered nature of vocal editing. In Western commercial music production, technical roles such as producers or engineers are frequently dominated by men, and artistic or performative roles, *especially* vocalists, are dominated by women

[2] Digital Audio Workstations (DAWs) are software programs which 'cluster' various types of music production functions including recording, editing using multiple different kinds of musical notation, and sound synthesis. (See Kvifte 2010: 213). Most music production in the digital age is undertaken using a DAW. Some common DAWs include Apple's *Logic Pro X*, Avid's *Pro Tools*, Ableton's *Live*, Steinberg's *Cubase* and Image-Line's *FL Studio*.

[3] Throughout, terms in double quotation marks refer explicitly to terms which my informants used, and single quotation marks to other more generic sources.

(see Provenzano 2019b: 66). What gendered assumptions therefore come to the fore in vocal editing processes, and how can their presence be explained by understanding the voice and technology fetishism which underpins this culture? The latter three chapters of the thesis address these concerns by showing how subtly gendered assumptions about vocal skills and qualities come to the fore in the usage of different vocal editing technologies, contributing to the broader academic conversation about gender in Western popular music production which has been spearheaded by scholars such as Tara Rodgers (2010), Catherine Provenzano (2018, 2019a, 2019b), and Paula Wolfe (2021).

Before providing more context, this research contributes to existing musicological literature in two main ways. Firstly, since my primary aim is to define the Western commercial pop vocal sound and the way such a sound is produced, I build on existing research into vocal tuning practices in Western popular music, expanding their focus by providing a more holistic account of the cumulative digital processes which produce this sound, rather than focusing on individual technologies in isolation. Secondly, I address ethnomusicologist Eliot Bates' call for popular music studies to engage more fully with 'the tools, techniques, workflows, and critical vocabulary of recording production, and with the audible aesthetics of recorded sound' (2013: 17), by engaging in as much detail as possible with the technicalities and discourses of Western music production. Relevant literature is appraised both at the end of this introduction and extending through Chapter 1.

Context and definitions: vocal editing in a networked songwriting and production industry

The setting for this project is the transatlantic commercial popular music production and songwriting industry(ies) from the years 2018-2021. Although my fieldwork was largely based in the UK, some of my informants were US-based or worked internationally (both in-person and online). As I will explain shortly, the sharing of pedagogical information among producer-songwriters online also produced a transnational community of practice spearheaded by big name producers across Western music production. My contextualization is based on a dialogue between my readings of relevant literature on this topic and my first-hand experiences of the industry during fieldwork. I consider production and songwriting as subsidiaries of a broader 'recording' industry which includes other facets of music

recording and distribution processes including, for example, mastering, streaming, and the various business activities of record labels and music retailers, which are not given extensive treatment here. For ease of interpretation, I borrow a diagram from Geographer Andrew Leyshon (2000), who two decades ago conceptualized the recording industry as a network, with nodes which coalesce into overlapping groups focused on the 'creative', 'reproduction', 'distribution' and 'consumption' elements of the process. Of particular significance here are the interactions between production, songwriting and 'musical instruments and supplies' (here, plug-ins), all of which lie in the 'creative' network:

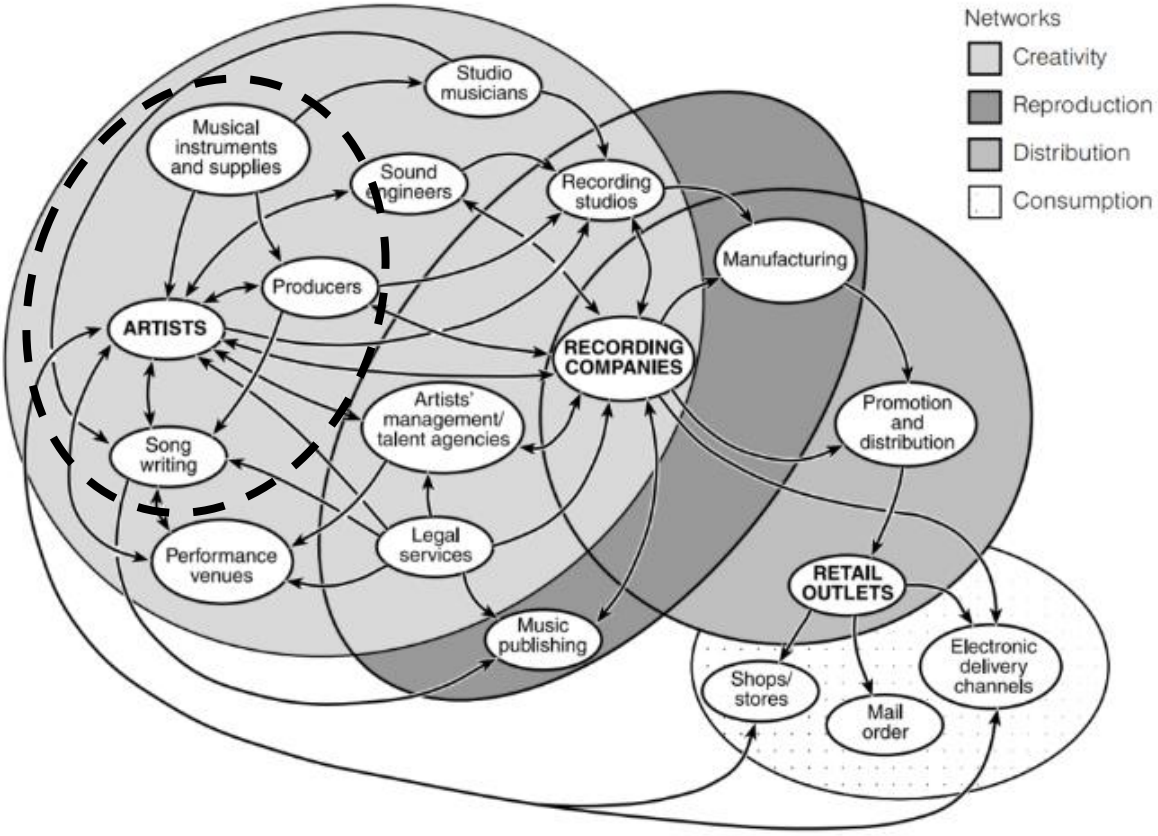


Figure 1: 'Musical Networks' by Leyshon, additional dotted circle added (2000: 61).

The significance of vocal editing in this industry emerges from a proliferation of shifting roles and terms in songwriting and music production, largely resulting from new capacities for distributed work in the digital age (ibid.). The collaborative work of producing a popular music track was previously confined to dedicated studio spaces, yet can now be achieved by small groups of individuals who work on the writing, recording and editing of a track remotely. Recording is increasingly more likely to take place in 'home studios' or 'bedroom studios' during the earlier stages of a track's inception, and projects are advanced through online communication and file sharing (see Théberge 2004, 2012). In this contemporary context, the separation between 'producers', 'songwriters' and 'artists' in this diagram does not quite hold. As many of my fieldwork informants would have attested during my ethnography, many who consider themselves music creators span some combination of these roles, usually working together with others to form the complete skill set required for the creation of a full track.

In this culture, *producers* are responsible for the technical management of the overall architecture of a popular music track, usually achieved digitally within a DAW (Figure 2), and *songwriters* for the inception and composition of a track including its vocal melody, harmony and lyrics, not necessarily including its technical aspects. In reality, the former is often integrated with the latter, because the documentation and development of musical ideas through recording is so foundational to songwriting in the 21st-century, and full drafts ("demos") of tracks are therefore usually completed with the assistance of a DAW (see Bennett 2011; McIntyre 2001). Max Martin is the most well-known example of a *producer-songwriter* of this kind (see Seabrook 2015). *Artists* are the public face of a released track, and they are not always present for the creation of one of their tracks from start to finish, but they often choose to perform active songwriting (and, to an increasing extent, production) roles. The term 'artist' often refers to a skilled vocalist, however sometimes producers and DJs also market themselves in a similar way to artists and claim a portion if not all of the public-facing credit for a track (e.g. David Guetta, Zedd, LAUV). Additionally, and missing from this diagram, is the newer role—*topliner*—a specific kind of songwriter who only provides vocal melodies and lyrics. Ester Dean is probably the most famous example. As journalist John Seabrook

has suggested, this role has been consolidated in Western commercial popular music over the past 20 years in tandem with a new kind of standardized songwriting technique he describes as the 'track-and-hook' method: this refers to a collaboration between a producer/track maker, who produces an instrumental part for a song within a DAW, and a topliner (or 'hook writer'), who then writes the melodies and lyrics 'on top of' this instrumental, usually orally through recording (see Seabrook 2015: 199-204).

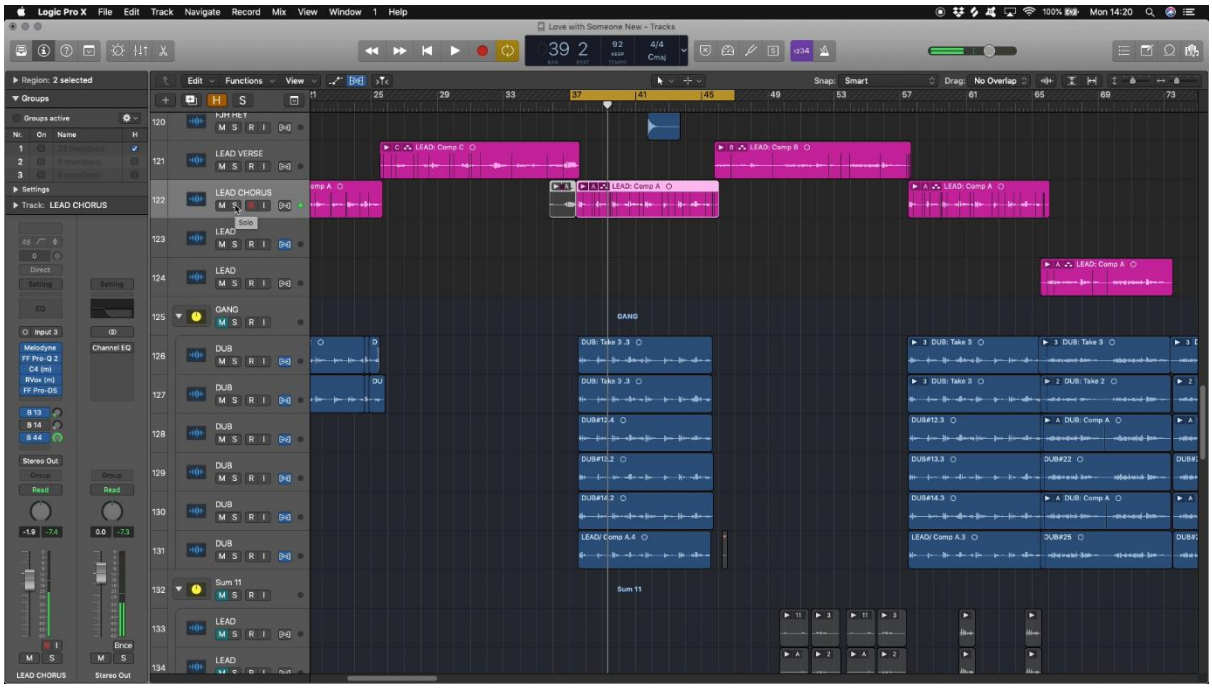


Figure 2: A producer-songwriter editing vocals within a Digital Audio Workstation (DAW), Apple's *Logic Pro X*. (Image from my informant, Chris). The pink audio tracks represent the lead vocal, alternating between the song's verse and chorus, and the blue tracks underneath are doubles and harmonies. Out of view are the other tracks in the session, which included synths, guitars, and drums.

Crucially, then, although the spheres of production, songwriting and artistry are somewhat overlapping, they are still mainly segmented into those who sing and those who do not: 'producer-songwriters', the subject of this dissertation, assume a largely technical role but are also involved in the songwriting in a non-vocal capacity, and they usually collaborate with topliners or artists who provide their voice as the missing link in the process. These producer-songwriters are self-employed and their work is project-based: they usually work in duos or small groups during the early parts of songwriting, then remotely and alone for later editing stages, and they tend to work with the same

people repeatedly when a good working relationship is established. Their aims are to build up a portfolio of projects and to gain work with increasingly reputable clients, leading to higher-stakes projects with higher earnings and recognition, and so on.

The combination of songwriting and production, oriented around the track-and-hook method, therefore represents a subsidiary or tributary industry within the recording industry as a whole, which trades in the production of vocals (see Figure 3). More specifically, there is both a consistent demand from major record labels for new artists to add to their rosters, as well as demand from established artists and their management who have already been signed by record labels for new tracks (or, more accurately, ideas for tracks) to release under that artist's name. The voice is a crucial currency within this economy, because recording companies will evaluate potential new artists based on demo tracks which not only illustrate the structure and melody of a song through a topline, but also importantly exemplify the unique character of an artist's voice. Additionally, it is common for established signed artists to put out "calls" or "briefs" to a network of producers, songwriters and topliners in order to source new material and, vice versa, it is common for aspiring producers, songwriters and topliners to "pitch" demos they have created to established signed artists and their management. In both cases the vocal part is usually the most identifying and important feature of a demo, and therefore determines whether it is bought by the artist. Adding to the process, if a track originally produced by some producer-songwriters and/or topliners has been bought by another more established artist, that artist's voice may be re-recorded and edited by other production professionals to replace the voice of the original producer-songwriter or topliner. Further, in the context of the track-and-hook method in which no track is complete without the perfect topline, it is common for established producers or artists to send voiceless instrumental tracks ("productions") to a list of topliners, and to choose the best vocal part from among the proposals received, or even to create an amalgamation of multiple submissions (Seabrook 2005: 200). Variations on this process are abundant, and this is the market within which my informants work. They could be working with an unsigned artist to add tracks to their portfolio in the hope that they will be signed; working with a topliner to write demos for pitching or to respond to a brief; or working with an established signed artist on a track they have either

acquired through their networks or written themselves to release. It follows that the voice is a driving force of this complex and fragmented labour process and a commodity within this economy, as toplines are continually bought, sold, and replaced by the voices of other topliners or artists.

This labour process extends and elaborates processes of Western record production which have been developing since the 1960s: Robert Toft (2010), for example, describes how the achievement of a 'hit' North American single in the late 1960s and early 1970s entailed several stages of versioning and re-release with multiple vocalists, and Phil Harding's (2020: Ch. 5.5) documentation of in-house writing, recording and production at PWL studios in 1990s London suggests that 'session vocalists' were commonly hired onto commercial pop projects in the late 1980s and 1990s to provide 'guide' vocal recordings for artists to copy in their final recordings of a track. At the start of the 21st century, and in the wake of the complete digitization of popular music's means of production, this labour process has become increasingly networked and transnational, and the specific role of the 'topliner' consolidated.

Also missing from Leyshon's diagram are the established industry terms 'mixing' and 'processing', which are fundamental to vocal editing practices, due to the more integrated role they increasingly play in the early stages of songwriting across Western popular music. 'Mixing' has been a core component of studio production since the 1960s and refers to the balancing and positioning of individual musical elements in a popular music track using a variety of techniques which manipulate their perceived prominence or spatial location (see Huber & Runstein 2018: 449). 'Processing' (short for signal processing) is an important part of mixing, and refers more vaguely to any over-arching digital modification that is applied to a single track in a mix (e.g. vocals, drums) (ibid. 403; see also Sterne & Rodgers 2011).[4] Historically, mixing was usually a separate, later stage in the production process, undertaken by a dedicated mixing engineer and oriented towards the track's final release (see Kealy 1979). Specialist mixing engineers are still very common in the industry, usually being hired to work on the later stages of professional projects, however the producer-songwriters I worked with

[4] Note that the term signal processing is defined slightly more specifically in science and engineering. Sterne & Rodgers suggest that it refers to 'the mathematical modelling of signals over time and through circuits, which then allows for their modification' (2011: 32).

also undertook substantial amounts of their own mixing and processing at early stages in the process of creating a demo, especially to the voice, to make the demo sound as good as possible in the competitive marketplace I have described. In addition, for high-stakes tracks, it is increasingly common for vocal editing specialists—sometimes termed *vocal producers*—to be contracted in specially to undertake the final stages of vocal editing and mixing/processing before the track is released. I use the term *vocal editing*, then, to refer to a holistic sense of the multiple ways the recorded voice is altered and refined throughout this long-winded process. My informants used this term in the same way to refer to all the work they would undertake on the voice in their own time after recording. Vocal editing permeates the production of a track: from its inception in songwriting (of which vocal improvisation and recording, rather than written notation, form a significant part), through the creation of a demo for pitch or otherwise, to the track's professional finishing (involving mixing and mastering) and release.

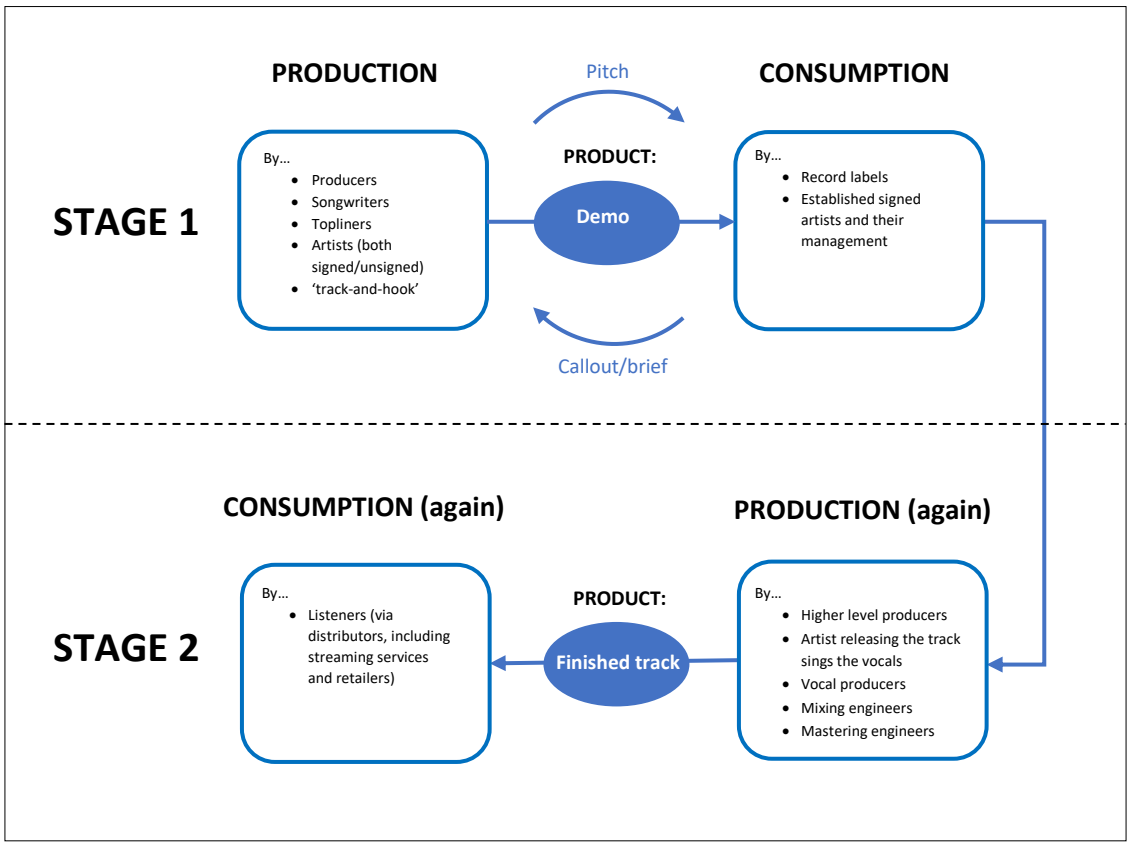


Figure 3: Simplified diagram of vital production → consumption stages in the Western commercial popular music industry, by the author. This is a generalized representation of what is usually a more iterative, multi-staged process dependent on individual projects.

Further, the increasingly distributed labour process of commercial popular music production is fertile ground for a separate industry to exert its influence and market its products: music technology, or what Leyshon's diagram calls the business of 'music instruments and supplies'. Paul Théberge (1997) wrote extensively about this industry in the late 1990s, illustrating how music creators were increasingly also consumers of technology, and my fieldwork showed how such consumption has developed and intensified into the early 21st century. In the context of vocal editing, the main types of music technology consumed by producer-songwriters are 'plug-ins': these are purchasable pieces of software with clearly defined functionalities which extend or enhance the audio editing functions already available in a DAW (Figure 4). As Jonathan Sterne has explained, in digital culture more broadly the term 'plug-in' can also mean 'add-on' or 'extension', referring to a piece of software which adds new functions to a host program without altering the host program itself (Sterne 2019b). Producer-songwriters will already be using a central DAW (e.g. Apple's *Logic* or Avid's *Pro Tools*) within which they undertake most of their recording, writing and production, and as they buy more plug-ins, they add to the functionality of this DAW, curating an expanding library of tools which can be reached in a couple of mouse clicks.

Importantly, and as I will show throughout the thesis in relation to vocal editing, the plug-in industry expands the reach of its products by exploiting the burgeoning online culture of informational exchange in the transatlantic music production industry. Specifically, my fieldwork showed how producer-songwriters increasingly turn to online sources of pedagogical information such as YouTube and Facebook to learn new production techniques and develop their practice, as well as to acquire new work. These online sources include video tutorials on YouTube shared by both individual producers and music technology companies;[5] Facebook groups aimed at the commercial pop recording and production communities, of which several of my informants were part (e.g. 'MAKE POP MUSIC: Pop Producers and Songwriters, 'Hey Audio Student', and 'Logic Pro X

[5] Online videos are cited throughout the thesis with reference to the YouTube channel or company webpage which uploaded them. All are detailed in the reference list (e.g. Alex Rome 2021).

Producers'[6]); online music production magazines (most notably Sound on Sound[7]); pedagogical blogs including those sponsored by music technology companies (e.g. producerhive.com, waves.com, musicianonamission.com, musicradar.com, izotope.com); and online music production forums (gearspace.com).[8] Such sources covered the same vocal editing techniques my informants were using, and employed the same kind of culturally specific language to describe the desired sound that the voice ought to achieve through carrying out such editing efficiently. I also knew from initial fieldwork that my informants had engaged deeply with the online pedagogical sources on YouTube, as this was their primary source of knowledge during their early years of interest in music production, and such DIY self-teaching usually preceded any formalized education or employment. Although my informants praised these online communities for their seemingly democratic, peer-to-peer nature, my fieldwork also showed how plug-in companies intentionally hijacked such communities to advertise their products, through the intersections between tutorial videos, targeted advertisements on YouTube and social media, and advertising partnerships with celebrity producers such as Chris Lord-Alge and Tony Maserati (see Bourbon 2020). In the case of vocal editing, the most influential plug-in companies include Waves Audio (sometimes referred to as just 'Waves'), Universal Audio, Antares, Celemony and FabFilter, and I show how the fetishisms underpinning vocal editing are both reflected and perpetuated by these companies' manipulation of their consumers. In these online spheres, the lines between pedagogies and commercial publicity and marketing are blurred, a phenomenon broadly characteristic of the internet in the 21st century.

[6] Links: https://m.facebook.com/groups/makepopmusic; https://en-gb.facebook.com/groups/heyaudiostudent/; https://www.facebook.com/groups/1611121552532875/.
[7] Link: https://www.soundonsound.com/
[8] Multiple written sources from these informal online fora are cited throughout the thesis using author names. Their provenance is made clear in the text as and when they appear, and all are detailed in the reference list (e.g. Computer Music 2013).



Figure 4: A selection of plug-ins used by one of my informants for vocal editing. These include an EQ and De-esser produced by a plug-in company called FabFilter (*Pro-Q 3* and *Pro-DS*), and two compressors produced by Waves Audio (*RVox* and *C4*) (Image from my informant, Chris).

Genre: 21st-century 'commercial' popular music?

My use of the label '21st-century commercial popular music' is intentionally broad. It refers to both a collection of musical styles whose outputs could belong to several similar genres, as well as to music produced within the specific industrial structure described above. The term 'meta-genre' or 'genre world' (see Brackett 2016: 1) may be most appropriate to describe this loose musical group, however in the interests of simplicity I tend to use the term 'genre' throughout. Here, I am also drawing on existing work by music theorist Asaf Peres (2016), who has attempted one of the only in-depth academic investigations of the aesthetic character of this genre.

I chose to focus on this kind of music for several reasons. Firstly, the *subtle* use of vocal editing processes was pronounced in this music, in contrast to other genres—most notably hip-hop and trap—where extreme digital modifications to the voice are used for aesthetic effect (see Brøvig-Hanssen & Danielsen 2016). In commercial pop, vocal editing processes are largely imperceptible to listeners untrained in music production, yet I also knew such processes were extensive and complex.

Further, in both academic and journalistic circles, the subtle, corrective use of tuning software—often described as 'AutoTuning'—has received substantial critical attention. This is because correcting vocal pitch is seen to undermine vocal authenticity and skill, provoking anxiety about how technology might be threatening our very sense of who we are, both auguring a 'posthuman' future (Prior 2018; Ragona 2013; Ledsham 2015) and amplifying gendered divisions of labour (Provenzano 2018, 2019a). I wanted to interrogate such dystopian accounts of vocal tuning and provide a more accurate account of the multiple, cumulative processes which produced the commercial pop voice.

Secondly, the aesthetic character of this music has received scant academic attention in musicology, particularly in music theory and analysis, even though there is a growing body of literature attempting to analyse the sound of popular music. Such literature tends to focus on other genres such as 20th-century rock and indie music, or electronic dance music, rather than the most commercialized 21st-century genres (Peres, 2016: 4). It is likely that this partially stems from the difficulties of defining genre labels like 'commercial pop' or 'mainstream pop', even though such terms are clearly culturally significant and do meaningfully group certain kinds of music together. I wanted to try and pin down the meaning of these terms, especially since it became clear during my fieldwork that such terms had meaning for my informants. Most of the professionals I spoke to seemed to understand what I meant by terms like 'proper pop*',* 'straight pop', 'commercial pop' or 'mainstream pop' alongside other genres such as 'indie', 'metal' or 'trap'. They also had a strong sense of how 'commercial pop' demanded a specific kind of vocal editing in comparison to these other genres.

Thirdly, the most commercialized popular music production processes in the West are heavily guarded. Owing to the shared intellectual property at stake in songwriting and production, as well as the intimate nature both of home and studio vocal recording setups, it is very difficult for outsiders to gain unrestricted access to, or detailed knowledge about these processes. In the highest-stakes commercial projects involving celebrity artists and producers, this kind of access is nearly impossible. The fact that the production of this music is so inaccessible to those interested in researching it may be another reason why there are only few studies in musicology and popular music studies which

address commercial Western popular music *production* processes in detail, and so many dealing instead with issues of reception. Given this context, I wanted to try and find more empirical evidence regarding the production of this music, particularly the production of voices which is especially guarded in these commercial scenarios.

Moving on to definitions, firstly, my use of the term 'commercial popular music' encompasses a more specific kind of music than that which is legitimized by institutions who measure popularity numerically, such as record charts and their relationship with Contemporary Hits Radio (CHR) stations across the West. This could have been one way of defining the scope of my project, but in the current climate this would include music which more clearly belongs to other genres, especially hip-hop and trap (e.g. Lil Nas X, Drake). These genres are intentionally excluded because their vocal editing techniques are very different to the type of music that I am focusing on. Indeed, one of the main characteristics of commercial pop is that it is defined relationally (see Brackett, 2016: 1-32), in broad contrast to these hip-hop derived genres.

Instead, my use of the term '21st-century Western commercial popular music' refers to a type of music with the following core features:

- This music **blends features from several dance-based, electronic popular genres** including dance-pop, R&B, electronic/electro-pop, and house (and derivative genres including mainly commercial house, commercial tech-house and tropical house).
- In contrast to the electronic music genres on which this music is largely based, **the voice is the most central feature of this music.** The instrumental parts of the music are produced mostly electronically, with layered collections of synthesized sounds forming a musical "bed" which supports the lead vocal part. The lead vocal is additionally supported by vocal doubles, harmonies and ad-libs which form a vocal "arrangement". The music is based around the interaction between these vocal and instrumental parts, which vary and develop between the different sections of a song (verse, pre-chorus, chorus, bridge etc).

- Common features of the **vocal style** or singing technique employed by vocalists include: the intentional exaggeration of the contrast between head and chest voice, sometimes used as a way of marking out different sections within a song (e.g. see Peres 2018); and the widespread use of a specific kind of vocal melisma, in which the starts and ends of certain sung syllables or words tend to gradually approach or glide towards or away from the intended melodic pitch, rather than hitting it exactly at the start of the syllable or word.
- **The voice is heavily edited, but in a way that remains subtle.** This contrasts with hip-hop genres, which frequently use extreme technological transformations of the voice to create strong aesthetic effects. The subtle editing is intended to iron out any perceived imperfections in a vocal recording; to augment the voice's foregrounded positioning within a track's "mix"; and to enhance certain perceived aesthetic qualities such as "brightness" and "presence".
- **This music is embedded in the commercial labour process of songwriting and production** which I have just described. Its creation is usually catalysed by, or aimed at, major (rather than minor) record labels and artists on their rosters. Put another way, these tracks are intended by their creators and editors to function as hits, whose hit status will nowadays be determined through internet-based metrics, such as a track's Spotify play count, rather than material sales (see Hodgson 2021: 8-9). Correspondingly, owing to the marketing activities of major record labels and modes of online circulation, tracks in this genre often reach very large global audiences.

The aesthetic centrality of the voice as well as the leading role of vocal editing in the production process is therefore one of the most defining features of this genre. Moreover, the main processes involved in commercial pop vocal editing—the focus of the latter three chapters—are the following:

- **Comping:** the process of splicing together selected portions of multiple recorded takes of a vocal performance in order to obtain what the editor perceives to be the best possible

performance. This is achieved using the waveform cutting, splicing and moving functions already available within a DAW.

- **Tuning:** the use of dedicated plug-ins—usually either Celemony's *Melodyne* or Antares' *AutoTune*—to correct the pitch contours of a vocal recording so that they conform more closely to a designated equal-tempered scale, as well as to occasionally enhance or bring out certain melismatic features of the voice (e.g. "curves" and "scoops") which are considered particularly stylistic. (Academic literature frequently uses the term 'pitch-correction' to refer to the use of such technologies, however the term "tuning" was more widely used by my informants and encompasses this more holistic sense of shaping the voice).
- **Equalization (EQ):** a technique for manipulating the relative prominence of selected bands of the frequency spectrum of an audio signal. Generally, in commercial pop production, lead vocals are "EQed" in a way which removes a large part of their low-frequency energy, and subtly emphasises their high-frequency energy.
- **Compression:** a technique for reducing the overall dynamic range of a sound signal, which in commercial pop vocal editing is used to increase the dynamic consistency of a voice by amplifying the quietest parts such that they reach a similar threshold to the loudest parts (i.e. make it more consistently loud). Compression is also conceived as a way of enhancing particular spectral qualities of a voice through their amplification, and the way this interacts with EQ.[9]

Apart from comping, these editing processes are achieved using dedicated plug-ins, which are "inserted" into vocal tracks within a DAW, and which apply modifications throughout the track automatically according to certain parameters set by the user inside each individual plug-in.

[9] These are by no means the only set of techniques involved in vocal editing. Other techniques such as reverb and delay, for example, are uniform components of vocal editing and contribute to the commercial pop vocal sound. I do not discuss these techniques in this dissertation, however, partly for space reasons, but also because such techniques manipulate the sound of a lead vocal in ways which are more external to its core sonic characteristics, related to the modification of the perceived space which the voice occupies, and occasionally creating rhythmic effects. During the vocal editing sessions I observed in fieldwork, my informants also spent much more time on tuning, comping, EQ and compression than they did on reverb and delay.

Furthermore, across my fieldwork, the cumulative application of these techniques in varying orders, and using varying numbers of plug-ins in sequence, was referred to as a "vocal chain". Transplanting the language and imagery of analogue mixing practices into the digital domain (Bell et. al. 2015), the "vocal chain" encompasses the series of plug-ins "inserted" into a vocal track, their order, and sometimes repetitions; as well as some additional plug-ins the track is "sent" through after these "insert" modifications have been applied (see Figure 5). In comparison to some of the other producers and engineers I worked with in less commercial genres, unique to my informants was the intense complexity and density of their vocal chains, which included multiple rounds of tuning as well as sometimes 7 to 10 different EQ and compression plug-ins in iteration, each enacting a very subtle modification to the voice. The hyper-detailed, compulsive nature of this practice was another general characteristic which led me to consider vocal editing through the conceptual lens of fetishism. This was a circuitous and obsessive process of refinement in which the ideal commercial pop voice was laboriously sought through an incessant building-up of very small adjustments, which my informants felt all contributed to a desirable final result. As I will explain later, the circular and synergistic quality of vocal chains is both escalated by, and directly benefits, the plug-in industry's ongoing design of new vocal editing technologies.

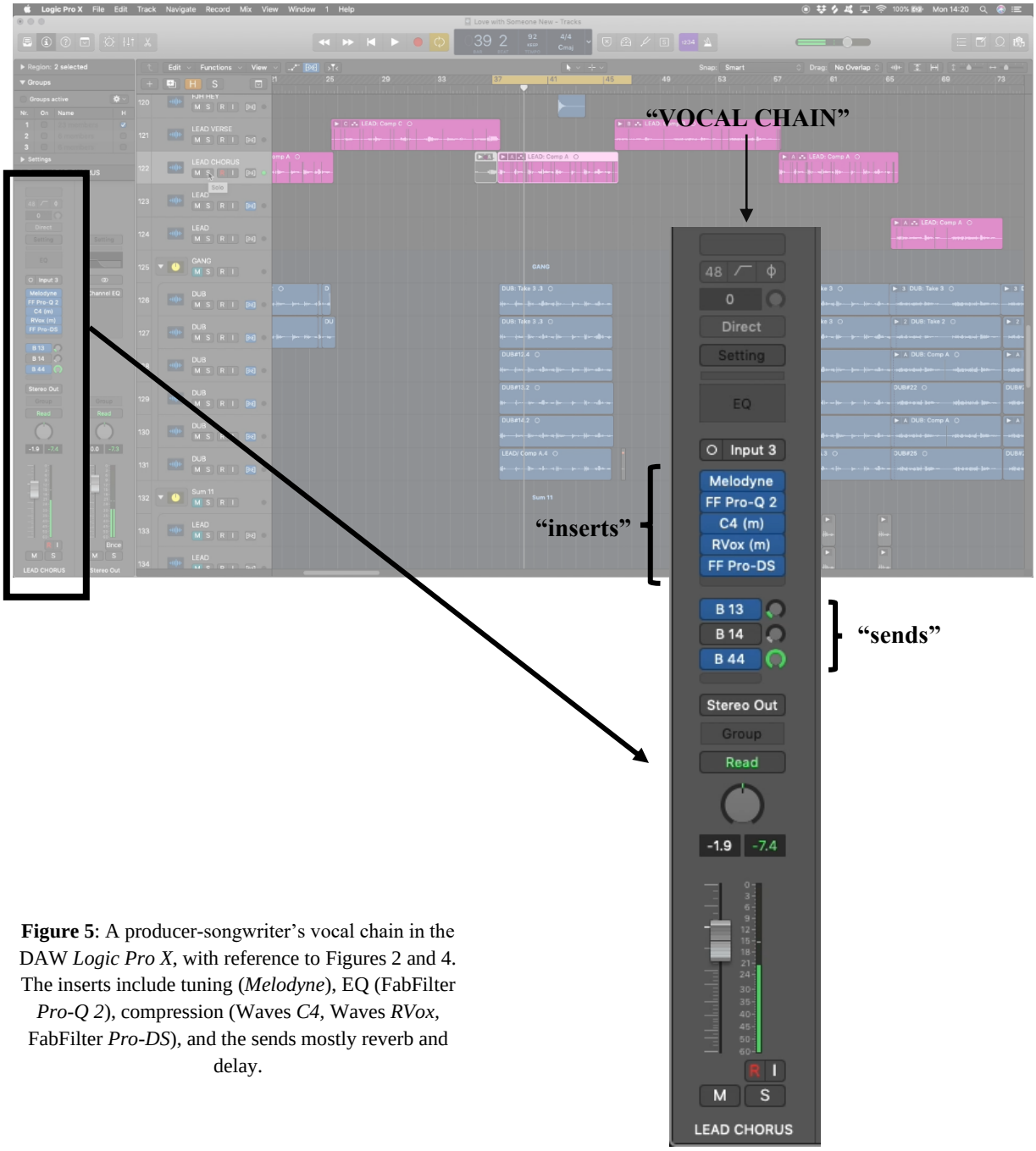


Figure 5: A producer-songwriter’s vocal chain in the DAW *Logic Pro X*, with reference to Figures 2 and 4. The inserts include tuning (*Melodyne*), EQ (FabFilter *Pro-Q 2*), compression (Waves *C4*, Waves *RVox*, FabFilter *Pro-DS*), and the sends mostly reverb and delay.

It is worth noting that this genre does occasionally use more extreme digital manipulations of the voice for aesthetic effects, including, for example, vocoding (using specially designed plug-ins such as iZotope's *VocalSynth 2*), octave doubling (frequently using a plug-in called *Little AlterBoy* by SoundToys), formant shifting, and the use of distorted vocal samples as instrumental-sounding melodic hooks. However, my analysis intentionally focuses on the most subtle, normalized vocal editing processes, rather than these extreme sonic trademarks. Again, this is because my core aim is to understand exactly how the overall sound of the commercial pop voice is achieved. The more subtle vocal editing processes are not just industry-standard procedures; they are defining aesthetic features of the music.

To give a sense of how this music sounds, some tracks exemplifying the commercial pop genre include 'Wish You Well' by Becky Hill & Sigala (2019), 'Don't Call Me Up' by Mabel (2019), 'Sweet But Psycho' by Ava Max (2020) and 'Levitating' by Dua Lipa (2021).[10] Taking the Becky Hill & Sigala track as an example, and describing the sound in non-technological terms, this voice is supported by an instrumental arrangement which blends features of house and trance. The lead vocal part dominates the track throughout, across two verses, two choruses, a bridge, and a final chorus repeat.[11] This lead voice is at times supplemented by other more subdued backing voices, especially as the music builds up towards the chorus, which are mostly sung by the same singer and either copy or harmonize with the existing voice (for example at 1'01" moving into the second half of the chorus). The vocalist mostly sings in a mix of head and chest voice in the verses, and especially gently at the start of the track, which contrasts with the fuller chest voice employed across the choruses. There is also a substantial melisma on the chorus lyric ('I wish you well', 0'45"), in which the first instantiation of the word 'well' (0'47") employs a very prominent pitch drift from B♭ to C, followed by further melismatic embellishments on the tail of the word and its repeat in the next iteration of the

[10] YouTube links provided here for ease of access: https://www.youtube.com/watch?v=lJOLxSaEffo, https://www.youtube.com/watch?v=9TQKyDD9Yig, https://www.youtube.com/watch?v=WXBHCQYxwr0, https://www.youtube.com/watch?v=TUVcZfQe-Kw

[11] This is a very simple explanation of the song's structure for the purposes of brevity. I direct the reader to Asaf Peres' detailed explanation of song structures in 21st-century popular music (2016: 126-182) for a more nuanced overview of contemporary 'verse-chorus' structures like the one heard in this song and the way they are defined by modern sound production.

phrase. The tuning of the vocal melody is consistently accurate throughout, relative to equal temperament, although the vocal style also entails that some core melodic notes are preceded by subtle glides upwards from a marginally lower pitch (e.g. at the lyric just described). The voice maintains a consistent volume despite what sound like substantial changes in performance energy at key structural points, for example, when contrasting between the second half of verse one (0'30"-0'44") and the beginning of the chorus (0'45"). The vocal timbre is consistently bright and intense relative to the instrumental elements of the track, and it has additional husky or breathy qualities at certain points, especially in verse 1, which contrasts between a breathy airiness in the first part (0'15"-0'29") and a more powerful husky sound in the second part (0'30"-0'44"). The various dimensions of this vocal sound will be explored more fully throughout the thesis, in combination with deeper explanation of how the different technologies across the vocal editing process contribute to its constituent characteristics. A statement summarizing the overall commercial pop vocal sound these techniques produce, in relation to a track by one of my key informants, is provided in the conclusion, on pages 294-296.

Methodology

My research employed multi-sited ethnography to gather qualitative data about vocal editing. The way my ethnographic explorations played out in practice were far more unconventional than I had planned, however, owing to the fragmentation of the songwriting and production industry and the effects of the Covid-19 pandemic. I tell the story of how this ethnographic fieldwork unfolded, offering valuable insights into the way this industry is unique and differs from its older, more institutionalized counterparts.

Initially, my plan was to probe existing networks of UK music industry contacts to try and find studio settings to use as bounded field sites (see Hammersley & Atkinson 2007; Born 2018). I quickly realized, however, that this access would be impossible without sanctioned membership in such sites, such as via temporary employment (e.g. see Bates 2015), study, or a very close contact. The initial contacts I made—mostly in-house producers and engineers employed by studios—though happy to give me short periods of time to chat about vocal editing, were unenthusiastic about my

intentions to observe vocal recording and production. Often older, experienced men working across genres, these producers asserted that vocal recording and editing in studios, where the vocalist was typically present, required carefully cultivated management of a tense social situation to "coax" the best performance out of a nervous singer. This was frequently compared to healthcare practitioners' exercise of good "bedside manner" as a way of facilitating genuine and unproblematic interactions with patients (2018B[12]; White 2006; see also Atkinson 1997: 98). One engineer even called it "Derren Brown-ing"[13] the vocalist (2019F). Arguably, these ideas are generational relics, founded on assumptions about the emotional volatility, fragility and irrationality of vocalists which are obviously gendered.[14] Still, it was made clear to me that these were delicate situations which could easily be disrupted by the presence of an outsider.

Nevertheless, several of these contacts did allow me to observe some sessions, during which I learnt a considerable amount about vocal recording and editing conventions. More revealing, however, was that the most commercialized genres I wanted to study were seldom recorded or produced in these studio settings. This prompted me to seek alternative contacts using both social media and snowball sampling (see Ritchie et. al. 2014: 129) from existing informants. Using these new contacts, I began to gain access to commercial pop projects which had been conceived by disparate groups of individuals in collaboration, each with their own recording and editing setups. Accounting for this project-based industry structure, I gradually took a more individualized approach to the ethnography, centring on separate producer-songwriters and their working lives rather than places or sites of production. Around 50 individuals floated in and out of the purview of my study, with whom I conducted a mixture of informal meetings and semi-structured, longer interviews (these are all listed in Appendix 1). I also undertook some of my own recording and songwriting work in the university studios at Oxford in collaboration with an existing pop vocalist contact, as a way of

[12] All formal fieldwork engagements are listed in Appendix 1 and referenced in the format YEARX (where X is a letter or letters).
[13] This refers to the British TV personality Derren Brown whose reality shows centre around him hypnotizing participants into performing certain actions or changing their beliefs. The implication here is that producers should covertly manage the vocalist's psychology in a way which leads them to perform better without realizing it.
[14] See Provenzano (2019a: 95-104) for similar discussions regarding the gendered nature of vocal labour in pop recording scenarios.

familiarizing myself with songwriting processes and digital vocal editing techniques (2019S, 2019T). Some contacts, however, were willing to offer more time and energy to the project than others, mostly in the form of repeated interviews and observational data. This culminated in a focus on three key informants:[15] Jason, Chris and Oliver. These informants spanned multiple levels of career-progression as well as stylistic emphasis within the commercial pop genre, providing a neat overview of the vocal editing practices I wanted to analyse. At this point I also gathered transcripts and notes from the internet sources I mentioned on p. 23, based on what I had already discovered. These resources functioned as a secondary source of ethnographic data which allowed me to compare between my informants' practices and those that were represented as archetypal in online pedagogical contexts, and the frequent similarities provided evidence for the huge role played by online cultures of pedagogical exchange in the development and consolidation of commercial pop vocal editing.

Despite the increasing focus on key informants as the research progressed, my contact with the wider group of producer and songwriter contacts I had developed extended across the project, and several of the most generative interviews conducted with these contacts happened during the time I was undertaking the richest periods of observation with Jason, Chris and Oliver in 2020. This broader base of informants therefore formed an indispensable part of the study, both prompting my focus on these key informants and providing very important contextual information along with accounts of alternative vocal editing approaches. In particular, the producer-songwriters I call Robin, Michael, Ryan, Diana, Fred, Tim and Magnus offered rich perspectives which I draw on throughout.

Before I introduce my key informants, there is more to say about the impact of Covid-19 on this ethnography, which was surprisingly both negative and positive, owing to the unique character of this industry. When the pandemic started, the progress of the ethnography was uncertain, given that my strategy for creating reliable contacts depended on informal in-person meetings, and opportunities

[15] The term 'key informant ethnography' circulates widely in the methodological literature, however it is frequently discussed as if it a were research method in its own right (Tremblay 1957; Marshall 1996; van Haitsma 2009)—more concise and directed than a fully-fledged ethnographic project—rather than considered as an organic outcome of the ethnographic process, which it usually is (e.g. see Whyte 1943). My focus on key informants arose naturally from my exploration of the industry rather than being an initial strategy for limiting the scope of the project, as well as being complemented and contextualized by my wider fieldwork online as described earlier.

for gaining observational data were lost. My informants began working from home, however, which generated a new and unexpected source of field data. I asked several informants if they could take screen recordings of any vocal editing work they were undertaking, as well as, if they were able, talking as they worked about what they were doing. Many were happy to oblige, as it provided a change from the monotony of working at home, and the power of their modern computers meant that screen recording did not disrupt their regular working habits. My key informants, in particular, sent several videos covering multiple vocal editing projects. This kind of observational data provided a window into the day-to-day working lives of these producer-songwriters without the social awkwardness of in-person observation, and was far easier to analyse than the field recordings I had taken in-person. I could play, pause, rewind, and take screenshots of their work in-progress, and I could compare their editing procedures with the real-time voiceovers. This illustrated both how my informants understood the work they were doing, as well as which portions of their practice they felt were worth verbalizing and which were not. This situation therefore revealed just how much of the work of commercial popular music production was already suited to remote-working scenarios. It also highlighted a potentially novel form of ethnomusicological data in the form of screen recordings of the DAW, which are a fruitful opportunity for ethnographic analysis of popular music production given the tense social situations which surround such work in-situ.

Indeed, I had several conversations with producers during this period which revealed much about the changing nature of this industry and the central role of vocal recording and editing within it. They explained how, although collaborative songwriting with other vocalists, topliners or artists worked better in-person, routine vocal editing was unaffected, as the producer-songwriter usually conducted this alone after writing and recording anyway. What's more, the increased capabilities of video-calling as well as the availability of software especially designed for remote musical collaboration (e.g. Splice, Pibox and Avid Cloud Collaboration for Pro Tools) made it possible for collaborative songwriting processes to happen online in some cases. If not, many of the vocalists these producer-songwriters worked with could record and send vocals to them for editing. The greater

individualization of this process facilitated my observation of the editing stages, as the physical presence of a vocalist in a studio session was no longer of any concern.

With this background in mind, my first key informant was Jason, a "topliner, vocal producer, mix engineer and producer"[16] based in London. I contacted Jason through an administrator of a Facebook group and we first met in October 2019 at a contemporary multi-functional workspace in central London where Jason occasionally hired a small recording room for his work. We began chatting about his background and my study, and this meeting spawned several observational opportunities at both this space and his own home, as well as a huge number of screen videos he sent me during the pandemic covering multiple projects across 2020 and 2021. Jason got into music through playing in a pop band in his teens, which sparked an interest in music production. He taught himself the rudiments of production using the internet, which led to a stint of recording and producing tracks for a music library. Through doing this work his interests turned to commercial pop production and songwriting and he completed a master's course in production[17] where his interest and skills in vocal editing peaked. When we met, he was around 30 years old and developing an increasingly prestigious portfolio of songwriting and production work. This included both pitching demos to established artists in collaboration with other topliners and artists with whom he had continuous working relationships, as well as working as a vocal editing specialist for an established producer-DJ duo. Some of the tracks that Jason had production credits on had Spotify play counts running into the millions. His description of himself as a "vocal producer" reflected the large amount of time he spent editing vocals in comparison to the instrumental portions of tracks (which were usually the preserve of the artist or DJ who owned or bought the track). At the time we met, Jason's ongoing production projects were numerous and he was able to make a very comfortable income from working in music full-time. He would regularly complete entire songs in one or two day-long writing sessions with topliners or artists, and he was able to finish vocal editing projects on existing tracks in two to three days, or even a matter of hours, for which he charged between £300-£2000 depending on the size of

[16] These are Jason's own terms, provided on one of his social media profiles.
[17] I refrain from including the exact titles of any university courses to maintain the anonymity of my informants, because the specificity would easily reveal the names of the institutions which offer them.

the job. Jason was very frank about his intentions to work on only the most commercial projects—"I'm not trying to be an artist […] I'm not interested in cool, I just want to get some hits" he clarified in one interview (2019X). Through such work, he aimed to increase the size of his portfolio and network as quickly as possible. He also had an amazing ability and drive to theorize at length about his musical work. He had a deep interest in the science of acoustics, the physiology of the voice, and Western music-theoretical principles, and was constantly trying to integrate these into his vocal editing work. I also mention him first because, across my fieldwork, he was by far the most forthcoming informant and provided me with vast quantities of data. I have tried to balance this by not drawing any conclusions in my writing which are based on Jason's data alone, however it would be impossible for me to deny that many of the avenues I pursued throughout the project were sparked by his material.

My second key informant was Chris, a producer, songwriter, artist and university tutor of a similar age to Jason, based in a major northern city in England. I contacted Chris through another songwriter whom Chris had taught. We first met in early 2020 and, owing to distance and the pandemic, only ever spoke over video call. Like Jason, Chris had developed an interest in production in his teenage years, which led to a formalized undergraduate qualification in music production (although he described how most of what he learnt about production came from the former period of self-teaching, and his university education was mainly a way to enhance his reputation). Chris differed from Jason, however, as despite working in commercial pop, his main artistic interests lay in a more indie-alternative style of popular music. Given that music production and songwriting were his main source of income, he felt unable to concentrate solely on his preferred genre as his commercial songwriting and production projects were more reliable and generated the most income. Consequently, Chris' working life was split across multiple ventures. His work in commercial pop involved close collaboration with aspiring individual artists who were paying him on-spec to work on their projects, and with whom he worked repeatedly due to trust built up over time. He also worked on his own projects in the more indie-alternative genre where he took a more dominant artistic role. Lastly, he worked intermittently as a tutor at a local university, giving courses specializing in

songwriting, production and music business skills. All the observation I did with Chris concentrated on a commercial pop track he had helped to write and produce, working with a relatively unknown, although developing, artist (Audio 2). In contrast to Jason, Chris' work in the commercial genre was doing reasonably well but wasn't quite having the same reach online (although this wasn't really one of Chris' major concerns or measures of success in his work). The material I gathered from Chris was therefore slightly different to Jason's because he clearly found commercial pop vocal editing quite menial and tedious: he often talked about the techniques in quite a satirical, detached way, and demonstrated a much deeper interest in aspects of vocal editing and mixing he considered more profoundly artistic and applicable across all popular music production. Chris' material was therefore very useful for interrogating which characteristics were specific to commercial pop vocal editing and which were conceived as more general or foundational to popular music production.

Finally, my third key informant was Oliver, a much younger producer-songwriter and vocal producer, recently graduated from a university in the south-east of England where he completed an undergraduate course in music production. I met Oliver through the producer and university educator whose email I mentioned at the start of this introduction (Oliver was one of the "vocal processing addicts" this contact was referring to). We met in late 2019 at a café in the university, where he brought his latest Apple MacBook and state-of-the art headphones to show me some of his recent productions. During this first meeting he described how, exactly like Chris and Jason, he had developed an interest in music production at an early age and began to teach himself on the internet using informal pedagogical resources such as those found on YouTube. This informed his decision to study music production, following which he was attempting to develop his production and songwriting career by working on a variety of projects in collaboration with other ex-student artists and vocalists. My observation with Oliver, which was conducted remotely during early 2020, focused on one of these projects (Audio 3). Since Oliver was at an early stage in his career, music was not his only source of income, and he was reliant on additional employment to support his living costs. In addition, Oliver's projects were in a more specific musical style within the sphere of commercial pop—he felt they belonged to 'commercial house' or 'commercial tech-house' genres, which he

understood as subgenres of commercial pop containing recognizable house and tech-house features, including a 'four-to-the-floor' beat and a specific palette of synthesized sounds. Oliver's vocal chains tended to be a lot longer and more complex than Chris and Jason's, and he also discussed how he had picked up a lot of vocal editing techniques from the internet and from university without thinking about them too much. Oliver therefore seemed like a target demographic for the plug-in industry: a classic Gen-Z 'bedroom producer', he had an intense desire to push his post-university career forwards and therefore professionalize his vocal editing skills as quickly as possible, so he was highly likely to buy any new products he thought might give him more of a professional edge.

Regarding data analysis and interpretation: after I had gathered several screen-recordings and began to focus more solely on these three key informants who had provided me with the most data, I watched their videos carefully, creating transcripts and taking screenshots. I also asked them to send me audio files of their released tracks, as well as of both the edited and unedited voice in isolation of other parts of the mix, so that I could listen to the sound of voice in more detail at moments where I wanted to understand the editing more closely. I then began to group together core common themes and trends in vocal editing across the informants, with reference to earlier interviews, revising these themes in light of lengthy follow-up interviews which were undertaken in order to clarify aspects of their process in more detail. Much of this work involved corroborating to what extent their practice as it was represented in the videos—both verbally and visually—matched up with the way they had described it to me in conversation, as well as disentangling in follow-up interviews the kinds of terms they were using to describe certain sonic features of the voice or technological processes. I also triangulated this data with the knowledge I had gained from both other informants and online pedagogical sources to differentiate between practices that were specific to commercial pop and those which seemed to be preached as general practice across Western audio engineering and production.

Perhaps regrettably, all of my key informants are male, however this reflects the existing gender imbalance in the industry, with production-focused roles overwhelmingly dominated by men and vocalist/artist roles dominated by women. Nevertheless, throughout my fieldwork, I met and spent time with several female producers, engineers and songwriters, including one female vocal

editing specialist, Diana (2019U). I had intended for such relationships to generate further observational opportunities, but these contacts, like many others, dried up after the first meeting as the informants either became too busy, lost interest in the study, or most likely weren't prepared for the more involved observational work I was asking of them. Regarding gender issues, in my interpretation of this data I am therefore only able to comment on how gendered assumptions about the voice or the role of vocalists are espoused by male producer-songwriters, but there is scope for future research to provide an account of how such assumptions may be reproduced and/or differ in the case of female music production and songwriting professionals.

In order to paint a picture of the commercial pop vocal sound produced by these individuals, three central tracks are referred to throughout, each of which constitute a project completed by Jason, Chris and Oliver respectively. I have full written consent from all three informants to discuss the production of such tracks here. Audio 1 is a track with vocals partially edited by Jason, 'All I Got'; Audio 2 is a track with vocals edited by Chris, 'Fall in love with someone new'; and Audio 3 is a track with vocals edited by Oliver, 'Dangerous'. (Importantly, in the case of Audio 1, although part of the writing of the song and the preliminary stages of vocal editing were undertaken by Jason, for the final released track provided here the vocals were re-edited by another producer. This means that the vocal sound is not the result of Jason's techniques alone.[18]) A small number of other audio tracks (4 to 7) are included with the thesis to illustrate specific points, and they are referenced throughout the text as and when they appear.

Methodological literature review: understanding Western popular music production through technology and aesthetic discourses

The methodology adopted in this research is informed by a blending of various literatures across ethnomusicology, anthropology and popular music studies. My incorporation of this interdisciplinary literature aims to account for the relationship between the wider musical culture surrounding vocal editing and the sound of the actual voices edited by these music production professionals.

[18] I mention this at Jason's request as he felt it was important not to take any credit for the other producer's techniques.

Firstly, Eliot Bates' (2010, 2013, 2016) studies of the studio production of Anatolian folk music in Turkey have similar methodological aims. Bates (2013) suggests that popular music studies' combination of approaches from two main academic encampments—cultural studies and sociology, vs musicology and music theory—has resulted in methodological deficiencies across the discipline. In particular, Bates argues that popular music studies' roots in cultural studies and sociology are to blame for the discipline's failure to engage with discourses and technicalities of sound recording, even though they are fundamental to how popular music is produced in most industrialized societies. Consequently, Bates argues that the discipline has neglected to pay the right kind of attention to the sonic aesthetics of popular music (ibid. :17). Similarly, he criticizes how several musicological attempts to analyse the sounds of popular music draw significantly on the language of Western classical music theory (e.g. Moore 1997), mostly resulting from the way the design of undergraduate music courses across Western universities has, historically, tended to consider this kind of music-analytical training as a fundamental component of musicological education (and, by extension, implying that this is a universal way of understanding and elucidating musical aesthetics) (Bates 2013: 17).[19] I think there are numerous reasons for these deficiencies, including: the steep learning curve of sound engineering; financial and social barriers to entry; the continual institutional intertwining of popular music studies with either musicology or cultural studies/sociology; and a lack of facilities, funding and time in the few university departments where popular music is studied. I agree, however, with Bates's suggestion that greater engagement with the theories and methods of anthropology (i.e. ethnography) would ameliorate the situation, given that anthropology endeavours to both delineate and interrogate the discourses and practices which are unique to any given culture (especially in the case of cultural production: the discourses and practices which define the creation of specific kinds of cultural products).

Turning more closely to the issue of discourses, Bates (2010) also highlights the ways music production processes are defined by the various terms, often unique to particular cultures, that music

[19] See also Tagg 1979. I should note that in recent years several Western music departments have taken significant steps towards changing this status quo.

creators use to specify their aesthetic goals. In Anatolian folk music production, for example, the terms ‘parlak’ and ‘büyük ses’ (which Bates loosely translates as ‘shine’ and ‘big sound’ respectively) embody sonic ideals for Bates’ informants, and specific production techniques are used to attain such ideals. More specifically, probing earlier literature which precedes Bates’ approach, ethnomusicologist Steven Feld and musicologist Paul Théberge have both discussed how discourse about musical aesthetics can be interpreted as a verbal representation of the concepts underlying the way any given culture understands, appreciates and makes their music. These ‘aesthetic discourses’ (Born 1997: 483, see also Born 1995: 18-20) are usually built around metaphors relating to concepts or phenomena existing in other domains of everyday life, such as the use of aquatic metaphors by the Kaluli people of Papua New Guinea (Feld, 1981). Feld suggests that such aesthetic discourses should be considered fundamental to the ‘music theory’ of any culture. Since the term ‘music theory’ is usually understood narrowly in musicology to refer to the theoretical discourse of Western classical music composition, the term is seldom mobilized more widely to refer to ‘a conceptual system of cultural principles that scaffold cognition, production, and interpretation of patterned sound in a social setting’ (1981: 45). As Feld has argued:

> Ethnomusicologists have uncritically accepted “music theory” as a special accomplishment of the West that allows “us” to analyze “them.” I have spoken with scholars who insist that nonliterate people, who do not write music, cannot have “theories of music.” One text goes as far as claiming “…in primitive music a scale does not exist in the mind of the native musicians, so a musicologist must deduce it from the melodies” (Nettl 1956:45). […] Statements like the one above glorify the position that music theory is an artifact of studying “the notes”; this creates a […] lack of concern with theory as a social phenomenon (1981: 44).

Feld’s ideas suggest that attempts to extrapolate universally applicable theoretical structures from music reflect a lack of awareness that musical theories are ‘social phenomena’ with their own associated discourses, rather than abstract systems existing innately in music which can be decoded. Further, Théberge, in his pioneering study of the Western music technology industries in the late 1990s, highlighted how musical theories are usually intertwined with modes of musical production, which involve various kinds of technology (1997: 155-185). In the case of Western classical music, for example, the development of staff notation in the context of institutionalized music making from the 14th to 19th centuries resulted in a valorization of the composer and the score or ‘work’ as the locus

of aesthetic innovation in classical and romantic music, and this situation was instrumental in creating a conception of music (among Western upper classes) as an art form based on 'detailed calculation of technical means' (ibid.: 179). By extension, such a deeply rooted conceptualization of music itself lay the foundations for music theory as an academic discipline, where it was conceptualized as a method of understanding or revealing such technicalities across many kinds of music, rather than the product of a specific cultural setting. Building on the general flavour of Théberge's work, Bates' arguments suggest an alternative approach, which accounts for the fact that the development of music technologies across the 20th century and their associated practices and discourses are essential to the creation of popular music. Such a perspective also accounts for the fact that, intertwined with such technological developments, popular music production centres on the manipulation of timbre and space, alongside other musical parameters such as melody, harmony and rhythm, which are given greater attention in Western classical music theory and analysis. It is these parameters (timbre and space) and other sonic subtleties, such as continuous rather than discrete pitch contours, that are least well captured by Western classical approaches given their dependence on Western staff notation.

Even further, Théberge suggested that the increasing embeddedness of digital technologies in Western music production had led popular music practitioners to develop a richer aesthetic discourse about their music, centred on an ambiguous ideal of 'sound' to which their music-producing activities were increasingly oriented (188-209). Théberge argued that this discursive proliferation paralleled a change in the means of music production, since musicians' refigured status as 'consumers of technology', along with the increasing rationalization of music production processes opened a progressively expansive world of sonic possibilities. To add a further layer of interpretation to Théberge's work, he implies that the aesthetic discourse of popular music production has become more abundant and vibrant with the advent of digital music production technologies. The ability of such technologies to provide music creators with 'any sound they can imagine' (the title of Théberge's text) led them to develop a complex vocabulary to describe such sounds, which is fundamental to how they conceptualize and organize their musical activities. In the case of commercial pop vocal editing, not only has the increasing proliferation of digital vocal editing technologies amplified the complexity

of this discourse, but it has also developed in tandem with the already complex and evocative discourse used to describe the human voice—a theme I explore more fully in Chapter 1. To quote Théberge at length:

> The ideational aspects of [popular and non-western music] are intimately bound to the very processes of music-making, to the instruments employed, and to the characteristic verbal constructs—often oblique and metaphoric in nature—used to describe musical activity. […] Musicians and studio engineers have developed a virtual "theory" of sound through the deployment of metaphoric expressions and binary oppositions that, first, define their experience of sound in meaningful ways and, second, help organize their sound-making activities […] [a] vast array of paired terms, such as fat/thin, warm/cold, wet/dry, clean/dirty, organic/processed, are used to describe fundamental aesthetic values through which sounds in a given context are assessed... (Théberge 1997: 162; 207-208, citing Feld 1982: 163-165 and Feld 1981).

Some other foundational characteristics of an aesthetic discourse of Western popular music production are outlined by anthropologist Thomas Porcello, whose writing on discourse about sound in North American recording studios suggests that notions of sonic purity and the fidelity of recorded sounds are valued by North American recording engineers. This is reflected in the variety of metaphorical terms they use to refer either to an undesirable diffuseness of sound (e.g. 'blurred', 'distant', 'breathy', 'muffled') or the over-modification of a performed sound through recording and production (e.g. 'blanketed, 'boxy', 'brittle', 'thin') (Feld, Fox & Porcello 2004: 327). He also more generally highlights the recurrence of metaphors synaesthetic to touch and sight in this discourse, which are highly codified (e.g. 'wet', 'dry', 'deep', 'bright', 'round') (ibid. 324). Jonathan Sterne and Tara Rodgers (2011) have also highlighted how metaphorical constructs such as rawness ('raw' meaning unprocessed), decay (e.g. in relation to the ceasing of a sound), and moving water (e.g. 'wave', 'current', 'channel', 'flow') are fundamental to the discourse surrounding digital signal processing in the West—a core component of most audio-technical disciplines linked to the structure of electronic circuitry. By extension, such constructs trickle down into vocabularies of Western popular music production.

In a more contemporary context, Théberge (2013) has also discussed the increasingly important role of digital *visualizations* of music and sound in informing Western popular music production processes. Théberge explains how DAWs represent music visually in the form of horizontal layers of musical parts and a vertical grid representing the passage of musical time (ibid.: 2;

see also Kvifte 2010: 214). On a more precise level, the user primarily interacts with sound waveforms, visual representations of sound which show up on the horizontal 'tracks' and can be manipulated by the user at will. Most importantly, Théberge argues that the various visualizations of audio in the DAW result in a 'qualitatively different form of material engagement with sound' (ibid. 9) in comparison to earlier sound recording and production with hardware where such digital visualizations were less important. Such visualizations provide the ability to both zoom in and zoom out on portions of sound and arrangements of sounds in their entirety, affording increasingly precise modifications to individual sounds or timbres as well as the manipulation of larger structural features of tracks (ibid.: 9-12). In the case of commercial pop vocal editing, the singing voice is visualized in different ways across all the technologies mentioned throughout this dissertation, and the way such visualizations inflected my informants' practices was a recurring theme throughout my fieldwork.

Despite my broader focus within the literature so far, there are some studies which address both vocal editing and the analysis of vocal timbre. Two recent PhD theses directly address vocal tuning techniques in the context of North American music production and engineering using ethnographic techniques. Catherine Provenzano (2019a) uses ethnographic interviewing to argue that the use of tuning software across Western popular music production has resulted in a decoupling of pitch from the labour and skill involved in popular music singing, and correspondingly an increase in the value of vocal 'emotion' as a key marker of vocal labour, both in popular music production and its reception by Western listeners. Owen Marshall (2017) explores ethnographic data from two music studios, using conversation analysis techniques to account for the edited voice as a 'collective sociotechnical accomplishment', and similarly to outline the aesthetic currency of vocal 'emotion' among studio engineers. Recent publications by Provenzano also develop her arguments about the gendered character of vocal labour and its intersection with vocal tuning technologies (2018, 2019b). Jonathan Sterne (forthcoming) is also preparing a chapter on the cultural history of tuning software. Other sources contemplate the transformative 'posthuman' potential, both utopian and dystopian, of various digital modifications of the voice in popular music over the past half a century (e.g. Auner 2000; Weheliye 2002; Ragona 2013; Bell 2015; Young 2015; Prior 2018). Some popular music

studies sources also propose more comprehensive methods for the analysis of singing voices, accounting for the link between vocal timbre and the physiology of the vocal tract (Heidemann 2016), as well as ways to identify the use of music production technologies on popular music voices (Malawey 2020). These literatures perform a more subsidiary role, however, because my analysis attempts to define the culturally specific aesthetic of the Western commercial pop vocal sound and to fully account for the way such an aesthetic is produced. This has required a more holistic perspective on vocal editing practices and their cultural positioning, rather than a focus on either vocal sound alone or any individual technology involved in the process of vocal editing. Indeed, other than a book by Victoria Malawey (2020: 134, 140) which only addresses the topic briefly, I am yet to encounter any musicological literature which fully discusses the role of EQ and compression in creating popular music voices.

Lastly, I also draw on practice-based literature including both textbooks designed as pedagogical materials for students of music production and audio engineering extending from the late 1990s to late 2010s (Gibson 1997; Case 2007; Hodgson 2010; Oswinski 2013; Moylan 2014; Izhaki 2017; Huber & Runstein 2018) and publications from, or associated with, the *Journal on the Art of Record Production*, a recently-formed academic journal which promotes the interdisciplinary study of record production and whose contributors are predominantly music production practitioners (Zagorski-Thomas 2007, 2010; 2014; Taylor 2017, 2018; Marrington 2017; Bourbon 2020; Harding 2020; Thompson & Harding 2019; Bennett 2011; McIntyre 2008; Savage 2011; Hepworth-Sawyer et. al. 2019; Frith & Zagorski-Thomas 2012). I also consult sources from scientific journals, mostly relating to audio engineering and acoustics, when discussing EQ and Compression (e.g. Deruty & Tardieu 2014; Välimäki & Reiss 2016; Moore 2012, 2020; Moore et. al. 2016, 2017; Hermes 2020). The textbooks, in particular, were useful for providing general background about the kind of music production techniques my informants were using, as they have a tendency to claim that there are universally good techniques or approaches to music production and mixing, and can often be quite dogmatic about the way certain things should be done. These literatures were seldom specific to the commercial genre I focus on due to their attempts to universalize, so they mainly provide a deeper

ethnographic background to commercial pop vocal editing practices, although there were points in some textbooks where techniques specific to commercial genres were mentioned (e.g. Senior 2019: 8.1.3; Gibson 1997 throughout).

Chapter summary

The structure of the thesis is as follows:

Chapter 1 explains the conceptual frame of fetishism and how it emerged from my fieldwork. Theoretically, I draw on the Marxist notion of commodity fetishism and Freudian erotic fetishism, both of which have similar conceptual foundations, and have been re-interpreted in contemporary contexts by anthropology and film studies respectively. Then, I explain how voice fetishism and technology fetishism were present in my fieldwork, each of which contains two different types of fetishism which parallel my theoretical delineation of the term. In the commercial pop context, voice fetishism is firstly present in a quasi-erotic sense, linked to ideas about the voice's representation of inner human identity and physicality, yet it is also increasingly common for lead vocal lines to be fetishized as commodities. Technology fetishism, slightly differently, consists of two forms of commodity fetishism only, embodied in two kinds of plug-ins used across vocal editing. I call these 'analogue emulations' and 'digital novelties', and the commodity fetishism of each has differing characteristics. These kinds of fetishism of voice and technology—each of which has two main sub-fetishisms—mutually intensify one another in a variety of ways: a situation which plug-in companies use to their advantage.

Chapter 2 concentrates on the aesthetic discourse my informants engage during vocal editing. I show how a particular selection of metaphors, underpinned by erotic voice fetishism specifically, coalesce into an over-arching metaphor I call 'vocal foregrounding', which drives vocal editing. The use of terms such as "up-front", "forward", and "present", illustrates how the aesthetic aims of vocal editing are to generate an exaggerated sense of proximity, and corresponding intimacy, for the listener, who hears the voice within the space of a stereo "mix". When explaining these metaphorical terms, producer-songwriters also allude to the auditory experience of voices in real-world situations of

proximity, a conflation of the aesthetic with the real which I suggest is borne out of erotic voice fetishism. I also address how my informants' use of the more ambiguous terms "tone" and "dynamics" engage voice fetishism more generally, encompassing features of the voice which they believe are special or unpredictable.

Chapter 3 focuses on tuning and comping, along with several related techniques used in the initial stages of vocal editing which are intended to refine the voice and prepare it for processing with EQ and compression. The chapter primarily argues that digital visualizations of the voice during this stage (especially pitch-contour and waveform visualizations) objectify the voice in a way which further impels the fetishism of the voice as a commodity, divorced from the labour performed by a singing body. I also show how producer-songwriters' dialogues with vocalists during this stage of vocal editing encourage a kind of possessive fixation on the objectified voice as a way of preserving their power over it—a gendered phenomenon which I interpret through the lens of erotic voice fetishism. Additionally, these two kinds of voice fetishism interact with the commodity fetishism of tuning technologies which are valued both for the capacities they provide to manipulate visual forms of the voice and their elusive sonic signatures. Broadly, the visual objectification of the voice in tuning and comping also motivates producer-songwriters to work with the voice visually rather than aurally.

Chapter 4 addresses equalization (EQ), expanding the theme of realism introduced in Chapter 2 by showing how my informants conceptualize the technique as a way of preserving and enhancing the voice's most "real" qualities, even though it aims to exaggerate high-frequency energy in the voice to contribute to vocal foregrounding. I argue that this discourse of realism in commercial pop vocal EQ is motivated by a close interaction between the erotic kind of voice fetishism and the commodity fetishism of 'para-graphic' EQ plug-ins, which together make producer-songwriters believe that EQ is an especially powerful, realist technique for manipulating the voice. The discourse of vocal realism in EQ also fuels its description using surgical metaphors, which both reflect and magnify the gendered power dynamics of vocal editing. Additionally, Chapter 4 shows how plug-in companies capitalize on this realist approach to EQ by developing "intelligent" EQ plug-ins (*Pro Q-3*,

Soothe and *Gullfoss*) which claim to interpret vocal frequency spectra in ways informed by the science of auditory perception.

Chapter 5 shows how, in contrast to EQ, commercial pop vocal compression is powered by the commodity fetishism of digital emulations of revered analogue compressors (the *1176* and *LA-2A*). This commodity fetishism is underpinned by ideas about the imagined sound lent to vocals by plug-in emulations of these compressors, as well as the mystification of vocal compression as an inherently complex and elusive technique in Western music production discourse more broadly. In addition, the interaction between this commodity fetishism and the fetishism of the voice—particularly relating to its "dynamic" qualities—prompts the use of compulsive, iterative compression processes. I also show how the plug-in company Waves Audio capitalizes on the commodity fetishism of these vocal compressors through its marketing partnership with celebrity engineer Chris Lord-Alge, further demonstrating how the masculine tone of analogue commodity fetishism is carried over into the digital domain.

Altogether, following the establishment of the core theoretical framework of the thesis and the contextualization of vocal editing's cultural and discursive characteristics in chapters 1 and 2, the latter three chapters zone in on the integral digital processes of vocal editing, aiming specifically to portray a sense of how such processes each contribute to the commercial pop vocal sound (responding to research question 1), and additionally to interrogate how their deployment by producer-songwriters is shot through with culturally distinct interpretations of the voice and technology (responding to research question 2), which I present through the conceptual lens of fetishism. Each chapter also paints a subsidiary picture of the plug-in industry's role in driving such processes.

Chapter One: Contextualizing Voice and Technology Fetishism in Vocal Editing

When I first met Chris, we talked about one of his favourite vocalists (2020F).[20] One of the reasons he enjoyed working with this vocalist so much was because he thought her voice was especially unique. Much of his vocal editing processes were therefore oriented towards preserving this uniqueness—in his words, to "stay true" to her voice. He devoted a lot of time and effort to recording the voice, and edited these vocals subtly and sparingly, to "get it exactly right" and cultivate a "raw feeling". During further discussion, however, Chris also explained how he felt obliged to use certain vocal editing processes to situate this voice within the commercial pop style specifically, in a way which contradicted his desire to "stay true" to the voice's characterful origins. This included "heavy compression" using multiple digital compression plug-ins in iteration, to achieve a "pop-py, crisp, clean sound". "It's hard for me to get away from that style of vocal editing, because I like her sound being really compressed", he justified. Even though Chris prioritized retaining the unique character of this voice, he felt that certain more extreme editing processes were necessary and aesthetically desirable.

Similar situations abounded throughout my fieldwork, and they highlighted two seemingly antithetical trends: producer-songwriters valued individual voices for their unique sonic identities, which they tried to preserve as far as possible; and at the same time these voices were being transformed by cumulative chains of digital plug-ins, which were constantly being extended and re-ordered. A central challenge throughout this project, therefore, was finding a way to account for the simultaneous presence of these two contradictory ideas within the culture. Early in the research I considered proposing that the former was being subsumed by the latter, and that voices were becoming dehumanized or decoupled from singing bodies in producer-songwriters' work. However, the sound of this music suggested that there was something deeper and more complicated going on, as despite the potential disconnect between the labour of singing and vocal editing in production,

[20] The audio example from Chris discussed in subsequent chapters (Audio 2) does not include the vocalist in question.

commercial pop tracks seemed to be totally oriented around the singing voice and its sound. In addition, a sense of respect for the unique skills cultivated by vocalists/artists was always very high among my producer-songwriter informants, and, more broadly, the continued successful reception in the West of commercial artists with voices perceived in informal press discourse as particularly distinctive or radical (e.g. Billie Eilish, Lana del Rey)[21] suggests that the idea that singing voices sometimes just 'have' that 'special something' which makes them unique persists in spite of the various editing techniques used to make them conform to a specific commercial style.

Given this context, and accounting for existing literature, I felt that the simultaneous presence of both ideas more accurately represented an intermingling of two prevalent ideologies existing in Western popular music culture, which I call the fetishism of the voice, and the fetishism of technology. The significance of these two kinds of fetishism in commercial pop vocal editing is that they mutually intensify one another: expressed in the simplest form, the heightened attention that producer-songwriters pay to the sound of the voice means that an increasingly large number of increasingly expensive plug-ins are used to edit it, further fuelling the level of detail at which vocal editing operates. More specifically, producer-songwriters feel that they are cultivating the special sonic qualities of voices—their uniqueness—even though, paradoxically, the strength of this aim encourages them to implement an increasing number of digital refinements to their sound, refinements which modify this uniqueness in a variety of ways. In parallel, the growing subtleties of this process encourage producer-songwriters to curate and expand their libraries of digital vocal editing technologies. This feeds back onto the focus on vocal uniqueness, since vocal editing involves a greater number of technologies and takes a longer amount of time than the editing of other musical elements, bolstering producer-songwriters' assumptions that there is something about the voice which makes it special and important. This situation directly benefits transatlantic plug-in companies, who develop and market plug-ins that target specific aspects of both voice and technology fetishism.

[21] See Robinson (2017) for an example of this kind of press rhetoric surrounding the uniqueness of certain contemporary commercial pop vocalists.

This chapter explains my use of fetishism as the overarching theoretical frame for this thesis, both defining the characteristics of these two mutually influencing fetishisms and introducing my fieldwork more fully by showing how these ideas gradually emerged from observation of my informants' vocal editing practices. To begin, I will outline the conceptual roots of the term: I start from early anthropology's use of the word to describe the religious power thought to inhere in material objects; then I move through Karl Marx's theorization of commodity fetishism and its subsequent development by contemporary anthropology; and finally I address the erotic definitions of the term, beginning with the psychoanalytic theory of Sigmund Freud and deepened in the context of contemporary feminist film studies. Then, the chapter is split into two central parts according to the two dimensions of the concept I propose: first voice fetishism, then technology fetishism, both of which have their own two components. In the case of the voice, firstly, an erotic kind of voice fetishism underpins the aesthetics of commercial Western popular music genres more generally, linked to ideas about the voice's representation of a whole human identity or body, and the way such ideas support the aim to centralize and foreground the voice within the spatial environment of popular music mixes to generate an impression of physical intimacy. These ideas are reproduced in commercial pop vocal editing, however it is also increasingly common for lead vocal lines (or "vocals") to be fetishized as if they were commodities, and to be conceptually separated from the labour of vocal performance—yet despite this, the former kind of voice fetishism persists in the culture. I interpret technology fetishism, in contrast, through the lens of two different kinds of commodity fetishism only, which relate to two types of vocal editing plug-ins. 'Analogue emulations' are prized for the important yet elusive sonic qualities they lend to the voice, and 'digital novelties' are valued for the imagined sophisticated technical control over the voice that they allow. Figure 6 represents this theoretical framework in tabular form for clarificatory purposes. Throughout, existing literature is put in dialogue with discussions of my field data to show how voice and technology fetishism, and their respective component kinds of fetishism, are rooted within the culture of vocal editing.

		Types of fetishism in vocal editing	
		Voice fetishism	**Technology fetishism (plug-ins)**
Types of fetishism in theory	**Commodity fetishism (Marx)**	• "Vocals" as commodities, hiding labour of performance (vocalist identity/body)	• 'Analogue emulations' • 'Digital novelties'
	Erotic fetishism (Freud)	• Voice represents fundamental qualities of whole human identity/body • Supports vocal foregrounding in pop mixes, to create sense of intimacy	

Figure 6: Tabular summary of the intersections between theoretical and ethnographic uses of the concept of fetishism in the thesis.

1.1. Defining fetishism

The term 'fetishism' abounds across the contemporary humanities and social sciences, although its precise meaning fluctuates widely depending on discipline. This thesis employs two related definitions of the term fetishism which derive from Marxism on the one hand and psychoanalytic theory on the other, although these definitions directly inform one another and sometimes overlap. Even though the term is popularly understood pejoratively as an abnormal erotic attachment to non-erotic objects, the varied history of the term across the past three centuries reveals how it is much more conceptually rich and complex than this narrow understanding. Indeed, critical theorist Emily Apter (1991: 3) has argued that the concept is theoretically productive precisely because of this historical fluctuation in its usage, and intellectual historian William Pietz has suggested that the contemporary humanities and social sciences have benefitted from trying to 'interrogate the real power of fetishes' (2003: 313) rather than dismissing them as socially constructed abnormalities. In Pietz's words, 'can something important be learned by identifying the strange power exerted by

sacred ritual objects, by seductively desirable commodities, by objects of perverse sexual fascination, and by deeply moving artworks?' (ibid.). Here, I use the concept to interrogate the strange power of both the voice and technology in commercial pop production. Before expanding on the two interpretations of fetishism which inform this aim, I will briefly address the term's history, which has been most fully outlined by Pietz (1985, 1987, 1988, 2003).

Etymologically, the English word 'fetish' is derived from the Portuguese *feitiço* which, in the 18th century, was a patronizing term used by European colonizers to refer to material objects which were deemed to be of religious significance to West African natives (Pool 1990; Pietz 1985). Gradually, the term was incorporated into very early anthropological research conducted with such communities. In early precursors to the anthropology of religion, for example, the term referred to material objects of adoration or worship, in the context of theorizing the reasons behind their significance (e.g. De Brosses 1760; cited in Pool 1990 & Pietz 1985). Paralleling this early anthropological usage, enlightenment philosophers such as Kant criticized those who were prone to fetishism as it represented, to them, an unenlightened incapacity for 'disinterested judgement' in favour of privileging material objects and subjective desires (Pietz 2003: 308). In late 20th-century anthropology, the term was still being used in certain studies to connote material objects in specifically West African traditions that local peoples felt contained some supernatural or magical power (see Gell 1998, cited in Born 2005: 16-18; MacGaffey 1977, 1994), although there was some acknowledgement that the colonial, demonizing origins of the term may be problematic (see MacGaffey 1998: 217). Anthropologist Wyatt MacGaffey, for example, provided accounts of the cultural and religious significance of 'nail fetishes' or *minkisi*—sculpted vessels or figurines that contain substances thought to be spiritually powerful (e.g. soil, relics) and the sacred power of which may be activated by the driving-in of nails (see Wilson 2020)—to the BaKongo religion of Western Zaire, describing a single 'fetish' as a 'container' (1990: 49) which holds 'a spiritual entity, a personality from the land of the dead, present in a material body, by its own choice or otherwise, but not restricted to that body' (1994: 125).

In the 19th century, Karl Marx used the concept as a way of criticizing capitalist economic systems. Specifically, in the opening chapters of *Das Capital* (1867), Marx appropriated the more holistic definition of the term from early ethnological studies of religion (Pietz 1985: 11) to develop his theory of 'commodity fetishism'. This is a phenomenon which Marx theorized was both an important by-product and a fundamental driving force of capitalist economies. Marx defines the term 'commodity' very broadly as 'an object outside us, a thing that by its properties satisfies human wants of some sort or another' (1867: 1),[22] and claims that capitalist systems (in contrast to pre-capitalist or socialist systems) are propelled by the accrual of wealth in the form of 'an immense accumulation of commodities' (1867: 1). Consequently, under capitalism, commodities are primarily valued according to how much they can be exchanged for (what Marx calls 'exchange-value'), which is an abstract quantitative measure, more so than their inherent utility (what Marx calls 'use-value'). The phrase 'commodity fetishism', coined by Marx, is a metaphor which attempts to describe how, for consumers in capitalist economies, the valuing of commodities according to their exchange value above all else means that they acquire a magical quality because they are severed from the conditions of their own production. Evoking the use of the term fetishism in the earlier 19th-century studies of non-western religion, Marx described how this magical quality exists because the value of the commodity is understood as being part of the form of the commodity itself rather than the product of real labour. Most importantly, this means that the experience of a commodity becomes a kind of imaginary substitute for social relations, as the form of the commodity conceals the social relations of its production, and such relations therefore come to be represented as if they were relationships between commodities rather than relationships between people:

> The mysterious character of the commodity-form consists therefore simply in the fact that the commodity reflects the social characteristics of men's own labour as objective characteristics of the products of labour themselves, as the socio-natural properties of these things. (Marx 1867: 165).

Put another way, in consuming commodities, people imagine or fantasise that they are somehow accessing or experiencing the social and human relationships for which the commodity is a

[22] Marx's definition is different from the understanding of the term in contemporary Western business as a resource which appears to have a relatively uniform monetary value regardless of the source of its production (e.g. see H.T. 2017).

substitute—human relationships that are in fact inaccessible or missing. In this way, the fetishized commodity becomes a source of solace for the alienation that Marx believed was widespread under capitalism. Marx uses numerous examples of material commodities to illuminate this process and the different forms it takes. Coins, for instance, are very obvious globally fetishized commodities due to the way their incorporation into systems of currency results in their widespread conceptual equation with specific exchange value as opposed to the labour processes which fashion them from metals. To be clear, Marx's theories also imply that commodity fetishism is a key factor underpinning commodification, which refers to the more general process whereby an object's exchange value gradually supersedes its use value in capitalist societies (see Hearn 2017). Marx's claim that commodities are fetishized is, more broadly, an attempt to criticize what he saw as capitalism's expanding commodification of everything in Western society, including immaterial concepts: 'the things which until then had been communicated, but never exchanged, given, but never sold, acquired, but never bought—virtue, love, conscience—all at last enter into commerce.' (Marx 1847, cited in Hearn 2017: 44). Following Marx, early 20th-century radical leftist theorists associated with the Frankfurt School—most notably Theodor Adorno—argued that capitalism's encroachment on cultural products turned certain types of art and music into especially fetishized commodities, deceiving the masses and harming the potential for revolutionary collective consciousness (Pietz 2003: 312; see also Adorno 1938).

Marx's theorization of commodity fetishism and its adoption by the Frankfurt school have been expanded and nuanced by contemporary anthropology, and I draw my first interpretation of the term 'fetishism' (i.e. commodity fetishism) from this work. In particular, various strands of contemporary anthropology have developed the idea that the more abstract concepts Marx implies (e.g. 'virtue, love, conscience') are persistently commodified under capitalist economic systems and, by extension, are 'fetishized' in the Marxist sense.[23] Linguistic anthropologists, for example, have interrogated how forms of language and behaviour become commodified under certain situations,

[23] To be clear, contemporary anthropology still retains a strong focus on material commodities (see Appadurai 1986; Tsing 2015) although anthropological explorations of more abstract commodities are more relevant to this dissertation.

such as the appropriation of specific linguistic registers as markers of local authenticity in advertising (Del Percio & Duchene 2012) or the utilization of regulated speech and behaviour in service-based roles (Kockelman 2006; Hochschild 1983; see also Shankar & Cavanaugh 2012). Paul Kockelman, for instance, explains how villagers from Chicacnab in Guatemala were trained by Western volunteers to perform formulaic and standardized speech customs in English and Spanish (e.g. identification, greeting, parting) and behaviours (e.g. pace length, punctuality) in their roles as local guides for Western eco-tourists, which Kockelman interprets as a commodification of the labour of basic speech and behavioural acts for consumption by visitors. In this context, Kockelman interprets commodity fetishism as the 'unjustified ascription of agency' (ibid.: 90) to these immaterial commodities: specifically, the villagers' conduct appeared to the eco-tourists to be a powerful testament of the positive effects of eco-tourism on local life, obscuring the laboured reality of such speech and behaviour (ibid.: 78). Importantly, as Shalini Shankar & Jillian R. Cavanaugh (2012: 361) suggest, the voice also has a key role to play in the commodification of language and behaviour, because it is often assumed to signify certain invisible characteristics of human beings such as identity (including race, gender etc.) and physiology, which are frequently conflated. Similarly, medical anthropology has long discussed how bodies, body parts and bodily functions are commodified across global medicine. For instance, Lesley Sharp (2000) provides an overview of the commodification of the body and its parts—including, for example, the market for cadavers for dissection—implying that the fetishism of body parts is a key part of this process (ibid.: 292). Further, Taussig (1980) used the concept of 'reification'—an expanded interpretation of Marxist commodity fetishism by Georg Lukàcs (1923)—to describe how North American medicine in the late 1970s fetishized the signs and symptoms of disease as purely biological and physiological processes, unhelpfully obscuring the social and institutional relations of medical treatment and disempowering patients, who were incapable of understanding the nuances of their conditions and how they were managed. Broadly, these approaches inform my analysis of vocal editing, which entails both the commodification of the voice (a form of communication produced by a person) in certain cases, as well as the commodification of software programs (plug-ins) which exist immaterially in the digital realm and which objectify and manipulate the voice in various ways.

The second way I interpret the term fetishism relates to the use of the concept in psychoanalytic theory, which builds on the legacy of Sigmund Freud's use of the term. Notions of sexual fetishism as an erotic fixation on an impersonal object, body part or quality had appeared in psychological literature as early as the 1880s, again building on the early ethnological usage of the term (Pietz 2003: 310). This definition informs popular understandings of the term fetishism and persists in contemporary psychological medicine. In the case of the latter, fetishisms of this kind are defined as '[the] experience [of] sexual arousal from objects or a specific part of the body which is not typically regarded as erotic' (APA 2013), and are considered abnormal, even pathological disorders if they cause distress to the fetishist (ibid.). Freud's use of the concept is broader, however: within Freud's assumed universal model of the Oedipus complex (see Freud 1910), he suggests that the presence of fetishism among his male patients was a manifestation of the repression or 'disavowing' of the sudden realization of physiological sexual difference between his father and mother ('castration anxiety'), with the fetish object becoming a substitute for the maternal phallus: an idea that the boy knows is impossible yet which he somehow also clings onto in the form of a fetish (Freud 1927). Freud also claimed that common objects of fetishism—including fur, feet and underwear—all seem to relate partly to the imagined location or shape of a maternal phallus in the context of an overall body (ibid.: 3).

Although the clinical applicability and universality of Freud's ideas about erotic fetishism have been significantly criticized, particularly in feminist literature (e.g. Irigaray 1974; Kristeva 1980; cited in Pietz 2003: 315), it is this more abstract concept of fetishism repressing or covering-over a reality, as well as the notion of a fragmented body part or object standing in for an overall body or person, which inform my definition of the term along these lines. This definition also implies that erotic fetishism maintains a desired illusion in the face of a reality which provokes anxiety, as a strategy for controlling or mitigating this anxiety. Indeed, the concept was used by late 20th-century feminist film studies scholar Laura Mulvey (1975) to interpret the way female characters were represented in mid-20th-century narrative cinema. Of particular relevance here, Mulvey develops the concept of 'fetishistic scopophilia' to refer to a particular, masculine mode of relation to female

figures in film. She conceives this fetishism in Freudian terms, as a response to or compensation for the castration anxiety inherent in the female figure, which 'builds up the physical beauty of the object [the woman], transforming it into something satisfying in itself […] hence the cult of the female star' (1975: 14, clarification added). Notably, her subsequent analysis of films by Josef von Sternberg suggests that the concept applies not only to the female figure in its entirety, but also to fragments of the female body which are isolated through camera techniques (e.g. the face). In her words: 'The beauty of the woman as object and the screen space coalesce: she is no longer the bearer of guilt but a perfect product, whose body stylised and fragmented by close-ups, is the content of the film and the direct recipient of the spectator's look' (ibid.).

These conceptual foundations of erotic fetishism are significant in the context of a thesis about the recorded voice, and often recorded female voices specifically. This is firstly because recorded voices in popular music tracks are decontextualized from the bodies which produce them, and yet, as I will explain in the next part of this chapter, a whole physical body and identity is always imagined when such voices are listened to—just as this kind of fetishism entails the fragmentation of specific body parts or objects which relate to an imagined whole person. More generally, both Freud's and Mulvey's interpretations of the concept focus on the masculine fetishism of women or features of women, and erotic fetishism therefore already has gendered undertones. As I will illuminate in the next section, and the subsequent chapters, commercial pop vocal editing aims to exaggeratedly foreground the (usually female) voice within an imagined space, and this is intended to generate a sense of intense intimacy between a listener and a vocalist, by creating an illusion of spatial proximity to the vocalist's body. Although not a literal manifestation of erotic fetishism, this recalls the same themes, because of the way a specific part of a person (a voice) is fragmented from the whole person and yet seems to stand in for the whole body and identity of that person (the vocalist) in relationship to a listener.

Indeed, and following on from my earlier suggestion that software objectifies the voice, the term 'objectification'—which Shankar & Cavanaugh (2012: 365) define as 'a process by which nonobjects are given object-like qualities, involving the externalization and materialization of

meaning and value'—is a key feature of both commodity fetishism and erotic fetishism. Lukàcs (1923) also uses a similar term in his development of Marx's ideas, where he describes how commodity fetishism's covering-over of social relations by the commodity form gives the commodity a 'phantom objectivity'—although he later uses the more specific term 'reification' as a way of describing this process outlined by Marx:

> The essence of commodity-structure has often been pointed out. Its basis is that a relation between people takes on the character of a thing and thus acquires a 'phantom objectivity', an autonomy that seems so strictly rational and all-embracing as to conceal every trace of its fundamental nature: the relation between people. (Lukàcs 1923: n.p.)

More broadly, in the medical context Sharp discusses how a more general sense of objectification is intrinsic to the commodification of the body and its parts, involving the 'transformation of persons and their bodies into objects of economic desire' (2000: 293), as well as the conceptual fragmentation of certain body parts from the body as a whole. Objectification is also a key feminist term, broadly defined as 'the seeing and/or treating of a person, usually a woman, as an object' (see Papadaki 2019). Feminist literature has also drawn attention to the ways objectification often involves the reduction of a person to solely their body or body parts (ibid., citing Langton 2008: 228-229), which are themselves objectified when fragmented from a broader whole (e.g. breasts). Not insignificantly, acts of reducing or fragmenting the female body are also a core theme addressed by social studies of cosmetic surgery and gender (see Basalmo 1992; Dull & West 1991: 62-64). With these interpretations in mind, objectification is a term I use predominantly in Chapter 3 of this thesis to explain the ways the voice is represented visually in tuning and comping and their related processes, and the way such visual objectification directly impacts the kinds of editing my informants do as well as the way they view the (female) vocalist's labour throughout the process. The notion of objectification is also extended implicitly throughout Chapter 4, which addresses the gendered implications of producer-songwriters' use of surgical metaphors in vocal EQ.

Across these varying interpretations of the term fetishism, a common theme is that the form of the fetishized object conceals or stands in for something else, or a larger whole, which is in some way inaccessible, indefinable or hidden. In anthropological studies of West African 'fetishes', this

external thing was the supernatural or transcendental entity which formed the focus of indigenous sacred activities, and which was therefore considered incapable of itself taking a material form: the term literally described a fetish's magical powers to make such spiritual beings material. Marx's commodity fetishism plays on these connotations of the term, partially using it as a metaphor for explaining the conditions of capitalism but also implying that the centrality of commodities in capitalist economies shares characteristics with this kind of religious fetishism. This is because commodity forms come to embody or stand in for social relations which are hidden; an act of concealment which ultimately functions to maintain the power of capitalist economic systems. Commodity fetishism can therefore be interpreted both as a kind of false consciousness or denial experienced by individuals living in capitalist societies and an explanation for why commodities (like technology, and the voice) can matter so much to people. Contemporary anthropology accounts for this process in the case of more abstract commodity forms such as language, behaviour, and the body. The Freudian theory of erotic fetishism implies that fetishisms both repress and embody fundamental experiences from early life, yet more importantly involve a fixation on fragmented body parts, objects or behaviours which stand in for an imagined whole person—a fixation which functions to deny or control an anxiety that the fetishist has about the power and agency of this person. This position is explored more deeply in late 20th-century film studies relating to cinematic representations of women.

These intersecting definitions and metaphorical uses of the term fetishism inform my use of the term to understand the central and mutually influencing fetishisms of both the voice and technology in the culture of commercial pop vocal editing. The following sections of this chapter illuminate the precise dynamics of these two kinds of fetishism, as well as providing broader context for the existence of each in Western popular music culture. In the case of the voice, both commodity fetishism and erotic fetishism come to the fore, and in the case of technology, I analyse two differing forms of commodity fetishism, where these theories are further developed in relation to gender and power.

1.2. Contextualizing voice fetishism in vocal editing

Several key ideas about the voice comprise the voice fetishism unique to commercial pop vocal editing, evoking aspects of both erotic fetishism and commodity fetishism. For example, in relation to the former, Chris's discussions at the start of this chapter showed him praising a voice for its uniqueness or character that should be preserved in vocal editing. On other occasions during my fieldwork, the talent for the right kind of singing for commercial pop—voices which not only fit the required style but also were easy to record and edit—was conceived as an immutable characteristic that certain singers just "had" and others did not. Contrastingly, in other cases, the vocalist's body and identity did not enter discussions at all during vocal editing. This was specifically the case when editing a voice that had been sung by a topliner for the purposes of pitching a demo, as if the eventual sale of this vocal line meant its original singing body was unimportant. More specifically related to the sound of the genre, producer-songwriters also consistently explained how their vocal editing processes aimed towards carving out a central and foregrounded position for the voice within each track they were working on, to create a sense of the voice being up-close and personal with the listener: an idea which I interrogate more deeply in Chapter 2. To expand and define these themes and the ways they relate to the concepts of fetishism I have outlined, I will first turn briefly to continental philosophy and literary studies, as there are several relevant authors whose theories about the human voice resonate with some of these ideas.

Contextualizing erotic voice fetishism in Western popular music culture

Numerous scholars in philosophy and literary studies have explored the idea that the human voice is special because, apart from the association with linguistic communication, each human voice has a unique sonic and affective signature which differentiates it from other voices. These approaches are preoccupied with defining the essence of this elusive quality. The main example of such an approach—which, not coincidentally, is cited frequently across music studies—is Roland Barthes' essay 'The Grain of the Voice' (1972). Barthes' central thesis is that the essence of the human voice, when approached sonically and not including linguistic specification, is the 'grain'. Barthes provides

anecdotal evidence from his own experiences of listening to two classical singers singing the same song (Charles Panzerá and Dietrich Fischer-Dieskau) to claim that certain singers possess a grain when others do not, meaning that the sounds of the latter are easier to describe in words. The quality of the 'grain' is broader than just vocal timbre: there is a physical aspect to it as well, as he claims it is inextricable from the body of any performer, and therefore can be found in instrumental as well as vocal music (although its presence in the latter is clearly more pronounced and important to Barthes). To quote, 'the 'grain' is the body in the voice as it sings, the hand as it writes, the limb as it performs' (ibid.: 188). The 'grain' therefore refers to this non-linguistic quality of the voice rooted in physiology that resists precise definition, yet that is simultaneously the defining feature of any voice. More recent philosophical takes on the voice are provided by Adriana Cavarero (2005), Mladen Dolar (2006) and Steven Connor (2000, 2014), all of whom extend Barthes' legacy by imbuing the voice with a similar kind of enigmatic character, suggesting that it eludes meaning in a way which makes it important and special. Dolar's book *A Voice and Nothing More* (2006), for example, defines the voice as indefinable, arguing that its essence lies in extra-linguistic elements: "[the voice] is what does not contribute to making sense. It is the material element recalcitrant to meaning […]." (15). Steven Connor's work focuses on the idea of the voice having non-referential traces which can be separated from specific linguistic meaning, represented in both ventriloquism (Connor 2000) and non-linguistic vocalizations such as sobs, hums, and stutters (Connor 2014).

Given the definitions of fetishism I have just provided, I would contend that such philosophies share characteristics with erotic fetishism, the object of such fetishism being the human voice. They present the voice as an entity which contains a transcendent, ineffable quality that cannot be fully accessed or explained—a quality which is reminiscent of vocal timbre but somehow also supersedes it. They also maintain that this quality has the ultimate power to define the essence of any individual voice, and therefore to distinguish it from others. More pertinently, they conceive the voice as a special fragment of a larger body—a fragment which is elevated in importance due to their assumptions that it has powerful capacities to characterize the most profound qualities of this body.

Furthermore, they generalize across all human voices, assuming that these ideas are universally applicable and cross-cultural.

The latter observation leads me to adopt a critical stance towards the way such approaches evoke erotic fetishism, which informs many of the points I make in this thesis about the way producer-songwriters conceptualize the voice. This stance follows recent work by Nina Eidsheim, who has interrogated the universalizing tone of these widespread ideas about the voice being somehow ineffable or exceeding meaning, especially the way assumptions about vocal timbre underlie such ideas. In Eidsheim's 3-part argument, she proposes that: 1. the assumption that vocal timbre somehow defines the true essence of voices, superseding linguistic communication, is a pervasive ideology about the voice across the Western world; 2. this assumption frequently entails a corresponding assumption that the voice provides an elusive point of access to some form of interior human essence, such as true identity or essential physical uniqueness; and, most importantly, 3. the connection made between 1. and 2. readily results in essentializing or stereotypical interpretations of vocal timbre that see it as inextricable from particular social identities (e.g. race and gender), and that link these social identities to essentialized physiological attributes (Eidsheim 2008, 2014, 2018, 2019).

As an example of the enduring significance of such assumptions in Western popular culture Eidsheim (2018, 2019: Ch. 3) discusses the mid-twentieth century career of singer Jimmy Scott, whose 'ambiguous' vocal timbre impacted his success due to marketing challenges resulting from his voice's straddling of multiple binary social categories ('white and black, male and female, young and mature' (Eidsheim 2018: 141)). In an attempt to problematize these conventions, Eidsheim (2014) has shown how vocal timbres are culturally shaped and learned rather than innate to individual singers: in the case of classical vocal pedagogy, for example, she illustrates how a huge amount of pedagogical labour goes into the cultivation of a particular kind of stylized vocal timbre over time, and yet classical vocal pedagogues still frequently assume that singers from certain racial backgrounds are endowed with certain kinds of timbral abilities or limitations based on their physiology (ibid.). Vocal timbre can also be imitated: Eidsheim describes how 7-year old Angelina Jordan gained fame across

Norway in the mid-2010s for her ability to convincingly imitate the voice of Billie Holiday (2019: chapter 5). Holiday's vocal timbre has been frequently described by fans as emerging naturally from her difficult life (ibid.: 155), and yet the capacity for imitation proves that such assumptions about explicit links between vocal timbre and identity or physiology are flawed. Although Eidsheim does not use the term 'fetishism' in her arguments, I believe that a notion of voice fetishism, predominantly comparable to erotic fetishism, is one way of conceptualizing the root of the problem she outlines: in these scenarios, the timbre of the voice stands in for a whole person or body which is assumed to have certain physical and identity characteristics, even though such characteristics cannot always be reliably linked to vocal timbre. The sound of the voice, fragmented from the body which has produced it, therefore obscures the complexities of its creation, and listeners instead believe that there is a stable relationship between a sound of a voice and a particular type of person from which it emanates. Further, this naturalized act of inflating the connotative power of the voice can be interpreted as a strategy for creating a sense of stability in the face of the ontological ambiguity that is experienced on hearing a voice in isolation from its body.

I mention Eidsheim's theories because my fieldwork resonated with her statements regarding assumptions about the innateness of vocal timbre in complex and subtle ways. Very generally, ideas about the power and innateness of vocal timbre seemed to underlie beliefs about the preservation or enhancement of vocal realism which guided commercial pop vocal editing practices: these ideas are detailed more fully in Chapters 2 and 4. In addition, my informants' tendencies to romanticize the uniqueness of voices could obscure their assumptions that specific vocal qualities were gendered. This issue comes to the fore especially in Chapter 4, which discusses the gendered assumptions about vocal "harshness" in EQing. More generally, the tendency for my informants to consider individual voices as naturally gifted to vocalists rather than the product of skilled training additionally resonated with Eidsheim's ideas. Moreover, Eidsheim's approach suggests that one potentially unethical consequence of this type of voice fetishism is the failure to acknowledge specificity and difference in voices, which may result in an additional failure to recognize discrimination on vocal grounds (see Baugh 2003).

I would also claim that additional literature in popular music studies illustrates how similar ways of relating to the voice have underpinned Western popular music reception throughout the 20th century. Several popular music scholars have—in now classic studies—discussed the central role played by the singing voice in the reception of recorded popular music in the West, which relies on a mediated relationship between an artist or 'star' persona and an audience (Frith 1996; Middleton 1990; Dibben 2006, 2009). Others have suggested that the voice plays a key role in defining the perception of 'authenticity' (Dibben 2009) which many have argued is fundamental to the way popular music is received by Western audiences (Moore 2009; Grossberg 1992). The voice's role in the perception of authenticity is also highly gendered (Mayhew 1999; Dibben 1999; Blackburn 2020). Anton Blackburn (2020), for example, analyses the reception of artist Lana Del Rey's early 21st-century vocal career to suggest that the assumed inextricable links between vocal characteristics and singers' 'real' personalities or identities are upheld more strongly in the case of female vocalists.

Some work in popular music studies, I would argue, reinforces this erotic kind of voice fetishism through its hermeneutic interpretation of the aesthetics of popular music. Richard Middleton, for example, links the salience of the voice in Western popular music directly to Barthes, drawing attention to its 'profundity' and 'sensuousness' in a very similar way to Barthes' descriptions of the 'grain':

> Popular music is overwhelmingly a 'voice music' […] The pleasure of singing, of hearing singers, is central to it. […] What needs noting here is not only the melodic (musical) side of singing but also the *vocal* side: the *voice*, in its commonly understood significance as the profoundest mark of the human. An unsounding human body is a rupture in the sensuousness of existence. Undoubtedly this is because vocalizing is the most intimate, flexible and complex mode of articulation in the body... (1990: 261-262)

Middleton's notion of the 'vocal side' of singing, in particular, directly follows Barthes' fetishism of the voice's 'grain', as something existing intrinsically yet mysteriously inside the voice. His statement also invokes notions of pleasure, sensuousness and intimacy which resonate with the concept of erotic fetishism, and singles out the voice as an especially versatile and rich site of human action and communication ('mode of articulation') relative to the rest of the body. Further, and paralleling Eidsheim's argument, Jonathan Dunsby (2009) has critiqued how Barthes' 'grain' has been

interpreted fairly literally across musicological scholarship, as referring to features of the voice which are not directly related to speech or language, e.g. timbre, or some notion of the body or physical identity of a singer residing in their voice. This literalism has led some popular music scholars to theorize that certain technological mediations or manipulations of the voice in popular music sound unsettling because they disrupt the 'grain' (e.g. Ragona 2013, Prior 2017). Middleton continues to this effect, additionally suggesting that the continued importance of the voice in popular music counteracts this disruption:

> Significantly, technological distortion of voice-sounds (through use of a vocoder, for example) is far more disturbing than similar treatment of instrumental playing (which is regarded usually as a logical extension of manual performance). [..] in the context of a widespread 'dehumanization' apparent in bourgeois art, the centrality of the voice in popular music - attesting to the presence of a human body - is a mark of a certain 'humanizing' project' (1990: 262).

Again, such interpretations of the significance of the voice evoke erotic fetishism because they assume that something like Barthes' 'grain' exists inside or underneath the voice which, although impossible to pin down exactly, is somehow unsettled or 'distorted' through technological mediation. Middleton's statement also concentrates on extreme, clearly audible 'distortions' of the voice, failing to acknowledge that the technological mediation of the voice across popular music is often much more subtle. How technologically 'distorted' does the voice have to be before its 'grain' is truly disrupted? Generally, I would also argue that such accounts of the voice in popular music studies (others include Malawey 2020 and Turino 1999), although engaging with general notions of listeners' reception of popular music very deeply, do not fully acknowledge the fact that these ideas about the voice are not universal, and are quite specific to Western popular music cultures. The end of this chapter contextualizes this specificity more deeply with reference to ethnomusicological literature covering alternative treatments of the voice in non-Western cultures.

Turning more directly to recording and post-production, previous research has outlined how popular music production professionals in specifically commercial genres use production techniques for foregrounding and centralizing lead voices in stereo mixes. Such an aesthetic aim also shares qualities with erotic fetishism. To explain: several popular music studies scholars have outlined the

spatial dimensions of this centralization, arguing that stereo localization techniques in popular music mixing intentionally figure the voice as a dominant component of popular music tracks (Dockwray & Moore 2008), however some go further to suggest that this intentional aesthetic centralization of the voice specifically aims to generate a sense for the listener of intense intimacy and closeness with the vocalist, especially in the case of headphone listening (Dibben 2013; Tagg 2019). (Although such approaches focus mainly on stereo localization techniques such as panning—whereas this thesis addresses how the centralization of the voice involves several processes which manipulate the mix space in different ways (e.g. EQ)—these arguments are nonetheless relevant). Philip Tagg, for example, terms this trope the 'monocentric mix', suggesting that it evokes notions of confidentiality:

> [The monocentric mix] not only implies that the singer is, as usual, the central 'reference point' of the piece but also that she has had her mouth placed nearer the listener's ear, not only by proximity to the recording mike but also by the relative volume accorded to the main vocal channel(s) in the final mix. This technique creates an actual or imagined distance between two persons (the singer and the listener) which is that of an intimate and confidential monologue (or potential dialogue). (Tagg 2019: 48-49).

Nicola Dibben (2013) builds on Tagg's ideas to argue that the exploitation of space is intended to create an illusion of physical proximity, and corresponding intimacy, between listener and vocalist, which promotes the listener's perception of the authenticity of a vocalist's identity as well as helping to construct a 'clear author-image' (ibid.: 110) for the song—both of which are very important for the way Western commercial popular music is marketed within a celebrity system. More precisely, she argues that such production norms both result from and reinforce a 'dominant reception ideology' of Western popular song recordings that they are 'expressive performances addressed to a listener' (ibid: 107). Eric Clarke (2013) has also discussed how, in certain popular music tracks, the manipulation of the voice in a perceived space can be interpreted as a manipulation of the listener's relationship to the singer's subjectivity (see also Dibben 2006). In one analysis of a track by Goldfrapp, Clarke shows how recording techniques exaggerate imagined intimacy between the listener and the singer's subjectivity by overtly increasing the sense of the voice's physical proximity to the listener through close miking (Clarke 2013: 106). Obviously, there are also cases across Western popular music genres where the voice is consciously de-centralized using stereo localization techniques for aesthetic effect (e.g. see Dockwray & Moore 2008; Lacasse 2000: 139-140), however, as Serge Lacasse explains,

these tend to be experimental cases which have more often occurred in indie and alternative genres throughout the 20th-century, whereas the commercial style I am focusing on tends not to make such experimental moves. Instead, the voice is continuously and exceptionally foregrounded and centralized in the spatial mixes of commercial popular music tracks using vocal editing techniques.

This aesthetic custom can be interpreted as exhibiting qualities of voice fetishism, specifically in the erotic sense, because, as Dibben, Tagg and Clarke all suggest, the voice becomes a proxy for an imagined intimate relationship between a vocalist and a listener. Following the same mechanics as erotic fetishism, the voice is a fragmented part of an imagined whole person and body: it becomes a substitute for particular assumptions about that whole body, and correspondingly a source of heightened focus in listening. The exaggeration of the sense of spatial proximity which defines the production of voices in Western popular music magnifies this focus, ensuring that the voice not only connotes a body but that this body occupies a virtual position of extreme proximity to the listener's ears and head, with corresponding associations of intense physical closeness, intimate communication and shared psychological space. Moreover, Dibben's writing suggests that this trend is quite a subliminal, naturalized 'ideology' of popular music reception—something that listeners are not necessarily consciously aware of, yet that exerts considerable force in their interpretation of this music. What's more, the idea that there is this virtual intimate relationship being constructed between the vocalist and the listener also suggests that the allure of the voice is amplified due to the impossibility of such a relationship ever materializing.[24]

This aesthetic trend of foregrounding and centralizing lead voices can be linked to developments in technology throughout the 20th century. Simon Frith (1996), for example, has pinpointed the introduction of 'crooning' in the 1950s as a key marker within this history. Crooning signalled both a new vocal aesthetic within popular music as well as a new kind of performance technique linked to the affordances of microphone technology. A style of singing represented by

[24] I mean this in a highly metaphorical sense, however imagined psychological relationships with the mediated personae of celebrities among fans have in fact been discussed at length by media studies researchers Donald Horton & Richard Wohl (1956).

early-to-mid-twentieth century male singers including Frank Sinatra and Bing Crosby, crooning emerged in tandem with the microphone's increasing capacities for the amplification of soft and 'close' (ibid.: 187) sounds that could not usually be heard in public performance. Crooners' soft, relaxed singing style capitalized on the ways the microphone provided this ability to 'hear people in ways that normally implied intimacy' (ibid.), and they developed a specific way of singing into the microphone to enhance this effect. This involved the singer moving in and out of range of the mic according to the volume and intensity of their singing, as well as suppressing excessive consonant sounds and breaths (see also Lockheart 2003). Notably, owing to this amplification of the intimate qualities of singing, crooners acquired a certain erotic allure among their fans, and the BBC even banned crooning from its airwaves for a time due to the perceived impropriety of its seductive capabilities (Baade 2011). As Frith suggests, the erotic connotations of crooning were widespread because of the way voices were, and still are, assumed to link directly to the physical bodies of vocalists: 'we assign [recorded voices] bodies, we imagine their physical production' (1996: 196). Later chapters of this thesis also imply that the fundamental sonic aims of microphone technique related to the maintenance of a continuous volume, and the supressing of certain mouth sounds have filtered down into commercial pop vocal editing (e.g. I discuss "de-essing" in chapter 5, pp. 278-280). Simon Zagorski-Thomas has also linked this 'exaggeration of intimacy' to historical shifts in technologies for music listening from group scenarios to more independent listening using headphones (2007: 204). In a more general sense, the idea of a voice standing in for an imagined human body was fundamental to the early successes of sound reproduction technology, which was initially praised for its ability to preserve the voices of the dead alongside its functionality for music listening (Sterne, 2003: 292; Stanyek & Piekut, 2010: 16). Arguably, the development of microphone technology is one of the initial driving forces of erotic voice fetishism in 20th-century Western popular music, preceding vocal editing: as Frith suggests, the microphone has 'draw[n] attention to the place of the voice in music, to the arrangement of sounds behind and around it, as [it] allows the voice to dominate other instruments whatever else is going on.' (1996: 188)

Voice fetishism in commercial pop vocal editing: erotic voice fetishism vs commodity fetishism of vocals

One of the most striking features of the culture of commercial pop vocal editing is the way the erotic type of voice fetishism I have up until now been describing seemed so persistent in the face of an increasingly fragmented labour process that encouraged producer-songwriters to treat the voice as a commodity. Put another way, in commercial pop vocal editing there were two kinds of voice fetishism existing simultaneously, paralleling my theoretical delineation of the term: one that was structured more like the erotic kind (which reflected the embeddedness of this kind of fetishism of the voice in Western popular music culture), and another which was more like commodity fetishism, and that seemed to reflect newer developments in the structure of the production and songwriting industries.

Simple examples of the former could be found in producer-songwriters' discussions of some of the voices they edited which they felt were particularly distinctive, like in Chris' description at the beginning of this chapter. Throughout my observation with Chris, his discussions of this voice were inextricable from the identity of the vocalist in question, and he spoke about this voice as if it were inexplicably aesthetically effective despite the vocal editing processes required to bind it to the commercial pop style. In another similar example, Michael—a very successful producer-songwriter who I spoke with briefly—described how one of his most popular commercial tracks had involved working with an artist who had what he thought was an incredibly powerful and expressive voice (2020E). Like Chris, his editing procedures aimed towards a "raw, exposed vocal", and yet Michael was also very up-front with me about the fact that the development of such a fetishized vocal sound still involved a huge amount of "processing and fine tuning" to achieve the sound that is "wanted, or expected" by listeners to the commercial style.

My conversations with one vocal editing specialist—who I will call Diana—provided another example of how the kind of voice fetishism which assumes deep links between the sound of the voice and the body or identity behind it was a motivating force of commercial pop projects, even though the increasing demarcation of vocal editing specialist and topliner roles entailed a detailed focus on the sonic form of the voice in isolation from all other parts of the process. To explain, Diana was

frequently working with high-stakes artists at the final stages of refining a track for release. The artist would usually remain in dialogue with Diana throughout the editing process, keeping a watchful eye over the edits she made. This contrasted completely with some of the other vocal editing specialists I interviewed, who were also working with high-stakes artists but usually had little dialogue with them, or with the producer-songwriters involved more deeply in the project. Diana felt very strongly that her job demanded a deep sensitivity to the skills, artistry and attitudes of vocalists, and correspondingly a very cautious approach to vocal editing, to avoid any friction which may result from Diana trying to edit their voice in a way they did not agree with (2019U). In Diana's case, therefore, the vocalist behind the vocal was always at the forefront of her mind during her vocal editing work. In the end, this prevented me from undertaking any further fieldwork with her because she felt it may disrupt relationships with her clients.

Notably, it became clear from my long conversations with Diana—additionally corroborated by discussions of the topic with other informants (2018B, 2018C, 2018G)—that these anxieties resulted from both the vocalist and the vocal editor's awareness that those listening to the edited voice would assume that such a sound was a direct representation of the vocalist's performance, and that any perceptible hallmarks of vocal editing would disrupt this assumption, encouraging the listener to view the vocalist as inauthentic and/or incapable in some way. Not only this, but Diana's vocalist clients clearly felt some desire to exert control over the products of her vocal editing work—their performances—even though the labour of vocal editing was an important and necessary contributing factor to the sound of their voice on a released track. There was a sense that the vocalists wished that the link between the edited voice and their performance could be true and direct, and that their control over Diana's process was therefore an attempt to obscure the role of vocal editing as far as possible in the resulting sound. Diana's strategy for the careful management of these situations validated and was very sensitive to these anxieties. Most importantly, the importance Diana placed on mitigating such anxieties in her vocal editing illustrated the sheer force of beliefs that link the sound of the voice to the identity and physicality of a singer—a key hallmark of erotic voice fetishism—in the culture of commercial pop vocal editing.

Nonetheless, moving to the second kind of fetishism of the voice I have identified, the way some vocal editing specialists spoke about their work suggested that they treated the voice like a commodity. Importantly, however, the commodity fetishism of the voice occurred only in two instances: 1. when the voice being edited was that of a topliner, and not an artist (i.e. when the vocal line was in fact a commodity to be sold); and 2. when vocal editing specialists were brought in on a pay-per-project basis to undertake vocal editing only, instead of being involved in the broader songwriting and production of a track. In one example of the former kind, Jason sent me a video of him editing the voice of one of his topliner collaborators for a demo they intended to pitch to a label, and he overlaid this video with real-time discussions of his process intended to teach this topliner how to edit her own vocals (2020N). At the time, given my own unconscious assumptions about the personal nature of the voice, I was struck by the extent to which Jason talked to his colleague about her voice in a completely depersonalized way, analysing its sound and pointing out its flaws directly to her. However, it was clear from my discussions with Jason that this kind of collaboration was totally normal in this instance, where the topliner was providing Jason with a large quantity of vocal recordings which infrequently remained tied to the tracks in question because they were often replaced with the voice of an artist who would eventually release the song.

In an example of the second kind of vocal commodification, one vocal editing specialist who Michael frequently hired on his projects—whom I will call Ryan—described how he undertook vocal editing jobs for a variety of producer-songwriters in his professional network, essentially entering the process of creating a track at the "cleaning-up" stage, after the voice had been both recorded and initial edits done (2020K). Ryan nearly always completed this work remotely, and the vocal in this preliminarily edited stage provided him with "raw material" which he then "went over with a fine-toothed comb" before sending it back to the client. In addition, my key informant Oliver described how he acquired similar kinds of on-spec vocal editing jobs through 'Vocal Kitchen' (vocalkitchen.com), a company which collects ready-made commercial pop vocals-only tracks and makes them available for purchase (2019BB). Advertised as 'vocals to-go', these vocal lines are bought by producers and DJs to incorporate into their own productions as ready-made, sometimes

royalty-free samples (Figure 7). In these instances, the voice—or more accurately the "vocal"—is an entity whose editing is geared completely towards its eventual sale, either as a released track or as one of Vocal Kitchen's ready-made samples, and any links to the identities or bodies of singers fade into the background.

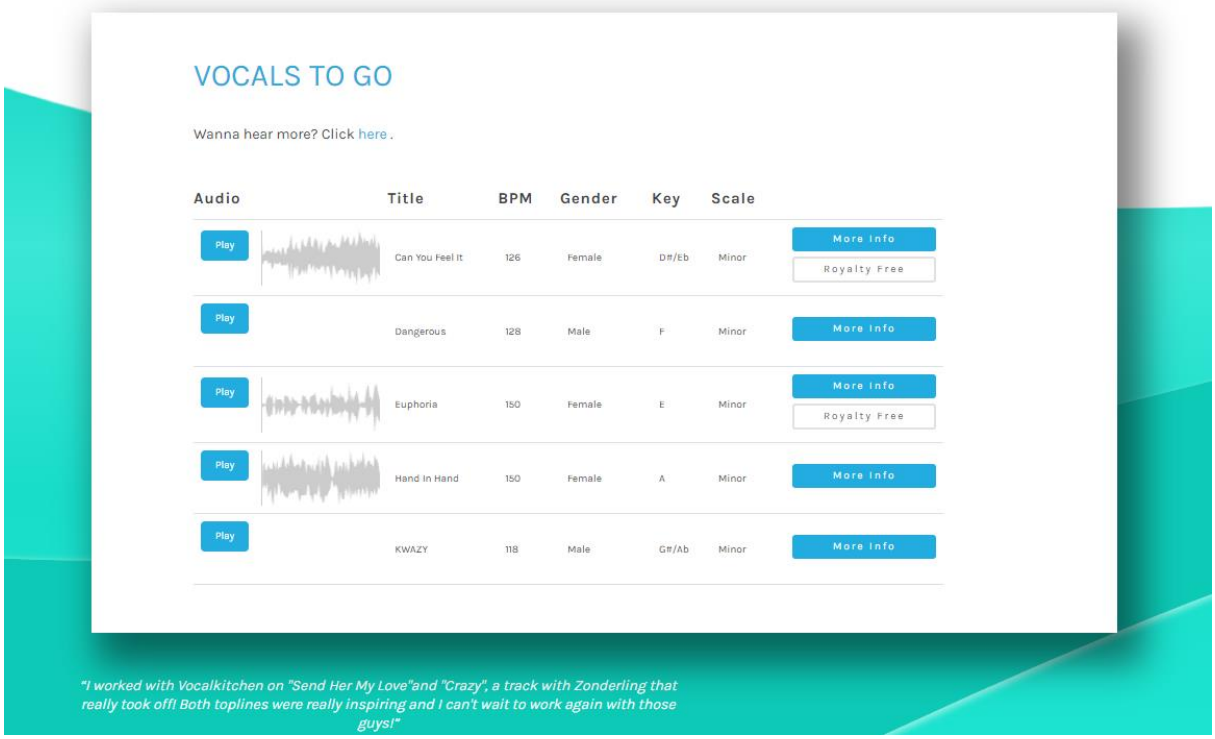


Figure 7: Screenshot from the homepage of www.vocalkitchen.com, accessed 17 Feb 2022.

In cases like this, the voice was prepared as a commodity for onward sale, and approached solely in relation to its sonic form. Here, vocal lines were created, bought and sold, occasionally in ways that were completely decontextualized from any track or identifiable vocalist—Chapter 3 delves more deeply into the ways this commodity fetishism of the voice impacts portions of the commercial pop vocal editing process. For now, this contrasted completely with Chris' and Michael's discussions, but the key difference here is that the vocalists they discussed were artists and not topliners, and they

were fully involved in the production and songwriting of their respective projects. Put more succinctly, in commercial pop vocal editing there were two sides to voice fetishism, yet both were in force throughout the various stages of vocal editing in the commercial pop labour process: on the one hand, voices were elevated as powerful means of indicating an imagined vocalist's whole identity and physicality, yet, on the other, they were also increasingly commodified in the Marxist sense, especially in the case of topline vocals and pay-per-project vocal editing specialists. In many ways, the fact that the former persists in relation to artists despite the increasing presence of the latter says a lot about the staying power of the broader voice fetishism characteristic of Western popular music I have outlined, and which recalls themes of erotic fetishism.

Finally, to address an imbalance in this section, my informants did show how trends of foregrounding and centralizing the voice were intensified in the case of commercial pop, closely linked to the ideas about intimacy which I have just explained. However, I will save the fuller exploration of this theme to the next chapter, as it is so foundational to the culture that it deserves a longer exposition. Moving on, these examples of voice fetishism in commercial pop vocal editing, which blend both notions of erotic fetishism and commodity fetishism, lead neatly into my explanation of technology fetishism in vocal editing which is focused more solely on the fetishism of plug-ins as commodities.

1.3. Contextualizing technology fetishism in vocal editing

Alongside voices, the second central topic of my fieldwork was the selection of plug-ins producer-songwriters were using to edit them. In particular, one of my key findings was that the same assortment of plug-ins were being used to edit the voice across all of my key informants working in commercial pop. These included, most centrally: either *AutoTune* or *Melodyne* for pitch-correction; FabFilter *Pro-Q* for EQing; and software emulations of the hardware compressors Teletronix *LA-2A* and Universal Audio *1176LN* Limiting Amplifier (hereafter referred to as *LA-2A* and *1176*, pronounced "eleven seventy-six") for compression (Figure 8). Some other relevant plug-ins which my informants used occasionally do enter my discussions at points, however the use of this set of plug-ins

for routine vocal editing was an absolutely defining feature of the practice, and many of my informants (including all three key informants) incorporated these plug-ins consistently in their vocal chains. Although these plug-ins could also be used to edit other musical elements, it was clear that they had acquired reputations among music production professionals for their efficacy as vocal editing plug-ins. This section considers how this homogeneity in plug-in usage across commercial pop vocal editing reflects a commodity fetishism of digital technology in two main forms: firstly, through the fetishism of plug-in emulations of analogue hardware, which producer-songwriters believe lend elusive sonic characteristics to the voice; and, secondly, through producer-songwriters' beliefs that the newest digital plug-ins will give them advanced technical capabilities to edit the voice in an intelligent, scientific way. Like voice fetishism in this context, then, technology fetishism also contains two contrasting subsidiary forms of fetishism, although in this case both are forms of commodity fetishism rather than erotic fetishism.

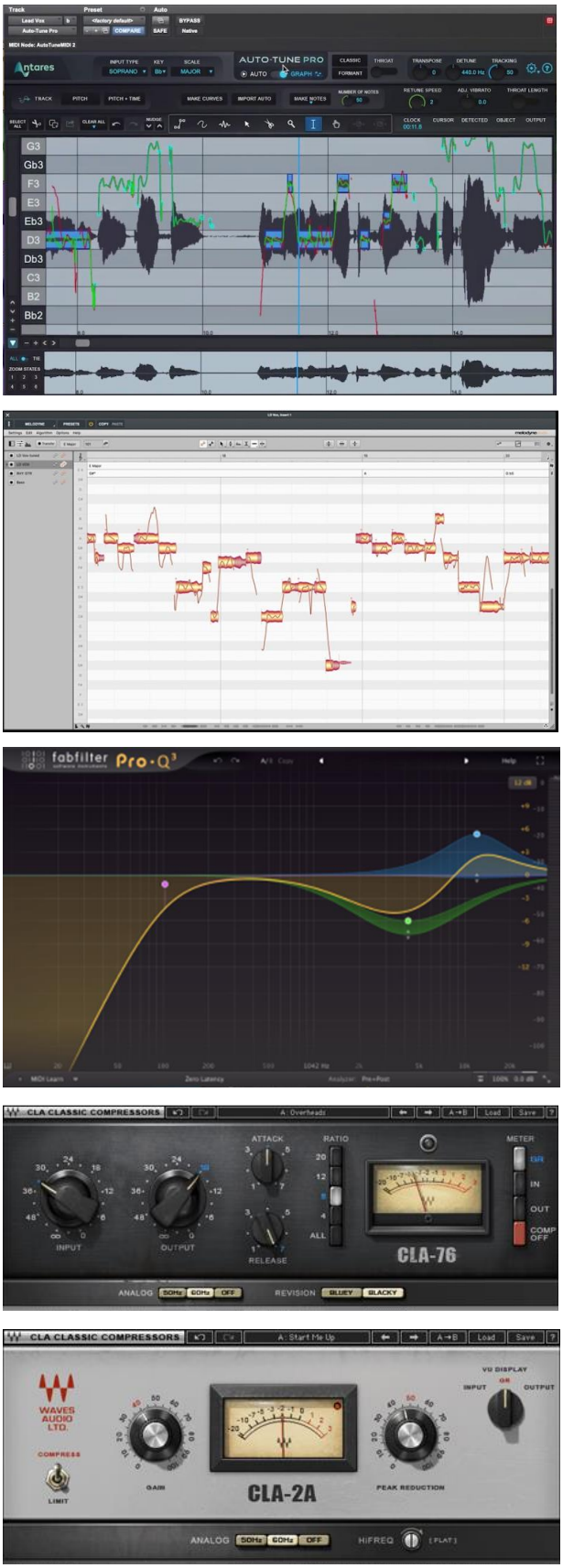


Figure 8: Introducing the central plug-ins of the thesis: *AutoTune*, *Melodyne*, FabFilter *Pro-Q 3*, *1176*, *LA-2A* (sources for the same images are given in later chapters: pp. 154, 155, 234, 258, 259).

Before delving into some of the theory behind my claims about the commodity fetishism of digital audio technologies, I will provide some background about the main music technology companies who produce these plug-ins, as well as saying a little about the way their products are distributed across the market for vocal editing technologies. Another detail to mention is the way these plug-ins are acquired by producer songwriters: they can either be purchased outright for permanent ownership (ranging anywhere from £10-30 for cheaper plug-ins with limited functionality to £100-200 for the newest state-of-the-art software) or rented by paying monthly subscription fees (through a third-party platform such as Splice or via packages offered directly by plug-in companies).

The two largest companies operating in this arena—producing also the widest variety of plug-ins—are Waves Audio Ltd. and Universal Audio Inc. Waves was founded in 1992 and began to build its reputation among music production professionals for high-quality, easy-to-use, cheaply available plug-ins starting with the *Q10* Equalizer (which Waves claim on their webpage was the first commercially-available audio plug-in in history, see Waves 2022a) and the *L1* Ultramaximizer. Today, Waves' product catalogue includes over 200 plug-ins which perform a variety of functions across the spectrum of music production. These include software emulations of the *1176* and *LA-2A* compressors which are part of a 'signature collection' sponsored by celebrity engineer Chris Lord-Alge. Universal Audio (UA) has been around for a lot longer in comparison: founded in 1958 by Bill Putnam Sr. (a legendary figure in Western music production and audio engineering, see Bushnell & Ferry 2011), the company was known throughout the latter twentieth-century for its production of audio engineering hardware. Following a series of mergers and acquisitions, the company was re-established in 1999 by Putnam's descendants and started to re-issue some of its earlier successful hardware models (see Robjohns 2001; Fuston 2020) as well as to break into the software market by emulating such analogue hardware in plug-in form (Universal Audio 2022; Sterne 2019a). In a broad sense, UA had a reputation among my informants for a more premium line of plug-ins in comparison to Waves, specifically including their highly regarded emulations of the *1176* and *LA-2A* compressors.

The other most relevant companies covered in this dissertation are Antares, who develop *AutoTune*, Celemony, who develop *Melodyne*, and FabFilter, who develop *Pro-Q*. These companies

are smaller, and their activity is focused on these individual flagship plug-ins which have been particularly successful. In contrast to Waves and UA, therefore, their success in the plug-in market results from their targeting of much more precise niches within the practice of music production by developing these plug-ins with very specific functionalities. Antares and Celemony, for example, have both been well-established since the early 2000s, and their product offerings still centre respectively around the tuning plug-ins *AutoTune* and *Melodyne*, which music production professionals predominantly associate with vocal editing.[25] FabFilter was established in 2002 yet their reputation has only solidified since the early 2010s, owing largely to the success of the *Pro-Q* EQ plug-in which was released in 2011, followed by *Pro-Q 2* in 2014 and most recently *Pro-Q* 3 in 2018 (Inglis 2014, 2017). I provide a fuller discussion of this plug-in in Chapter 4. A few other smaller companies develop even more specialized flagship plug-ins which are used in commercial pop vocal editing—including, among others, oeksound, SoundTheory, SoundRadix and sonible—however, discussion of these companies is also confined to Chapter 4 as the plug-ins in question are EQ-based.

One other marginally relevant company to mention is iZotope Inc., who have acquired their reputation primarily for their development of plug-ins for audio restoration (see Théberge 2013), and over the past decade have diversified their activities to include the development of plug-ins which condense a variety of audio processing techniques previously undertaken separately into a single plug-in. Their plug-in *Nectar*, for example, condenses the various stages of a commercial pop vocal chain into one, including tuning, EQ, compression and other effects, which occupy separate modules within the plug-in (Figure 9).[26] This plug-in was seldom used by any of my informants, however, as it was considered amateurish and aimed at "bedroom producers" (2019A) who were yet to learn music production fully and simply wanted to achieve passable pop vocals as fast as possible, so I do not dwell on this plug-in in any more detail. Nonetheless, iZotope's approach is notable for the way it rationalizes what are taken to be the most important components of commercial pop vocal editing and presents them all in one easy-to-use, streamlined interface for the user. Their plug-in *VocalSynth*

[25] I direct the reader to the PhD theses by Provenzano (2019a: Ch.2, Ch. 3) and Marshall (2017: Ch. 2) for fuller expositions of the histories of and philosophies behind these two companies.
[26] https://www.izotope.com/en/products/nectar/features.html

(Figure 10), which provides a comprehensive suite of vocoder-like effects, is also worth a brief mention, as it has been used much more widely than Nectar over the past 5 years in commercial pop production. However, my fieldwork similarly confirmed that this plug-in was only really used as an isolated sound effect for providing a bed of distorted vocal harmonies under a lead vocal[27] rather than being a core component of routine vocal editing work, and so I do not discuss it further. Nonetheless, some of iZotope's audio restoration plug-ins are mentioned briefly in Chapter 3 as my key informant Oliver used them quite frequently.



Figure 9: *Nectar 3* interface. This user is using 4 modules in the vocal chain (row at the top), with EQ currently open. (https://www.adsrsounds.com/product/software/izotope-nectar-3-vocal-processing-software/?gclid=Cj0KCQiAjc2QBhDgARIsAMc3SqSEWz9vQqD0p0UqaO-haYxlLYgqVxu9Lg_HiE4RFMHIhvX4uXV_K2oaAngbEALw_wcB)

[27] This plug-in, or a sound which resembles it very closely, can be heard on the recent tracks: 'i'm so tired…' by lauv (2019), 0'07"-0'26" (https://youtu.be/boUl0jPZMdg?t=7); 'The Middle' by Zedd ft. Maren Morris & Grey (2018), 0'29"-1'03" (https://youtu.be/M3mJkSqZbX4?t=36); and 'Delicate' by Taylor Swift (2017), 0'03"-0'23" (https://youtu.be/tCXGJQYZ9JA?t=3).



Figure 10: iZotope *VocalSynth 2* interface. The shape in the middle provides the user with an abstract visual representation of how the settings they adjust within the plug-in manipulate the sound of the voice. (https://www.izotope.com/en/shop/vocalsynth-2.html).

Two forms of plug-in commodity fetishism: analogue emulations and digital novelties

My informants' use and discussion of these plug-ins in vocal editing illustrated subtly contrasting kinds of commodity fetishism, mirroring two dominant styles of plug-in design across these companies. On the one hand, "analogue emulations" (including Waves' and UA's *1176* and *LA-2A* plug-ins), are plug-ins which attempt to recreate both the visual and sonic characteristics of well-known analogue hardware in software form. As Sterne (2020) describes, these plug-ins model the inner workings of analogue hardware digitally to try to impart the same kind of sonic effects, even though the exact mechanisms by which this is achieved computationally are seldom made public or obvious to consumers. They also utilize the common (although increasingly outmoded) trend in software design to create skeuomorphic interfaces—these retain the visual design cues of hardware in an attempt to make usage more instinctual (see Hayles 1999: 13-18; D'Errico 2016: 102-105). On the other hand, and increasingly more popular in contemporary music production, there are what I refer to

as ‘digital novelties’ (such as FabFilter *Pro-Q*, *AutoTune*, *Melodyne* and the iZotope plug-ins). These abandon skeuomorphism in favour of slick, minimal, interfaces, moving towards the newer trend of ‘flat design’: spearheaded in the early to mid-2000s by Apple (Turner, 2014), this design trend integrates a more broadly computer-oriented silicon-valley style aesthetic into the audio industry. Sonically, they also claim computational efficiency, technical intricacy and sometimes ‘intelligent’ functioning in their marketing, even though, like the analogue emulations, the inner algorithmic workings of such plug-ins are hidden from users. These two kinds of plug-in design are therefore as much visual as they are sonic, and both dimensions are key to the specific flavour of commodity fetishism in each case. Firstly, analogue emulations promise to create elusive sonic effects through their usage, yet their mechanical-looking interfaces obscure the exact ways such effects are created, meaning that the user comes to associate the effects with the forms of the plug-ins themselves, thus translating a similar kind of commodity fetishism that exists in relation to analogue equipment into the digital domain. Secondly, digital novelties promise the user an unparalleled degree of technical mastery over sound, which is apparently facilitated by the plug-ins’ sophisticated, science-based design. Like the analogue emulations, however, such rhetoric hides the exact ways these plug-ins function from users, instead encouraging them to associate the quick and easy attainment of very high sonic standards with the modern, silicon-valley-style look and feel of these plug-ins. In both cases, these plug-ins—as consumable commodities in the Marxist sense—present their values and efficacies as if they inhere within the visual and tactile forms of their digital interfaces, making the inner workings of such plug-ins invisible to producer-songwriters, who see them predominantly as means to ends rather than impactful agents in the culture of commercial pop vocal editing. To be clear, the stark contrast between the different kinds of commodity fetishism relating to analogue emulations and digital novelties comes to the fore in chapters 4 and 5 especially, and plays only a contextualizing role in chapters 2 and 3.

To expand on the former kind of commodity fetishism, several musicologists have highlighted the pervasiveness of this fetishism across both Western and non-Western audio production and reception cultures, particularly describing how it is defined by the notion of an

ineffable ‘sound’ that such technologies lend to audio through recording and post-production. Louise Meintjes (2003: 71-108), for example, has shown how South African studio practices in the 1990s were characterized by beliefs among both studio engineers and musicians that studio hardware enacted magical, indefinable effects on sound, compounded by dense technical vocabularies as well as metaphors of interiority, privacy and space travel. In a similar way, an article by Marc Perlman (2004) discusses how North American audiophilia discourse exhibits very strong beliefs that particular—often expensive—technologies have elusive, subtle, sonic effects that can only be detected by connoisseurs. Théberge’s discussions of the digital musical instrument industry in the 1980s and 1990s attest to the prevalence of such believed effects, which Théberge terms ‘imputed characteristics’, following political scientist William Leiss (1997: 139-140, citing Leiss 1976: 82). Music analyst Megan Lavengood has also interrogated the importance of the abstract notion of sonic ‘warmth’ to the fetishism of analogue equipment, specifically synthesizers, in comparison to their digital counterparts: this relies on the assumed physicality of the sound’s production, and the flaws and ‘microfluctuations’ in the sound which result from this, and are additionally affected by the sound’s environment (2017: 106-107; 2019). In the case of analogue equipment specifically, there is also a visual and tactile dimension to this fetishism, which is extended into the digital domain when such tactility is emulated in software interfaces. The visual complexity of such interfaces contributes to this sound-oriented commodity fetishism by hiding the inner workings of analogue technologies, therefore making their effects seem indefinable and magical. Importantly, though, what all these writers allude to is a specific kind of commodity fetishism which exists in relation to technologies for music recording and production in both analogue and digital domains: the notion of an elusive yet desirable ‘sound’ to which popular music production professionals strive, and which these technologies claim to provide, functions as the magical or indefinable ‘thing’ underpinning this commodity fetishism. The forms of the technologies themselves conceal the mechanisms of such sonic effects.

Théberge’s discussions of the advent of digital sequencers and synthesizers in the late 20th century further suggest that this fetishized notion of ‘sound’ both arises from and reinforces the

commodified status of these digital technologies within the rationalized music production industry that Théberge describes. Put another way, musicians place an increasing premium on the idea of 'sound' and the power to produce such a 'sound' which, crucially, can only be achieved using an increasing number of digital technologies such as synths, which are bought ('consumed') in growing quantities. In one of Théberge's examples, he outlines how an offshoot 'cottage' industry developed alongside the production and marketing of synthesizers, which surrounded the production and sale of the individual synth sounds themselves (what Théberge calls 'programs' but what are today usually called 'patches') (1997: ch. 4). In this scenario, the sound-oriented commodity fetishism intensified to the point where not only the technologies but the synth sounds themselves became commodities. Although this scenario is a direct precursor to the contemporary market for sample packs and Virtual Studio Technology (VST) presets (e.g. Native Instruments' 'Kontakt', see Terren 2021) rather than plug-ins, the sound-oriented commodity fetishism is still important. Théberge's work implies that the contemporary music technology industries have extended and naturalized the qualities of such commodity fetishism of analogue audio technologies into the digital domain, to their commercial advantage.

Jonathan Sterne (2019a, 2020) has recently suggested that sound-oriented commodity fetishism is fundamental to the production and consumption of analogue emulations in the contemporary plug-in industry. He explicitly uses the term 'commodity fetishism' to describe this phenomenon, and goes back to Marx to flesh out this term more fully in a contemporary context. In particular, he draws on Marx's term 'phantom objectivity' to describe the idea of 'sound' underpinning the commodity fetishism of audio technologies: this term describes the autonomy or 'thing-ness' that a commodity seems to have, or more accurately that quality of the commodity which obscures the social relationships which have produced it. Crucially, Sterne argues, the concept of timbre[28] is therefore key to the fetishism of digital music technologies, because it is the parameter of

[28] Although Sterne uses the term 'timbre' here, it is worth noting that the elusive sonic quality being sought by producer-songwriters in my fieldwork was not reducible to just 'timbre', and instead encompassed notions of vocal "tone" and "dynamics" as well—culturally specific terms which I will analyse in chapter 2 (from p. 126 onwards).

sound which most clearly distinguishes sounds from other sounds—it gives a sound its distinctness in a way which other parameters such as pitch and rhythm do not. Here I would extend Sterne's argument by suggesting that the fact that timbre is frequently mystified in Western musical discourse as the most indefinable or inaccessible parameter of sound (see Fink et. al. 2018; Dolan & Rehding 2020; Walden 2019) contributes to this fetishism. Such mystification intensifies in the context of the voice. In my fieldwork, this kind of fetishism of analogue emulations was most clearly evidenced in relation to compression plug-ins, and the subtleties of this fetishism are discussed more fully in Chapter 5.

The commodity fetishism of digital novelties is slightly different, with ideas about the imagined sound such plug-ins engender receding into the background. Journalist Michael Terren has drawn attention to the increasing replacement of interface skeuomorphism in the audio plug-in industry with this new 'uniform, technocratic style' (2021: n.p.), which Sterne additionally describes as 'flat, easy-to-render, Bauhaus-inspired design' (2020: 165). Terren uses iZotope's *Vinyl*—one of the company's earliest plug-ins released in 2001 which is intended to add 'Lo-fi' crackle and hiss effects to audio in a manner resembling vinyl records—as a case study illustrating this shift. An interface clearly resembling an analogue mixing desk complete with a chrome body and screws was updated in 2020 to remove most of this skeuomorphism in favour of a plainer black interface with minimal sliders (Figure 11). According to Adobe, 'flat design' and 'material design' are terms in common parlance among software designers which refer to newer styles of user interface (UI) design that move beyond skeuomorphism in this way: 'flat design' removes 3D elements completely, along with any stylistic features which reference the real world, favouring minimalism, simple typography and bold colours; and 'material design' is a softer approach which 'uses mostly flat elements but with 3D touches' (Babich, 2020: n.p.). iZotope's new version of *Vinyl* is totally flat, and the FabFilter *Pro-Q, AutoTune* and *Melodyne* interfaces are similar in character (see Figure 8). Sound engineer Andrew Bourbon (2019: 211-212) has similarly drawn attention to this shift, describing how recent innovations in plug-in design have proliferated away from the exact emulation of analogue hardware towards both the simplification of controls and the development of tools for 'non-traditional' audio

processing including tuning, spectral noise reduction and manipulation, and 'creative processing', considering iZotope a market leader in this regard. There is little discussion of the sonic dimensions of such a shift, suggesting that the sleek and minimal visual character of digital novelties is one of the main reasons for their appeal.



Figure 11: iZotope's *Vinyl* before and after 2020 (https://splice.com/plugins/1847-vinyl-vst-au-by-izotope, https://www.izotope.com/en/company/press/2020/vinyl-update.html).

Terren argues that such a shift represents a 'plug-in gentrification, an industry transitioning from homespun curiosity to humourless, platformist expansion', with his term 'technocratic' implying that these plug-ins are part of a broader societal agenda of technical and scientific control (2021: n.p.). These are powerful ideas, however there is a second, less loaded explanation for the increasing predominance of these plug-ins across the industry. As Théberge shows (1997: 58-69), the burgeoning phenomenon which economist André Piatier (1987/88) called 'transectorial innovation' had a huge role to play in the development of the digital musical instrument industries in the 1980s and 1990s—this concept refers to the intertwining of separate kinds of innovation across different capitalist industries.[29] At that time, the music technology industries were intertwined with the consumer electronics industries, especially in the context of digital audio production's increasing dependency on

[29] Théberge paraphrases Piatier thus: 'the phenomenon in which innovations developed to meet the needs of a specific industrial sector come to play an important role in the creation of new innovations and commodities in formerly unrelated industries' (1997: 58).

the availability of powerful digital signal processors (DSPs). The development of some of the most successful synths and samplers of the 70s and 80s by E-mu, Roland and Ensoniq, for example, was dependent on the development of microprocessors by electronics companies Intel, Motorola and Zilog (Théberge 1997: 59-60). Théberge also claims that transectorial innovation is complemented by transectorial marketing: in particular, he explains how during this time promotional strategies from the computer industries were imported directly into the marketing of new audio products, such as the audio company Ensoniq's borrowing of the early marketing slogans for the Apple Macintosh in the mid-1980s (1997: 65). Clearly, a similar thing is happening here: these plug-in companies are directly borrowing contemporary design principles from other technological industries which design software (which not coincidentally were also initiated by Apple), as well as using these to inform the tenor and aesthetics of their marketing, imbuing these plug-ins with a similar hi-tech, computerized aura. Indeed, the newer interface of *Vinyl* has much in common with the contemporary user interfaces of Apple and Android smartphones (when used in 'dark mode'), and comparing iZotope's brand image with Apple's (e.g. through their internet landing pages[30]) reveals striking similarities, including large sans serif fonts, white and chrome colour schemes, modular layouts and very simple product descriptions with little contextual information. Moreover, such design principles are also just as likely to be capitalizing on user familiarity as the older skeuomorphic designs: newer generations of producers and songwriters are increasingly less likely to be familiar with complex analogue audio hardware and much more likely to be fluent with the flexible, slick and modular interfaces of smartphones and modern-day Apple computers.

Regardless of the societal implications of such a shift, however, in my fieldwork both kinds of plug-in were used frequently, and this can be explained by understanding the different kind of technology fetishism which underpins digital novelties in contrast to analogue emulations. These plug-ins have an aura of scientific objectivity and aesthetic neutrality: their minimal, unfussy interfaces, alongside their marketing slogans, which frequently claim that the plug-ins process audio using algorithms which are based on scientific research, make users feel like such plug-ins endow

[30] https://www.izotope.com/ and https://www.apple.com/uk/, accessed 25 May 2022.

them with cutting-edge technical power to edit audio as sensitively as possible. iZotope advertises its recent edition of *Nectar 3 plus*, for example, as a 'vocal assistant' which 'analyz[es] the content of your vocal performance and generat[es] an effects preset form-fitted to the nuances of your singer and your music.' (iZotope 2022). When discussing why he used FabFilter *Pro-Q 3*, my informant Chris also expressed a lack of understanding about the plug-in's operation but a belief in its supreme effectiveness nonetheless, conflating its 'algorithms' and its visual appeal:

> It probably boils down to something boring like algorithms… it just sounds really good! [laughs]. It looks beautiful as well… (Chris, 2020F)

Certain plug-ins, including those by iZotope, also appeal to recent trends in the adoption of artificial intelligence (AI) and machine learning across the tech industries in their marketing of such plug-ins, claiming that their inner workings can 'understand' and process audio 'intelligently', even though the extent to which AI techniques are actually incorporated into the design of such plug-ins is never explained fully. For example, plug-in company sonible describe how their 'Smart EQ' 'intelligently' 'analyses spectral information' and how its 'algorithms' help to achieve a 'well-balanced sound' with the user only having to set a few simple preferences, therefore providing them with 'an ideal basis to take further creative steps' (sonible, 2022). Underpinning this commodity fetishism is the assumed cutting-edge power of these plug-ins and the imagined technical mastery that such software will give producer-songwriters over sound, promulgated by the companies' reinforcement of such plug-ins' algorithmic complexity combined with their sleek and simplified interfaces which are strongly connected to the aesthetics of 21st-century computer software more generally. Chapter 4 discusses this kind of commodity fetishism more fully in relation to FabFilter *Pro-Q 3* alongside other 'intelligent' EQs that were used by my informants.

Plug-in commodity fetishism and the online cultures of home music production

My fieldwork showed how the culture of online self-teaching and home production in commercial pop provided a fruitful landscape for plug-in companies to exploit these kinds of technology fetishism in the marketing of their products. To provide some deeper context: in the past 20 to 30 years, owing

to the increasing affordability and accessibility of both music hardware and software, it has become more and more common for music production and recording to happen in home studios, and even in producer-songwriters' bedrooms, rather than dedicated state-of-the-art studio spaces (Théberge 2012; Tomaz de Carvalho 2012). Paralleling this shift has been an increase in the number of young, aspiring music production professionals who begin their careers by acquiring their own tech and teaching themselves using materials available on the internet, rather than acquiring skills and knowledge through formal pedagogy or through on-the-job work. As I have explained, my key informants undertook these formal types of learning later in life at specialized universities after this informal self-teaching, as a way of formalizing and enhancing skills they had already acquired through the internet. (Notably, none of my informants had undertaken any internships or apprenticeships in studios, even though some of the older producers and engineers I spoke to described how this used to be a common path of entry into the profession). In this scenario of atomized, do-it-yourself learning, it is commonplace for individuals to freely share pedagogical information and advice with each other in various online spaces—including through the creation of YouTube tutorials, on user forums (e.g. gearspace.com), and on Facebook groups (e.g. Make Pop Music).

All my key informants referenced YouTube, in particular, as the pedagogical resource from which they had drawn the most information over the course of their careers. There are a huge number of music production tutorial channels on YouTube, many of which are run by a single producer who superimposes a video of themselves onto a screen recording of their DAW whilst they explain their music production techniques in a 'vlog'-like style (Figure 12). Notably, this was the format in which my key informant Oliver chose to create his screen videos of vocal editing to share with me, further illustrating the influence of YouTube tutorials on his practice (2020Q). Some examples of successful YouTube tutors for commercial pop include Austin Hull, Warren Huart, Devvon Terrell McLeod, Alex Rome and Andrew Huang, all of whom have over 100,000 subscribers.[31] The vast majority of these YouTube tutorials are available to watch for free, as tutors with the widest audiences will be

[31] Here are links to their respective YouTube channels: https://www.youtube.com/c/MakePopMusic, https://www.youtube.com/c/Producelikeapro, https://www.youtube.com/c/DevvonTerrell, https://www.youtube.com/c/AlexRome, https://www.youtube.com/c/andrewhuang.

able to generate a stream of income through the platform's pay-per-view function, however some tutors advertise further, more comprehensive, paid content which their viewers can access through external webpages (e.g. https://www.makepopmusic.com/). In this context, several of my informants claimed that it was common for younger professionals like themselves to be more forthcoming with sharing information about their work in comparison to older professionals, offering this as one explanation for why they wanted to participate in my research (2019X; 2020E; 2020K). This 'Peer 2 Peer' economy, in which a 'pay-it-forward' attitude towards learning the ropes of production seemed common, was praised by my informants for its seemingly democratic nature—for how it seemed to be allowing more people to follow and develop their interests in music production without formalized education or employment. Vocal editing specialist Ryan, for example, felt that 'music' was 'more open nowadays', and believed that earlier production professionals had a tendency to keep specialized information guarded, linked to their dependence on studio hardware. This contrasted with the contemporary open sharing of knowledge on forums in the digital age (which, he claimed, was likely to be even more open among experienced professionals such as himself because of their career stability) (2020K).

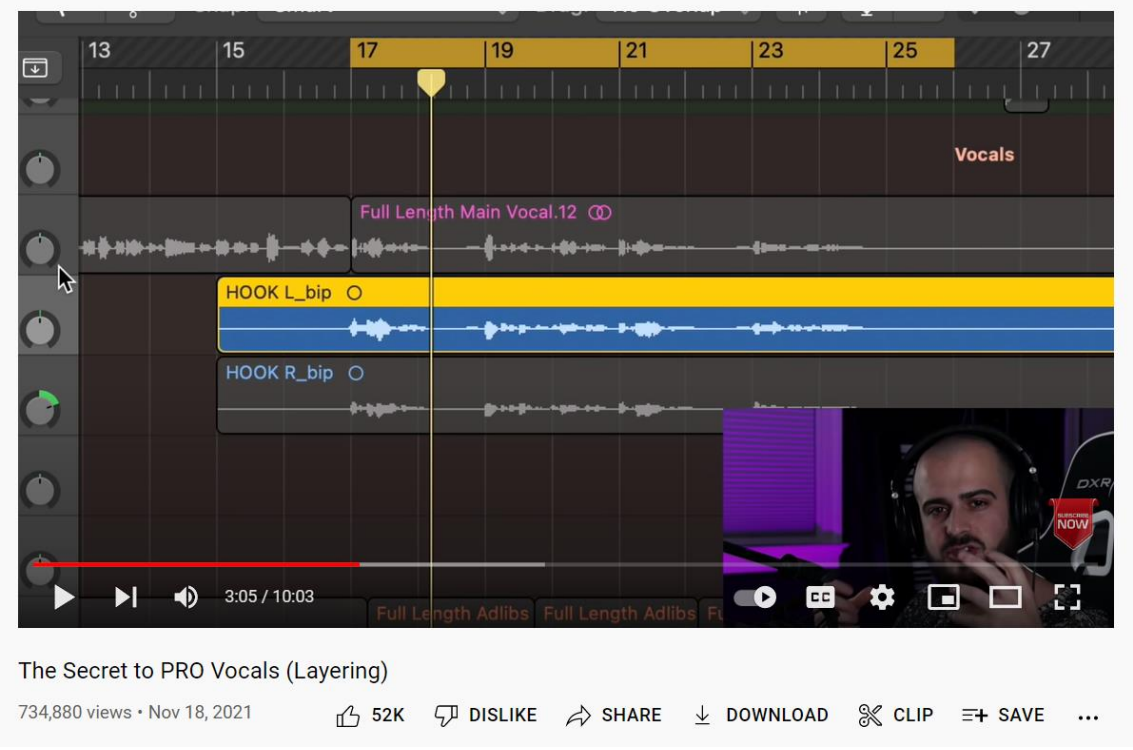


Figure 12: Screenshot from a tutorial video by Alex Rome (Alex Rome 2021: 3'05").

Despite this dominant attitude that technological developments have contributed to a more collaborative, participatory music production culture, I observed how the plug-in industry was able to use such an open landscape of information sharing to its advantage. As I began to immerse myself in such online scenes—through membership in Facebook groups, browsing audio forums, watching and subscribing to YouTube tutorial channels, and subscribing to plug-in company mailing lists—I began to receive a sudden influx of targeted advertisements across my social media accounts as well as YouTube, encouraging me to buy several of the newest plug-ins including, among others, Waves' *CLA Epic* (released in 2020) and Antares' *AutoTune unlimited* (released in 2021). Some screenshots from these adverts are presented below in Figures 13 and 14, to give the reader an idea of the kind of commodity fetishistic imagery and dialogue they use: the CLA advert emphasizes an abstract notion of the 'sound' offered by the plug-in as well as presenting visuals which resemble audio hardware yet which are clearly digital constructions; the AutoTune advert presents the technical operations of the plug-in as indispensable to a producer's creative process. It became clear to me that, through the data-sharing capacities which exist in the age of surveillance capitalism (Zuboff 2019), these companies were directly exploiting the proliferation of pedagogical informational exchange and collaboration online, in an attempt to naturalize the presence of their plug-ins in such a culture and to capitalize on their commodity fetishism. One conversation with my informant Chris, for example, drew attention to the efficacy of targeted ads in building a reputation for the latest plug-ins: for one of the newest plug-ins he had recently integrated into his vocal chain, he could not remember at all when or where he had heard about it first, only having a hunch that he may have seen it somewhere on Facebook (2020T).

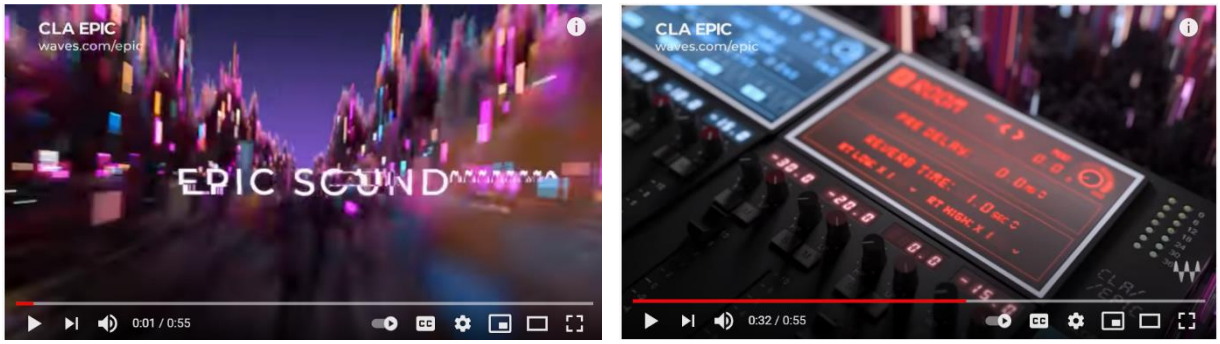


Figure 13: Two screenshots from the advertisement video for plug-in suite *CLA Epic*. In the second screenshot, the text is flickering and unreadable, intended to give an abstract impression of analogue hardware. (Waves Audio 2020b)



Figure 14: Screenshot from an advert for *AutoTune unlimited* (Auto-Tune / Antares Audio Technologies 2022)

Plug-in commodity fetishism and gender

A final dimension of the commodity fetishism of plug-ins arising from my fieldwork was its masculine tone. Broadly, this quality magnifies the questions of gender which run throughout this thesis. Several writers have discussed how Western studio production discourse fetishizes audio technologies in the context of the profession's gender imbalance. A recent PhD thesis by Alex Annets (2015), for example, claims that 'gear fetishism' (i.e. audio technology commodity fetishism) functions as a particular way of performing masculinity in audio production cultures, at the same time that it covertly maintains masculine hegemony in the profession despite widespread claims about the

'democratizing' potentials of new audio technologies. A similar issue permeates Théberge's work (1997): he describes how the tone of 1990s music magazines continually assumed a level of social and political neutrality in discussions of hardware and technical specifications even though readerships were usually more than ninety percent male. Drawing attention to the commodity fetishistic qualities of this covert masculine hegemony, he writes: 'with technology receiving top billing […], all social ills disappear into the glossy surface of the technical objects themselves' (ibid.: 123). Notably, the well-established audio technology discussion forum 'Gearspace' has only recently changed its name from the previous 'Gearslutz', in response to complaints from the community regarding the name's misogynistic undertones (Rogerson 2021). This name had been left unchanged since the forum's inception in 2002 however, and the fact that such a change only occurred nearly twenty years later is perhaps testament to the enduring masculine coding of gear fetishism across Western audio cultures, including commercial popular music production. In commercial pop vocal editing, this kind of masculine gear fetishism is further combined with the references to computer culture embodied in the design and marketing of digital novelty plug-ins: an external culture which is also very obviously coded masculine.

More specifically, in the UK context, the masculine tone of such gear fetishism intersects with the demographic realities of music production pedagogy, and the gender exclusion such demographics perpetuate. Georgina Born and Kyle Devine (2015), for example, showed how music technology degrees across the UK in the mid-2010s were overwhelmingly male dominated, in striking comparison to traditional music degrees, which were more balanced. Such demographic disparity also intersects with class, as music technology degrees were, on average, populated by people with a lower social class profile than those of traditional music degrees, whose social class profiles were much higher. The reasons Born & Devine provide for such disparity include: the embeddedness of music technology within a wider cultural association between masculinity and technology which precipitates the exclusion of women from such domains from an early age (ibid. 147-148); the way gendered prejudices enter into school classrooms through teacher interactions, instrument choices and the character of classroom spaces (ibid.: 148-150); and the situation of music technology within the

broader epistemological history of sound which various researchers have analysed in terms of a dialogue between 'hegemonic' masculinity and 'subordinate' femininity' (ibid.: 151). As implied by the contextualization I provided in my methodology section, my key informants' professional lives have all been formatively shaped by this setting. More generally, my analyses therefore suggest that the commodity fetishism of technology in vocal editing is interlinked with the masculine coding of music production professions and may even intensify it.

1.4. Voice and technology fetishism in other cultures

To indicate, however, that the mutually intensifying relationship between voice fetishism and technology fetishism which saturates Western commercial pop production is far from universal, I turn briefly to review studies of production in other settings and cultures. Firstly, regarding voice fetishism in studio production specifically: in Meintjes' study of the production of Zulu pop genres in South African studios, she briefly discusses how the front-line voices of the Mbaqanga genre, and especially their timbral qualities, seemed to figure very little in day-to-day studio discourse, despite the key contribution of these voices to the sound of the music. The reason for this, she explains, was that the important role of the voice in Mbaqanga had been naturalized and taken-for-granted, and singing was also conceived as women's work and therefore of minor importance in broader production matters (2003: 251-254). In contrast, as this thesis illustrates, discourse about the voice and its sound was ubiquitous in commercial pop production scenarios. Nonetheless, the gender dimension of Meintjes' discussion is similar in some ways, as the gendered nature of vocal labour (largely female) in comparison to the labour of vocal editing (largely male) was similarly naturalized and infrequently discussed during vocal editing in my fieldwork, even though the voice continually figured as an important topic of discourse in these settings. In a slightly different vein, the voice was not an object of discourse at all in Bates' Turkish production culture, clearly because Anatolian folk music is largely instrumental. However, the methods employed by engineers to exaggerate certain qualities of lead instrumental timbres, especially EQ (Bates 2010: 98-100), have some commonalities with commercial pop vocal editing. Broadly, these examples reinforce my claims that the centrality of the voice in commercial pop production is strongly linked to a deeply rooted ideology of voice fetishism

in the West which is more present than in the South African example, and parallels the way that solo instruments are treated in Bates' Istanbul studios.

Secondly, anthropological work which interrogates the role of the voice in musical cultures more broadly shows how aesthetic ideals about how the voice ought to sound, and the cultural ideas underpinning such ideals, differ from the Western commercial pop context. Nicholas Harkness (2013), for example, writes about how the performance of Western art vocal music in South Korea prizes an aesthetic ideal of "cleanliness" which is inextricable from the practice's intertwining with evangelical Christianity. Here, Harkness suggests, the sound of the voice is valued for the way it is construed as a direct link to religious knowledge and experience—with 'cleaner' voices more effective at establishing this link—whereas voice fetishism in commercial pop settings relies on more abstract assumptions about the links between the voice and the physical body or identity of the singer. Amanda Weidman (2021: chapter 5) has also problematized how the vocal aesthetics of 'husky' and 'raw' became increasingly desirable in South Indian female playback singing for cinema, in comparison to earlier aesthetic ideals of 'clearness' and 'sweetness'. The earlier aesthetic was strongly linked to the meek characterization of female roles in films, and the newer fashion only slightly changes the flavour of the gendering of these voices, veering towards an equally gendered sexualization of vocal performance. In dialogue with these examples, Western commercial pop vocal editing's aims to aesthetically foreground the voice within the "mix", and the foundations of this aesthetic ideal within a historical voice fetishism across Western recorded music, become clearer.

Finally, regarding the commodity fetishism of technology, Gavin Steingo (2018) has accounted for the ways a highly idiosyncratic culture of non-professional electronic music production in Soweto, South Africa, does not exhibit the same kind of commodity fetishism as Meintjes' commercial studio production in 1990s Johannesburg, or contemporary commercial music production in the West. Steingo describes an informal music-making practice facilitated by the in-person circulation of mp3s using digital storage devices such as memory sticks or hard drives, in combination with computers and laptops which are often old and prone to breakage. Hardware is often altered haphazardly to facilitate processes of musical exchange; and failures in the poor South African

electrical infrastructure frequently leave permanent audible artifacts in the mp3s (ibid.: 556-557). Linked to the poverty of Sowetan residents, Steingo characterizes this scenario as a 'a patchwork system of technical bricolage in which parts are absent, substituted, exchanged, scrambled, or mismatched' (ibid.: 558), a situation far messier, more flexible and less organized than Western commercial studio production contexts. In Steingo's words, the 'fetishism of the invisible inner workings of technologies' (ibid.), a key characteristic of the commodity fetishism of sound technologies which characterizes production cultures like Meintjes', is absent, because the bricolage-like culture of Soweto means that amateur musicians frequently take apart and recombine technological parts themselves, and additionally there are no 'closed and integrated technological systems' established by manufacturers (e.g. Apple) to reify this commodity fetishism. Here there is no mystification of the sound proffered by analogue technologies, nor the imagined technical control that sleek digital plug-ins will give music creators over the sounds they are working with. Correspondingly, there is less potential for such commodity fetishism to be exploited by corporate music technology companies or for such technology fetishism to intersect with fetishism of the voice.

Conclusion

This chapter provides important background for the mutually influencing voice and technology fetishisms which define commercial pop vocal editing. To summarize my main key points: firstly, my utilization of 'fetishism' as a conceptual frame for this thesis draws on both Marxist commodity fetishism and Freudian erotic fetishism. These differing, although complementary, theorizations of the term structured my explanations of voice and technology fetishism in vocal editing: the former blends features of both erotic and commodity fetishism, and the latter only encompasses commodity fetishism.

I claimed that voice fetishism in vocal editing is, on the one hand, defined by the assumed link between the sound of a voice and the identity and physiology of an imagined vocalist (a link which is often made through the concept of vocal timbre), so that the latter is experienced as being conjured up by the former—an idea which popular music studies shows has deeper historical roots in

Western popular music production and reception. This fetishism, closer to the erotic kind, is manifest further in the aesthetic trend of foregrounding and centralizing the voice in commercial popular genres, which production professionals believe creates an imagined intimacy between vocalist and listener through an implied physical proximity to the vocalist's whole body. Alongside this, the increasing commodification of "vocals" in the songwriting industry through toplining encourages a commodity fetishism of the voice in which the sound of such voices seems to be valued separately from the singing bodies which produce them (an idea taken up more deeply in Chapter 3). These two kinds of voice fetishism seem to contradict each other, yet the former persists despite the increasing presence of the latter. This can be partly explained by accounting for the different kinds of vocalist roles within the labour process of songwriting and production (artist vs topliner), but it is also perhaps further evidence for the sheer strength and embeddedness of the former way of thinking about voices across Western commercial popular music culture.

Next, I suggested that the commodity fetishism of plug-ins is another defining feature of commercial pop vocal editing. 'Analogue emulations' claim to lend an elusive 'sound' to the voice, and 'digital novelties' appear to give the user an extraordinary technical power to modify the voice according to objective and aesthetically neutral sonic principles, even though, in both cases, the inner workings of such technologies are seldom made available for users to understand. The plug-in industry reinforces these two kinds of commodity fetishism through their marketing strategies, which covertly exert their influence within the online communities where producer-songwriters share information with one another. In this way, the structure of commodity fetishism which Marx developed in relation to material commodities—a hiding of the social relations of labour behind the form of the commodity—is reproduced in the digital realm, covertly benefitting those who produce these software products. Moreover, the persistence of this split between two different kinds of plug-in used in vocal editing gives the outward appearance of great diversity and freedom of choice for producer-songwriters developing their vocal chains, however my analyses in this chapter and later suggest that such diversity in itself is a phenomenon which works to the advantage of the plug-in industry by creating an endless market for new plug-ins. Arguably, using the concept of commodity

fetishism also shows how producer-songwriters' beliefs in the efficacy of both kinds of plug-in are founded on fundamentally similar structures: their technological allure is embodied in the visual appearance and usability of their interfaces, drawing the user's attention away from the actual make-up of such plug-ins and instead encouraging them to take their efficacy and convenience for granted.

More broadly, the similar conceptual foundations of voice and technology fetishism mean that each amplifies the other. Threaded through the remainder of this thesis are examples from my fieldwork of such mutual amplification, although I will summarize its general procedure now. Primarily, since the voice has been historically fetishized in the West and in Western popular music specifically, a huge amount of labour is devoted to vocal editing in commercial pop genres, and my informants felt that the voice was the most difficult and time-consuming component of a track to edit. In particular, widely circulating ideas among producer-songwriters about the sound of the voice being particularly complicated and special meant that they were attracted by certain plug-in companies' claims that their most cutting-edge plug-ins either transformed the sound of the voice in a way that was sensitive to its integral features, or enhanced it with an elusive analogue charm. This, in turn, meant that producer-songwriters used a wider variety of plug-ins to edit the voice in comparison to other elements of a track. In addition, as later chapters illuminate in more detail, the different kinds of technology fetishism I have explained were made particularly extreme in the context of vocal editing, since the technologies make visible or modifiable specific sonic features of the voice (e.g. continuous pitch-contours, high-frequency energy, dynamic consistency) which, given the existing voice fetishism, were perceived as particularly important and elusive qualities in need of precise refinement with specific techniques.

The balance of differing kinds of plug-ins across vocal editing reflected this further and structures the following chapters: plug-ins used for vocal compression were mostly analogue emulations, EQ plug-ins mostly digital novelties, and tuning plug-ins also digital novelties (through they shared some characteristics with analogue emulation commodity fetishism). In the case of tuning and comping (and their related techniques), the increasingly complex and malleable visual representations of the voice offered by various plug-ins and DAW features encouraged an extremely

detailed focus on the voice's pitch and timbre contours, increasing the commodity-type voice fetishism by separating the sonic form of the voice from the singing vocalist's labour. In the case of EQ, the use of digital novelty EQ plug-ins combines with the glorification of vocal timbre to produce beliefs that the technique is a powerful and quasi-scientific way of preserving and enhancing the "real" voice. And finally, in the case of compression, the use of plug-in emulations of revered hardware compressors is fuelled by producer-songwriters' existing beliefs about compression technologies being mysterious and incomprehensible, as well as their assumptions about the unpredictability of the voice. In a more general sense, the plug-in industry's promulgation of both kinds of commodity fetishism of technology also feeds back onto the voice fetishism by encouraging my informants to spend more time and use even longer and more complex chains of plug-ins for editing the voice: this idea extends throughout the thesis, although it is developed most fully in Chapter 5. The next chapter delves into the aesthetic results of this synergistic fetishism more deeply, focusing on the voice fetishism present in the aesthetic discourse of vocal editing.

Chapter Two: Up-front and Central, Up-close and Personal: Voice Fetishism and the Aesthetic Discourse of Vocal Editing

> If your song has vocals, you generally want them to sit up-front in the mix. Right in the listener's face. But on their own, without the right processing, most vocals won't sound like that. They are simply too dynamic of an instrument. (Recording Revolution 2016)

> The first [characteristic of the "ever present pop vocal"] would be a sense of lyrical, flowing dynamics. You'll want to have a naturalistic vocal that's allowed to grow from loud to soft organically […] It is not uncommon, when trying to get that in-your-face, sparkly pop sound, to run a vocal through two compressors […] In terms of EQ, a top-end boost is often employed to give you that immediacy [and] closeness […] (Messite (iZotope contributor) 2018)

The above quotes encapsulate some of the evocative language commercial pop producer-songwriters use to describe the voice during vocal editing. Emerging from such language, their overriding concern is that the lead vocal parts are foregrounded within the perceived space of the "mix", as a way of both isolating the voice from other mix elements and creating a listening experience in which the voice is felt to be "close", "up-front" and "immediate". This chapter analyses this aesthetic discourse about the voice, claiming that it is built on the erotic kind of voice fetishism described in Chapter 1, and preparing the ground for the way such voice fetishism interacts with the commodity fetishism of technology which I describe more deeply in later chapters. I make three main points: firstly, I show how metaphors such as "close", "up-front", "forward" and "present" coalesce into an over-arching metaphor I call 'vocal foregrounding' which drives the practice of vocal editing, and which is defined by assumptions about the voice's representation of a body that are germane to erotic voice fetishism. Secondly, and building on this first point, I suggest that erotic voice fetishism underpins producer-songwriters' allusions to the experience of voices in real-world situations of proximity when explaining these metaphorical terms. Finally, I claim that two other generic terms in this aesthetic discourse, "tone" and "dynamics", embody qualities of erotic voice fetishism in more general ways when they are used to describe broad characteristics of voices which are considered particularly special or unpredictable.

My use of the term 'aesthetic discourse' rather than just 'discourse' alone builds on the work by Feld, Théberge and Bates discussed in the introduction. To recap, this work highlights how the

language music creators use to describe musical sound mediates their conceptualization of their musical aesthetics. Like these authors, I use the term 'discourse' rather than simply language or vocabulary because it implies something more powerful and active than these terms. Following the Foucauldian use of the term, 'discourse' denotes a combination of linguistic forms, practices, technologies and institutions which come together to form a dominant system of knowledge. Discourse both embodies and produces this system, as well as sustaining relations of power within it, and the functioning of discourse tends to be made invisible through its ubiquity (see Born 1997: 498, citing Cousins and Hussain 1984: 77-97). The term 'aesthetic discourse' implies an ideological system—both produced and reinforced through language—which structures and motivates the production of cultural products according to a specific set of aesthetic ideals (see Born 1997: 483). The role of linguistic metaphors in scaffolding aesthetic discourses is important here: since music has no physical existence and tends towards the nonrepresentational, it is common for discussions of music across cultures to employ metaphors to connote certain musical features—metaphors which are necessarily derived from phenomena external to the music (see Born 1995: 18-20). To elaborate further on Feld's work (1981, 1982), he explains how the Papua New Guinean Kaluli use water-related metaphors as one way of describing the structure of their music, its creation and performance. In particular, several metaphors about musical structure coalesce around terms related to 'waterfall', including the notions of a waterfall flowing into a 'waterpool' denoting the act of musical composition, waterfall 'flow' describing rhythmic pacing in song, and water 'rushing forth' indicating strong vocal technique (1981: 28-38). Feld suggests that the use of these metaphors is rooted in water's status as a 'core natural image with broad mythical and everyday cultural salience both materially and ideationally' (1981: 44) for the Kaluli. This example highlights how groups of metaphors which are related by certain over-arching associations ('waterfall'), or chains of associations, can group together in their usage to describe music. These metaphor clusters constitute an aesthetic discourse which impels the creation of music in a way which relates to external cultural phenomena (Born 1995: 19).

In the case of commercial pop vocal editing, producer-songwriters' aesthetic discourse is dominated by a group of metaphors which encompass the notion of foregrounding the voice within a space ("up-front", "central", "close", "forward"), alongside the related term "presence". These metaphors are both rooted in, and arguably reinforce, the erotic voice fetishism underpinning this practice in a variety of ways. Namely, such metaphors spotlight and isolate the voice within an imagined performance space that is shared with the listener, conceptualizing it as a special fragment of an imagined, phantom-like performing body, whose existence within that space is implied by the hearing of its voice. They also express the desire for this implied body to be brought as close to the listener as possible within this space. The assumption that the voice powerfully denotes the manifestation of a singing body or identity—an idea which I have argued is fundamental to erotic voice fetishism—therefore underpins aesthetic aims in commercial pop vocal editing to foreground the voice, which more deeply relate to the cultivation of a sonic illusion for the listener of exaggerated physical intimacy with the vocalist. In addition, the attainment of vocal states to which these metaphors were applied was often described by my informants through comparison with the psychoacoustic properties of hearing voices in real-world situations of proximity. This illustrated further how the aesthetic discourse was founded on beliefs that the voice powerfully signifies the spirit of a physical body and indicates particularly sensuous or desirable qualities of that body which are amplified or experienced as intimate and exclusive when such a body comes closer. This chapter therefore introduces an important theme of realism to this thesis, which is extended more deeply in Chapter 4.

Even though my discussions of voice fetishism in the previous chapter also highlighted beliefs in the voice's capacities to index the social and cultural identity of a vocalist as well as their physical body, there seemed to be a particularly physical aspect to this aesthetic discourse which privileged notions of proximity between bodies and glossed over the identity characteristics of voices, suggesting that ideas about identity were informing the editing of these voices more tacitly, hidden behind these over-arching aims of vocal foregrounding. For example, given that the voices I discuss are largely female and their editors male, the experience of implied physical intimacy on which these

dominant metaphors in the aesthetic discourse are based is clearly gendered. This lends further strength to my interpretation of the phenomenon through the lens of erotic fetishism, as the act of inflating the "presence" of the voice in a way which exaggerates the impression of physical intimacy between two bodies is done by a male fetishist (in a position of relative power) to a fetishized female voice (in a position of relative disempowerment).

Lastly, my discussions of the term "tone" in this chapter also connect with recent research by Robert Fink et. al. (2018), who have called for greater engagement with the related terms 'tone' and 'timbre' and the way their deployment within aesthetic discourses varies across popular music cultures. Fink suggests that these terms have been historically marginalized in popular music studies literature due to their perceived ambiguity, even though ideas about tone and/or timbre are often important defining features of popular music 'communities, genres, and artistic personas' (Fink 2018: 3). Instead, Fink claims that the discursive power of these terms should be fully interrogated, focusing on what purposes specific definitions of the terms 'tone' and 'timbre' serve in popular music cultures resulting from the ideological values that underlie them (ibid.).

The structure of the chapter is as follows. Section 2.1. is in three parts: firstly, I historicize the metaphor of vocal foregrounding by showing how it was gradually consolidated in the discourse and practice of Western popular music mixing throughout the 20th century. Secondly, I draw on fieldwork data to show how metaphors of vocal foregrounding—specifically "up-front", "close", and "forward"—are extended and exaggerated in the aesthetic discourse of commercial pop vocal editing, particularly relating to notions of real-world situations of proximity. Thirdly, I analyse the related notion of "presence" in similar terms, additionally highlighting its broad and polysemic function within this aesthetic discourse. Section 2.2. is subsidiary to the core theme of vocal foregrounding, focusing on two more commonplace terms in the aesthetic discourse: "tone" and "dynamics". I show how these terms evoke erotic fetishism in a more general sense, encompassing ambiguous, shapely aspects of voices and embodying ideas about the voice being a special, complex and unpredictable sonic entity in comparison to other musical elements of commercial pop mixes.

2.1. Foregrounding the voice in the aesthetic discourse of vocal editing

Vocal foregrounding in 20th-century commercial popular music mixing

The quote from Recording Revolution[32] which opened this chapter includes an exemplary use of the term "up-front" in the aesthetic discourse of commercial pop vocal editing. The terms used in this quote illustrate how the concepts underpinning this discourse are defined by ideas or assumptions that an imagined listener is situated directly in front of, or inside, the popular music "mix" space, evoking an intimate theatrical scenario (see Lacasse 2000, Moylan 2012). I term this predominant metaphor vocal foregrounding as the aesthetic aim embodied by this metaphor is for vocals to occupy as foregrounded ("up-front") a position as possible relative to the listener ("right in the listener's face") who occupies an ideal middle point within this space. Before addressing how this metaphor emerges in commercial pop vocal editing, I will outline how it is derived from the way the voice has been historically foregrounded in Western commercial popular music mixing practices more broadly.

Since the development of modern stereophonic sound in the 1930s (see Théberge et. al. 2015: 1-15; Ouzounian 2021; 61-83), followed by the inception of overdubbing (Stanyek & Piekut 2010: 21) and multitrack recording techniques in the 1950s which permitted increasingly greater isolation, combination, and control over individual sound sources (Théberge 1989: 100; Downs 2016; Clarke 2013), the notion of the "mix" has been at the forefront of popular music production discourse and practice. The "mix" refers to the perceived bounded environment of the popular music track within which individual musical elements are positioned and "balanced" during the process of "mixing". In more practical terms, mixing engineer William Moylan defines the "mix" as both a creative (or craftlike) and technical act which involves 'pull[ing] together all of the instruments and voices, and provid[ing] them with appropriate sound qualities and sound relations to best support the musical materials and the story or message of the song.' (2014: 413). In a more abstract sense, the "mix" is a virtual space: sound studies scholar Jacob Downs has described it as 'a merged plurality, an integrated

[32] This is a YouTube channel run by American producer-songwriter Graham Cochrane which uploads pop music production tutorials. See chapter 1 pp. 90-91.

constellation of ‘captured’ sounds forming a cohesive yet messy singularity.’ (2016: 1). Others have also discussed how contemporary popular music mixes in the latter half of the 20th century were notable for the way they used the spatial potential of how the mix is eventually diffused in stereo by headphones or loudspeakers to deconstruct conventional physical or realist performance environments, rearranging the positioning of sound sources into impossible spatial arrangements (Théberge et. al. 2015: 9; Doyle 2005).

Several popular music studies scholars have highlighted the importance of the “mix” space—and the way musical elements are separated and positioned within it—to the sound of contemporary popular music tracks more generally (e.g. Doyle 2004, 2005; Lacasse 2000; Zak 2001). Allan Moore and Ruth Dockwray, for example, have developed analytical tools for the study of sound spatialization in popular music mixes (Moore 2005, 2009, 2010, 2012; Moore et. al. 2009; Dockwray & Moore 2008), as well as to determine historical trends in mix spatialization, leading to their theorization of the ‘sound-box’ (Dockwray & Moore 2010) as a method for the transcription and analysis of recorded rock songs from the 1960s and 70s. The ‘sound box’ is a way of visually conceptualizing the four-dimensional ‘image of a virtual performance’ (ibid.: 181) created by a stereophonic mix (Figure 15).

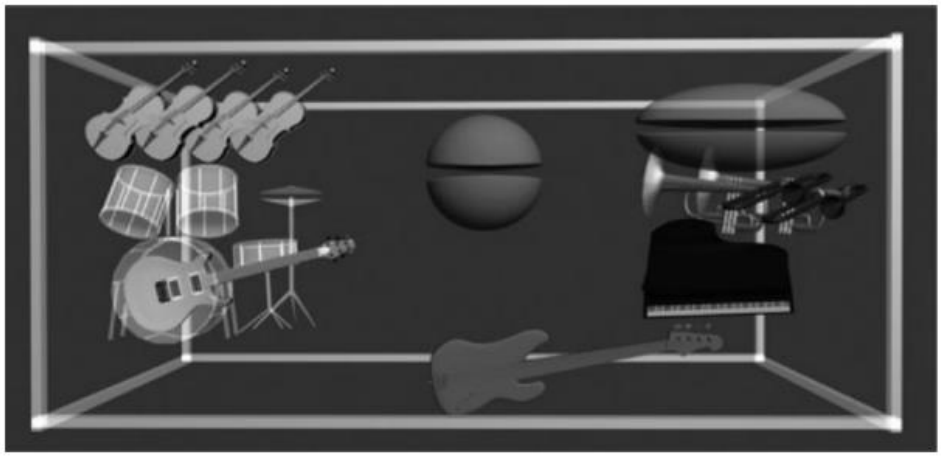


Figure 15: Sound box transcription for Brenda Holloway’s ‘You’ve Made Me So Very Happy.’ The lead vocal is represented by the sphere in two halves in the centre of the box. (from Dockwray & Moore 2010: 190).

In terms of *how* different musical elements are manipulated and positioned within the multiple dimensions of the "mix" space, practice-based literature about mixing documents these technicalities at greater length than these popular music studies approaches. The "mix" space can be delineated into horizontal, vertical and depth domains (or x, y and z axes), and the positioning of elements along these axes can be manipulated by adjusting different sonic parameters (Gibson 1997: 9-20; Hodgson 2010: 149-187; Izhaki 2017: Ch. 7: Marrington 2017: 8). Recordist Jay Hodgson (2010: 157), for example, theorizes that the 'three-dimensional canvas' (149) of the mix consists of a 'horizontal plane', a 'proximity plane' and a 'vertical' plane (Figure 16). More specifically: the horizontal plane (similar to the notion of the 'stereo field') can be manipulated using panning techniques (e.g. the 'pan pots' available in most modern DAWs (Figure 17)); the proximity plane using a combination of volume (e.g. compression, gain/level adjustment), frequency (EQ) and reverberation processing to create the illusion of depth or perspective as well as to manipulate the illusion of the size of the mix environment (see Doyle 2004, 2005); and the vertical plane using predominantly frequency balancing (with EQ), also partially exploiting the perception of pitch as a vertical space (Hodgson 2010: 149-187). Of course, these dimensions are also mutually influencing and blend together somewhat in perception—the collective result is that the listener perceives a three-dimensional mix-space populated with different elements, some of which may be more isolated or more perceptually salient than others.[33] Dockwray and Moore's 'sound box' is simply a different way of representing the general feel of such perceptual illusions diagrammatically. The perception of spatial arrangements like these is fundamental to Western popular music mixing practices.

[33] Clarke (2013: 93-95) addresses the psychoacoustic dimensions of the perception of the mix space more precisely, drawing on Brian C. J. Moore's (2003) work in the psychology of hearing to explain how: intensity as well as time and phase differences between the ears specify left/right positioning; spectral properties can specify all three of elevation, front/behind position and distance; signal-to-noise and signal-to-reverb ratios specify distance; echo and delay specify distance and volume or size or space, and reverberation time signifies volume of space. Importantly, Clarke draws attention to the fact that left/right localization is easier to explain in terms of the psychology of hearing than are the front/back, up/down and near/far dimensions, which depend on a combination of spectral properties, sound intensity and reverberation.

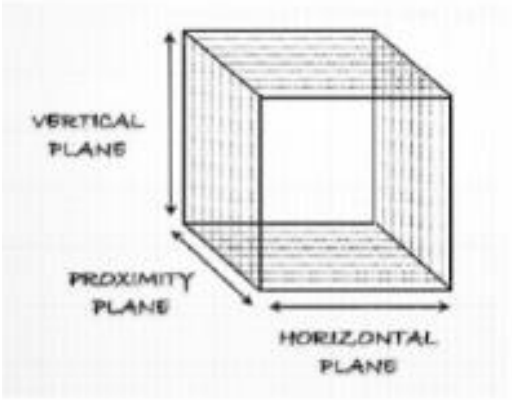


Figure 16: Hodgson's diagram of the three 'planes' which exist within a mix (2010: 156).

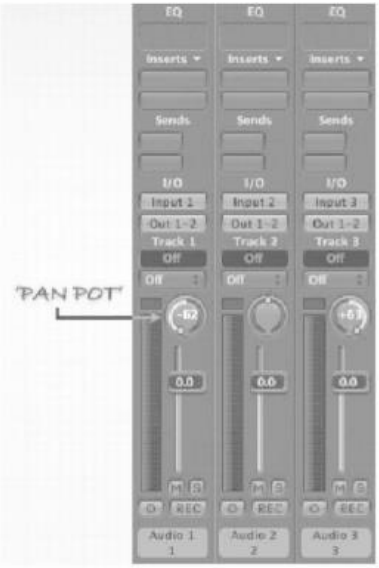


Figure 17: Hodgson's screenshot of a virtual mixing console, with pan pots noted (2010: 163).

Popular music studies scholars have also highlighted how, since around the 1960s, mixing techniques—especially in contemporary commercial popular music genres—tend to foreground and centralize the voice as much as possible within the mix space. Such foregrounding and centralizing of the voice can be partly linked to conventional pop-rock band performance arrangements, where the vocalist (singing into a microphone) was likely to be positioned front and centre, however, the consolidation of this spatial arrangement in mixing has become quite extreme, exceeding the mere imitation of performance setups. Dockwray and Moore, for example, have historicized this aesthetic trend using their 'sound box' to illustrate how a normative positioning of sound sources in rock music had developed by 1972 (the 'diagonal mix', visible in Figure 15), preceded by a clear shift towards placing the lead vocal track in the centre of the mix around 1966-1969 (2010: 187). Théberge also noted this aesthetic trend in 1989, pointing out how, during the 1960s, multi-track recording processes gradually led to a solidification of a mixing approach which foregrounded and centralized lead vocals, which were 'guaranteed exclusive right to one of the four precious tracks and subsequently given 'stage centre' in the final mix' (Théberge 1989: 103). Further, Dibben adds another social element to

the interpretation of these mixing trends, utilizing Edward T. Hall's (1963) theory of proxemics (Figure 18) to argue that production norms which place the voice up-front and centre are an attempt to construct a perceived intimate relationship between a vocalist and a listener. At this point it is also necessary to point out the difference between the perception of mix spatiality in headphones in comparison to speakers: the visual representations and theories I have described most obviously concentrate on the latter, where the listener is situated at an observational vantage point, in-front of the mix space—at the 'sweet spot', perfectly positioned at the centre of the stereo field (see Théberge et. al. 2015: 24-27)—whereas in headphone listening scenarios, the listener is more likely to perceive themselves as occupying a more central, immersed position 'inside' the mix (or even occupying an interior bodily space (see Downs 2021)). Particularities notwithstanding, the voice always occupies an up-front and central position in the mix in both scenarios. More significantly, this popular music studies work shows how this foregrounding and centralizing has been intensifying since the 1960s.

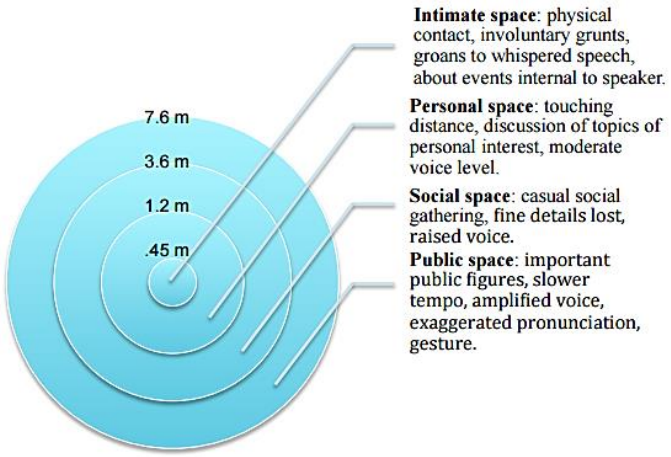


Figure 18: Dibben's extension of Hall's 'rings' of interpersonal distance which form the foundations of proxemic theory. The centre represents the location of the subject (/listener). (From Dibben 2013: 114).

Mixing textbooks and audio production sources go into even further detail about vocal mixing practices, noting the tendency towards mixing lead vocals which "sit" up-front and central, as well as more precisely outlining the mixing concepts underpinning such aesthetic ideals. Hodgson, for example, claims that lead vocals ought to act as an 'anchor point' within the mix space: they are 'almost without exception panned to the *front and centre*.' (2010: 160). Paul White from Sound on Sound magazine also claimed in 2012 that 'one notable contemporary production fashion is to keep vocals sounding relatively *forward*', linking this idea to vocals sounding 'free from obvious reverb', and 'retain[ing] the illusion of minimal processing' (White 2012: n.p.). An additional goal in Western popular music mixing, documented by audio engineer Kristen Hermes (2020), is the attainment of vocal 'clarity', which is related to the spatial considerations of mixing just explained, although slightly more specific to the separation or isolation of the lead vocal from other elements in the mix, and can be achieved via both spectral, spatial and intensity-related techniques.

Notably, none of these sources draw any attention to the genre-specificity of these kinds of aesthetic aims, unlike the popular music sources, usually just using terms like 'contemporary production', or 'pop production', which suggests that such trends are—among production professionals in the early 21st-century—considered to be pretty much universal or normative across Western popular music (even though this is unlikely to be the case). One exception is David Gibson's 1997 book 'The Art of Mixing' (and corresponding 1990 video[34]) which suggests that 'out front' vocals—towards the front of the depth plane—are most common in what he calls 'middle-of-the-road music, like Barry Manilow and Frank Sinatra', which is closer to the commercial specificity I am describing (1997: 82-83) (Figure 19). These sources illustrate how a voice (or lead vocal track) which is made to sound as "up-front", as central, and perhaps also as clear or isolated, within the mix as possible, is an aesthetic ideal in Western popular music mixing, especially in the case of commercial genres. Such sources also indicate how this aesthetic aim of foregrounding the voice has been naturalized as a fundamental and routine procedure in mixing.

[34] A link for ease of access: https://www.youtube.com/watch?v=TEjOdqZFvhY (TVHomeStudio 2012).



Figure 19: Visual representation of an 'out front' lead vocal in 'middle-of-the-road music' by Gibson (1990: 1:46:57)

To summarize, throughout the 20th century, commercial mixing practices have tended towards foregrounding the voice: they position the voice at the vanguard and centre of a virtual stage, in front of which the listener is positioned as close to the voice as possible. Such trends recall my discussion of the erotic dimensions of voice fetishism in Chapter 1 owing to the ways the voice is employed as a proxy for the construction of an imagined intimate physical scenario between a vocalist and a listener—a construction which is facilitated by attempts to replicate the sonic experience of physical proximity between bodies through sound mixing methods. This strengthens my claims that such fetishism of the voice is integral to the aesthetics of Western popular music more generally, and provides further context for the embeddedness of such fetishism in commercial pop vocal editing.

Vocal foregrounding in the aesthetic discourse of 21st-century commercial pop vocal editing

Throughout my fieldwork, my informants' extension of the metaphor of vocal foregrounding derived from mixing was primarily demonstrated in their use of terms such as "up-front", "forward", "close", "sitting well" and "sitting on top" to discuss the desirable results of vocal editing processes. More pertinently, many of these examples show how my informants couched these metaphors in terms of attempts to replicate the psychoacoustic experience of hearing a voice in real-world situations of physical proximity to a vocalist. In comparison to the practice-based literature about mixing just discussed, where the aim of vocal foregrounding seems to be largely interpreted as an aesthetic practice, these 21st-century producer-songwriters were more radically equating this aesthetic aim with notions of real experience, suggesting that vocal editing was an attempt to reconstruct or exaggerate features of this experience. This indicated that the erotic voice fetishism behind the metaphor of vocal foregrounding was embedded and normalized in commercial pop vocal editing, as ideas about the real-life experience of bodies in proximity seemed to be mixed up with ideas about the virtual construction of such an experience in my informants' explanations of their vocal editing goals. Put another way, ideas about bodies in proximity seemed incredibly prominent in my informants' use of these metaphors in comparison to these more generalized Western popular music mixing trends. Moreover, this suggested that the erotic voice fetishism underpinning the practice was quite extreme, as the link between the voice and a close physical body had been so naturalized to the point where aesthetic goals of vocal foregrounding were discussed as though this relationship was playing out in real life.

For example, in one of my most revealing discussions with Jason, he described how all his vocal editing work aimed to cultivate vocals which sounded "forward" and "close". In one statement, he explained how he felt "forwardness" was one of the defining aesthetic goals of commercial pop vocal editing:

> "I think pop vocals are all about having them as dry as they can be… like as *forward* as they can be." (Jason, 2019Z, italics added).

This statement showed how ideas about foregrounding the voice in the mix space were fundamental to the aesthetics of his vocal editing, even though here Jason was discussing a specific method for manipulating reverb on vocals in a way which disguised the sound of reverberation in the context of the whole mix. The term "forward" most obviously illustrates this, but his use of the term "dry" is also significant. In the context of talking about reverb, the terms "dry" and "wet" have codified meanings in music production—representing a continuum from a sound source untreated with reverb to one heavily treated with reverb. However, as Victoria Malawey (2020: 127) has recently argued, the terms have potential to be read more generically as describing the degree to which a listener is able to perceive the technological mediation of a sound, with "wetter" sounds more recognizably mediated. Regardless of the level of generality implied in Jason's use of the term, it was clear that the perceptual illusion of "forwardness" for him, was achieved by a variety of vocal editing methods that aimed to conceal the traces of their own application—that is, to maintain an illusion of realism, in which the aim was to trick the listener into feeling as though their experience of the voice in a track was as close as possible to a pure, unadulterated representation of how the voice had been recorded in proximity to the microphone. In this instance, such an illusion also connoted the perception of the voice sounding in a small and intimate space rather than a large and resonant one.

In the same session, Jason used the term "close" to describe the desirable effects of vocal EQ, also linking the concept to maintaining an illusion of realism. He claimed that the achievement of a "close" sounding vocal was linked to the way voices are generally perceived at differing levels of interpersonal distance in "real world" situations, additionally using the terms "clarity" and "presence" which are juxtaposed with these ideas:

> "This very top part [of the frequency spectrum] is important. The high frequency stuff above like 8-9kHz—that's all the detail that makes it sound *close*. If someone's far away, that's the part that dies away quickly in real world situations. You need to make sure that's there, […] to generate the overtones up [t]here to get that *clarity* and *presence*." (Jason, 2019Z)

This clearly illustrated how Jason's use of the metaphor of vocal foregrounding was linked to the way he believed voices functioned acoustically in real-life situations of proximity. Notably, although it is a somewhat informal scientific theory, Jason may be correct about the way the perception of sound

transforms with distance.[35] I discuss this issue more deeply in Chapter 4, however, as it more directly pertains to the use of EQ in commercial pop vocal editing. Here, the secondary point is more important, which is the way Jason's ideas about closeness are an example of how commercial pop vocal editing exploits and targets what producer-songwriters consider to be features of the way the voice is perceived in "close" real-world situations (high-frequency energy) and to exaggerate these features to make the voice's foregrounding in the mix seem more extreme. Jason's metaphors "clarity" and "presence" also contribute to the over-arching vocal foregrounding metaphor through their spatial connotations: "presence"—an important term in the aesthetic discourse discussed fully in the next section of this chapter—embodies a sense of shared physical space and attentiveness to stimuli in one's immediate surroundings, and "clarity" connotes distinctness and poise, which here refers to the isolation and spotlighting of a voice within a space.

A quote from a pedagogical article about vocal mixing by mixing engineer Dan Zorn at plug-in company Waves illustrates similar themes. The article opens by attempting to define the 'psychoacoustics of the high-end', alluding to scientific universals by introducing the concept of equal-loudness contours (otherwise known as Fletcher-Munson curves) derived from acoustical science in an attempt to define the most sensitive frequency areas in human hearing:

> The amount of high end in a vocal determines *how far away the listener will feel from the vocalist.* Vocals with less highs will always sound further away, while vocals with a present high end *confront the listener more closely.* (Zorn 2020, italics added)

Similar links between the metaphor of vocal foregrounding and ideas about how the voice is heard in real-world situations of physical proximity were invoked by my informants in relation to compression, which, producer-songwriter Robin discussed in terms of the way it amplified "quieter" vocal sounds such as breaths:

> [on the topic of parallel compression] So instead of bringing the louder bits down, you're bringing the quieter bits up. So that's why if you listen to a pop record you really hear the breaths. (Robin, 2020H).

[35] See Clarke (2013) and Monson et. al. (2014).

Compression will be discussed in more detail in Chapter 5, but for now, Robin's ideas explain the unique effects of compression in commercial pop vocal editing in terms of the way they exaggerate sonic features of the voice that would be heard more distinctly in close quarters. Amplified breathing noises also have clear erotic connotations which arise from the association between hearing the details of breathing and physical proximity to another human body.

In slight contrast, Chris's discussions of his vocal editing techniques incorporated metaphors of the voice being placed at varying degrees of both depth and elevation within the mix, including not only "jumping forward" but also "sitting on top", "elevated" or "sitting well" in an imagined space. These metaphors showed how Chris wanted the voice to occupy a privileged position within the mix, across multiple dimensions of the virtual space, suggesting that the metaphor of vocal foregrounding was quite broad for him, encompassing notions of height and situatedness as well as foregrounding. More importantly, Chris's use of these additional metaphors illustrated how notions of realism in the aesthetic discourse were only invoked in relation to the notion of foregrounding:

> [That word is] an important part, a strong part, so I'm sort of throwing everything at it—big EQ boost, big reverb, and that kind of makes it jump *forward.* (Chris, 2020G, italics added)

> (Q: Why did you use this [Scheps 74] EQ on the voice in his mix?) I'm guessing that within that mix it probably felt like [the voice] still wasn't quite sitting on top of things. So generally on the Scheps 74 [EQ] I'll add like […] a little boost there just to push it forward and out the way again. And the [higher] band on there as well, again just elevates it above everything. It's crazy actually that EQ—that band—it's literally like 'boomph' [he makes an onomatopoeic sound, like something thudding against a flat surface, or like an engine starting up]. (Chris, 2020S, italics added).

Chris's second quote, in particular, implies that the metaphor of bringing vocals "out of the way" or "on top", or "elevating" them above the rest of the mix, is referring to a similar impulse as the idea of vocals being pushed "forward". It seemed like Chris simultaneously conceptualized his vocal editing as aiming towards making the voice occupy a foregrounded position in the proximity (or depth) plane of the mix as well as the uppermost position on the vertical plane. These slightly different metaphors used by Chris emphasise how allusions to realism in this aesthetic discourse are specific to the notion of foregrounding, since he did not make any reference to replicating how the voice was perceived in real-world situations where the sound of the voice is elevated. This suggests that the link my

informants made between foregrounding metaphors and notions of realism resulted from such metaphors' entanglement with erotic voice fetishism: ideas about a vocalizing body existing in close proximity to a listener, and the connotations of intimacy which correspond with this spatial arrangement, have deeper roots in the practice in comparison to concepts of elevation or placement.

In later discussions, Chris also implied that although vocal foregrounding was one of the main aims of vocal editing, it ought to be moderated or facilitated by attempting to fit the voice more seamlessly into the overall environment of the mix. To explain this, Chris used a different set of metaphors related to painting, which also drew on ideas about simulating a live performance in which all the musicians are occupying the same space. This provided further context for how spatial metaphors were fundamental to the practice of vocal editing and formed a more generic metaphorical foundation in the aesthetic discourse out of which vocal foregrounding metaphors emerge. I quote Chris at length because, although not directly related to the dominant themes of this aesthetic discourse, this passage was remarkable for its use of spatial metaphors to describe the conceptual underpinnings of vocal mixing:

> (Q: What do you mean when you talk about getting the vocal to "sit well" in the mix?)
>
> I guess it's like… sounding like it's in the same environment as everything else. So for example, if you have a fully mixed track, and then you plonk on a dry vocal on top it's going to sound like 'the vocal lives *here* [hand gestures to one side], and the track's like *here* [hand gestures to the other side]'. So I guess the whole vocal mixing thing is like sculpting it to the world that you've created. You create a sonic landscape, right. I guess it's like a painting—you create the background, and then do you just draw a stick man on it, or do you look at where the sun is in the background and draw the shadow accordingly? It's just like getting it to live in the same world. Like… [a] similar space. And then it's like, if you slip with your paintbrush you've got a big black line. Trying to make sure it's all homogenous. So I guess when it's sitting right it's like it belongs, because it's frequentially [sic] and dynamically sitting in a place where it feels natural. It feels like it could almost be a performance. If it's a band track it sounds like they're in the same room. That's it essentially, you're creating like, rooms, spaces… I think with that track [Audio 2] it doesn't sound like a band anymore but it sounds like it's in some conceptual space, at least. (Chris, 2020S, asterisks providing emphasis).

Chris's statement paints the notion of getting a vocal to "sit well" as an attempt to generate a real or natural-sounding mix space, within which the vocal blends in, as a necessary counterpart to vocal foregrounding which attempts to bring the voice "forward" within this space. This ambiguity is an additional feature of the aesthetic discourse: the idea that vocal editing aims to simulate or even

exaggerate features of real-world scenarios of proximity, yet that this accentuation must be controlled and situated within the broader context of a mix. This balance drives the use of both vocal foregrounding metaphors and more general metaphors of placement or situatedness within a space ("sitting"). Here, for example, the most consistent feature of Chris's discussions was the idea of the "mix" being a "conceptual" space within which the voice "sits", which clearly links to the aesthetic ideals of mixing I have described.

Lastly, to compare across fieldwork and internet data, several popular tutorial videos on YouTube espoused these vocal foregrounding metaphors, using them to describe similar vocal editing procedures to those outlined by my informants here. This indicated that my informants' use of vocal foregrounding metaphors was bound up with their positioning within a transatlantic, online culture of music production, and suggested that such metaphors proliferate widely within this culture. For example, the quote from Recording Revolution (2016) which opens this chapter is taken from a video titled 'Mixing Vocals To Sound Upfront', and details a specific set of techniques involving the use of EQ to boost high frequency energy along with the use of two compressors in sequence to increase the overall volume of the voice to make it more consistent—techniques which will be discussed in much more detail in Chapters 4 and 5. The aim is ultimately 'to get every syllable, every vowel, every part of the vocal sounding up front' (ibid.). Similarly, a video by Waves Audio (2016) titled 'How to Help Vocals Stand Out AND Sit in the Mix' encompasses the same metaphors relating to the balance of vocal foregrounding with vocal mix placement to those just discussed by Chris, detailing another complex sequence of EQ and compression modifications to achieve this aim. Further, a video by tutorial channel Green Light Sound (2019) adopts a related metaphor, outlining a similar set of EQ and compression techniques which aim to "get the vocal to *cut through* a dense mix". Claiming that the techniques are frequently used by celebrity pop mixing engineer Andrew Scheps, this metaphor of "cutting through" embodies a notion of the voice competing within other elements in the mix space in the absence of vocal editing, which enables it to break away from this "dense" backdrop and instead occupy an isolated, primary zone within the space. Finally, a video uploaded by mixing tutor Michael White (2014) titled 'How to Pull A Vocal Forward in a Mix' uses almost all of the terms I have

mentioned thus far, or variants, within the first minute of the video ("forward", "stand out", "in-front", "on-top", "present").

There is small discrepancy relating to technical terminology to mention here. Despite the clear privileging of proximity to an imagined listener which this aesthetic discourse shows, there is a further term which circulates in the discourse of music production more broadly—the "proximity effect"—which refers to a specifically undesirable sonic feature arising from vocal recording processes. This is a term associated with microphone usage, and refers to the increase in low-frequency energy which occurs when a sound source is recorded too close to a microphone: approximately 1-foot, according to Huber & Runstein, and 'often most noticeable on vocals' (2018: 124). It is a curious and slightly ironic feature of this aesthetic discourse that, even though creating the impression of vocal proximity is a clear aim, there is also an association between close-miking and the generation of an unwanted sound in the voice associated with proximity. The details of how the impression of proximity is negotiated by modifying the frequency content of the voice using EQ will be taken up more fully in Chapter 4 but, for now, it is simply important to bear in mind that when I use the terms proximity or proximate I am not referring to the "proximity effect" and am instead getting at the more generally aesthetically desirable sense of proximity that my informants seemed to be aiming for when they talked about their vocal editing processes.

In short, vocal foregrounding was an established guiding metaphor in commercial pop vocal editing, and my informants' engagement with this metaphor as an aesthetic ideal was visible in their use of other metaphorical terms related to the privileged placement of the voice within a space, primarily through concepts of distance, depth, and proximity ("forward", "up-front", "close") figured in terms of real-world sonic experiences, and, to a lesser extent, elevation and a more general sense of situatedness ("sitting on top", "sitting well") to describe the aesthetic aims of vocal editing. These metaphors cluster into an over-arching metaphor of foregrounding the voice within a space, which is founded on an aesthetic aim to construct a virtual intimate relationship between two bodies that is redolent of erotic voice fetishism. Moreover, the link my informants make between the metaphor of vocal foregrounding and *real* situations of proximity is testament to the potency of erotic voice

fetishism in this aesthetic discourse. I now turn to a related metaphor in the aesthetic discourse—"presence"—which is given its own analysis as it merges ideas of space and intimacy in a slightly different way.

"Presence"

The term "presence" was incredibly common in my informants' aesthetic discourse about vocal editing, although it was used in relation to a wide variety of qualities sought during the vocal editing process, and its meaning seemed to vary widely in comparison to other vocal foregrounding metaphors which consistently related to spatial conceptions of the "mix". Here, I suggest that the ubiquity and polysemy of the metaphor of "presence" in the aesthetic discourse of commercial pop vocal editing results from its more direct engagement with the qualities of erotic voice fetishism which underpin the aesthetics of vocal foregrounding. More specifically, the term "presence" evokes the notion of a spirit or force that is felt or implied but cannot be detected. It also has a relational quality, encompassing notions of 'being present' with others or with an environment, an event etc. In contrast to the spatial foregrounding metaphors ("up-front", "forward", "close" etc.), the term "presence", when used to describe the voice, therefore more explicitly relates to the construction of an illusory intimate relationship between listener and vocalist which is at the heart of the aesthetics of commercial pop vocal editing. Given that the term powerfully captures this core aesthetic and cultural value, it performs a catch-call, wildcard-like function in the aesthetic discourse, deployed extensively to describe various desirable effects of the vocal editing process.

Before analysing my informants' uses of the concept, it is worth mentioning that the term 'presence' is known to circulate widely in music production and audiophilia discourses (see Milner 2009: 129-182), and is also a topic of discussion in a branch of practice-based research into immersive media (see Lombard & Ditton 1997). Although these definitions were not invoked directly by my informants, they reproduced similar ideas in their explanations of the term, meaning that such definitions provide important context for its usage in this aesthetic discourse. Music journalist Greg Milner, for example, claimed that the concept was foundational to audiophilia and high-fidelity

fanaticism during the 20th-century (2009: 129-182). He also suggested that the value of the concept in audiophilia discourse intensified substantially following the inception of stereophonic sound in the 50s, and the corresponding move in studio production towards the creation of 'impossible spaces' in mixes rather than the reproduction of realistic performance environments:

> If the audiophiles had a conceptual hook on which they hung their obsession, it was the search for "presence," often described in terms of the living-room-versus-concert-hall dialectic. John Urban defined presence as simply "the aural illusion of being in the same room with the performers." Herbert Brean in *Life* called it a "hard-to-define quality" that could "transport the listener to the concert hall." A recorded sound with presence did more than just capture the music perfectly. It captured the sound of music made in a specific space. The audiophile had to feel the sound flowing over him and bouncing off the walls, as though he were hearing it live. Presence signaled a return to the Victor/electrical-era ideal—record the room, not just the music—and a final refutation of the Edisonian belief that recording should only document the sound of music as heard in a flat, non-reverberating utopia. (ibid. 139).

Notions of heightened or constructed realism come to the fore in this definition of 'presence', specifically encompassing the creation of an 'illusion' of shared physical space and of the direct, unmediated experience of live performance. Similarly, the International Society for Presence Research—an institution with a practical focus in informing how creators of immersive media can enhance viewer or listener experiences—define the term as 'a sense of 'being there' in a virtual environment and […] *an illusion of nonmediation* in which users of any technology overlook or misconstrue the technology's role in their experience' (International Society for Presence Research, 2021: n.p.).[36] Further, presence researchers Theresa Lombard and Matthew Ditton (1997: n.p.) provide a simple introduction to some of the defining 'aural presentation characteristics' which contribute to the perception of 'presence': these include 'sound quality' (in a simplistic sense 'high fidelity sounds' are judged to be more 'realistic' and therefore more likely to contribute to a sense of 'presence'), and 'dimensionality' (where the spatial characteristics of sound, e.g. in stereo, are manipulated in a way which replicates a kind of 'aural realism' in a virtual space). Generally, the

[36] The full definition is provided at International Society for Presence Research (2000: n.p.): 'Presence (a shortened version of the term "telepresence") is a psychological state or subjective perception in which even though part or all of an individual's current experience is generated by and/or filtered through human-made technology, part or all of the individual's perception fails to accurately acknowledge the role of the technology in the experience. Except in the most extreme cases, the individual can indicate correctly that s/he is using the technology, but at *some level* and to *some degree*, her/his perceptions overlook that knowledge and objects, events, entities, and environments are perceived as if the technology was not involved in the experience. […]'

characteristics of ‘presence’ addressed in these definitions evoke similar themes of realism to vocal foregrounding.

In both cases, the term comprises a notion of heightened or selective realism, achieved through painstaking technical means, relative to an assumed loss of fidelity or realism which happens through technological mediation. The term is therefore paradoxical: the aim is to create an experience which appears to be as unmediated as possible, but the achievement of this state requires complex and subtle technological modifications. Even though these definitions do not relate to voices specifically, these connotations mean that term acquires greater potency in relation to the voice because, since voices are usually strongly associated with the identity or humanness of a vocalist (and, as I have discussed, this idea is one of the foundations of erotic voice fetishism), a “present” voice not only involves an aural illusion of realism, or of occupying a non-mediated space, but also an aural illusion of a non-mediated or immediate relationship with another human being. This further fuels the all-encompassing usage of the term in the aesthetic discourse of commercial pop vocal editing: it refers to a whole variety of sonic traits which frequently index the voice being in an immediate or direct relationship to the listener, as well as just generally sounding as distinct and dominant in the mix as possible. Indeed, across music production the term presence is not limited to “vocals” (e.g. see Gibson 1997: 93 for an example of its usage in relation to drum parts), however its use to describe the recorded voice is particularly ripe for analysis because of the notions of immediacy associated with the term which come to connote intimacy in relation to the voice.

The quote from plug-in company iZotope which begins this chapter, for example, is taken from an article which attempts to taxonomize genre-specific vocal editing styles, the fifth of which is labelled ‘the *ever-present* pop vocal’, and most directly addresses the vocal editing style this thesis focuses on (Messite 2018). Defined against other styles such as the ‘buried lead’ in atmospheric rock, progressive rock, and progressive metal genres, and the ‘harmonically distorted rock vocal’, the ‘ever-present pop vocal’ is outlined in terms of the vocal editing techniques used to create it, involving compression to achieve that ‘in-your-face, sparkly pop sound’, parallel compression for ‘warmth and

solidity[37]' and EQ to create 'immediacy' and 'closeness'. These statements from iZotope illustrate how "presence" in commercial pop vocal editing discourse is an overarching term which connotes many other subsidiary desirable aesthetic qualities, many of which relate to the aesthetics of foregrounding which I have described ("in-your-face", "immediacy", "closeness") but which also include other buzzwords which relate more to the spectral content of the voice ("sparkly", "warmth"). The latter terms will be discussed later in this thesis because they relate more directly to the specific technologies referenced (EQ and Compression, which are the subjects of Chapters 4 and 5), but important here is the way "presence" in this aesthetic discourse seems to be conceived as a holistic, felt quality existing in the voice, achieved when a variety of vocal editing techniques come together in the right combination. Like Milner suggests, it appears to be an elusive, 'when you know, you know', sort of quality, and these ideas of ineffability or intangibility also recall themes of fetishism.

My informants' understanding of the term also invoked a similar sense of the general feeling of vocal "presence" as well as using the term to refer to identifiable sonic qualities to do with the spectral content of the voice, a general sense of the lead vocal arrangement taking up substantial space within a mix, and the qualities of certain sung vowels. The flexibility of the term showed how it was a fundamental aesthetic value in commercial pop vocal editing, yet one that was also indexed by many different vocal qualities and vocal editing technologies. Oliver and Chris, for example, during observation and interviews, used the term to describe adjustments to the spectral content of the voice using EQ and saturation plug-ins:

> [I'm using a] Pultec [EQ, adding a high-end boost] just for a bit of presence (Oliver, 2020P)

> [I'm adding] a bit of saturation again [using the plug-in *Saturn* by FabFilter]. It adds a very light distortion. Just makes it a little bit more present. (Chris, 2020G)

In this context, "presence" was conceived as a subtle modification or enhancement applied to the voice's frequency content, and the way Chris and Oliver described it almost evoked the subtle addition of flavour in cooking. Indeed, I observed the term "to taste" being used in internet material

[37] Article states 'stolidity' but this is almost certainly a spelling error.

when referring to production modifications made at the producer's discrepancy (e.g. 'add EQ to taste', see SoundOracle 2018: 20'54"), which evokes similar cooking metaphors. However, the spectral modifications Oliver and Chris use to achieve "presence" are nonspecific: for Oliver, it is a high-end boost, and for Chris, it is the addition of saturation and/or distortion. Both techniques are intended to create different, although partially related, effects on the way the voice is perceived by the listener: in these production circles, a high-end boost is thought to create an exaggeration of high-frequency energy which emphasizes certain features of the voice associated with performance in close or intimate spatial zones. Saturation—(a subtle form of harmonic distortion applied digitally to the frequency spectrum of a sound, associated with the desirable distortion created by analogue equipment)—is intended to generate a similar sense of intimacy by creating a subtle inharmonicity in the voice not dissimilar from 'vocal fry'.[38] More generally, "presence" is used to describe both techniques because they both involve manipulating the spectral content of the voice to lend the vocal performance an exaggerated sense of proximity to and intimacy with the listener. This shows how the term was understood by my informants as quite a general aesthetic concept or idea related to proximity and intimacy, the attainment of which was gestured towards via a variety of vocal editing methods.

Incidentally, Oliver also used the term to describe the effects of compression, combining it with size terms like "big" and "massive", the term "well-rounded" which suggests a sense of boundedness or finitude, as well as the proximity term "in-your-face":

> So a lot of the tracks I'm working on at the moment are like tech-house rollers and things like that, which keep their energy up the whole time. […] The vocals are *big* and *massive* and there I am definitely, definitely […] compressing because I want that vocal to be *well-rounded, present and in your face*… stuff like that. (Oliver, 2020R, italics added)

Here, and building on the previous definition which understood "presence" as linked to the impression of proximity or intimacy achieved via spectral means, Oliver's language associates the term with an overall impression of the vocal's size or dominance in the mix. A "present" vocal is "big", it has

[38] See Danielsen (2010: 143-146) for a lengthy discussion and definition of this term and its association with a gendered kind of intimacy.

"energy" and is "well-rounded"—but the term "in-your-face" is also used, reiterating concepts of proximity and closeness (though perhaps not intimacy to the same degree). In other words, Oliver's description portrays an understanding of the voice filling-out and controlling the stage-centre of the mix, and "presence" is conceived as a key feature of this dominance. The elision of the term "presence" with all these other terms which evoke a notion of the vocal as the protagonist of a popular music mix again illustrates how the term is polysemic: it is a catch-all term which sets off a chain of related connotations ("in-your-face", "big" etc.), and which describes an ambiguous, though desirable, result of the vocal editing process.

In a slightly different vein, Jason used the term to differentiate between two different takes of the same word—('when')—during comping. The more "present" word, for Jason, was the word which had a brighter[39] sounding 'e' vowel, and he chose to include this comp instead of the one with a darker 'e'. He also elided the term "present" with the term "crisp", which connotes a whole variety of other concepts including firmness or rigidity as well as freshness and cooler temperatures:

> So that [word] seems a little bit more like *present*… it's kind of *crisp* sounding… compared to… [plays back the other take, with a darker 'e' vowel]. It's very subtle, but a small difference. (Jason, 2021A, italics added)

Jason's definition therefore made a link between the term "present" and the spectral quality of certain sung vowel sounds. Fink (2018: 11) made a similar link in his attempt to define the term "presence" in popular music cultures, suggesting that the sonic correlates of the term lie in the location of a sound's 'spectral centroid', which is further linked to the perception of 'brightness' in a sound. Citing a psychoacoustic study of timbre by Stephen McAdams et. al. (2006), Fink suggests that the importance of "presence" in the perception of timbre is linked to a more fundamental facet of human communication—the perception of vowel shifting in speech (Fink 2018: 11). It is notable that Jason's understanding of the term coheres so well with this preliminary academic theorization, but it should be clear from the other examples in this section that this is not the only way "presence" is understood

[39] To clarify, I am referring to the use of the terms bright and dark which are codified terms often used to describe particular vowel formations in singing and speech according to the sound produced by different formants in the voice (e.g. see McAdams et. al. 2006).

in the aesthetic discourse of commercial pop vocal editing. Again, this illustrates the amorphousness and multidimensionality of the term: in certain contexts, it is linked to precise acoustic correlates which index other perceived qualities such as brightness, forwardness, proximity and intimacy, but in other contexts it appears to be a more general term used to allude to such qualities without pinpointing any specific sonic features.

In a more ambiguous example, Oliver seemed to separate the terms "presence" and "warmth", suggesting that although the vocal he had produced in Audio track 3 was "present", it did not necessarily achieve a sense of "warmth". Even though the quote from iZotope at the start of this chapter seemed to elide the two terms, considering "warmth" as a subsidiary of "presence", this didn't seem to be the case for Oliver. And, in quite a revealing exchange, when I asked Oliver to define what he meant by "warmth" he reprised some of the terms he had previously used to describe "presence":

> I dunno I think this vocal is like, was nice and… present. But I think some vocals – my vocals could probably be a little bit warmer. (Q: Can I ask what you mean by warm?) I dunno, I think these sound very poppy like very bright, and stuff like that. Which works… but… I wanna try and maybe have a bit more like fullness and well-roundedness… if that makes sense. (Oliver, 2020R).

Again, this exchange illustrated the way the "presence" metaphor refers to a general sense of a voice occupying a foregrounded, distinct and dominant position within the mix, as well as being close to and intimate with the listener, rather than relating to more specific sonic qualities.

Considering my claims about the concept of fetishism more generally, the common usage of the term "presence" in this aesthetic discourse also evokes the notion of a mystical or otherworldly entity whose "presence" is strongly felt but cannot be seen. Although my informants did not explicitly use the term in this way, they seemed to convey a sense of strong aesthetic importance and gravitas every time they said it. Yet it also only ever described very amorphous and suggestive qualities of the voice—qualities which encompassed many ideas and connotations. Just like voice fetishism depends, in the erotic sense, on the unconscious assumption that the voice stands in for an imagined whole body or person behind it, the notion of "presence" seemed to stand in for the qualities of aesthetically successful voices, yet the qualities it referred to were miscellaneous and obfuscated.

In essence, this aesthetic discourse prioritizes "presence" as an ideal state for lead vocals to reach after several cumulative vocal editing processes, yet the term is more of a holistic or gestalt[40] aesthetic ideal rather than a term with precise sonic connotations. The term's wide-ranging, multi-purpose usage in the aesthetic discourse relates to the way it is both a key aesthetic ideal and a key ideological term linked to erotic voice fetishism through the way it powerfully brings together notions of intimacy, immediacy and realism. Arguably, this gives it a kind of buzzword status in this aesthetic discourse, where it can stand in for multiple kinds of desirable lead vocal sounds. It may also be the case that the sheer ubiquity and powerful polysemy of "presence" in this aesthetic discourse amplifies the perceived gravitas of ideas about intimacy and realism associated with the vocal foregrounding metaphor more broadly, magnifying the underlying erotic voice fetishism in the process. Moving on, the terms "tone" and "dynamics" also underlie voice fetishism in a similar way through their intentionally ambiguous and suggestive qualities.

2.2. Vocal "tone" and "dynamics": voices are special and unpredictable

The terms "tone" and "dynamics" used in relation to vocals—terms which circulate widely in Western discourse about music more generally—had specific and complex meanings for my informants. In this aesthetic discourse, "tone" and "dynamics" were vague terms which gestured towards felt, over-arching, shapely aspects of voices, yet did not quite define them precisely. This vague or intentionally ambiguous use of the terms can be explained in a variety of ways: they have a wider historical use in Western popular music production and reception discourses (see e.g. Gibson (1997) on dynamics, and Fink (2018) on tone), and they also encompass quite holistic sonic qualities given that various parameters of sound interrelate in complex ways in the case of human voices (pitch, timbre, loudness). It would be very easy to dismiss or skim over such terms due to their perceived insubstantiality. However, in this section I will show how the use of these terms in commercial pop vocal editing reflects a more general kind of erotic voice fetishism linked to the way the voice is

[40] To be clear, I use this term in the way the Cambridge Dictionary (2022) defines it: 'something such as a structure or experience that, when considered as a whole, has qualities that are more than the total of all its parts'

perceived by producer-songwriters to be a special, complex and unpredictable sonic entity in comparison to other mix elements. Moreover, my discussion of "tone" in particular begins to illustrate how such voice fetishism interacts with and strengthens the commodity fetishism of technology in this culture. Each subsection begins with a contextualization of each term in Western popular music cultures more generally, followed by a discussion of the use of the terms in commercial pop vocal editing which integrates fieldwork data.

Vocal "tone"

The concept of 'tone' has, across the 20th and 21st centuries, held significant currency in Western popular music cultures. In a more general sense, outside of this cultural niche, the term can connote a broad sense of quality, feel, or style, or more specific phenomena: in linguistics, for example, it can refer to the use of pitch as a lexical function in various languages; in music theory, to particular intervallic relationships; and in the visual arts to colour, brightness or hue. In Western popular music specifically, and as Robert Fink, Zachary Wallmark and Melinda Latour have theorized (2018), the term functions as an expanded version of the related term 'timbre'. Both terms broadly indicate an idea of sound 'quality', and they suggest that 'timbre' in popular music discourses usually refers to the more scientific notion of a musical parameter that is not pitch, rhythm or loudness, relating more directly to the spectral make-up of a sound (defined by spectral, temporal and spectrotemporal aspects as well as transients), whereas "tone" refers to a more ambiguous cultural or aesthetic idea—a term to which 'ideological values' (ibid.: 1) about sound are often ascribed.

For instance, the term often underpins popular music marketing and advertising strategies as well as acting as a proxy for other social and cultural issues such as race or gender. Using the example of the concept of electric guitar 'tone' in hobbyist and professional magazines, Fink et. al. show how the notion—which evokes some mysterious or elusive, ideally attainable quality of guitar sound—was foundational to rock guitar players' perceptions of authentic guitar performance, and, as such, was and still is fundamental to the way guitars and accessories are marketed as well as the relationship that guitar players have with their instruments (ibid.: 1-9). Later in the same book, Catherine Provenzano

(2018: 159-182) links the vocal tone(s) created by the plug-in *AutoTune* to a particular politics of gender and race, where certain 'tonal' expectations are demanded of black men and white women respectively. Expanding the guitar example to popular music more generally, the writers in this volume paint a picture of the 'pursuit' of tone in popular music production as 'relentless'—a 'quest' for an elusive sonic status to which popular musicians strive but which can never quite reach due to the fluidity of the concept (Fink 2018: 1). Paralleling the research in this collection, I interpret the term "tone" in the aesthetic discourse of commercial pop vocal editing as a further reflection and reinforcement of erotic voice fetishism as well as a concept which makes the mutually interacting voice and technology fetishisms which define this culture more apparent, as here it stands in for both a unique, indefinable quality of a voice and the elusive ("tonal") effects that certain plug-ins are thought to give to such a vocal "tone".

Preceding some more detailed examples, one general finding from my ethnography was that producer-songwriters consistently referred to vocal "tone" rather than "timbre" when discussing the modifications made to the voice using digital plug-ins throughout vocal editing. At first, I thought that their use of the term "tone" was unnecessarily imprecise, and that they had meant to say "timbre". However, the quotes I discuss in the following examples show that the reason they do not use the latter term is precisely because it is *too* precise. The term "tone" is all-encompassing, bringing together not just vocal timbre but pitch, "dynamics" and identity as well. Talking about "tone", for these producer-songwriters, is a way of acknowledging several qualities or beliefs they have about the edited voice all at once: it reflects an interpretation that vocal pitch, timbre and volume/dynamics are all interrelated in complex ways; that the plug-ins used throughout vocal editing both affect the tone of the voice and have their own identifiable tonal qualities; and that voices have unique identities or characters that cannot be solely pinned on timbre. Indeed, in a discussion with my fieldwork informant Chris (2020S), he described the term "tone" as a "broader version of the term timbre" including not just the spectral identity of the voice but also accent, diction and "character". This intentional vagueness surrounding the term "tone", I would suggest, both supports and promulgates erotic voice fetishism, because it reinforces the idea that popular music voices have ineffable or

elusive qualities that aren't clearly identifiable and yet are very important for the way such voices are appreciated and interpreted by listeners.

In several instances, my informants conceived of "tone" as an innate vocal quality unique to each individual singer—a special and consistent feature of the voice which defined their vocal style and marked its individualism and character:

> [on the topic of EQ] The main thing as well is that a lot of it is to do with *tone* […] And some people just have that kind of voice, which can help. (Michael, 2020E)

> [Her] voice is very characterful, and I think you'd want to like stay true to that *tone* […] [segué into discussion about conservative use of tuning to preserve this character] (Chris, 2020F)

Thinking back to my discussion of Nina Eidsheim's work in Chapter 1, the use of the term "tone" in this context shares characteristics with erotic voice fetishism because these producer-songwriters seem to understand particularly identifiable or characterful vocal "tones" as qualities which singers just *happen* to possess naturally or innately, in spite of the influence that their particular cultural or aesthetic milieux—or, perhaps more significantly, the producer-songwriters' interpretation and modification of their voice within a particular genre context—may have had on the way their vocal "tone" is heard by listeners. In the case of both quotes, both informants seem to display a kind of reverence and utmost respect for the voices being discussed: it is as if the vocalists have been 'gifted' with these indefinable yet distinctive and "characterful" vocal tones, which, ironically, are modified throughout the vocal editing process. Michael's assertion that "some people *just have* that kind of voice", for example, arose in the context of a discussion about how best to apply EQ to voices in the commercial pop genre—in particular, the use of the technique to exaggerate high-frequency energy (discussed further in Chapter 4). The idea that some people might "just have" a voice, or vocal "tone", which already possesses some of these qualities is therefore a fallacy because the focus on such qualities clearly arises from the specific aesthetic context in which Michael is working. Yet phrases like Michael's—which suggest a belief in an innate vocal "tone" which exists outside of this aesthetic context—were still frequently used by my informants.

Similarly, Chris's quote discussing his favourite vocalist collaborator who I mentioned in Chapter 1 illustrated how he was aiming to gear his vocal editing processes towards preserving the "true" vocal tone of this singer. More specifically, Chris felt that one way of achieving this preservation was to not undertake too much vocal tuning, as this might adversely affect the "character" of the voice by adjusting certain tonal qualities too much (this phenomenon will be discussed more fully in Chapter 3). Such interpretations conceive of vocal tone as a gift or special ability: a unique, yet indefinable, marker of the singer's identity which the vocal editing process should be careful to preserve, or otherwise enhance via technological modification. In turn, this suggests that ideas of vocal "tone" are endowed with power to define the identity and personhood of vocalists—a key marker of erotic voice fetishism.

In a slightly different way, producer-songwriter Oliver implied that a specific kind of vocal "tone" achieved through the modification of the voice using extreme pitch-correction was a defining feature of the contemporary "R&B" genre. He described how, in most of his vocal editing, he preferred to use the plug-in *Melodyne* to correct pitch in as subtle a way as possible, unless working in the "R&B" style (his term) where he would use *AutoTune* as it lent a more recognizably 'AutoTuned' "tone" to the voice:[41]

> I only use *Melodyne*, unless it's for R&B and they want their vocal *tone,* like, the R&B sort of sound [then I would use *AutoTune*]. (Oliver, 2019BB).

Oliver's discussions showed how the labelling of different kinds of vocal "tone" can occasionally act as a differentiating feature of certain genres, particularly when such a "tone" has a strong link to certain plug-ins or plug-in brands (e.g. *AutoTune*). The idea that a particular plug-in might be responsible for creating a specific kind of vocal "tone" rather than the presumed innate qualities of the voice itself links back to my discussions of technology fetishism in Chapter 1. Here, the sonic effect of a certain plug-in brand is mythologized as having an identifiable "tone" that it gives the voice in comparison to other plug-ins (e.g. *Melodyne*) which perform very similar modifications yet which do

[41] I expand on the differences between these plug-ins further in Chapter 3, pp. 151-158.

not possess the same unique "tonal" qualities. Indeed, there is a particular model of the plug-in—*AutoTune 5*, released in 2006—which was mythologized by one of my informants for the particular "tone" it gave vocals produced around that time (2019V). Internet sources show how this 'classic *AutoTune* sound' is consciously sought by some vocal producers, who sometimes opt for using *AutoTune 5* in particular contexts on purpose, even though there have been several newer iterations of the plug-in released since then.[42]

Finally, alongside his understanding of "tone" as a marker of unique vocal character or identity, Chris also claimed that vocal "tone" could be defined by a given singer's accent, or the interplay between their accent and the accent required of the commercial pop genre:

> I think tone can be a lot of things as well. Again, it becomes really abstract but like it can be in someone's accent, for example. [The vocalist Chris was working with], she's [European]. But her speaking voice is very American. So her American accent has a big impact on the tone, irrespective of the timbre of her voice… in the way that she will say certain syllables, the way it sounds when she says certain words. Even down to the choice of words. For example, with an American accent, some words rhyme that wouldn't rhyme in an English accent. It's the choice of words that they choose to use because of the sound of their voice, I think [that] really plays into the tone. (Chris, 2020S).

Chris's description of vocal "tone" as a quality "irrespective of timbre" is further testament to its perception as a fundamental yet elusive quality in vocal editing circles. For Chris, the quality of "tone" existed outside or above established musical parameters (e.g. timbre), blending multiple dimensions of the voice including features which make up accent including pacing, vowel quality, rhyme, diction etc. Again, I would suggest that this holistic, gestalt usage of the term can be interpreted as a kind of voice fetishism, because the concept stands in for several different qualities of the voice that, although somewhat related to the term 'timbre', resist precise description. Parallels can be drawn here between my informants' uses of the term "tone" and Barthes' use of the term "grain" to refer to an imprecise quality of the voice, broader than timbre, that resists description and yet simultaneously functions as a defining essence of a body or person. As I argued in Chapter 1, this

[42] Reviews of *AutoTune Pro* (the newest iteration of the plug-in) discuss how its 'classic mode' incorporates the old *AutoTune 5* algorithm as a way of satisfying this demand. (see https://www.sweetwater.com/sweetcare/articles/auto-tune-quickstart-guide/. https://www.soundonsound.com/reviews/antares-auto-tune-pro).

logic shares qualities with erotic fetishism, because it inflates the value of a single feature of a human identity or body, taking it to indicate more fundamental features of this larger whole.

In summary, such examples, alongside the fact that the term "tone" supersedes the term "timbre" in this aesthetic discourse, illustrate how the imprecision of the term reflects and fuels the erotic fetishism of the voice (as well as the commodity fetishism vocal editing plug-ins, to a lesser extent) on the part of these producer-songwriters. More specifically, the term "tone" is used in multiple guises to refer to multiple, distinct features of the voice, including: an elusive quality which defines a unique vocal style or character; the sonic identity lent to a voice by a certain type of plug-in which defines the sound of a genre; or the uniqueness of a voice which results from certain features of speech or accent. The commonality between all these qualities is that "tone" is conceived as a kind of 'special something', a subtle, indescribable quality that the voice just *has* and is fundamental to the way it sounds, but that at the same time is difficult to describe.

Vocal "dynamics"

Similar to the term "tone", producer-songwriters had their own, polysemic definition of the term "dynamic" or "dynamics" in relation to the edited voice. More specifically, they used the term to refer in a very broad sense to any kind of temporal change or variation in level within the vocal sound, alongside the more restrictive sense of variations in volume. This broad usage of the term is common across Western popular music mixing and production discourses more generally, and contrasts significantly from, say, the use of the term 'dynamics' in Western classical music cultures to refer mainly to relative loudness or volume levels. More significantly, I will suggest that this all-encompassing usage of the term plays a particularly important role in the erotic voice fetishism which I am arguing underpins this aesthetic discourse. Owing to the voice's volatile and unpredictable sonic characteristics in comparison to other elements in the popular music mix—(not only do popular music voices fluctuate widely in volume but this is inextricable from corresponding changes in timbre (or "tone") which are additionally linked to pitch and register changes)—my informants considered the voice more "dynamic" than other instruments, and they paid considerable attention to the finessing of

vocal “dynamics” during vocal editing. Before addressing field data I will define this culturally specific usage of the term “dynamics” more precisely.

The definition of “dynamics” taken from the Sound on Sound glossary page highlights the very general nature of this term in music production circles:

> Dynamics: a way of describing the relative levels within a piece of music. (Sound on Sound 2022).

Comparison with a definition of the term taken from Grove Music Online—a definition which represents how the term is understood in Western classical music practice and Historical Musicology—illustrates the much vaguer use of the term in Sound on Sound, which refers only to the relative ‘levels’ within a ‘piece of music’ rather than incorporating any more specific references to changes in loudness or volume, or to the aesthetic significance of the term:

> Dynamics: The intensity of volume with which notes and sounds are expressed. In the 20th century dynamics came to be seen as one of the fundamental parameters of composition which function interdependently to create musical meaning and structure. (Thiemel 2001).

The term ‘level’ does have a specific meaning in science and engineering, but only as a logarithmic measure of relative quantity (Thompson & Taylor 2008: 28). In acoustics, it is usually used as a way of measuring specific dimensions of sound intensity including pressure, power and exposure, measured in decibels (dB) (BSI, 2020: 6-8). However, these more specific understandings of the term level (involving distinctions between terms like power and intensity) are not covered in the Sound on Sound glossary nor were they ever invoked by my informants, which suggests that the term ‘level’ has acquired a more vernacular usage in relation to the term “dynamics” in music production cultures, encapsulating more generic changes in the perceived intensity or power of a sound. Notably, I found it very difficult to source any clear definitions of the term ‘level’ across either mixing textbooks or the Art of Record Production literature.

In fact, a book by mixing engineer David Gibson published in the late 1990s (referenced earlier in relation to his ‘bubble’ diagrams of mix spaces) captures this holistic understanding of the term “dynamics” in music production more accurately, in a way which mirrors the way my informants used the term. Gibson’s informal theorizations are not ground-breaking, but they are remarkable for

the way they attempt to characterise the aesthetics of mixing (across genres) in an over-arching, totalizing way, using simple, cutesy descriptions. (Indeed, Gibson's work has acquired a somewhat mythologized, meme-like status in YouTube music production pedagogy circles thanks to a video from the late 90s where he summarizes the content of his book on camera (TVHomeStudio 2012) (Figure 20)). Gibson's understanding of the term "dynamics" is strongly linked to his conception of mixing more broadly, and reads as follows:

> The art of mixing is the way in which the dynamics we create with the equipment in the studio interface with the dynamics apparent in music and songs. When we speak about dynamics, we are not talking about the common terminology used for volume dynamics; we are not talking about changes in loudness. We are talking about changes in intensity. […] A dynamic in music is any-thing that you get out of music. Music touches us in just about every aspect of our lives, and however you relate to music is, of course, valid. There are millions of dynamics discernible in music that affect us theoretically, emotionally, physically, visually, psychologically, physiologically, and spiritually. (Gibson, 1997: 76).

Gibson's definition of dynamics supports my argument about the more holistic nature of the term in this culture because it appeals to a more general sense of 'intensity' in music (a term which Gibson does not define), rather than loudness. His idea about "dynamics" in music applying to 'just about anything you can get out of music' is perhaps slightly far-fetched, but nonetheless significant for the way it implies a conception of "dynamic" musical qualities as any sense of transformation or fluctuation in a musical track which generates some sense of excitement or response in the listener. For Gibson, "dynamics" are a more general quality of movement or excitation in music strongly linked to the mixing process, and this is a much broader conception of the term than "common terminology for volume dynamics." In subsequent passages of the book, Gibson even expands this gestalt conception of the term further by linking it to an informal theorization of the different "dynamics" people perceive in music (i.e. different kinds of change and transformation)—including among others 'feelings and emotions', 'common structures found in the world', 'the workings of the brain', 'music theory', 'physical reactions', 'visual imagery and abstract colours and imagines', 'bubbles' , and 'spiritual connotations' (Gibson 1997: 76-78)—suggesting an even more expanded (perhaps extreme) conception of the term dynamic in comparison even to my informants. In addition, note the appeal to visual metaphor which he includes in the quote below—'the dynamics that people

see in music'—implying that the term has these kind of cross-modal, all-encompassing characteristics.

> So now you see that music can evoke a wide range of possible dynamics in people. They are as varied as people and life itself, and they are all valid. The recording engineer should look for and be sensitive to the dynamics that people see in music. The recording engineer's job is to create musical dynamics with the equipment that reveals or enhances the magic people find in music. (Gibson 1997: 78).



Figure 20: Screenshots from 'The Art of Mixing' by David Gibson (TVHomeStudio 2012).

Despite their simplicity, Gibson's theorizations are important because they illustrate a similarly expanded conception of the term "dynamic" to that used consistently by producer-songwriters in my fieldwork. As an initial example, the ending of the quote from Recording Revolution which opened this chapter highlights the way the term "dynamic" is applied to vocals in this way. To repeat:

> You generally want [vocals] to sit up-front in the mix.[…] But on their own, without the right processing, most vocals won't sound like that. They are simply *too dynamic of an instrument.* (Recording Revolution 2016)

The concept of dynamics in this context is intended to evoke a general sense of dynam*ism* or of the volatile qualities of the voice rather than to suggest changes in volume or loudness. To elaborate: by calling the voice 'too dynamic of an instrument' this quote means that the way the singing voice fluctuates throughout a song—in all three of volume, timbre and pitch which are all interrelated and mutually influence one another—makes it a complex and difficult task to modify it into the required aesthetic standard ("up-front"). Indeed, the video by Recording Revolution provided at the same source goes on to claim that the voice's status as a "dynamic instrument" should be the guiding principle for vocal editing, encouraging a careful and calculated approach to working with the natural fluctuations in the voice throughout a song.

One exchange I had with my informant Chris exhibited a similar understanding of the term "dynamics" as a term associated with a general sense of movement and dynamism across parameters rather than simple changes in loudness or volume. In a similar way to previous quotes, Chris's description used vivid metaphor and onomatopoeia, as well privileging the voice as particularly remarkable or special within his editing practices more broadly:

> Thinking about the word dynamic, now, would be kind of… not so much loudness but like movement in loudness? So for example if a track's like, '1,2,3,4, AAAAAAAA [he shouts and holds an 'ah' sound with consistent energy] … end', that wouldn't be very dynamic. But if it's the introduction, building up to the verse, and the chorus comes, and then you come back down for the second verse… it's this kind of movement. And the sort of spectrum of *size* and *width* is I think what I would mean when I say dynamic. [...] So for example if I say a *big* chorus, it doesn't necessarily mean loud. It probably means *dense* frequency-wise, *wide* spectrum-wise, *deep* reverb-wise, and loud volume wise. […] So I guess those kinds of processes and those shifts in size would be what I'd say you'd want to achieve to have a kind of dynamic mix. I guess it's about movement, right? If it's dynamic it's about moving, rather than just being static.

> [Q: But how do vocals fit into all this? How do the dynamics of a vocal track fit into a mix like that in comparison to say, a piano track?]
>
> Those two examples are interesting, cause I think pianos can be almost as dynamic as voice. I think any acoustic instrument in general, actually, has the ability to be just as dynamic—maybe not quite—but very close to as dynamic as a voice. But I think with voice you have things like lyrics and breath—stuff that's more 'frequential'—that really helps the dynamic as well. Like if you can hear a really soft breath, instantly that's going to make that voice feel like it's right in your ear. Whereas you don't have breaths on a piano. Well, you do have pedal sounds …
>
> [Q: So I guess voices are perhaps more unpredictable?]
>
> Yeah 100%. *You kind of know what you're getting when you press a C on a piano, but there are so many ways you can sing a C.* (Chris, 2020S, asterisks to show emphasis and italics added).

In a broad sense Chris's understanding of dynamics as "movement in loudness" echoes aspects of Gibson's definition of the term in terms of 'intensity', whilst also providing a more expansive sense of the term using a variety of spatial and shape-like metaphors. Chris's language shows a messy elision of various kinds of metaphors in mixing when it comes to the term dynamics: he uses terms to do with size (big/small), width (wide/narrow), space or directionality (depth/left/right), density (linked to frequency) alongside the more general sense of movement and morphology over time. Although the metaphor groups that Chris is using are specifically related to the conception of a "mix space" which is unique to the culture of music production, it is significant that *all* of these metaphors are invoked by Chris in relation to the term dynamics, as such a holism illustrates the gestalt nature of the term—it is clearly a very important and loaded concept in mixing, yet it has this broader understanding beyond the notion of loudness which is multifaceted or polysemic.

More specifically, Chris's discussion of the specific dynamic character of the voice illustrated how he perceived vocal "dynamics" to be special and interesting to work with in comparison to the "dynamics" of other mix elements such as the piano. The idea that what he describes as the unique "frequential" (frequency-related) and lyrical content of the voice ("lyrics and breath", presumably inclusive of other mouth sounds associated with the diction of certain consonants or words) can, in their fluctuation throughout a track, contribute to the "dynamic" character of the voice, clearly shows how this expansive conception of the term "dynamic" that I am describing becomes even more expansive in relation to the voice. In addition, Chris's statement that "there are so many ways you can

sing a C" in comparison to the piano where "you know what you're getting", suggested that he believed the voice to be slightly more unpredictable and therefore to have more creative potential in comparison to other mix elements. This also unambiguously illustrated that Chris considered the voice to be sonically important and special in the context of his mixing processes more generally.

In summary, like the term "tone", "dynamics" plays an important role in the aesthetic discourse of vocal editing as a purposefully holistic term referring to an ambiguous yet important quality of the voice that must be treated carefully during editing processes. Specifically, the term "dynamics" in this discourse evokes a general idea of movement, fluctuation or change in intensity both across a popular music mix and within the sound of the voice itself. Alongside "tone", such terms reinforce the erotic voice fetishism that underpins this aesthetic discourse, because they indicate my informants' sense that the sonic qualities of the voice are elusive, complicated, unpredictable, and composed of fluctuations in several musical parameters (volume, timbre/"tone", pitch) all at once. The use of these terms also shows how my informants think that these special qualities mark out the voice as particularly aesthetically generative and ripe for modification in comparison to other elements of commercial popular music mixes.

Conclusion

This chapter has illustrated how the aesthetic discourse of commercial pop vocal editing is driven by erotic voice fetishism. A variety of metaphors coalesce to form this aesthetic discourse. Firstly, and most primarily, terms such as "up-front", "forward", and "close" (and some related equivalents including "sitting out front" and "on-top" etc.) encapsulate a desire to overtly foreground the voice within the "mix" space. I call this over-arching metaphor 'vocal foregrounding', and its predominance within the aesthetic discourse indicates that one of the most important aesthetic goals of vocal editing is to make the listener feel like they are in virtual proximity to the vocalist. Moreover, this dominant metaphor is founded on the mechanisms of erotic voice fetishism, which assume that there is a direct link between the voice and an imagined physical body. In turn, this fuels the connotations of physical intimacy between bodies which characterize vocal foregrounding metaphors alongside the notion of proximity, and which producer-songwriters interpret in terms of the emulation of real-world situations

of heightened closeness. The related term "presence" is deployed ubiquitously in the aesthetic discourse because of the way it directly embodies these fundamental ideas. In addition, the more ambiguous terms "tone" and "dynamics" engage erotic voice fetishism more generally: they encompass features of the voice which producer-songwriters believe are special or unpredictable, and stand in for a whole variety of vocal qualities which are difficult to describe precisely due to the way the voice consists of multiple sonic parameters (pitch, timbre, loudness) which interrelate in complex ways in comparison to other mix elements.

A new and important theme of realism therefore emerges from my account of this aesthetic discourse, which I will return to in Chapter 4 of the thesis. This analysis suggests that strength of notions of realism in this discourse result from the embeddedness of particular ways of relating to the voice—which I describe as erotic voice fetishism—in the culture of Western popular music production more broadly, and which are amplified in the case of commercial pop vocal editing. Namely, since such fetishism depends on a concept of the voice being attached to or representing a body, ideas about the sonic properties of real-world interactions between bodies begin to guide the discourse (and practice) of vocal foregrounding in certain instances, in ways which arguably solidify this form of voice fetishism in the culture by making it seem natural and normal rather than specific to the Western commercial pop context. Notions of heightened or selective realism also come to the fore here (and are deepened further in Chapter 4), as vocal foregrounding sometimes involves exaggerating particular features of the voice associated with real-world situations of proximity (e.g. high-frequency energy, breaths), in order to make the impression of proximity more obvious or extreme. Again, the notion of erotic voice fetishism provides a useful way of contextualizing these tendencies: the voice is considered a powerful indicator of or link to a physical body, but here it is also fragmented from that body and certain qualities of it are emphasized, particularly those which are associated with a particular (intimate, sensual) way of relating to that body in proximity.

The characteristics of this aesthetic discourse are both reflected in and solidified by the sonic aesthetic of the commercial pop genre more generally. Listening to Audio Track 1, for example, particularly when using headphones, the voice occupies an exaggeratedly foregrounded and privileged

position, dominating the sonic "landscape" (to use one of Chris's terms) of the "mix" which supports it, alongside functioning as the main structuring feature of a track's narrative (the progression from verse, pre-chorus, to chorus, etc.). I admit that my suggestion that this discourse has a more performative or influencing dimension on the genre itself is based on my analyses provided in this chapter as well as my immersion in the production culture over the four years of this research project, rather than on any rigorous cross-comparison between the state of the commercial pop aesthetic discourse at certain points in history and the aesthetic character of the genre. I do believe, however, that the voice is afforded an increasingly foregrounded position in current commercial pop tracks and that such an exaggeration or foregrounding is intensifying. The conclusion to this thesis will muse on the nature and significance of this aesthetic more deeply.

Finally, this analysis prepares the ground for the subsequent three chapters of this thesis which—in the process of accounting for the roles played by comping, tuning (and related techniques), EQ and compression in commercial pop vocal editing—describe more precise features of this aesthetic discourse relating to these technologies specifically. For instance, there are some terms which play huge roles in the aesthetic discourse that I have chosen not to cover in this chapter, as they are only used in relation to these specific parts of the vocal editing process (e.g. "emotion", "harsh"). In addition, this aesthetic discourse becomes more complex in relation to EQ specifically and, as a result, Chapter 4 discusses a host of more specific terms and the ways producer-songwriters relate them to particular points on an imagined archetypal vocal frequency spectrum (see pp. 204-218). Chapter 4 also expands more fully on the theme of realism, focusing on the way producer-songwriters view the exaggeration of high-frequency energy in voices as a way of emulating the real experience of hearing them. Tuning, comping and compression, in contrast, have slightly different and perhaps less evocative discourses surrounding their usage. In the case of tuning and comping (and related techniques), this is because these technologies provide highly malleable visual representations of the voice already which make the practice a largely visual one, somewhat removing the need for verbal descriptions or analyses of vocal pitch and timbre. Not coincidentally, the visual objectification of the voice performed by these technologies—which forms the focus of Chapter 3—also primarily serves

the commodity-kind of voice fetishism, meaning that ideas relating to the voice's indexing of an intimate singing body recede into the background somewhat when these technologies are used. For compression, the perceived complexity of the techniques and plug-ins involved mean that producer-songwriters seldom describe the sonic effects of compression plug-ins using precise terminology, instead upholding certain 'legendary' plug-ins for their abilities to produce such effects and dramatizing their actions on the voice using violent terminology (e.g. "slamming"). These themes will be explained more fully in due course, but, most importantly, the terms described in this chapter which come together to produce the overarching metaphor of vocal foregrounding, alongside the important general terms "tone" and "dynamics", are foundational to this culture of vocal editing and frequently surface throughout the rest of the thesis.

Chapter Three: Comping, Tuning and the Visual Objectification of the Voice

My key informants started the vocal editing process by comping and tuning the voice to prepare it for further processing using EQ and Compression. Comping involves working through several recorded takes to piece together the 'best bits' of a vocal performance, and tuning (or pitch-correction) is intended to sculpt the pitch contours of a voice so that they conform more closely to a designated scale. These processes were part of a broader stage of refining the voice, which also included some other complementary techniques including timing, manual gain automation ("clip-gain") and audio restoration. Each involved zooming-in visually on the constituent parts of a voice, painstakingly working through vocal tracks at a microscopic level of detail. This chapter claims that comping, tuning and these related techniques visually objectify the voice, in a way which primarily strengthens the commodity fetishism of the voice, although additionally engages erotic voice fetishism and both kinds of commodity fetishism of sound technologies. The visual objectification of the voice prompts producer-songwriters to reduce the voice to a visual form, separated from its production by a vocalist, however their necessary dialogues with vocalists also encourage a kind of possessive fixation on the objectified voice as a way of preserving their power over it. The commodity fetishism of the voice also fuels the commodity fetishism of tuning and comping technologies, which blends aspects of both digital novelty and analogue emulation commodity fetishism through the valuation of such technologies for the sophisticated capacities they provide to control the voice in visual form and, in the case of tuning plug-ins, their unique sonic effects.

An important finding emerging from this analysis is that, in some cases, the interpretation of the voice as a visual form displaced listening in the vocal editing process. In other words, resulting from their increasing familiarity with these visual representations, my informants frequently made rapid adjustments to the voice by looking at rather than listening to it, illustrating even more radically how the visual objectification of the voice motivated the practice. Through close analysis of these moments, it also became apparent that specific characteristics of the commercial pop vocal sound were influenced by such visual objectifications, including the "sliding" characteristics of vocal pitch

contours, the regulation of the loudness of individual consonants and vowels within lyrics, and the expansion of a lead vocal's dominance within a mix through the layering of vocal doubles and harmonies.

In a theoretical sense, this chapter therefore interprets the increasingly important role of software visualizations in contemporary music production—discussed in the literature review with reference to Théberge's (2013) work—as a form of commodity fetishism, of which the process of objectification is a significant part. Following Théberge, I highlight the inordinate significance of visual representations of sound in this practice, through analysis of precisely how producer-songwriters interact with these representations during vocal editing. However, my interpretation of these visualizations as a form of objectification extends Théberge's approach and makes it more specific to a vocal context, in which notions of human bodies and identities strongly inflect these complex ways of working with sound (more so than e.g. drums, synths etc.). Throughout my observation of this early vocal editing work, the voice was reduced to a series of linear contours and 'blobs' which were compulsively sculpted at incredibly high levels of precision, leading producer-songwriters to fetishize the visual form of the voice, which stood in for the singing vocalist and the labour of their performance.

More broadly, such commodity fetishism of the voice works to the advantage of the plug-in industry, as it complements the commodity fetishism of comping and tuning technologies which blends elements of both kinds of commodity fetishism of technology I discussed in Chapter 1. My informants prized certain plug-ins and DAW features (including *AutoTune*, *Melodyne*, playlist comping, "clip gain" and iZotope's *RX*) for the way their functionalities supposedly allowed for more sophisticated manipulation of these visual forms of the voice. Even though these plug-ins and DAW features are ostensibly digital novelties, my informants also occasionally conceived *AutoTune* and *Melodyne* in terms of their supposedly unique, identifiable sonic signatures: a belief which shares characteristics with the commodity fetishism of analogue emulations.

Additionally, although this chapter principally makes the case that visual objectifications of the voice in comping, tuning and related technologies advance vocal commodity fetishism, the visual objectification of the voice can also be interpreted as a key point of convergence between both the commodity fetishism of the voice and erotic voice fetishism. Namely, even though the identity or body of the vocalist seems to recede into the background during tuning and comping as a result of this kind of vocal objectification, the practice necessarily remains connected to the identities and bodies of vocalists through the relationships producer-songwriters must retain with their vocalist collaborators. Ultimately, this means that the vocal commodity fetishism entwined with these visual objectifications is a kind of temporary state during the most involved moments of tuning and comping (which nonetheless influences the practice substantially). In fact, my analyses at the end of this chapter—of instances where producer-songwriters were required to undertake or revise tuning and comping processes in ways which accounted for the artistic aims of vocalists—suggest that the notion of visual objectification of the voice can also be interpreted through the lens of erotic voice fetishism, specifically in relation to fixation as a form of control. Precisely, despite the vocalist's key involvement in the broader process of songwriting and production, producer-songwriters still seemed to treat the voice as if it was a special artistic entity which *they* primarily ought to control, isolating and elevating its aesthetic importance to the extent that its editing became almost possessive. Facilitated by visual objectifications of the voice, producer-songwriters not only downplayed the labour the vocalist had put into performance, but also attempted to enhance and perfect that vocalist's voice according to their own preferences. Not insignificantly, this kind of fetishism, which similarly thrives on the visual objectification of the voice, is also gendered.

The chapter is divided into two main sections. Firstly, I will provide some history and context for tuning and comping, also briefly addressing subsidiary techniques that serve similar purposes of preliminary refinement including vocal timing, manual gain adjustment ("clip-gain") and audio restoration techniques. Such histories address the roles played by plug-in companies Celemony and Antares, as well as the longer lineage of DAW development by companies such as Avid and Apple, in driving the visual objectification of the voice in vocal editing. Then, I will show how these

visualizations inform commercial pop vocal editing processes. In particular, Jason and Chris's interpretations of vocal style in the commercial pop genre were entangled with their particular uses of the graphical capabilities of tuning software (*AutoTune Pro*) and layered waveform representations of the voice in the *Logic Pro X* DAW, and the way such visual representations encouraged them to work with the voice visually rather than aurally. In the final two shorter sections I also address two more specific effects of the visual objectification of the voice in comping and tuning. These include the visual objectification of vocal "emotion", which in certain cases was conceived as a visual rather than an aural feature of the voice, and the way the visual objectification of voices inflected and sometimes intensified the gendered divisions of labour in collaborative songwriting scenarios involving male producer-songwriters and female vocalists. In general, this chapter begins to paint a picture of the multiple digital processes that combine to create the commercial pop vocal sound, illustrating how such processes are underpinned by the interrelationship between differing kinds of voice and technology fetishism, and portraying the gendered qualities of such processes.

3.1. Contextualizing vocal comping, tuning and related techniques

Comping

Comping—or, more accurately, 'track comping'—is a simple technique for cutting up and recombining audio tracks. A precise definition of the term is provided by plug-in company Celemony: 'Comping is the technique of selecting from multiple takes the best rendering of each passage and concatenating the chosen clips to obtain what, given the available material, you consider the optimal performance' (Celemony Software GmbH 2018: 18). The roots of contemporary comping techniques go back to the history of multitrack recording with magnetic tape in the late 1940s and 1950s: 'tape splicing' became a common practice whereby reels of tape would literally be cut up and stuck back together using glue (McMurray 2017: 30). As Mark Katz has argued, tape splicing represents an important historical marker in the history of sound recording more generally, as it introduced an

unprecedented degree of tactile manipulability to the recording process (Katz 2015: 42-42).[43] A more generalized history of the impact of tape recording on sound engineering is provided by Susan Schmidt-Horning, who suggests that magnetic tape initiated the possibility of delaying decisions about performance until after-the-fact of recording, contributing to 'fix-it-in-the-mix' approaches (Schmidt-Horning 2013: 186-187). Simon Zagorski-Thomas pinpoints the birth of comping at the invention of the Ampex Tape Machine in 1947 and suggests that its most important historical impact was the way it facilitated 'the ability to zoom in [on recordings] at a microscopic level' (Zagorski-Thomas 2014: 61).

In contemporary music production, such 'zooming-in' is made even easier by the visual representations of audio tracks in DAW interfaces, which evoke vertically layered stacks of horizontally rolling tape (Théberge 2013; Bell et. al. 2015). For comping specifically, most contemporary DAWs enable the sorting of different takes of the same track into groups or 'playlists', which can seamlessly be 'comped' into a whole without having to split audio recordings and drag them between tracks. In 2009, for example, the eighth version of *Pro Tools*—a DAW produced by music software company Avid—introduced this feature, which allowed users to group multiple takes of a lead part in their own 'lane', which sits under the main track almost like a drop-down menu (Thornton 2009). Previously, each take would have had to occupy its own track individually (Thornton 2005). This functionality enables the user to comp many takes very quickly by listening to each take in turn, selecting the parts they want to comp in by clicking and dragging across the take, and then simply pressing a button to add the selected portion to the main comp (Figure 21). Other DAWs such as Apple's *Logic Pro X* have similar functions. In addition, looping functions, which are also ubiquitous in most modern DAWs, allow users to speed up the process of isolating and selecting the desired portion of a track to comp in by looping a very small portion of audio (e.g. a 2 second phrase) and listening to each take of this portion in turn.

[43] Another relevant source here is Journalist Greg Milner's evocative discussion of the impact of tape on the history of recorded music, where he paints a picture of the medium as introducing a new kind of materiality and durability to the idea of recorded sound (Milner 2009: Ch. 4).

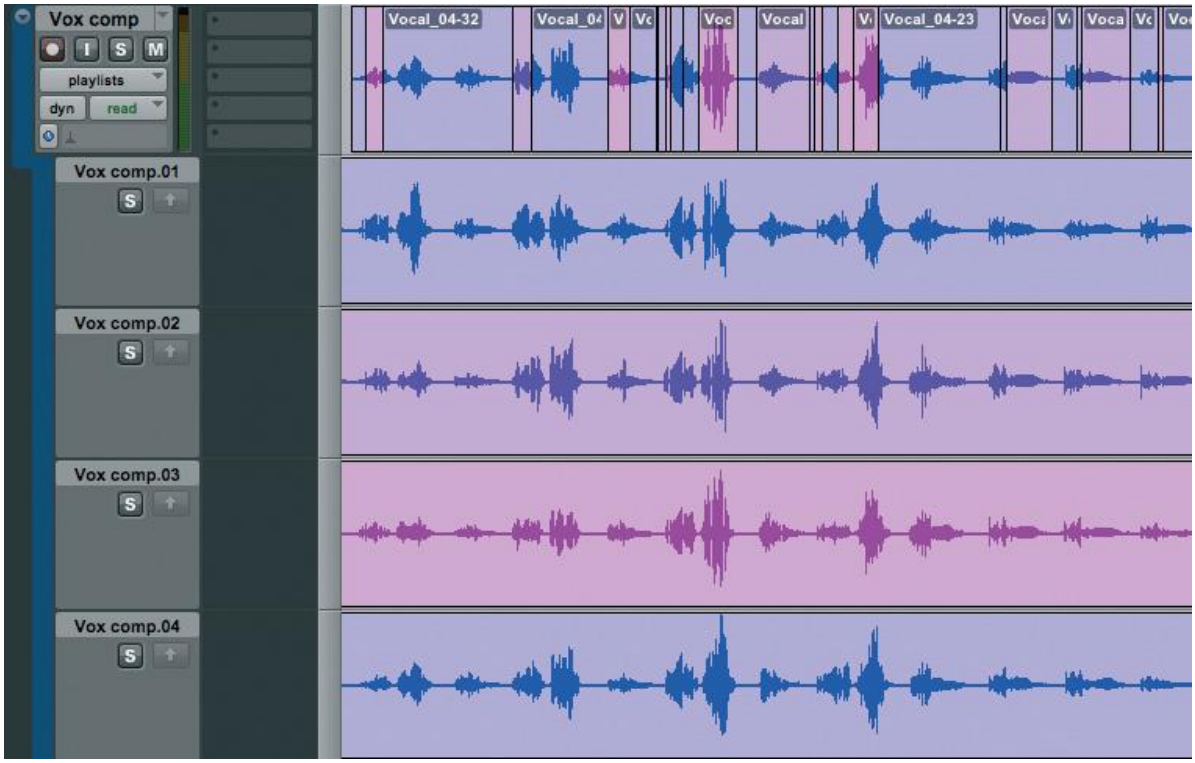


Figure 21: Comping a vocal in Pro Tools 8 using the playlists function. To quote: 'Regions comped to the master Playlist retain the colour of their source comp lane so you can see where they came from.' (from Thornton 2009).

These DAW functions facilitate quick and easy comping, which, in turn, makes it easier to comp at increasingly high levels of detail. This is specifically useful for vocal comping which, in my fieldwork, often involved concatenating a large number of takes—sometimes as many as 15 or 20—since vocals were considered the most important element of a track, as well as particularly inconsistent and unpredictable in comparison to other musical elements. My informants' comping approaches were noteworthy simply for the way they involved 'zooming-in' on the voice to an extremely high level of precision: occasionally, for example, comping sometimes three or even four takes into a single vocal syllable. At the start of the vocal editing process, most of my informants spent a hefty portion of time (anywhere from 1 to 2 hrs) "doing the comp". In addition, despite the fact that comping is obviously not solely done to vocal parts, it is notable that a lot of the articles about comping available in Sound on Sound magazine tend to mainly use vocal comping as an

example,[44] as it suggests that vocal comping is perceived among music production professionals as particularly complex or laborious in comparison to other forms of comping.

Alongside ubiquitous 'playlist' functions, contemporary vocal comping techniques using DAWs like *Pro Tools* and *Logic Pro X* also illustrate the importance of the *waveform* as a visual objectification of the voice in vocal editing. Huber & Runstein define 'waveform' as a 'graphic representation of a sound-pressure level or voltage level as it moves through a medium over time', going on to elaborate how 'a waveform lets us see and explain the actual phenomenon of wave propagation in our physical environment and will generally have the following fundamental characteristics: amplitude, frequency, velocity, wavelength, phase, harmonic content, envelope.' (2018: 47). Thinking back to Chapter 2, vocal editing waveforms therefore provide producer-songwriters with a useful visual estimate of vocal "tone" including key characteristics such as loudness, rhythm and diction, approximations of which can be detected through looking at the shape of waveforms. My informants were highly skilled in interpreting and manipulating waveform representations of the voice—a common skill across music production more generally. They observed how waveform shapes correlated with certain vocal features such as consonants, sustained vowel sounds and breaths. Percussive consonants at the beginnings of words like fricatives and plosives, for example, tend to appear as peaks on a waveform (described in audio engineering as 'transients'). Zoom tools in DAWs also enabled my informants to zone in on very tiny portions of vocal waveforms, to isolate miniscule selections of audio to cut out, or to make overlaps between sections as imperceptible as possible by precisely positioning crossfades (Figure 22). As Théberge has suggested, waveform representations in DAW workflows offer a unique 'form of material engagement with sound' (Théberge 2013), and with commercial pop vocal editing such an engagement entails both listening to and 'looking at' the voice, which, through the waveform, is represented visually as an abstract, colourful, blot-like structure.

[44] When searching for 'comping' on the Sound on Sound website, for example, 5 of the most-read 'techniques' articles use vocal comping as their example.

Lyrics: *Climb up the fen - ce, we go, we go,*

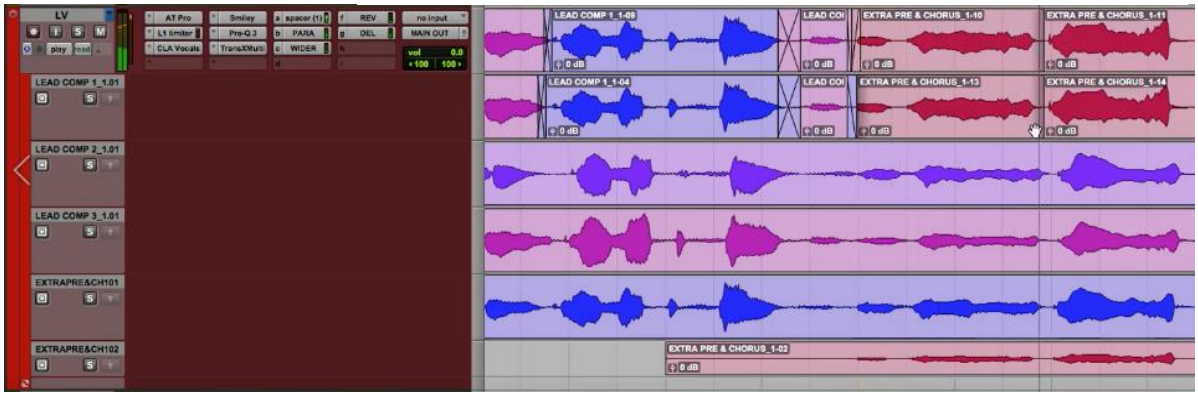


Figure 22: Complex comp in a track by Jason. The top track is the lead vocal. (2020J)

Tuning

In contrast to comping, tuning has a much more recent history, and the term also has much blurrier definitions. In a very broad sense, the term 'tuning' in contemporary music production refers to the use of digital 'pitch-correction' plug-ins such as Antares' *AutoTune* and Celemony's *Melodyne* to adjust pitch. This pitch adjustment is enacted either automatically, according to a user-defined scale (in equal temperament), or manually, by using a keyboard-roll interface which represents the pitch-contours of a sound as curves or shapes that can be directly manipulated. It is notable that the mixing and production textbooks I consulted as part of my fieldwork contained very little reference to digital pitch-correction plug-ins, and, when they did, they were only mentioned in passing, accompanied by warnings about their potential pitfalls if used too superfluously (Huber & Runstein 2018: 439-440; Case 2007: Ch. 10.2.3; Oswinski 2013: Ch. 11). Huber & Runstein, for example, describe how these plug-ins can automatically detect and correct pitch 'inaccuracies', which can be useful for 'fixing' the overall tuning of a track or undertaking 'manual micro-tuning' to correct 'out-of-tune' notes (2018: 439). My informants did not always use these plug-ins to 'correct' pitch, however, which makes it difficult to provide a stable definition of the term 'tuning'. In this chapter, I define tuning as the use of pitch-correction plug-ins both to 'correct' pitch as well as to sculpt vocal "tone", as my informants seemed to use tuning software for both purposes. Additionally, in contrast to other vocal editing

techniques such as comping, compression or EQ, tuning is much more exclusively and exhaustively used on voices. The uptake of Antares' *AutoTune* plug-in in the late 1990s was primarily focused on the software's ability to correct *vocal* pitch, even though it could also be used on other instruments (White 1997).

Tuning's ambiguous definitions in popular music production largely result from the technique's inseparability from the *AutoTune* brand name. In press discourse surrounding the topic of tuning, for example, it is synonymous with a sound effect created by the Antares-branded plug-in *AutoTune*, allegedly first used on Cher's 1998 track 'Believe' (Provenzano 2019a: 355). According to this origin story, the producers of 'Believe' tried to conceal their discovery of this effect, instead claiming in an interview that they used a Digitech Talker Vocoder pedal (Sillitoe 1999, cited in Danielsen & Brøvig-Hanssen 2016: 119; Dickinson 2001). In fact, the track's producers used *AutoTune* set to a 'retune speed' of zero: the 'retune speed' is a parameter in the plug-in which dictates how long the software takes to adjust a note to the 'correct' pitch (Danielsen & Brøvig-Hanssen 2016: 119; Provenzano 2019a: 162-164). When set to zero, this creates a jittery, 'gerbil-like' (Frere-Jones 2008), and exaggeratedly 'digital' vocal sound, as the software analyses the signal almost instantaneously, eliminating subtle transitions between pitches.

Contrastingly, in commercial pop, most of the vocal tuning is largely imperceptible to listeners not trained in music production, because tuning software is mostly used to make very subtle adjustments to vocal pitch contours manually, which seldom produces digital-sounding artefacts in the voice. Several academic and journalistic sources discuss this key difference: they contrast 'corrective' tuning, which refers to the more subtle use of the software to fix pitch discrepancies and 'smooth out vocal imperfections' (Provenzano 2018: 159) used frequently in commercial pop genres (e.g. Taylor Swift is the example given by Provenzano); with 'overt' tuning – which refers to the 'retune speed zero' effect, deployed more frequently in rap and hip-hop genres (Danielsen & Brøvig-Hanssen 2016; Ledsham 2015; Reynolds 2018). Provenzano refers to this effect as 'The Auto-Tune Effect' (TATE) (Provenzano 2019a: 4). The effect is also particularly associated with the early 2000s rap artist T-Pain and is sometimes referred to in press discourse as the 'T-Pain effect' (Parys 2020). Such terms were

used frequently by my informants to refer to a sound that they consciously tried *not* to achieve in their vocal editing. In this chapter I therefore focus predominantly on what these sources describe as the 'corrective' mode. This term does not fully capture the practice of commercial pop vocal tuning, however, as my informants did not feel that the only purpose of tuning was to 'correct' vocal pitch. As I will illustrate shortly, they used it more expansively to make subtle adjustments to both vocal pitch and timbre, including to smoothen continuous vocal pitch contours and to exaggerate or even create "slides" in the voice.[45] This was a highly stylized use of tuning oriented towards the production of a particular overall vocal sound, rather an act of 'correcting' the vocalist's performance, and could be described as vocal "tone"-shaping (with reference to Chapter 2). Indeed, I agree with Owen Marshall's denotation of a third type of tuning technique which straddles 'corrective' and 'overt' modes—'creative' tuning—which captures this expanded, "tone"-sculpting approach (2017: 96). Given that the distinction between 'corrective' and 'overt' tuning does not adequately describe the use of the technique in commercial pop, I therefore prefer to use the terms 'naturalistic' and 'non-naturalistic' to refer to these different tuning approaches, with the 'creative' tuning my informants employed occupying a position somewhere in between the middle and the naturalistic end of this spectrum.

During my fieldwork, the two most used tuning plug-ins were Antares' *AutoTune* and Celemony's *Melodyne*, and both facilitate vocal tuning by visualizing the voice in slightly different ways. Although existing literature has claimed that the usage of these two plug-ins in popular music production maps directly onto the binary between naturalistic and non-naturalistic approaches (with *Melodyne* used for the former and *AutoTune* for the latter, see Provenzano 2019: 184 and Marshall 2017: 78), both plug-ins enable all three kinds of tuning I have identified, and both were used by my

[45] Making sense of the interrelationship between vocal pitch and timbre in tuning requires additional clarification of the way vocal pitch and timbre are linked in acoustic terms. The most obvious way vocal pitch and timbre are interlinked concerns vocal register: put simply, vocal timbre changes depending on the pitch sung, and these changes can be varied for expressive effect. In tuning, vocal pitch and timbre are even more closely intertwined because adjusting pitch contours with too much deviation from the original signal may create artefacts in the timbre of the voice due to the transformation of vocal formants (see Heller 2012: 358-365). This is why vocal timbre can sound unfamiliar and 'posthuman' when heightened, non-naturalistic tuning is used (see Provenzano 2019a: 209, 355-356).

informants. Although the broad functioning of each plug-in is similar, both have slightly different interfaces, and the algorithms underpinning each plug-in are slightly different.[46]

AutoTune allows the user to apply pitch-correction in two modes: automatic, and graphical. The automatic mode requires the user to set several parameters which will dictate the process of automatic tuning: including key, scale, the range of the singer (soprano, alto, tenor, bass), as well as the 'retune speed'. The key and scale determine the set of notes to which input pitches are automatically 'tuned': at any moment, the 'target pitch' will be the note of the scale that is 'closest to the detected input pitch' (Antares Audio Technologies 2018: 18). As I have briefly mentioned, the retune speed governs the speed with which the input pitch is tuned towards the target pitch, thus directly influencing the tuning's degree of subtlety (Figure 23). *AutoTune*'s graphical mode, in contrast, allows the user to directly manipulate pitch-contours manually by re-drawing pitch-contour lines. Figure 24 shows the graphical mode in-use in one of the most recent iterations of *AutoTune*—*AutoTune Pro* (released in 2018)—which exhibits green, red and blue pitch-contour lines, with pitch on the y axis and time on the x axis. The red lines represent the detected pitch contours of the input vocal recording; the blue lines represent the editable 'curves' produced by the software ('output pitch contours');[47] and the green lines show modifications made to these curves by the user (Antares Audio Technologies 2018: 35). These pitch-contour lines are superimposed on top of the vocal waveform, which cannot be manipulated, but which can be enlarged visually as the user zooms in on the pitch contours. The user can also use the 'make notes' function to capture portions of the curve in boxes

[46] I direct the reader to Provenzano's (2019a: 105-255) discussions of the two plug-ins for an explanation of their algorithmic differences, as there is no space to cover this here.

[47] This difference between the red and blue curves may seem slightly redundant, but it is in fact a way of allowing the integration of some degree of automation into the graphical editing process. Once the user decides to start editing an input signal (red curves), they can use the 'make curves' function to produce blue curves, but they also have the option to apply some automatic tuning so that the blue curves adhere more precisely to the keyboard-roll *already,* before editing with the green curves (Antares Audio Technologies 2018: 16).

and move them up and down the y axis, which can speed up the process if less precision is required (Figure 25).



Figure 23: The auto-mode interface for *AutoTune Pro*, with scale and retune speed noted (Image from Antares Audio Technologies 2018: 11). The 'Humanize', 'Flex-tune' and 'Natural vibrato' parameters can also be adjusted, however they were not used by my informants.

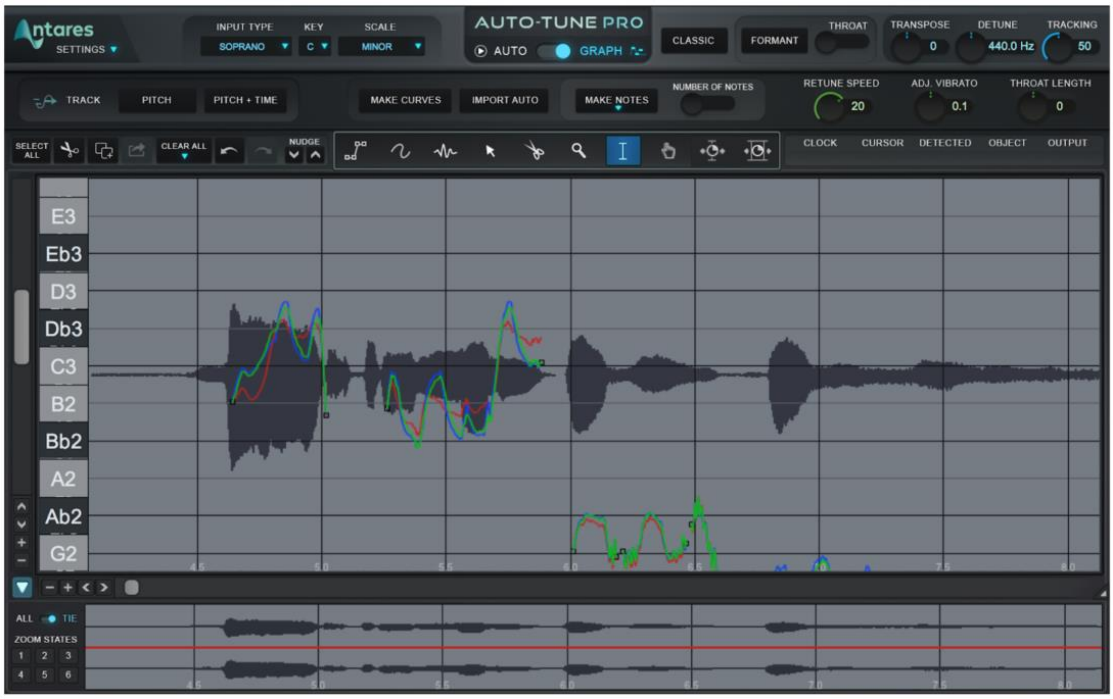


Figure 24: An example of a user manually tuning a voice using *AutoTune Pro*'s graphical mode (Image from Antares Audio Technologies 2018: 13).



Figure 25: An example like the above but including the 'make notes' function, which is represented by the rectangular blue boxes which are superimposed on the pitch-contour lines. (https://www.antarestech.com/wp-content/uploads/2019/12/es-atpro.jpg)

In contrast to *AutoTune*'s graphical mode, Celemony's *Melodyne* analyses the input signal and visually represents it as a series of orange 'blobs'.[48] According to Celemony, the blobs are intended to represent 'notes', not 'waveforms': they reflect the *Melodyne* algorithm's specific interpretation of a sustained pitch centre within a melody (Celemony 2020, cited in Provenzano 2019a: 198). Pitch contour lines also bisect the blobs, similar to *AutoTune*'s graphical mode. *Melodyne* also offers an automatic tuning mode, known as the 'Correct Pitch Macro' function (Celemony Software GmbH 2018), although this cannot be applied in real-time as an insert plug-in in the same way *AutoTune* can. The *Melodyne* interface, and the way it is integrated into the DAW workflow, is more solely centred around this similarly 'graphical' way of adjusting pitch-contours manually. During the tuning process in *Melodyne*, a user can move whole blobs with reference to the piano roll on the left-hand side (Figure 26); split up blobs into constituent parts (Figure 28); and use the 'pitch drift' and 'pitch

[48] 'Blobs' and 'notes' are terms Celemony use rather than my own (Celemony Software GmbH 2018).

modulation' tools to adjust the extremities of the oscillating pitch contour lines within individual notes (Figure 27).

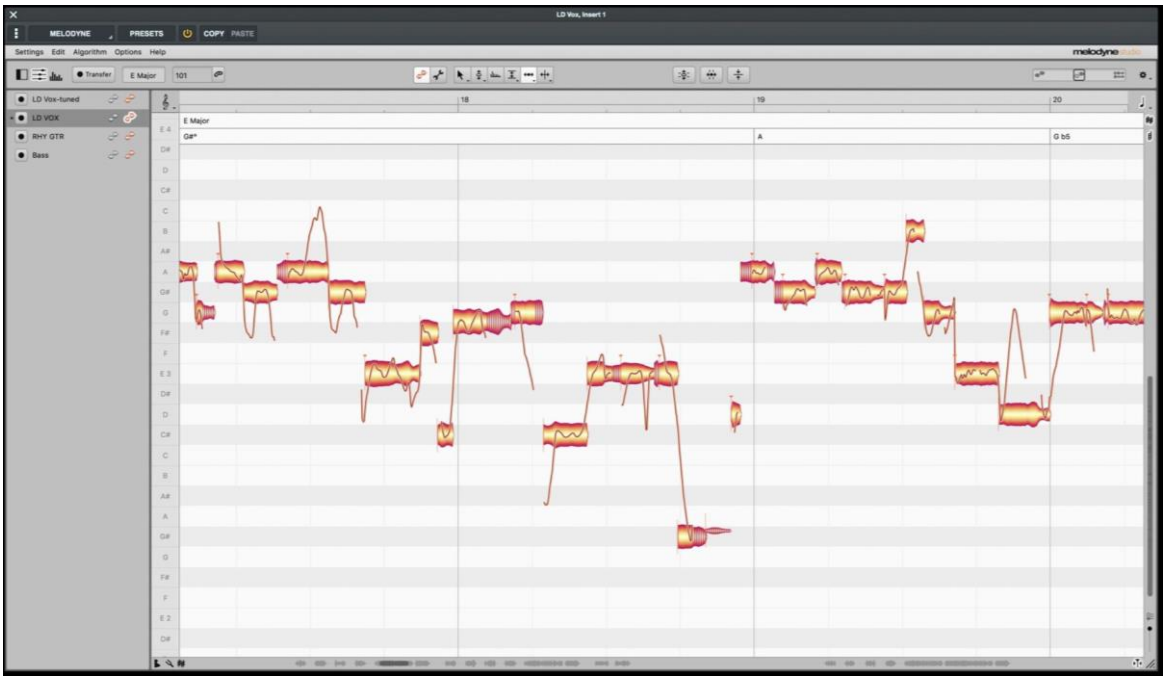


Figure 26: *Melodyne 5* interface. (https://www.mixonline.com/technology/product-of-the-week-celemony-melodyne-5)

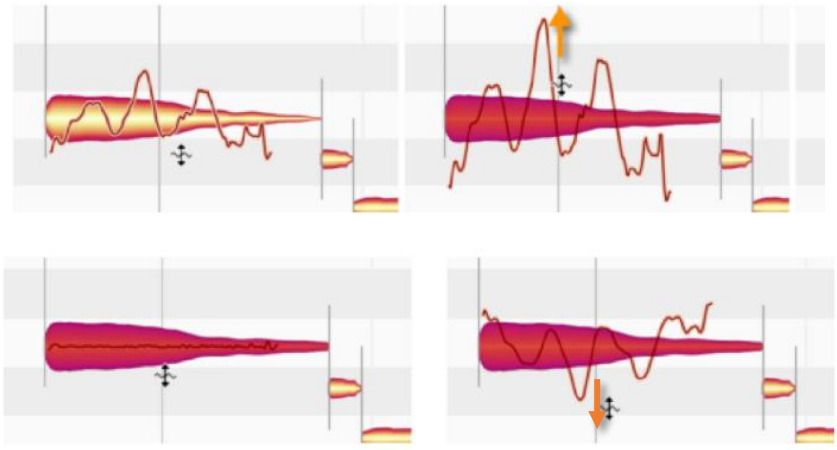


Figure 27: Using the pitch modulation tool in *Melodyne*. (Celemony Software GmbH 2018: 92)

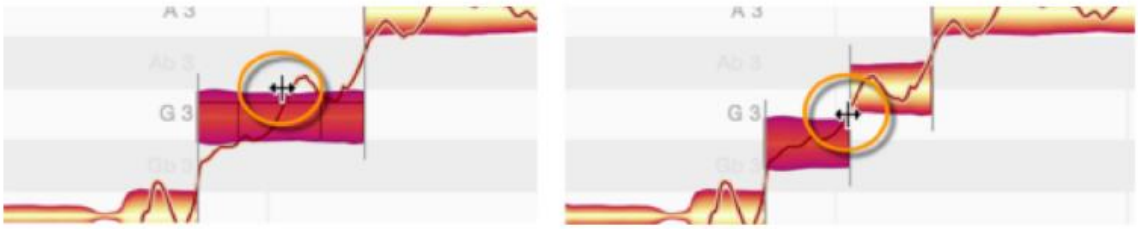


Figure 28: Using the note separation tool in *Melodyne*. This tool allows the user to split up notes and then drag the newly separated notes into new positions. (Celemony Software GmbH 2018: 108).

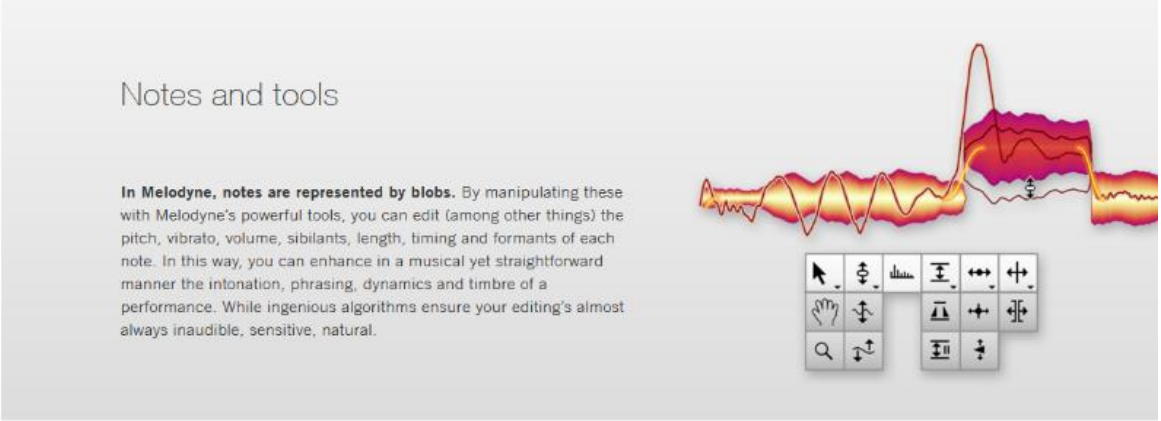


Figure 29: An example of how *Melodyne*'s promotional materials encourage a more holistic way of working with vocal sound, rather than just focusing on pitch. (Celemony 2020)

Most importantly, these plug-ins do not just enable pitch-correction. Their integration of multiple visual representations of the voice facilitates 'creative' tuning by allowing for a more holistic, visual approach to sculpting vocal "tone". In both cases, the plug-ins give the user a visual sense of how their modifications to vocal pitch-contours affect vocal waveforms. With reference to Figures 24 and 25, *AutoTune*'s graphical mode represents the voice as a series of linear pitch contours which the user is encouraged to trace over and re-draw in the process of tuning, as well as including the waveform in the background so that the user can see how the amplitude of the signal corresponds to the pitch-contour. The cursor can be used like a paintbrush to trace over the original voice and re-shape it. *Melodyne*'s interface, in contrast, visualizes the voice as a series of discrete 'notes' which the user is encouraged to pick up, cut up into pieces, and move around, as well as including the pitch contours

which can only be condensed or expanded within ‘notes’ and cannot be literally re-drawn. Both interfaces objectify the voice as a series of different shapes that can be gradually sculpted into a desired form. *Melodyne*’s interface encourages a more holistic idea of working with the voice as ‘sound’ by providing the blobs which represent neither waveforms nor pitch contours but rather a general sense of the voice’s ‘shape’ (Figure 29). *AutoTune* represents the voice slightly differently and not as holistically, instead combining two separate objectifications of the voice by superimposing pitch contours on top of the waveform.

Notably, across my fieldwork, appraisals of these tuning plug-ins also revealed how their commodity fetishism blended characteristics of both digital novelty and analogue emulation fetishisms. On average, my informants justified their allegiance to either *AutoTune* or *Melodyne* by claiming that their chosen plug-in provided more sophisticated or flexible ways of manipulating the visual form of the voice. For example, several preferred to use the newest versions of *AutoTune* in graphical mode for the kinds of manual, corrective alterations that the others usually associated with *Melodyne*. They argued that the *Melodyne* interface was unwieldy to use, and that *AutoTune* enabled more direct manipulation of pitch-contour lines (2019U, 2019X, 2019CC). *Melodyne* users similarly expressed aversion towards *AutoTune*, arguing that the graphical-mode interface provided less scope for manipulation (2019BB, 2020I, 2020F). However, some informants also alluded to the unique yet indefinable sonic signatures possessed by each plug-in, despite the fact that when used for naturalistic or creative tuning they produce very similar results. *AutoTune’*s association with the effect it produces at a retune speed of zero fuelled claims like these. Chris, for example, told me that he only used *AutoTune* if he wanted something to “sound like T-Pain” (2020F), and Michael described how he used *AutoTune* “lightly” as a “stylistic choice” for a unique project because he felt it gave the voice a “modern edge” (2020E). My discussions of Oliver’s claims that *AutoTune* provided a recognizable “R&B” sound, along with another informants’ continued adherence to *AutoTune 5*, in Chapter 2, illustrate similar sentiments (see pp. 130-131). Similarly, some of my informants justified their usage of *Melodyne* over *AutoTune* because they felt that it offered a particularly naturalistic “sound” to the tuned voice that *AutoTune* in graphical mode was incapable of producing (2019C; 2020K), suggesting

that *AutoTune* was more likely to produce "unnatural" sounding modifications (2019BB, 2020I, 2020F). Most importantly, both forms of commodity fetishism complement the vocal commodity fetishism enacted by the visual objectification of the voice in these plug-ins, by further solidifying producer-songwriters' intense focus on the form of the voice (both visual and sonic) separated from the context of its production. Moreover, the fact that a core characteristic of analogue emulation commodity fetishism was here present in relation to tuning plug-ins—technologies that have only ever been digital—suggests that this kind of commodity fetishism was being reproduced in the digital realm to such an extent that it occasionally encompassed digital novelty plug-ins as well as analogue emulations.

Timing

Alongside tuning and comping (which are the main techniques discussed in this chapter as they seemed relatively more time-consuming and creatively laborious in my informants vocal editing processes), several other techniques which employed a combination of DAW functionality and specialized plug-ins further illustrated my informants' tendencies to work with the voice in a variety of visual forms, suggesting that these plug-ins had a subsidiary role to play in the visual objectification of the voice in vocal editing. In my later analyses in this chapter, such processes also complement the comping and tuning techniques I discuss.

For example, vocal timing, like comping, largely involved using the DAW's tape-like interface to cut up and move portions of audio manually. All my informants stressed the importance of keeping vocals "in-time"—both by ensuring that the lead vocal followed the desired rhythm of the song's melody, and also by making sure that the ensemble between the lead vocal and doubles or backing vocals was as "tight" as possible (2020S, 2019Z). Usually, this involved manually splitting up portions of audio and lining them up with the tempo grid provided in most DAWs, or lining them up with the transients of existing vocal waveforms (e.g. backing vocals, doubles or a guide). (Figure 30). The practice of time-aligning tracks in DAWs is extended from the history of multitrack recording, where performances recorded asynchronously might have resulted in timing

discrepancies.[49] Most importantly, like comping, the way the DAW visualizes vocal sound plays a big role in vocal timing: waveforms allow the lining up of particular vocal features (e.g. transients) with the DAW's tempo grid, or with the same features in other vocal tracks (doubles, backing vocals). This functionality allows producers to manipulate vocal timing at hyper-detailed levels, sometimes without having to listen to the voice at all. Although fixing the timing of lead vocals was most frequently achieved in this manual way, some informants used *VocAlign* for time-aligning lead vocals with backing vocals, illustrating a further kind of visual representation of the voice in this production culture.[50] *VocAlign* works by analysing the peaks and troughs in the waveform of a guide audio (e.g. a lead vocal) and then using time-stretching algorithms to force an out-of-time 'dub audio' (e.g. a backing vocal) to match it (White, 2000) (Figure 31).



Figure 30: Complex manual timing of a very short vocal phrase against the DAWs tempo grid by Jason (2020J)

[49] For a richer contextualisation of the notion of timing as it pertains to other instruments and recorded music more generally, I refer the reader to work by Anne Danielsen (2019), Simon Zagorski-Thomas (2010), and the forthcoming monograph by Jonathan Sterne and Mara Mills. As of January 2021, this forthcoming monograph, supported by the Max Planck Institute for the History of Science, will be titled *Tuning Time: Sequences from the History of Time Stretching and Pitch Shifting.*

[50] Although my key informant Jason used the plug-in quite frequently, some other informants felt that the plug-in was too time consuming or too prone to error (2019CC, 2020K, 2020N).

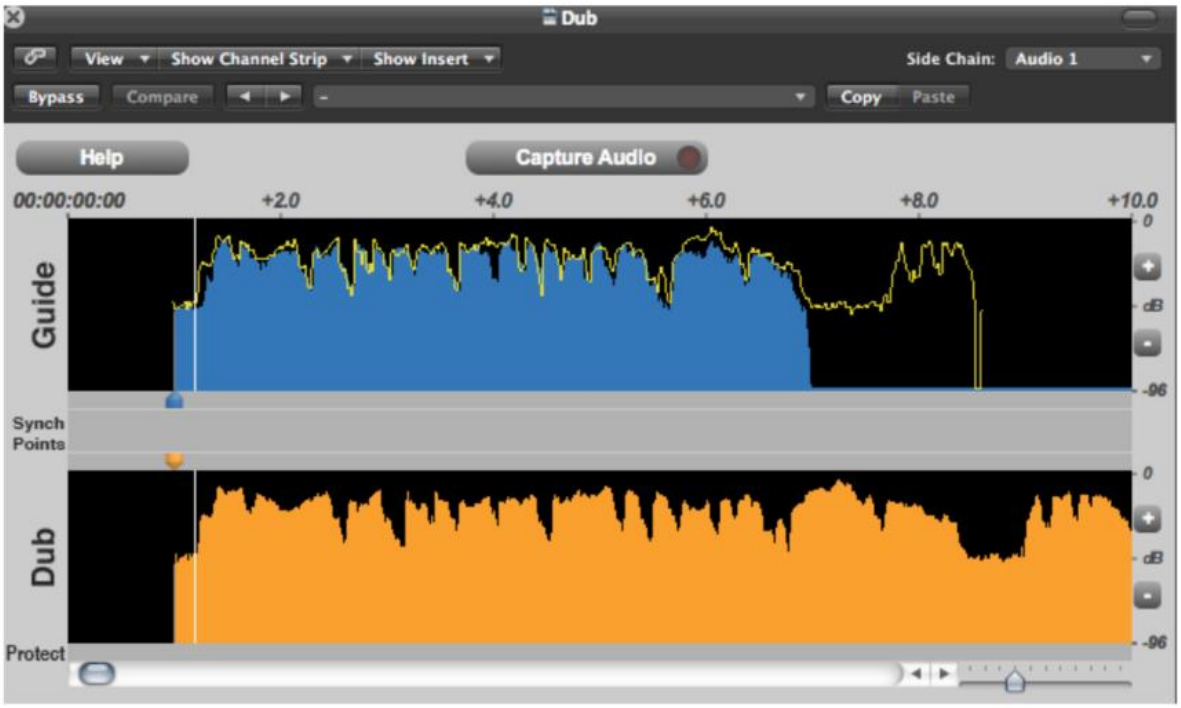


Figure 31: Using *VocAlign* to time-align one vocal track with another. The yellow trace over the guide vocal shows the proposed alignment (from Synchro Arts 2013: 12).

Manual gain adjustment ("clip-gain")

Another technique that several of my informants used—which took advantage of the visual objectification of the voice as a waveform—was manual gain adjustment, otherwise known as "clip-gain", which is a process similar to volume automation. As Peres describes, automation is a fundamental feature of all DAWs which allows the user to 'draw' in a curve across a track which, during playback, automatically alters a chosen parameter according to a specified unit (Peres 2016: 4-5). Although not quite the same as volume automation, "clip-gain" is a feature unique to the *Pro Tools* DAW which allows the user to draw curves across a track (consisting of lines and 'breakpoints') which directly control the track's gain levels (Avid 2018: 941)—in much simpler terms, the user can draw a fluctuating line across the track's waveform which manipulates its perceived loudness over time. As a user draws a clip-gain line on a vocal track, they can literally *see* the vocal waveform

transforming in real-time: perhaps a transient is reduced in size, or another portion of the waveform is expanded. As I will explain later in this chapter, such visual objectifications of the voice directly influenced what my informants' thought was most stylistic or appropriate in the commercial pop singing voice right down to the level of the relative loudness of individual syllables, consonants and vowels (which, in turn, encouraged them to make modifications to the voice visually rather than aurally based on these stylistic assumptions). I frequently observed "clip-gain" being used at an early stage in the commercial pop vocal editing process as a way of regulating or evening-out the gain levels of a vocal track. Most often, this process entailed breaking up the voice into constituent vowels and consonants which were either turned up or down. Figure 32 shows an example of "clip gain" on a vocal track in Pro Tools: you can see how, in the second image, the waveform is ever so slightly condensed according to the way the clip-gain line in the centre subtly rises and falls.

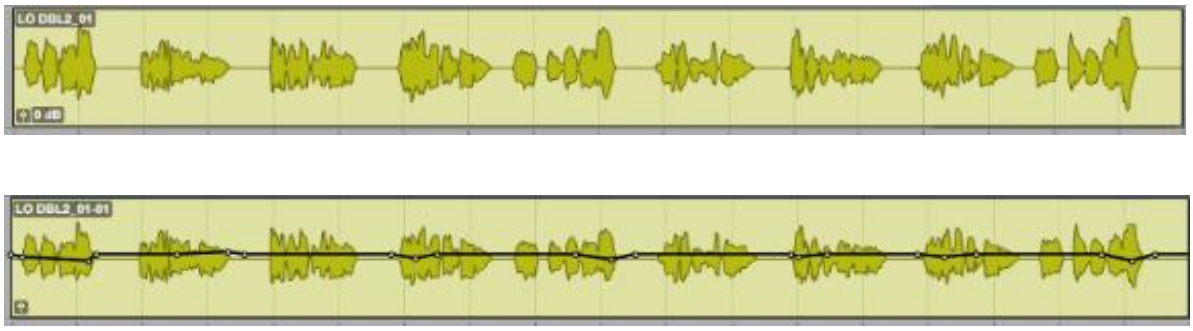


Figure 32: An example of a vocal phrase before and after Jason drew a clip-gain line to regulate the voice's loudness. (2020O).

The act of regulating or evening-out the dynamic range of a sound signal is usually associated with compression—indeed, compression is still a very important part of the commercial pop vocal editing process and will be discussed in much more detail in Chapter 5. However, the reason some of my informants opted to use clip gain *as well* as compression is because they believed it kept the dynamic range of a vocal in-check *before* all the other editing techniques were applied to the voice. This idea of heightened anticipation—a kind of obsessively preparing the voice for further processing—is another fetishistic quality of commercial pop vocal editing, and a theme I also interrogate more deeply

in Chapter 5. This is a compulsive tendency in which subtle, complex preparations for further vocal editing stages seem largely redundant or superfluous by the time later techniques which make the same modifications more extensively are employed. Arguably, the use of these intricate clip-gaining processes also suggests that the visual objectification of the voice has a role to play in driving these compulsive iterations which characterise the vocal editing process, because it visually highlights what is perceived as an inadequacy in the voice relative to the aesthetic aims of compression (too many peaks and troughs in the vocal waveform) at the earliest stages of vocal editing, which prompts the usage of clip-gain to visually anticipate the effects of compression.

Audio restoration tools

Occasionally, difficulties or mistakes in the vocal recording process resulted in unwanted sonic artefacts in the voice which my informants removed at the start of vocal editing with the help of dedicated audio restoration plug-ins. During my fieldwork, the market leader in this kind of software was iZotope, whose suite of audio restoration plug-ins *iZotope RX* was used and mentioned by multiple informants.[51] Such plug-ins do not perform such a consistent role in creating the commercial pop voice in comparison to, say, tuning software. They are worth noting briefly, however, for the way they visually objectify vocal timbre: Théberge (2013), for example, highlighted the remarkable way such plug-ins permit the direct modification of visual representations of audio spectra, using similar tools that exist in photo editing software like lassos and paintbrushes to 'edit' the spectrogram (ibid.: 12).[52] This is a process of visually rubbing out or scrubbing away unwanted tarnishes on the voice, visually objectified as a spectrogram. Figure 33, for example, shows one of my informants removing some unwanted "harshness" in this manner. These techniques were only used when something was felt to have gone 'wrong' during the recording process. For example, some of the reasons my

[51] I do not provide any fieldwork codes here because my informants' usage of the iZotope plug-ins is covered in a later section about gender in which I choose to anonymize the informants completely.

[52] In this paper Théberge is referring to the release of an earlier plug-in by iZotope—*IRIS*—which is in fact a synthesizer which makes use of similar spectral representation tools (Nagle 2015). The first RX plug-in suite was released in 2008 and contains a 'spectral repair' mode with a very similar interface, although it is designed for audio restoration rather than sound synthesis (Thornton 2008). *RX-6*, a radically updated version of this suite, was released in 2017 (Robjohns 2017a). Théberge's arguments are still very relevant here as it is clear both plug-ins visualize sound in similar ways even though they are for differing creative purposes.

informants decided to use audio restoration at the start of vocal editing included the gain being turned up too high during recording, resulting in clipping, and the choice of microphone resulting in an unwanted “resonance” that would complicate any further processing.

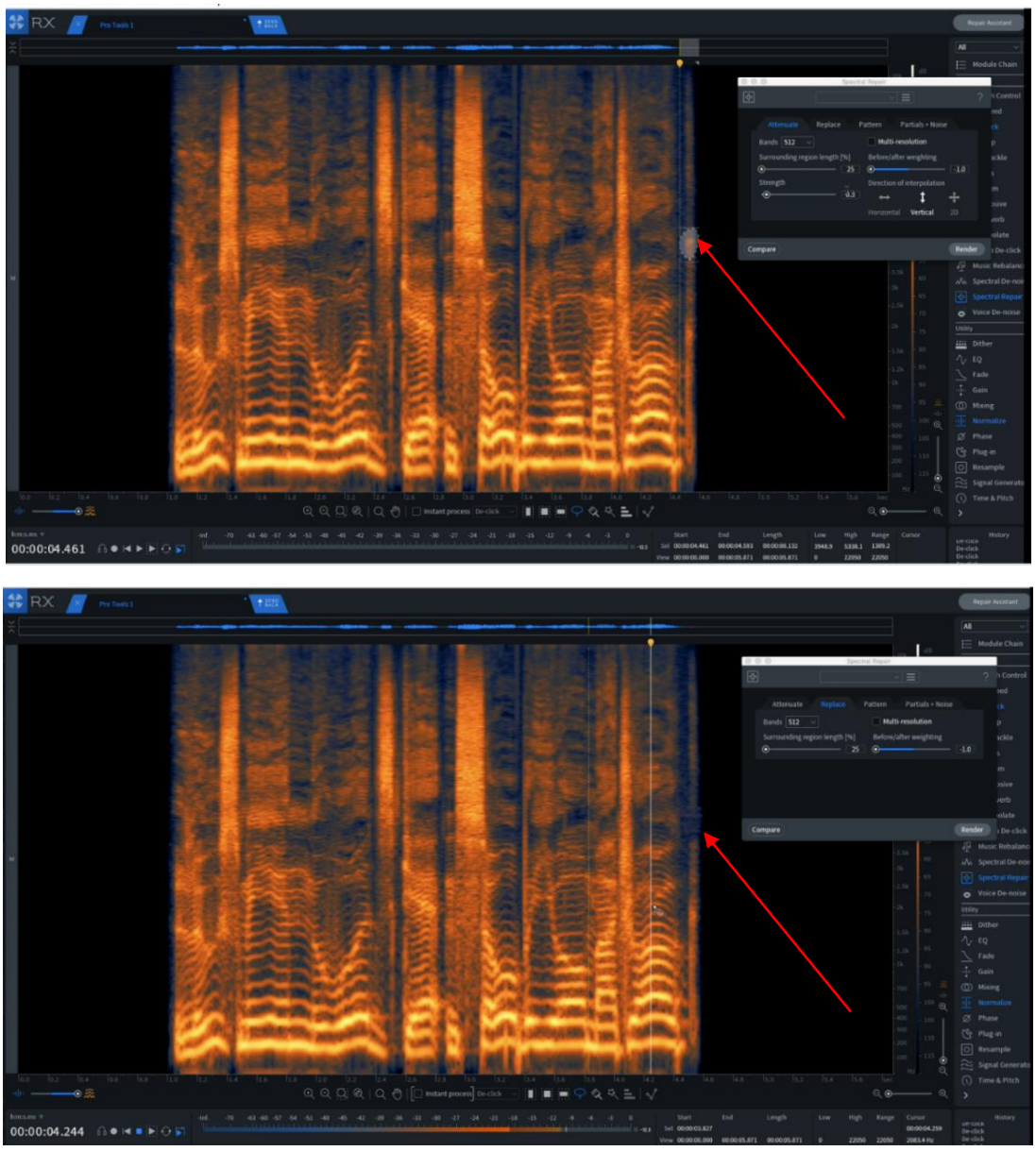


Figure 33: An informant using the ‘spectral repair’ function in iZotope *RX* to remove “harshness” – before and after. You can see how the tiny spot of light on the spectrogram is erased using the selection tool. (anonymized)

To summarize, this early stage of vocal editing—which included predominantly comping and tuning but also related techniques intended to refine the voice and prepare it for further processing—entailed working digitally with the voice represented in a variety of visual forms. Waveforms and linear

contours dominate these visual representations, and they function as visual silhouettes of the voice's complex pitch and timbre fluctuations. *AutoTune* juxtaposes both these forms of visual rendering, and *Melodyne*'s blob-like structures provide a more suggestive illustration of vocal pitch and timbre intertwined. Both vocal timing and manual gain adjustment ("clip-gain") harness the visual capacities of the DAW (which presents track waveforms in a layered fashion) to stack and synchronize vocal phrases and to regiment the undulations of vocal waveforms. The spectrographic representation of the voice in iZotope's *RX* permits the visual erasure of vocal deficiencies. All these depictions present the recorded voice to producer-songwriters in a visually objectified form, collapsing the complex sonic experience of listening to the voice, along with the skilled performance of a vocalist, into digital entities on a screen which can be manipulated at will. As I will demonstrate in the following analyses, such visual objectifications of the voice directly inform producer-songwriters' vocal editing techniques by encouraging them to work with the voice visually rather than aurally, as well as to construct over-arching theories about how the voice sounds in the commercial pop style based on this visual way of editing the voice.

3.2. Visual objectifications of the voice in comping, tuning and related techniques in commercial pop vocal editing

Visual vocal editing and its effect on interpretations of the commercial pop vocal style

The edits my informants made during the tuning and comping stage were based on habitual, repeated work with the voice represented in these visual forms. I observed how such visual representations of the voice in software not only facilitated their tuning and comping work at increasingly high levels of detail, but also encouraged a visual rather than aural approach to vocal editing, which informed what they thought were the most important aesthetic features of the commercial pop singing voice more broadly. Alongside illustrating these phenomena, this section implies that, through their feed-back onto my informants' conceptualizations of the commercial pop vocal style, such visual objectifications were therefore having a very strong incidental impact on the sound of the commercial pop voices these informants edited. Put another way, I suggest that through the mediating effect of these visual representations on my informants' interpretations of the commercial pop voice, the visual

objectification of the voice informs certain resulting features of the commercial pop vocal sound. To show this in more detail, I focus on data from Jason and Chris. Jason's tuning techniques showed how the visual representations of the voice in *AutoTune Pro* had led him to develop a theory that "sliding" vocal pitch contours were particularly "stylistic" for the commercial pop genre, and his work with vocal waveforms in the Pro Tools DAW during comping had also informed his ideas about appropriately stylistic vocal diction and pronunciation, relating to the loudness of individual vocal consonants and vowels. In addition, Chris's vocal editing was stimulated by the layered waveform representations of the voice during comping in the DAW: he focused on surrounding the lead vocal with several other intricate vocal layers, building up an overall vocal "arrangement" to create a "big" vocal sound which reinforced the voice's foregrounded position within the track as a whole. These sonic features—"sliding" or "gliding" vocal melisma, consistent diction, and thick lead vocal arrangements—are common across the commercial pop genre more broadly, which suggests that the influencing role of visual objectifications of the voice during this stage of vocal editing is powerful and widespread.

During interviews and observation, Jason repeatedly claimed that the commercial pop singing voice tended to "slide" around pitches rather than hitting them exactly, and described how his tuning approach intended to preserve and emphasize this "stylistic" feature of commercial pop singing. Importantly, this sensitivity to vocal pitch contours was a skill Jason had acquired over several years of recording commercial pop singers and using the visual representation of the voice in *AutoTune Pro*'s graphical mode to tune them. He began one tuning session,[53] for example, by explaining how his main aim was to preserve the "original" pitch-contours of the voice in *AutoTune* as far as possible (2020N). Jason characterized these pitch-contours by contrasting commercial pop singing styles with musical theatre: he described how one of the defining, and most effective, characteristics of "pop singers" was the way they frequently did "quick slides" up to notes, in contrast to musical theatre singers who, and I quote "always try to hit the note bang on, which sounds really… proper… and very

[53] To clarify, I have not been given permission by Jason to include the tracks from these other sessions with the thesis, so I only describe them and include screenshots from his process. However, the voices in these tracks sounded very similar to the example given in Audio 1.

boring, I think." (ibid.) Jason demonstrated this by using *AutoTune Pro* to draw in a musical-theatre-style pitch-contour, as an example of what commercial pop singers *didn't* do. Figure 34 shows this modification: the pitch contour of the original vocal phrase is represented by the red line and the "musical-theatre" modification made by Jason is the green line, with the change labelled in red (alongside a transcription in Western staff notation and overlaid with the lyrics, to give a clearer idea of how this phrase sounds). This figure demonstrates how Jason has removed the incline in the pitch-contour at the start of the word 'got', instead forcing the voice to reach the note A as fast as possible. In the original curve, the voice does not hit the A right away, instead "sliding" upwards to reach it about halfway through (to use Jason's terminology). This exchange revealed how *AutoTune Pro*'s visual objectification of the voice had directly informed Jason's ideas about vocal style, even leading him to develop a theory that voices tend to 'look' a certain way in the commercial pop genre in comparison to musical theatre—after all, "sliding" and "straightness" are clearly very visual metaphors. What's more, when I suggested this to Jason in a later interview, he believed that his repeated use of the software over time was likely to have influenced his theorization of the voice in this way (2021D).

Shortly afterwards, Jason went on to show how this phrase *ought* to be tuned for commercial pop purposes: by exaggerating rather than omitting the voice's "slides", as well as making the starts and ends of the "slides" conform more rigidly to the notes of the keyboard-roll on the left-hand side of the interface. Figure 35 demonstrates this process: you can see how, for the three syllables 'got' '-no' and 'si', all of which come to rest on an A, Jason exaggerated the "slide" upwards at the start of each note, and made each slide begin more clearly on a G (this is most visible on the '-no' of 'another'). The accompanying transcription in Western staff notation illustrates this in further detail. Jason conceived this as an act of preserving the singer's voice whilst also enhancing it to its fullest potential, claiming that "all you're trying to do is make the thing that you sang as good as it can be, but still be how you sang it" (2020N). Whilst tuning this phrase, he also revealed another totalizing conception about commercial pop vocal styles, hypothesizing that individual singers often had a "consistent size of slide" throughout their performances (ibid.). Further, on another occasion, Jason described how if a

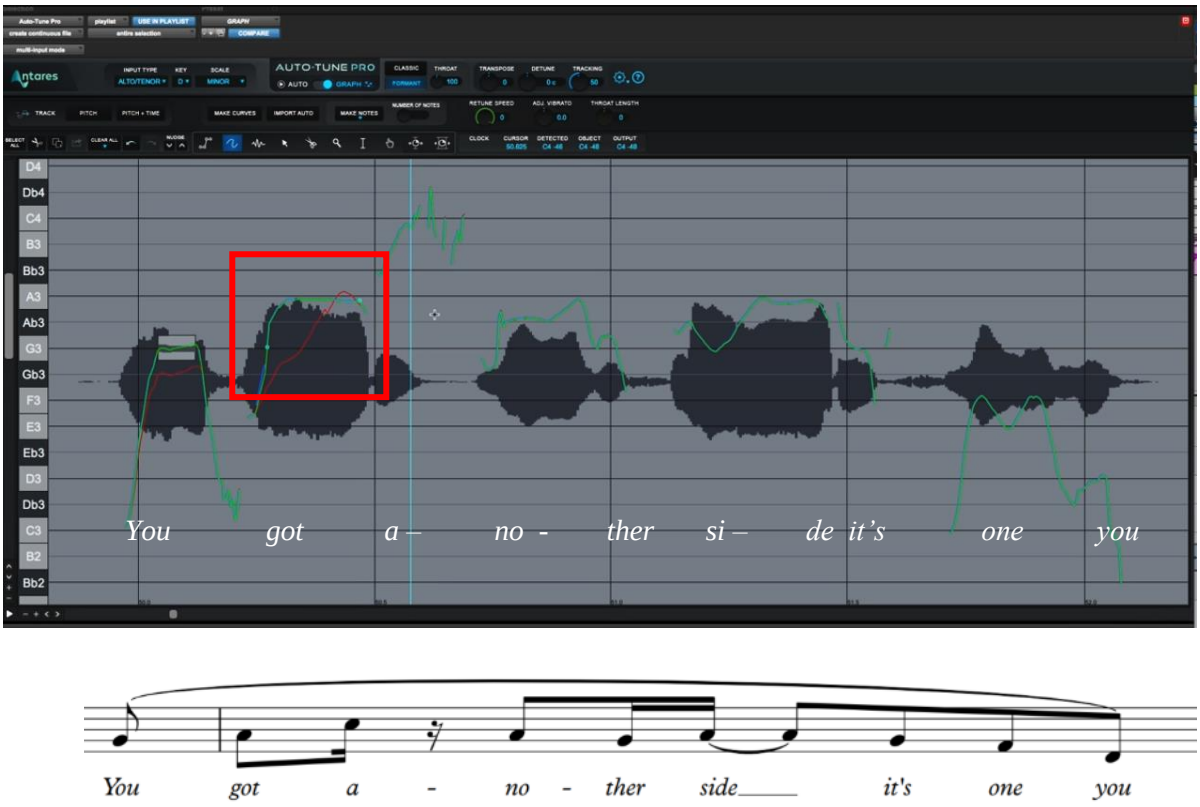


Figure 34: Musical theatre "straightness" labelled in red

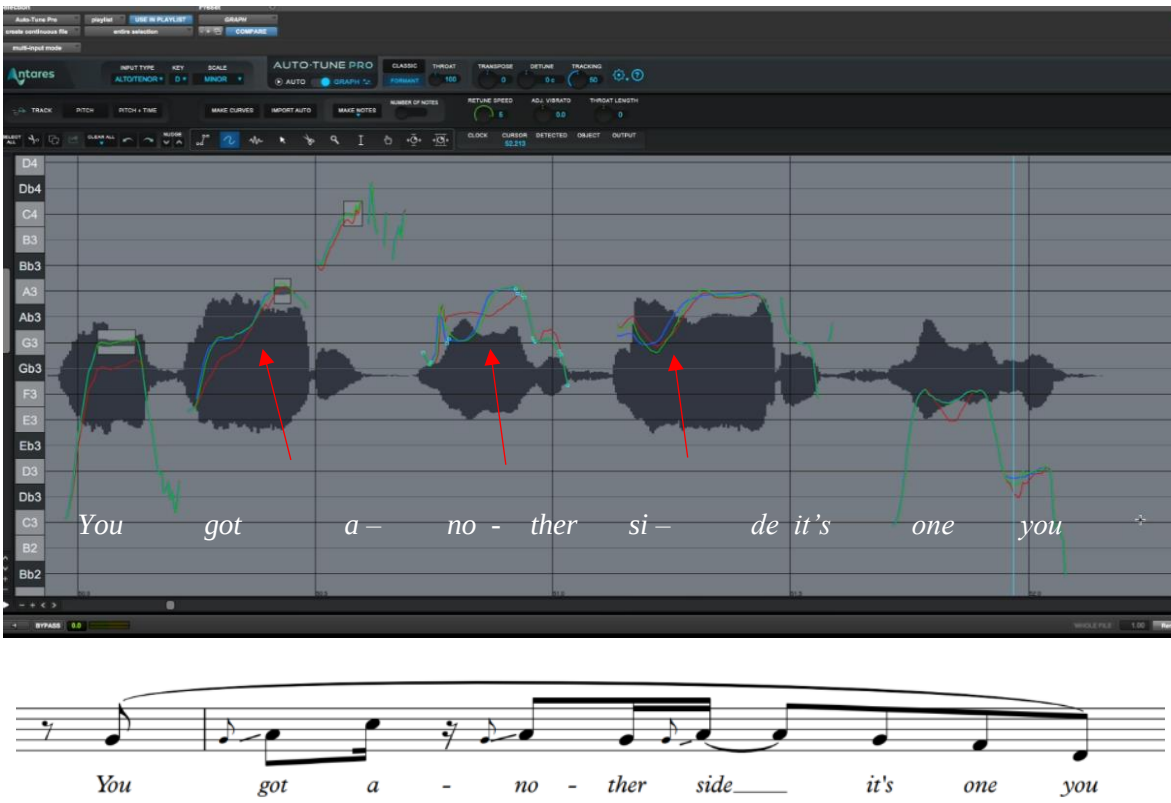


Figure 35: Slides labelled with red arrows.

vocal recording lacked a "slide" where it might be appropriate to have one, he would frequently "add one in afterwards" (2019Z). Clearly, Jason believed that the "sliding" voice was *so* stylistically appropriate and successful in this genre that he even used *AutoTune* to add "slides" when they weren't originally present. An example of one of these slides can be heard in Audio track 1 during the chorus at 0'48": you can hear how the vocalist slides on the word 'let' of the lyric 'let me give you all I got'.[54] Mainly, however this shows how Jason's ideas about "sliding" vocals—along with the way these ideas impacted the sound of his music via his tuning techniques—were reinforced or even catalysed by the unique visual affordances of *AutoTune Pro*'s graphical interface, and the way they allowed Jason to analyse and modify the pitch-contours of the voice at such a high level of detail. Indeed, Jason's implication in the later interview was that these ideas about voices "sliding" formed a unique underlying theory of his vocal editing practice that he had personally spent time curating through repeatedly looking at the voice in this way (2021D). In other words, he seemed to be trying to characterize the commercial pop vocal style so universally, and with such intricacy, in a way which relied entirely on a visually metaphorical understanding of the voice as a series of "sliding" curves.

Jason's interpretation of stylistic diction and pronunciation in commercial pop singing was also driven by waveform visualizations of the voice in comping and clip-gain processes. For example, Jason described how he would usually try to "turn up" and "turn down" features of the voice using clip gain, with the aim of getting a "consistently sized waveform" (2020N). In a different mixing session I observed, Jason discussed how he usually "turned up" some consonants in the voice more than others—including plosive consonants like Bs, Ds and Ps to make them sound "punchier", as well as Ms and Ns due to their quietness, "because [the shape of] your mouth sort of filters the sound" (2021B). Figure 36 demonstrates the waveform of this phrase before and after clip gaining, with the song's lyrics represented above: targeted consonants (Ps, M, Ls and Y) were turned up using the clip gain line, which made the corresponding points on the waveform expand, and the 'x' of 'fix' was turned down as Jason felt the sibilance was too loud. Jason's logic behind this process was that

[54] Unfortunately, I was unable to collect screenshots of Jason tuning this phrase as he forgot to video it, however comparing with this example showed how it was clearly a similar kind of sound.

making the gain of the vocal more consistent at this stage helped it to "cut through the mix" even more by the time it was compressed, even though at this stage the effects of the clip gain process could barely be perceived. Indeed, throughout our observation sessions, Jason would often do this kind of clip-gaining automatically, just by looking at the shape of the waveform and not listening to the voice at all. Jason's clip-gaining process therefore showed how his familiarity with these visual representations of the voice had, over time, led him to perform such clip-gain modifications visually, sometimes without even listening to the precise vocal features he ended up editing. This visual mode of vocal editing had also informed his tendency to generalize about commercial pop vocal diction in these terms, suggesting that certain consonants ought to be "turned up" more than others. More generally, the objectification of the voice as a waveform adjusted by a clip-gain line had enabled and encouraged him to sculpt vocal tracks at an incredibly high level of detail, right down to the nuances of individual consonants.

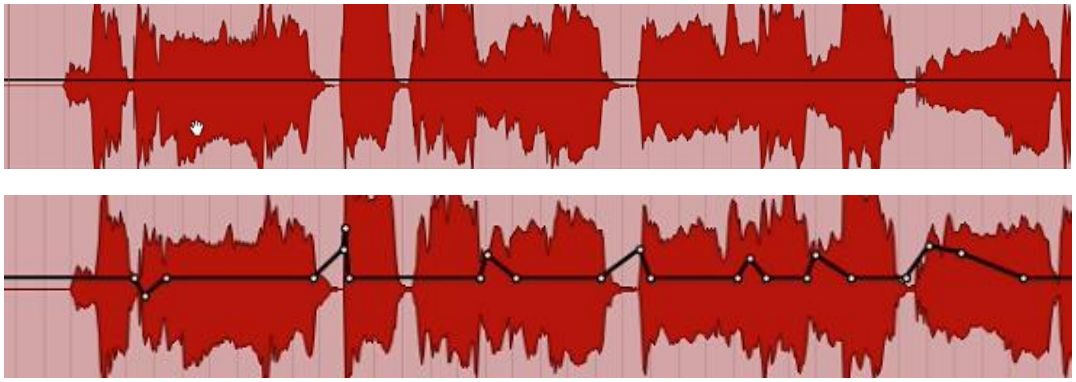


Figure 36: Before and after the clip gaining process. Consonants where clip gain modification has been applied are highlighted in red. Lyrics are 'Fix me up or cut me open, a love like you it's…' (2021B)

Similarly, Chris's comping processes were concerned with conceptualizing the lead vocal as one voice within several vocal layers: a theorization of the commercial pop vocal style which depended on the various layered and grouped waveform visualizations of the voice he worked with in the *Logic Pro X* DAW. For example, in one track I observed him producing (Audio 2), Chris's approach to polishing the voice revolved around fitting the lead vocal into a dense vocal "arrangement". Throughout interviews and observation, he repeatedly stressed his desire for "size", "energy" and "thickness" in the vocal sound (2020F, 2020H, 2020T)—which he conceived holistically as including not only the lead vocal but the large quantity of double, harmony and counter-melody vocal tracks which supported it (otherwise known collectively as backing vocals or "BVs"). This was a tactic he used for reinforcing the lead vocal's foregrounding within the mix. Chris's approach to comping the vocal was very much rooted in these ideas about layering and building up a "big" vocal arrangement, which contrasted with Jason's focus on refining miniscule features of the lead vocal. However these conceptualizations were also rooted firmly in visual metaphors, relating to the way the *Logic Pro X* DAW layers vocal tracks on top of one another, and permits highly detailed manipulation of vocal waveforms during comping including contoured fading and track grouping.

For example, when comping, Chris described how he wanted to "get a bit more creative" with the sound by taking a precise approach to editing the gaps between vocal phrases. The vocal arrangement at this point consisted of a lead vocal, 6 doubles in the same octave and 2 in the lower octave, and 1-2 harmonies. In the process of comping this arrangement, Chris painstakingly cut out the portions of audio in between vocal phrases and faded the beginnings and ends of each group of BVs. This fading was unique, however, because it was specifically tailored to the melodic content of individual vocal phrases. For phrases in the chorus that ended with a descending melodic contour, Chris applied long curved fades, and the phrases with an ascending contour instead ended with a short diagonal fade (Figure 37). Chris said that this was intended to "match the natural inflection of the voice". This is because the hook of the song's chorus—'fall in love with someone new'—falls on phrases 1 and 3, and the curved fade made the end of the hook sound louder for longer, whereas the second phrase was made to sound slightly more abrupt in contrast. In Chris's view, the effect of this

precise work was that the BVs followed the melodic, rhythmic and timbral contours of the lead vocal as precisely as possible, and he described this as a way of "maintaining the natural breathing" of the voice, conceived as a broader arrangement of vocal tracks. Listening to Audio Track 2, this feature cannot easily be perceived—it is very subtle—however this process showed how Chris conceptualized the early stages of vocal editing in terms of an arrangement of vocal layers and groups during comping, showing how his editing was led by a visual rather than aural sense of how the voice ought to sound. By working with the voice in a visually objectified form, he was able to build up dense vocal layers and structure large groups of vocal tracks.

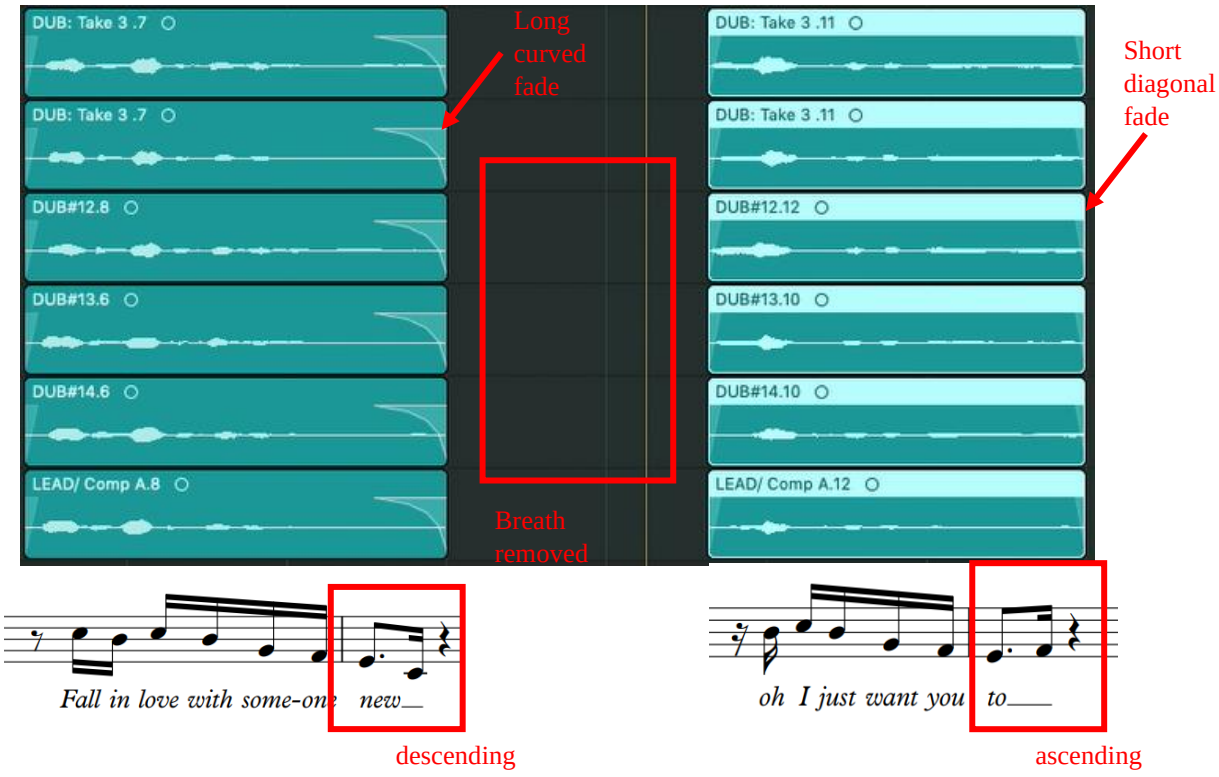


Figure 37: Chris's detailed approach to fading in the chorus of Audio 2

Despite his idea of "maintaining natural breathing", in the same set of phrases Chris also completely removed breaths to generate lyrical impact, in a way which was similarly informed by the visual layout of the vocal arrangement within the DAW. When timing the lead vocal, Chris decided to cut out every breath which followed the first statement of the hook, 'fall in love with someone new' (Figure 37). This was because: "By omitting the breath, it kind of gives the feeling that the phrase is definitive. Like it's an order." (2020H). According to Chris, omitting the breath provided a sudden contrast to previous phrases, and he argued that this marked the phrase as important and imperative. Clearly, the stark contrast Chris was getting at here was created by the sudden change in texture across the vocal arrangement: from several vocal doubles stacked beneath the lead vocal, to a complete absence of ambient vocal sound in-between vocal phrases. The starkness of this contrast was made even more apparent by the visual representation of the vocal arrangement in the DAW at this point, which was partitioned into neat groups representing vocal phrases, with nothing in between. This indicated that Chris's interpretation of contrast within the vocal arrangement was largely guided by the DAWs capabilities to arrange complex layers of vocal parts. It also implied that Chris's ideals of maintaining natural "breathing" and "inflection" described above didn't necessarily refer to physical aspects of singing. Instead, they were more metaphorical and, related to the visual layout of the vocal arrangement within the DAW, where the BVs and lead functioned together as an integrated entity, in which "breathing" (or the absence thereof) was synchronized across all vocal parts. This was further facilitated by the way the DAW allowed Chris to group vocal tracks together and edit them as units—which he referred to as 'GANG' vocals (Figure 38). In contrast to Jason, Chris's comping and vocal arrangement practices therefore seemed slightly differently guided by the visual objectification of multiple vocal parts as a series of layered waveforms. Chris aimed to cumulatively arrange these visual representations of the voice to give the impression of a "big" vocal sound. Indeed, similar to Jason's "sliding" vocals theory, Chris invoked an over-arching conception of the lead vocal being composed of "layers" as a guiding visual metaphor of his commercial pop vocal editing:

> A big thing for me is the layers. For a lot of projects, half of the tracks, like 70 odd tracks, will be the voice. Loads of layers. If you're looking at a chorus, for example, every voice will have like two double-tracks [panned left and right], and there will maybe be little stabs of harmony on certain words

> so every harmony we do we'll double it again. I think that degree of layering is how I get that poppy sound. (Chris, 2020F)

In summary, both examples illustrate how the visual objectification of the voice in software strongly informs commercial pop vocal editing practices and has resulting effects on the voices edited by these producer-songwriters. Over time, by working repeatedly with the voice in these visually objectified forms, producer-songwriters have come to rely on these visualizations as guides to how to achieve the sonic effects they are seeking in this genre. Occasionally, these visualizations were modified without any reference to the sonic effects such modifications created, indicating that a visual approach to vocal editing may be increasingly common in this practice. More broadly, these examples show how visual objectifications of the voice in comping and tuning and related technologies influence some of the key aesthetic characteristics of the commercial pop lead vocal style (outlined in the introduction to this thesis, pp. 27-28), through the ways they inform producer-songwriters' interpretation of what sounds most aesthetically appropriate or successful in the genre, including: the specifically melismatic quality of the voice, which here is achieved through an emphasis on the "sliding" characteristics of vocal pitch contours within individual words; its consistent loudness, which here is achieved through a selective approach to amplifying and quietening individual consonants and vowels within words; and its dominance within the track as a whole, which here is achieved through the careful alignment and "building up" of vocal doubles and harmonies. Despite the slightly different foci of the examples, contrasting Chris's track (Audio 2) to one of Jason's tracks (Audio 1) shows how the resulting sounds of the voice in Jason and Chris's released tracks are quite similar. (The voice in Chris's track perhaps does sound slightly more expansive and choir-like, supported by many vocal layers, especially in the later iterations of the chorus (e.g. 1'50"), but Jason's track has a similarly developed vocal arrangement as well as several vocal 'chops' and 'hooks' in the chorus which use formant modifications to achieve a robot-like sound.[55])

[55] The differing visual objectifications each informant focused on in the examples resulted partly from their different industry contexts: Jason focused more on vocal minutiae in these instances because he was usually doing the initial round of vocal "tidying-up" for the track or demo, which would eventually be bought by a DJ/producer or artist who would take Jason's work and develop it before it was released. In the case of Audio 1, this meant that a lot of the decisions about how to develop and build up the vocal arrangement were taken by the other DJ/producer, rather than Jason. In comparison, Chris's editing processes for Audio 2 were happening

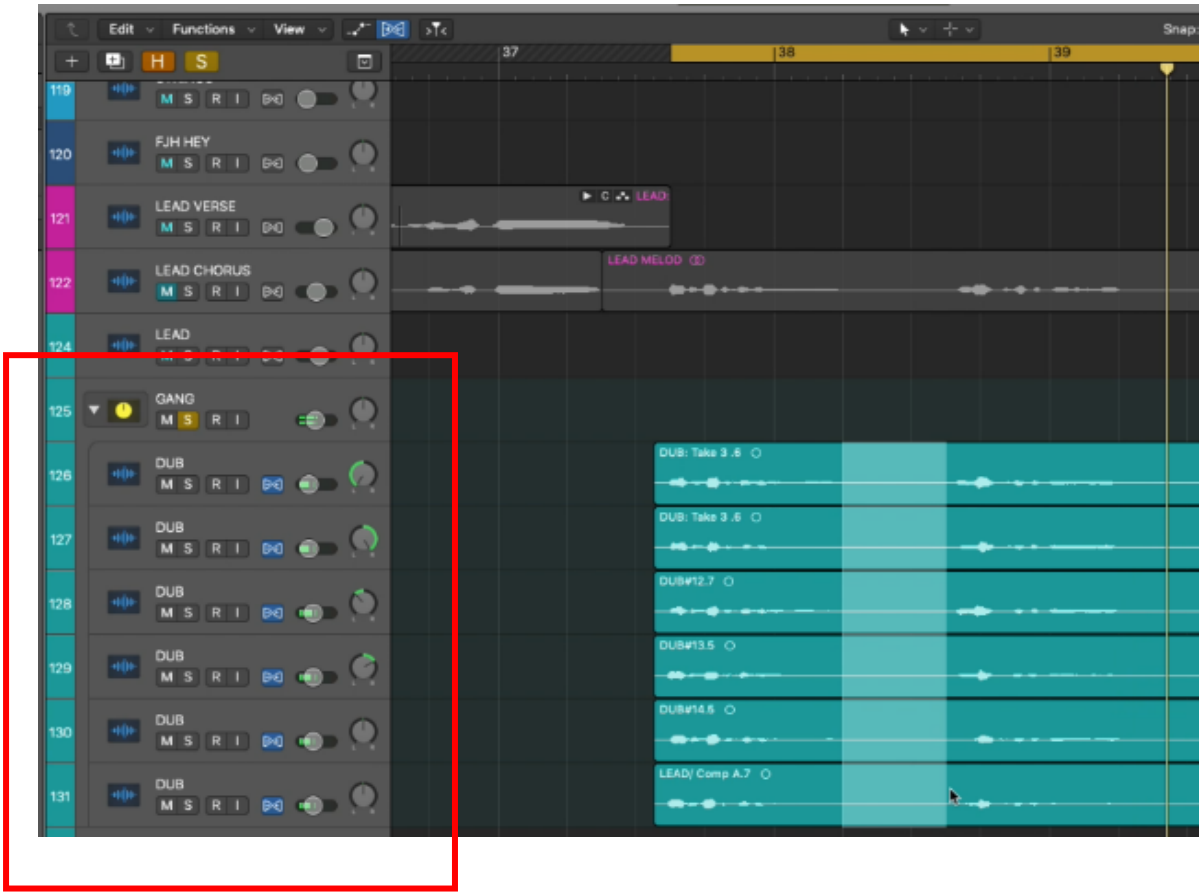


Figure 38: 'GANG' vocals (2020H).

The visual objectification of vocal "emotion"

Like Provenzano (2019a: 24) and Marshall (2017: 27-48, 96-100), I encountered numerous examples of producer-songwriters directly attempting to cultivate vocal "emotion" during vocal comping and tuning. Marshall, for example, provides an account of the value of the term "emotion" in the tuning practices of North American studio engineers, discussing both the importance engineers place on the "emotional" connections they nurture with their vocalist clients (2017: 43-45), as well as the steps they take to preserve the sound of vocal "emotion" in the tuning process. At one point, Marshall uses specific quotes to describe how, when using the correct pitch function in *Melodyne,* one producer associated the term "emotion" with the preservation of vocal imperfections and continuous pitch

closer to the track's release, and so were more oriented towards the positioning of the voice within the track as a whole.

contours (ibid.: 96-100). In my ethnography, however, there were also several examples of producer-songwriters attempting to *add in* emotion to the vocal performance if it wasn't already present. Most importantly, these additions of vocal "emotion" were understood in terms of specific characteristics of the visual objectifications of the voice provided by comping and tuning technologies. In turn, vocal "emotion" became a subsidiary commodified feature of the already commodified voice, achieved throughout the vocal editing process by making subtle changes to the visual shapes of the voice rather than being linked to the labour of vocal performance.

One example was producer-songwriter Robin, who, in a track which used slightly more overt non-naturalistic tuning in comparison to Jason and Chris's tracks, defined "emotion" in a voice as something which was disrupted by the "tell-tale robotic pitchshifting" sound of automatic tuning (2020R). This was a slightly anomalous example since the vocal editing in this track was not intended to sound as subtle as Jason, Chris and Oliver's tracks, however it provides a good framing for discussing which visual features of the voice producer-songwriters felt indicated "emotion" in this genre. Robin was attempting to emulate the sound of *AutoTune* in automatic mode at a retune speed of zero, but he achieved this effect manually using *Melodyne*. Robin's *Melodyne* edit window for the vocal phrase in Audio 4 is demonstrated in Figure 39: you can see how each orange 'blob' stays rigidly within its 'lane' in the piano-roll interface. Alongside this, Robin highlighted moments in his tuning process where he "manually adjusted the *glide* between notes to get rid of the tell-tale robotic pitchshifting, just to *emphasise the emotion* slightly." (ibid.: italics added) These points are represented by the red arrows in Figure 39, drawn by Robin. Although not precisely visible in the diagram, at both points Robin had adjusted these pitch-contour lines to make the transition between two notes smoother. Robin never explained exactly why he chose to emphasize the "emotion" at these points in the phrase—and, indeed, listening to this example in Audio 4 it is very difficult to hear any audible "emotion" at these points at all. However, the most notable thing about his approach is the way he defines emotion so clearly, as indicated by more "gliding" vocal pitch contours rather than discrete notes.

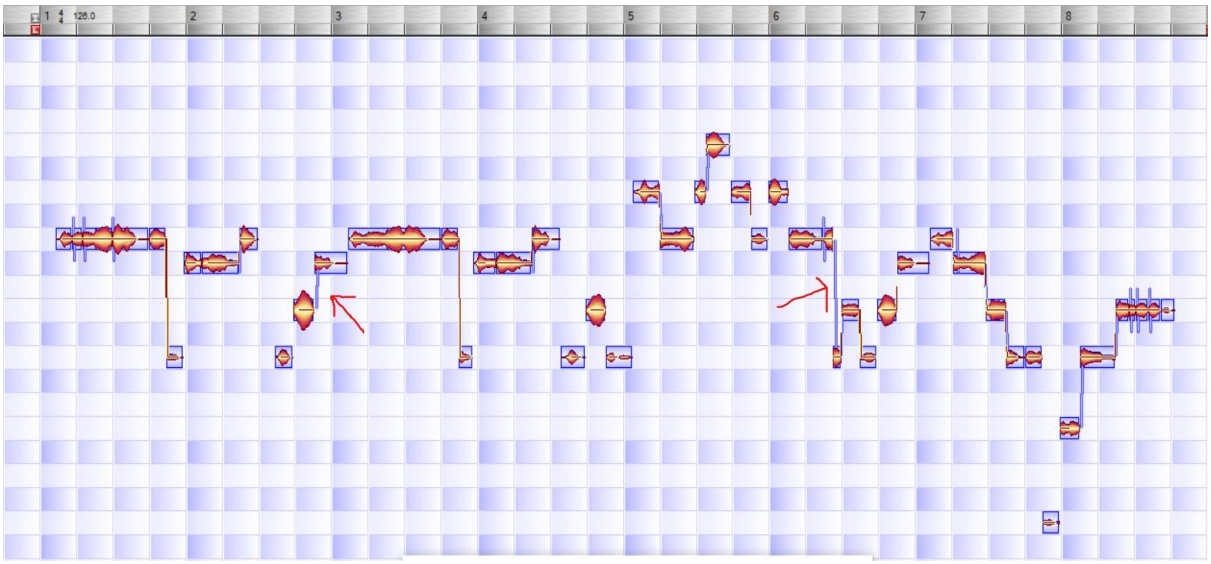


Figure 39: Robin's *Melodyne* interface, with his arrows pointing to his emphasis of "emotion".

The similarity between Robin's "gliding" and Jason's "sliding" vocals is not insignificant: both producers clearly feel that a voice which "slides" around pitches rather than rigidly adhering to a pitch-grid indicates a more "emotional" vocal sound, and this interpretation is couched in the terms of tuning software's visualization of the voice. For Jason these characteristics were described as "style", as well as "emotion" whereas for Robin it was just "emotion".

> Yeah I think, emotion and style… are linked to the way notes slide around as well… and the way that the volumes are edited [using clip gain]. So you can add a bit of emotion to things just with editing. (Jason, 2021D)

Despite the similarities in their discussions of the term above, Jason also related vocal "emotion" to specific comping and timing techniques, as well as the "sliding" of the voice during tuning. For example, he suggested that making the time to comp lead vocals at a very high level of detail could "add that extra 5% of emotion to the vocal":

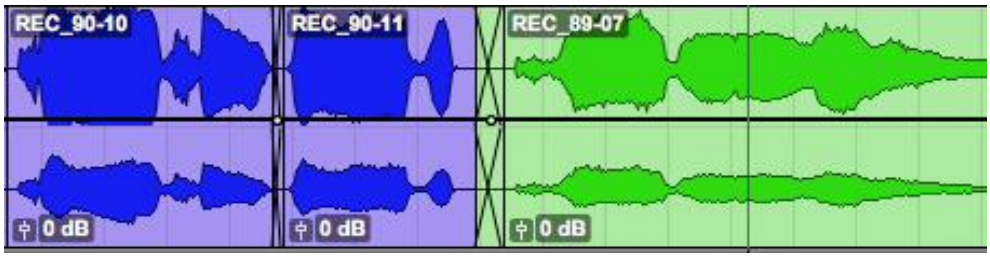
> It all depends really how consistent the vocals are. If they're exactly the same between takes, you can just use long sections, but if they're really different, then it's really worth separating out each word and picking the absolute best one, 'cause it just adds that extra like 5% of emotion to the vocal. (Jason, 2020N)

This notion of incredibly intricate comping—right down to the level of individual vocal syllables—making a performance sound *more* emotional is significant because it implies that vocal "emotion" is

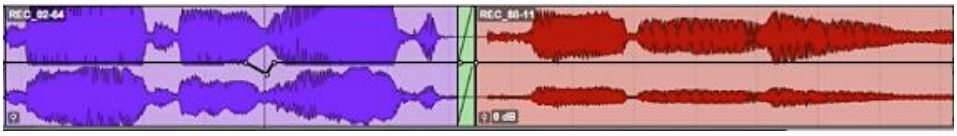
a feature that can be achieved and enhanced via the incredibly detailed manipulation of vocal waveforms. Furthermore, it implies that whatever "emotion" was present in the original performance is insufficient and needs to be enhanced according to a specific template of "emotion" which is derived from the representation of the voice in visual form.

Additionally, when editing Audio 1, Jason tried to enhance "emotion" in the voice by trying to fabricate a greater sense of "laid-back" pronunciation in one of the lead vocal phrases ('Don't you ever cut and run', Audio 1, 2'13"), using comping and clip gain (2020M). This involved painstakingly listening through to small portions of the phrase in each recorded take, and isolating the portions which he felt contained less prominent consonant sounds—which were smoother, lazier and more "legato" (in his words) (ibid.). Jason felt that this had consequences for the intelligibility of the words in the phrase, however—he wanted to hear a more definitive gap between the lyrics 'you' and 'ever', and to try and fabricate this while still maintaining the "laid-back", "legato" feel, he used clip gain to create a very brief drop in volume in between the two words. This comping process is shown in Figure 40. Again, this is an intensely detailed way of working with the voice which can barely be perceived in the final version of the track, however it shows how Jason's interpretation of vocal "emotion" depended on the ability to zoom-in incredibly closely on vocal waveforms, to the point where Jason felt able to modify detailed aspects of the vocalist's pronunciation. In comparison to Robin, this is a different definition of the term "emotion", referring more closely to Jason's ideas about stylistic diction and pronunciation which derive from the objectification of the voice as a waveform during comping, as opposed to pitch-contour visualizations.

Before:



After:



Don't you e - ver cut and ru - n

Figure 40: Comp for phrase 'Don't you ever cut and run', before and after Jason's edit (the second is zoomed out further than the first, so I've lined up the waveforms) (2020M).

Such interpretations of vocal "emotion" show how it is discussed and manipulated by these producer-songwriters solely in terms of the way the voice is visually objectified in tuning and comping software. Whether referring to the voice's "gliding" between notes, the level of detail with which the voice is comped, or the exaggeration of "laid-back" pronunciation, "emotion" becomes a tokenistic vocal quality which is sought throughout vocal editing by visual means. "Emotion" is also fetishized like a commodity in these cases, since its visual form in software takes precedence over any considerations of the vocalist's labour of generating vocal "emotion" in performance. There is also a controlling quality to these techniques, as the producer-songwriters take the labour of generating vocal "emotion"—a type of labour which seems more suited to the vocalist—into their own hands. I now turn to the controlling aspects of vocal tuning and comping and their intersection with the visual objectification of the voice in more detail, additionally considering their ramifications within the gendered nature of collaborative songwriting projects.

Visual objectifications of the voice and the gendered dialogues between producer-songwriter and vocalist

One of the central themes of recent scholarship on pitch-correction is the often-gendered power dynamic between a producer or engineer who edits the voice and the vocalist who provides a vocal performance to be edited. Provenzano argues that, in North American contemporary pop production contexts, this relationship is overwhelmingly gendered and problematically hierarchical, as the female vocalist's labour is de-valued by the male producer or engineer who takes the process into his own hands and puts his aesthetic decisions first (2019a, 317-339; 2019b). In this section I use anonymized case studies[56] from my fieldwork to show how visual objectifications of the voice in tuning and comping technologies inflect and sometimes intensify the gendered power dynamic of these supposedly collaborative dialogues. More specifically: on the one hand, the transformational powers of tuning and comping seemed to entail that vocalists were more active in the editing process than Provenzano suggests, providing feedback on producer-songwriters' work and requesting that their performances be edited in certain ways which took advantage of the visual affordances of such software. However, at the same time, such feedback was often disregarded by producer-songwriters, who were sceptical of vocalists' knowledge about the tuning and comping technologies in question. More importantly, producer-songwriters' access to visual representations of the voice during vocal editing work also prompted them to pick out what they perceived to be "errors" in the vocalist's performance which they would then correct, overriding the vocalist's performative authority—these were errors which they would be unlikely to pick up on were it not for the way the voice's visual objectification enabled them to edit it at microscopic levels of detail. This section also paints a picture of how the visual objectification of the voice blends characteristics of both vocal commodity fetishism and erotic voice fetishism: as the previous sections have shown, the visual objectification of the voice in these vocal editing technologies encourages producer-songwriters to work with these forms in isolation, with little concern for the vocalist's labour. Yet in these cases the sudden intrusion of the

[56] I fully anonymize these examples to protect the privacy of both my informants and the vocalists involved, given that this is a sensitive issue.

vocalists' artistic identity into the process prompted increasingly controlling sentiments over the voice on the part of these producer-songwriters and a policing of the boundaries of vocal editing in an attempt to ensure that it was preserved as an activity that only they could control, arguably leading to greater moments of fixation on the visual manipulation of the voice.

In one case, a vocalist had sent a producer-songwriter detailed feedback on a draft edit of one of her lead vocals. Even though the vocalist had less experience in vocal editing than the producer-songwriter, her feedback was very comprehensive, and related directly to specific words and syllables in her performance, which she suggested could be improved by the producer-songwriter through further vocal editing. In particular, she asked the producer-songwriter to make certain vocal phrases more expressive through comping in different takes, as well as to make her diction sound more relaxed. I was struck by the detail of the vocalist's feedback points as well as the ways they seemed to imply that seemingly difficult-to-quantify aspects of her performance such as expression and diction could be improved through further vocal editing. The quality of such feedback seemed to show a good general knowledge of post-production processes on the part of the vocalist as well as a desire for greater involvement in such processes. More generally, vocalists' desire for greater involvement in and knowledge of vocal editing is likely to be linked to the way their exchanges with producer-songwriters are consistently dominated by DAW session materials which include countless detailed visualizations of the voice in software. Indeed, some of my interviews with other informants discussed the increasing capacity and involvement of vocalists in production over the past 10-15 years, an involvement which is often spearheaded by vocalists who desire to teach themselves the ropes of production in their own time, as a way of taking back control of their voices during songwriting and production in the face of producer-songwriters' near-total possession of the technical power to edit them (2019I, 2019S).

In the context of this feedback, however, the producer-songwriter in question found it difficult to interpret the vocalist's demands, perceiving them as too vague, and occasionally even overruled her comments entirely. In one instance, despite the vocalist's suggestion that a lack of expressivity in her voice could be changed through slightly different comping, the producer-songwriter felt that this was

instead likely to be more of an issue with the song's "production" and "mixing", which could be rectified through the addition of greater reverb and distortion. To satisfy the vocalist's demands, however, I observed the producer-songwriter making changes to the comp "anyway"—just to create some kind of noticeable change that might accomplish her request, even though he thought such changes were redundant. This dialogue therefore seemed heavily gendered, evoking the classic misogynistic trope of women being over-emotional, 'difficult', and unreasonably emotionally demanding, with men believing that they consistently have to perform the complex and inconvenient labour of decoding what a woman is 'really' feeling in order to satisfy her. The producer-songwriter deflected the vocalist's feedback, deliberately choosing to assume that the vocalist was being unnecessarily vague and pedantic with her comments, and that her feedback reflected a lack of knowledge about editing processes—he felt the need to try and decrypt what she *really* meant by her feedback about expression and diction as a result. This case also suggests a deliberate, and gendered, policing of the boundaries of vocal editing on the part of the producer-songwriter: the vocalist's intrusion into the process is perceived as an inconvenience, or a transgression of the boundaries of the producer-songwriters' proprietary and carefully cultivated technical realm, the confines of which he attempts to uphold through this strategy of misdirection.

In another example, a producer-songwriter I observed was editing a lead vocal which had been recorded and sent to him remotely by a vocalist. Due to the remote working scenario, the vocalist had recorded herself in a way that wasn't up to the producer-songwriter's high standards: using a poor-quality USB microphone, and not using a reflection filter. This meant that, throughout the editing process, the producer-songwriter complained about the additional effort he *had* to put into tuning and comping the vocal, as well as undertaking a variety of audio restoration techniques, due to the vocalist's lack of recording expertise. In particular, her failure to use a "reflection filter" (otherwise known as a "pop shield") while recording—a mesh 'screen' frequently used in vocal recording across pop genres to reduce the prominence of plosive consonants when singing close to a microphone (White 2005)—as well as her "having her gain too low", led the producer-songwriter to use three iZotope plug-ins from the *RX* suite—'*Mouth De-Click*', '*Breath Control*' and '*De-*

Plosive'—to automatically reduce the prominence of the resulting "clicks", "pops", plosive consonants (Ps, Bs) and breathy sounds in her voice. Figure 41 shows these plug-ins in usage: all three automate these acts of repair according to very simple parameters set by the user in the interface (e.g. "sensitivity", "strength"), suggesting that such fixes have become routine in the practice. Moreover, this producer-songwriter felt that, had he been running the recording session, he would have made sure these problems did not arise. He also implied to me in conversation that, on average, the editing and production expertise of producer-songwriters substantially surpassed that of vocalists, and meant that he was more likely to "know what he was doing" had he been recording the voice. Although this dialogue was not explicitly gendered, it was clear that the producer-songwriter felt that the ideal scenario would involve him having the utmost control and power over the recording and editing of the voice from start to finish, and the absence of that authority in this context meant that he had to re-claim some of that power after the voice had already been recorded, without the vocalist's presence or knowledge.[57] Additionally, although during this session the producer-songwriter seemed to have been alerted to these deficiencies in vocal recording through listening, his attempt to clean up the issues involved not only using the iZotope plug-ins to perform over-arching modifications to "clicks", "pops" and breathy sounds, but also zooming-in very precisely on vocal waveforms in both the DAW and *Melodyne* to double check that such sounds had been sufficiently reduced in prominence, occasionally using comping techniques such as crossfading and sculpting the *Melodyne* blobs to soften such inaccuracies even further. The comprehensive visual objectification of the voice across these different types of software therefore prompted the producer-songwriter to notice and correct an increasing quantity of 'deficiencies' in the vocalist's recording, amplifying his frustrations. Situations like these also encapsulate the important and loaded role played by visualizations of the voice in what could be described as a tense and gendered conflict over who really has the authority and knowledge, as well as the brute power, to make changes to the voice, and whose sense of the ideal vocal aesthetic will ultimately prevail.

[57] See Provenzano (2019a: 331-339) for a lengthier discussion of similar themes in relation to what she calls 'non-consensual tuning'.



Figure 41: iZotope *RX-7* audio restoration plug-ins. Note the simplicity with which they allow these common vocal recording problems to be 'fixed' simply by using 2-3 sliders. Screenshots from fieldwork (anonymized)

The same producer-songwriter highlighted a very precise imperfection in the voice that he noticed with reference to its visual objectification in *Melodyne*, and which he felt he would have been able to prevent if he had directed the vocalist during the recording session. On a wordless syllable, the vocalist made a very specific "shaky" sound that the producer-songwriter disliked, and wanted to make "straighter": the syllable landed on an F#, but very quickly after hitting this note the vocalist reduced the power behind her voice resulting in a suddenly much breathier timbre and a raise in pitch towards a G. Notably, the producer-songwriter was alerted to this discrepancy since the *Melodyne* software had trouble translating this "shaky" vocal sound into visual form: instead of creating a neat, rounded 'blob' for the syllable, rooted to an F#, the software showed two blobs at both F# and G,

alongside a sudden fluctuation in pitch contour which resulted from the sudden change in timbre (Celemony Software GmbH 2018: 232). To achieve the "straightness" he desired instead, he ended up tuning-in the vocalist's move from F# up towards G when he would have preferred it to be one long F# (Figure 42 shows this change, highlighted in red). However there was little he could do with *Melodyne* to remove the wild anomaly at this point (and if he did try to remove it fully, it would likely have created unwanted digital artefacts in the voice). There was a disciplinary quality to this moment: the producer-songwriter felt like, if he had only been there, he would have made sure the voice was in check and prevented it from going awry. But the "shakiness" of the voice, along with the capabilities of *Melodyne*, meant that he was unable to put his aesthetic preferences first in the way that he wanted to. Furthermore, it was clear that the producer-songwriter had only been alerted to this small imperfection in her voice from *seeing* a visual representation of the voice's timbre and pitch contours in *Melodyne*. This implies that, in certain instances, producer-songwriters' capacities for manipulating the (female) voice during vocal editing are heightened by the way such voices are visually objectified in software. The producer also expresses the desire to control the situation from the earliest stage, and to not permit the vocalist to perform and record in her own capacity because of the risk of (visual) imperfections in the voice.

Despite the continual overriding of the vocalist's performance decisions in these contexts, however, these producer-songwriters did still maintain a strong level of respect towards the ways vocalists' skills in performance differed from their own expertise in editing. I provide these examples to contextualize the ways such gendered power dynamics were embedded in my informants' practices and largely unconscious. One producer-songwriter, for example, implied that the most successful, most "real" artists he worked with were those who gave the kind of detailed feedback described above: despite the time-consuming nature of implementing such feedback, he felt that the ability to provide it represented the depths of the female vocalist's performative expertise. He also suggested that vocalists and producer-songwriters were likely to have different priorities when it came to the editing of subtle vocal features, with vocalists more likely to make requests pertaining to performative expressivity and the communication of lyrics, and producer-songwriters more focused on the

subtleties of vocal pitch and timbre. The producer-songwriter from the second example was also consistently trying to qualify his complaints about the vocalist's recording with compliments about the specificity and value of her skill set in comparison to his own. He clearly valued the vocalist's labour, especially given the difficult context of remote working, and the fact that they were both working to a tight timescale and had several other projects to worry about at the same time. In general, both producer-songwriters were trying to prioritize their working relationship with the vocalists and to value their contributions, even though the mismatch in their skill sets occasionally frustrated them. Moreover, although most of the examples I include across this thesis involved a male producer-songwriter editing a female voice, a few of my informants did work with male commercial pop singers. There is scope, therefore, for analysis of how moments like these differ, if at all, across these two contexts, although the timescale for this thesis prohibited me from gathering any substantial data about the latter. These points are intended to add nuance to these discussions about gender, illustrating how it is important to bear the specifics of these working relationships in mind when considering how it is that the editor manipulates the voice, rather than simply arguing that the vocalist is always passive and disadvantaged in the process.

Moreover, most of my informants seemed to have some awareness that the intensely technical nature of their editing practices meant that they alone were responsible for their compulsive attention to such details during tuning and comping, and that such details were unlikely to concern the vocalist. Indeed, vocalists have never been required to tame the "plosives" or "clicks" in their own vocal performances—the demand for these plug-ins has likely developed as a result of the proliferation of visual objectifications of the voice in music production which draw greater attention to vocal "clicks" and "pops" represented in waveforms and pitch-contours, as well as the intertwining of plug-ins designed for both speech processing (e.g. for film dialogue and podcasting) and popular music production.[58] This perspective further supports my suggestion that the plug-in industry drives vocal editing practices through its strengthening of both voice and technology fetishism in the design of

[58] Some of iZotope's early products were mainly designed for use in broadcasting, for example. (See Sound on Sound 2008).

vocal editing plug-ins: here, such plug-ins and the differing ways they visually objectify the voice encourage producer-songwriters to focus on very specific vocal details throughout vocal editing—details that vocalists are seldom aware of during their performances.

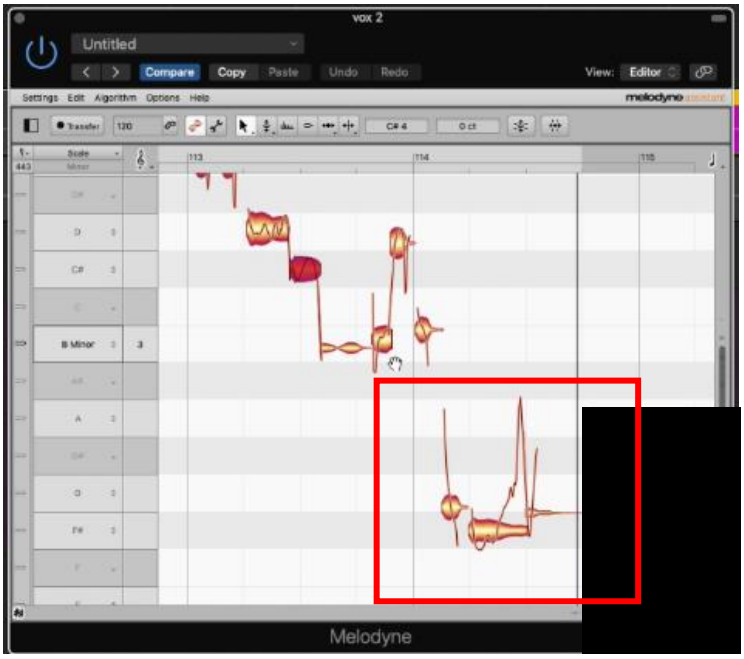
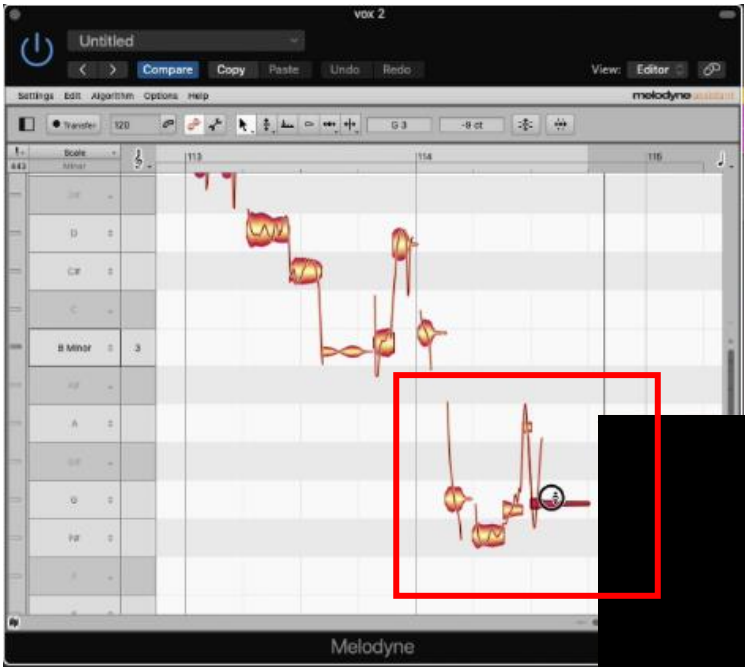


Figure 42: The anonymous producer-songwriter's Melodyne edit window, before and after "straightening" out the vocal syllable in question (the bottom right is obscured as there was a bit of his camera in the video which I couldn't remove as the two were overlapping)



On a side note, the continual respect afforded to the vocalists in these scenarios—respect which my informants were consistently keen to state clearly—may also have been somewhat driven by the fact that these producer-songwriters were required to maintain productive working relationships with artists: it was very important to them that such relationships were not disrupted as it would risk jeopardizing future work. This perspective is expanded in Marshall's thesis, where he claims that since tuning processes 'synthesise the combined labour of the engineer and the singer', the recorded voice in these contexts is 'entangled with an actual client to whom the tuner's intervention may ultimately be held accountable' (2017: 177). This is one of the main reasons why these examples are anonymized. It is also worth mentioning that, of course, any study which focuses on producer-songwriters' practices at such a high level of detail in this way is bound to give the impression that they have more agency and authority in the process than others (e.g. vocalists). However, this may be a skewed impression, and the tone of this analysis may have been different had the vocalists contributed more to my fieldwork data.

Conclusion

To conclude, this chapter has demonstrated how, in the preliminary stages of vocal editing, comping and tuning techniques are driven by visual objectifications of the voice in software. Comping and tuning are also part of a broader arsenal of preliminary vocal "tidying-up" techniques which similarly objectify the voice visually, including timing, manual gain automation ("clip gain") and audio restoration techniques. My informants Chris and Jason's familiarity with such visual objectifications of the voice informed their vocal editing techniques in ways which encouraged them to work with the voice visually rather than aurally, prompting: the cultivation of "sliding" vocal pitch contours, the selective amplification and attenuation of individual consonants and vowels, and the layering of groups of vocal waveforms to build-up "big" lead vocal arrangements. Such features constitute key aesthetic characteristics of the commercial pop vocal style. I also demonstrated the ways some of my informants attempted to exaggerate and emphasize vocal "emotion" during tuning and comping, showing how such a process is intensified by visual objectification, and ultimately results in a further commodity fetishism of this vocal feature. In addition, I provided anonymized ethnographic evidence

demonstrating how visual objectifications of the voice inflect the often-gendered collaborations between producer-songwriters and vocalists, frequently intensifying producer-songwriters' tendencies to disregard or question the labour of vocal performance.

More significantly, these discussions of the visual objectification of the voice across tuning and comping technologies are one example of how the interacting fetishisms of the voice and technology drive vocal editing practices. Primarily, the mediating role of such visual objectifications of the voice on this practice can be interpreted as a form of vocal commodity fetishism. During tuning and comping, the voice is objectified as a series of visual shapes: lines, curves and "blobs" that can be re-drawn, split apart, expanded and condensed. These visual objects stand in for the voice itself during the preliminary stages of vocal editing. They also begin to connote particular features of the voice ("emotion") that would usually be rooted in the performative labour of singing accomplished by vocalists. During vocal comping and tuning, the voice's visual objectification temporarily obscures the social relations of collaborative songwriting scenarios involving a vocalist, as well as the labour of the vocalist's performance, as the producer-songwriter simply sculpts the voice's visual forms whilst sitting at a computer in social isolation. In addition, my brief analysis of the balance between *AutoTune* and *Melodyne* users suggests that this vocal commodity fetishism intersects with and reinforces the commodity fetishism of plug-ins (incorporating characteristics of both analogue emulation and digital novelty commodity fetishism), prompting producer-songwriters to focus in even more detail on the different visual forms of the voice offered by these plug-ins. What's more, producer-songwriters' aesthetic ideals have been so strongly influenced by these forms of visual objectification that in some cases they begin to work solely with the visual form of the voice rather than being led by aural cues, illustrating in a more radical way how a form of vocal commodity fetishism guides the practice, to the extent that the visual voice takes precedence over the sonic (which should ostensibly be the most important focus of vocal editing). It may be the case that such commodity fetishism of the voice, which relies on familiarity with its objectified form in software, complements the broader commodification of topline vocals across the industry discussed in Chapter 1, where the voice is increasingly valued purely as a sonic form regardless of the vocalist and the

context in which they sang it. My discussion of the gendered dynamics of interactions between producer-songwriters and vocalists as they pertain to tuning and comping processes also suggested that, when producer-songwriters are confronted with (female) vocalists' necessary involvement in vocal editing processes, such commodity fetishism blends with a more erotic kind of voice fetishism which evokes notions of fixation and control.

The next chapter, about equalization (EQ), encompasses a different kind of intersection between voice and technology fetishism which defines the practice and which builds on Chapter 2, relating more directly to the erotic kind of voice fetishism, in which the sound of the voice stands in for ideas about the person or body behind it. One particular visual objectification of the voice does perform a subsidiary influencing role in EQ practices, however: as I will explain, para-graphic EQ plug-ins visually represent the timbre of a vocal track in its entirety and make it modifiable. The next two chapters will also consider how the plug-ins used in EQ and Compression, and the language with which producers describe their aesthetic functionality, occasionally encodes similarly hierarchical gender ideologies to those described here.

Chapter Four: "You Want it to Sound Real, Like a Person": Erotic Voice Fetishism, Para-graphic EQ Plug-ins and Realism in Commercial Pop Vocal Equalization (EQ)

The application of digital equalization (EQ) to the voice usually followed comping and tuning. EQ plug-ins allow a user to manipulate the relative prominence of selected bands of the frequency spectrum of an audio signal (Huber & Runstein 2018: 411). My informants "EQed" lead vocals in a way which removed a large part of their low-frequency energy, and subtly emphasised their high frequency energy. Linked to the aesthetic discourse I discussed in Chapter 2, this was a way of enhancing the impression of vocal foregrounding. Building on the theme of realism introduced in Chapter 2, however, vocal EQ had an especially strong discourse of realism surrounding its usage in comparison to tuning, comping and compression. Vocal EQ—undertaken by manipulating EQ "curves" within 'para-graphic' EQ plug-in interfaces—was conceived by my informants as a way of preserving and enhancing the singing voice's most perceptually important, most "real", qualities, even though it was clearly an aesthetic practice which involved creating a stylized impression of vocal proximity. EQ had somehow been naturalized in my informants' discussions of commercial pop vocal editing as having very little perceptible effect on the voice at all, or solely a functional one, aimed at making the voice sound more "real". When I asked Jason, for example, whether he thought vocal EQ might define the vocal style of the commercial pop genre more so than vocal tuning, he disagreed vehemently, clarifying:

> With the majority of people [i.e. listeners] in mind, you're going for *not* hearing the EQ. You want it [the voice] to sound *real*, like a person. (Jason, 2020Z)

In this chapter I claim that the strength of realist discourse in commercial pop vocal EQ results from a particularly powerful feedback loop between voice fetishism and the commodity fetishism of technology. In relation to the former, the technique provides the most direct capacity to manipulate the timbre of the voice in comparison to the other technologies involved in vocal editing. As I suggested in my theorization of erotic voice fetishism in Chapter 1—drawing on Nina Eidsheim's implicit criticism of the Barthesian legacy of conceptualizing the voice—timbre is the feature of the

voice which most strongly fuels erotic voice fetishism through beliefs that it indicates the identity and physicality of the voice's production more potently than other sonic parameters. Vocal EQ has therefore acquired a sense of power among producer-songwriters because it seems to provide access to this fundamental, interior quality of the voice. In the case of the latter, the assumed power of EQ to transform this core identifying feature of the voice also intersects with—and is intensified by—the commodity fetishism of digital novelty EQ plug-ins, specifically 'para-graphic' EQs. These plug-ins allow a user to manipulate an entire vocal track's timbre (frequency spectrum) through its visual representation as a graph-like curve. This presentation (or objectification) of vocal timbre as a flexible and malleable visual form combines with producer-songwriters' existing beliefs in the defining power of vocal timbre to make them feel like EQ gives them an inordinate amount of technical power to transform the voice. In turn, the technical aura of para-graphic EQ plug-ins fuels quasi-scientific descriptions of EQ which allude to universalizing ideas about human auditory perception, further strengthening the realist discourse surrounding the technique. In short, through its particularly pointed engagement with erotic voice fetishism through the implied physicality and individuality of vocal timbre, and the intersection of this belief with the commodity fetishism of digital novelty EQ plug-ins, EQ provides producer-songwriters with a particularly powerful illusion of mastery over the voice, prompting them to view it as a realist rather than aesthetic technique. Vocal EQ therefore provides a particularly apt illustration of the intersecting voice and technology fetishisms which drive commercial pop vocal editing.

Alongside these core arguments, this chapter also analyses how commercial pop vocal EQ's realist discourse obscures assumptions about the social identities of vocalists. This shows how such a discourse of realism helps to embed erotic voice fetishism in the practice, as stereotypical assumptions about the links between voices and an imagined social (here, gendered) identity are made to seem normal and natural. The chapter also implicitly continues the theme of visualization as objectification introduced in Chapter 3, as the appearance and functionality of para-graphic EQ plug-ins drive vocal EQing and its associated realist discourse, as well as encouraging similarly visual modes of editing the voice to comping and tuning. I also suggest that contemporary plug-in companies' appeals to

notions of "real" auditory perception in vocal EQ show how the plug-in industry plays an active role in sustaining the close interactions between erotic voice fetishism and the commodity fetishism of digital novelty plug-ins that motivates vocal EQing.

The chapter has four main sections which develop these ideas. First, I consider the broader history and context for EQ practices in music production, which informed my informants' EQ approaches. Second, I will explain how my informants' para-graphic EQ techniques were guided by beliefs that all vocal frequency spectra can be split up into portions within which desirable and undesirable vocal characteristics are located ("low end rumble", "body", "boxiness", "presence", "harshness", "sibilance" and "air"). My informants explained such features in terms of fashioning the voice into its most "real" or true state. The third section suggests that realist discourse in vocal EQ fuels tacitly gendered assumptions about the voice. In particular, my informants frequently invoked surgical metaphors to describe their functional, realist EQ approaches, making a direct mapping between digital vocal editing and physical interventions in the body, and implying a diagnostic, medicalized power dynamic between producer-songwriters and vocalists. Moreover, the diagnosis of vocal "harshness" in need of "surgical" removal was disproportionately applied to female vocalists. Finally, I contextualize three contemporary digital novelty EQ plug-ins: *Pro Q-3*, *Soothe* and *Gullfoss*. These plug-ins claim to provide easy-to-use EQ functions that are sensitive to the human auditory perception of the voice, both illustrating the pervasiveness of quasi-scientific allusions to realism in vocal EQ and suggesting that such ideas are being actively reinforced in the practice through the rationalization of particularly 'realist' EQ techniques by key players in the plug-in industry.

In a broader sense, this chapter draws attention to the extremely strong influence of this kind of aesthetic use of EQ on the voice (involving the exaggeration of high-frequency energy and the diminution of low-frequency energy) on the experience of listening to commercial pop tracks. Not only does it transform the timbral qualities of the voice, but it also attempts to solidify the voice's foregrounded and central positioning relative to the texture of commercial popular music tracks in their entirety. Yet despite the huge aesthetic impact of this style of EQing on the sound of the genre,

the technique has hitherto been ignored in academic literature about the voice in Western popular music (with the exception of recent work by Victoria Malawey (2020: 134), who discusses it only briefly). Indeed, it may be the case that practice-derived conceptions of EQ as a functional, objective procedure in service of vocal realism have contributed to the relative absence of critical appraisals of the technique in popular music studies.

4.1. Contextualizing vocal EQ

A brief history of EQ in audio engineering and music production

Definitions of EQ vary depending on whether they hail from acoustics, sound engineering or music production perspectives. Acoustic scientists Vela Välimäki & Joshua Reiss claim that equalization, in a very broad sense, can refer to any process where the frequency response[59] of a signal is altered or adjusted (Välimäki & Reiss 2016). More specifically, recording engineers David Miles Huber and Robert Runstein describe equalization as 'the *alteration* in frequency response of an amplifier so that the *relative levels* of certain frequencies are *more or less pronounced* than others' (Huber & Runstein 2018: 411, italics added). Sound on Sound gives a vaguer definition of an 'equalizer' which does not invoke the notion of frequency response, describing it as 'a device which allows the user to adjust the *tonality* of a sound source by boosting or attenuating a specific *range of frequencies*' (Sound on Sound 2022, italics added). For my informants, EQ usually referred to the technique of manipulating the relative prominence of certain user-selected areas of a frequency spectrum by making "boosts" and "cuts", usually using a digital EQ plug-in.

The term 'equalizer' is rooted in early 20th-century communications engineering. Over the course of the 20th century, the term shifted from referring to a more static, pre-defined technique of applying electrical 'filters' to sound signals as a way of enhancing their communicative efficiency,

[59] Broadly, 'the frequency response of a system describes the frequencies that can pass through that system.' (Ballou 2008: 27). Frequency response is usually used as a measure of how accurate or efficient an audio system is—e.g. a microphone or speaker with a 'flat frequency response' would be one in which the input signal was reproduced with no exaggeration or attenuation of any frequency band (Huber & Runstein 2018: 116). Equalization, in the most precise sense, is a way of deliberately adjusting the frequency response of a signal.

towards more flexible ways of manipulating frequency spectra in aesthetic contexts such as film and audio. The 'electric wave filter' was developed and patented by Glen Campbell at AT&T in the 1920s whilst working on ways to solve the problems associated with interference on telephone lines: in simple terms, a 'filter' would be added to a circuit as a way of attenuating a specific undesirable or intrusive frequency band from a signal and ensuring that the intended message could be more efficiently transmitted across the line without interference (Campbell 1922; Mills 2020). These filters made fixed modifications to frequency spectra, often represented as graphical shapes, the names for which still circulate widely in contemporary music production. A 'high-pass filter', for example, is one in which a specific frequency is set as a cut-off point, above which higher frequencies are allowed to pass through it and lower frequencies are attenuated. On a frequency/amplitude graph, this is visually represented as a steep ascending curve which plateaus upon reaching the user-defined cut-off point (Figure 43). Sound historian Mara Mills has recently suggested that filtering is a powerful metaphor through which the history of signal processing can be understood more broadly as defined by different kinds of mediation through which sound is shaped and enhanced: 'When you think about it, everything is a filter' (Smith 2007; cited in Mills 2020). Contemporary EQ techniques are part of this history of filtering.[60] Välimäki & Reiss, for example, suggest that early equalizer units were a response to a similar problem in telephone engineering whereby signals transmitted over long distances required correction or compensation so that the received sound matched the initially transmitted sound, hence the term's prefix 'equal' (Valimaki & Reiss 2016: 1), although it is not clear at what point in this history 'filters' became 'equalizers'.

[60] For more detailed discussions of the metaphor of filtering in the history of sound, I direct the reader to work by Melle Kromhout (2017, 2021) and Jens Papenburg (2020), whose work is cited by Mills (2020).

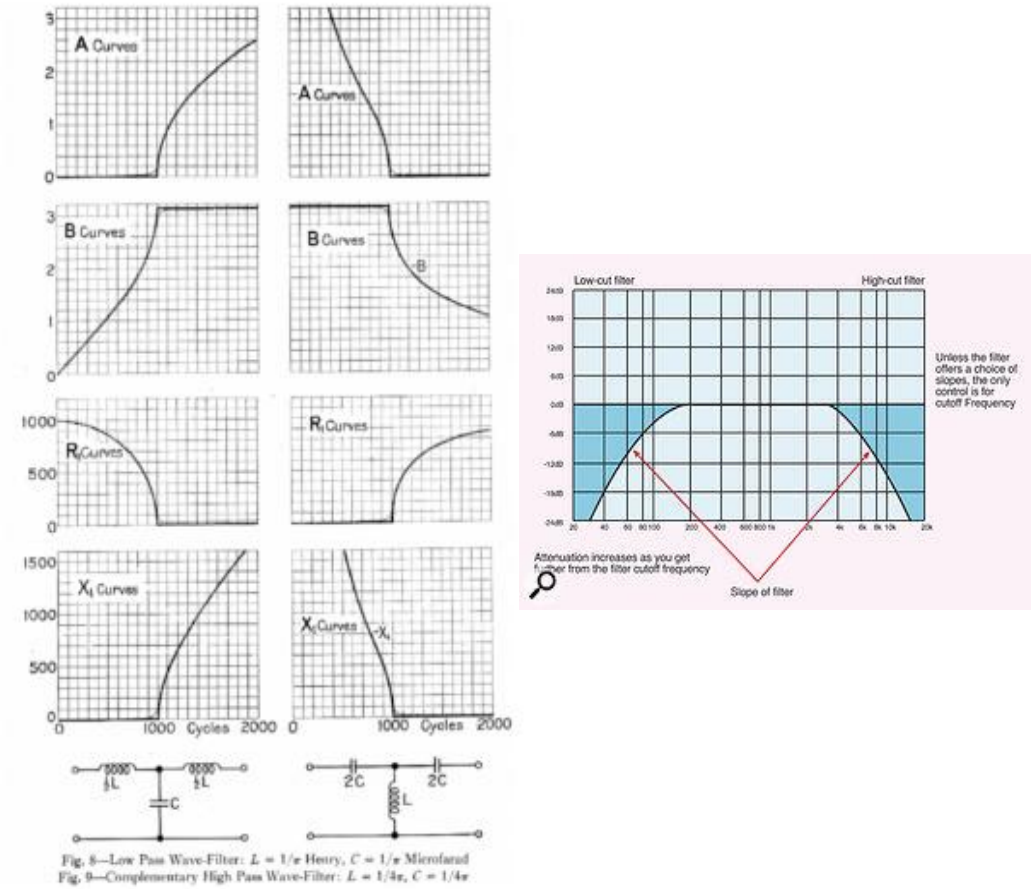


Figure 43: Glen Campbell's graphical representations of low-pass and high-pass wave-filters (the third row of graphs is most similar to contemporary representations) (Campbell 1922: 16), juxtaposed with a figure explaining 'low-cut' (=high pass) and 'high-cut' (=low pass) filters in EQ (Sound on Sound 2009). Note that depending on whether the term 'pass' or 'cut' is used, 'low' and 'high' will be switched around. This can sometimes be confusing when reading about EQ.

According to several audio engineering sources (Välimäki & Reiss 2016; Bohn 1990), the greater flexibility of contemporary equalizers can be traced back to John Volkman, who developed variable or adjustable equalization in the 1930s whilst working at RCA to improve the sound of film theatre playback systems. Sources regarding the actual content of Volkmann's work are difficult to find, however, supposedly, the name 'variable' arose from the fact that the user could select an individual frequency from a pre-defined set and perform 'boosts' (amplifications) or 'cuts' (attenuations) to it according to preference, rather than being restricted to pre-defined filter shapes (Välimäki & Reiss 2016: 132; Bohn 1990). Manipulating frequency spectra in a flexible way was also the goal of engineer George Massenburg at International Telecom Inc. in the 1970s, often credited with the

development of *parametric* equalization (Massenburg 1972; Bohn 1990; Välimäki & Reiss 2016: 140; Reiss & McPherson 2015: 99). Parametric EQs provide an even higher degree of user flexibility by facilitating the manipulation of three interlocking parameters: centre frequency, gain (or 'cut/boost') and bandwidth (or 'Q') (Reiss & McPherson 2015: 98). The centre frequency determines the frequency to be altered; the gain determines the increase or decrease in relative amplitude at the centre frequency selected; and the bandwidth describes the width of the region in which this cut/boost is enacted (Case 2007: 106-107) (Figure 44). The EQ plug-ins used in contemporary digital music production are all parametric EQs in some form or another, contrasting with *graphical* EQs which operate using slider controls where the centre frequencies and bandwidths are fixed (Bohn 1990), and which are seldom used in contemporary music production. It follows that the reason parametric EQs are predominantly used in music production is because of the sheer flexibility they provide when it comes to manipulating frequency spectra. This history encompasses a trajectory from the application of static pre-defined filter 'shapes' in early telephone engineering towards the ubiquitous use of parametric equalizers which allow music producers and engineers to manipulate frequency spectra at a very high level of detail. Parametric EQs turn frequency spectra into malleable objects.

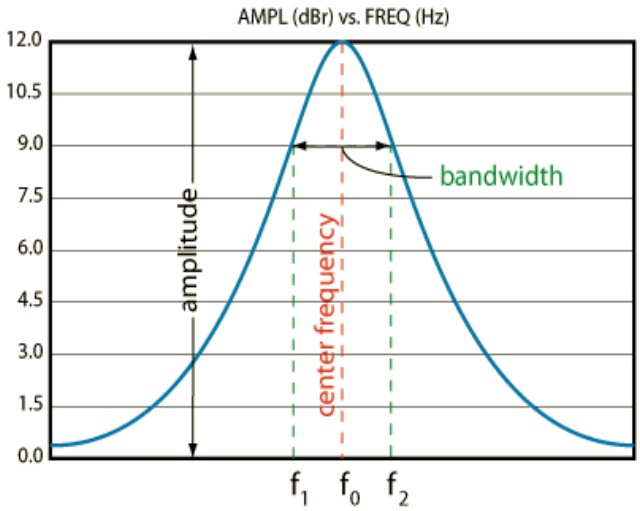


Figure 44: Diagram of parametric equalizer parameters, taken from Bohn & Rane (1982/87).

Digital EQ plug-ins: para-graphic and rotary interfaces

In contemporary popular music production, digital parametric EQ plug-ins generally adopt one of two different interface types, which allow users to manipulate frequency spectra using two contrasting visual formats. The first is a graphic or slider-based interface, consisting of a graph with frequency in Hz on the x-axis and amplitude in dB on the y-axis, usually with a centre line bisecting the y-axis at 0dB. My informants did not use any specific names for these interfaces, but engineer Dennis Bohn has termed them *para-graphic* EQs (Bohn 1988: 372). Para-graphic EQs allow the user to manipulate the frequency spectrum as if it were one continuous curve: the centre frequency and 'Q' can be selected by the user at any point on the central line at 0dB, then dragged upwards or downwards to create peaks and troughs which represent "boosts" and "cuts" (Figure 45). For the second interface type, I use the term *rotary* EQs, also following Bohn (ibid.). Rotary EQ interfaces consist of a set of rotatable 'knobs', typically separating the frequency spectrum into four or five generalized areas—e.g. lows, low-mids, mids, high-mids, and highs—including knobs for each area which correspond to some variety of the three parametric EQ parameters. Most pre-digital hardware parametric EQs adopted this kind of rotary interface, and many digital plug-ins which use this interface are software emulations of particularly revered or successful pre-digital EQs (Figure 46).

Importantly, and referring to my delineation of the two types of plug-in across the industry in Chapter 1, para-graphic EQ interfaces are digital novelties and rotary EQs are analogue emulations. Paralleling this difference, my key informants used both types of EQ to perform slightly different functions in vocal editing. Para-graphic EQs, in particular, were more foundational to my key informants' EQ practices, performing over-arching modifications to the voice at the outset of vocal chains, whereas rotary EQs were used occasionally for more explicitly creative boosting to increase vocal foregrounding. One informant made an interesting distinction between "functional" para-graphic EQ plug-ins and "characterful" rotary EQ plug-ins, with specific reference to FabFilter's *Pro-Q3* and the *Focusrite d2* respectively (2020D). Further, my informants seemed to spend a lot more time fussing over creating the most perfect para-graphic EQ "curves" than adjusting the knobs on rotary EQ interfaces. For these reasons, para-graphic EQs take centre stage throughout this chapter.

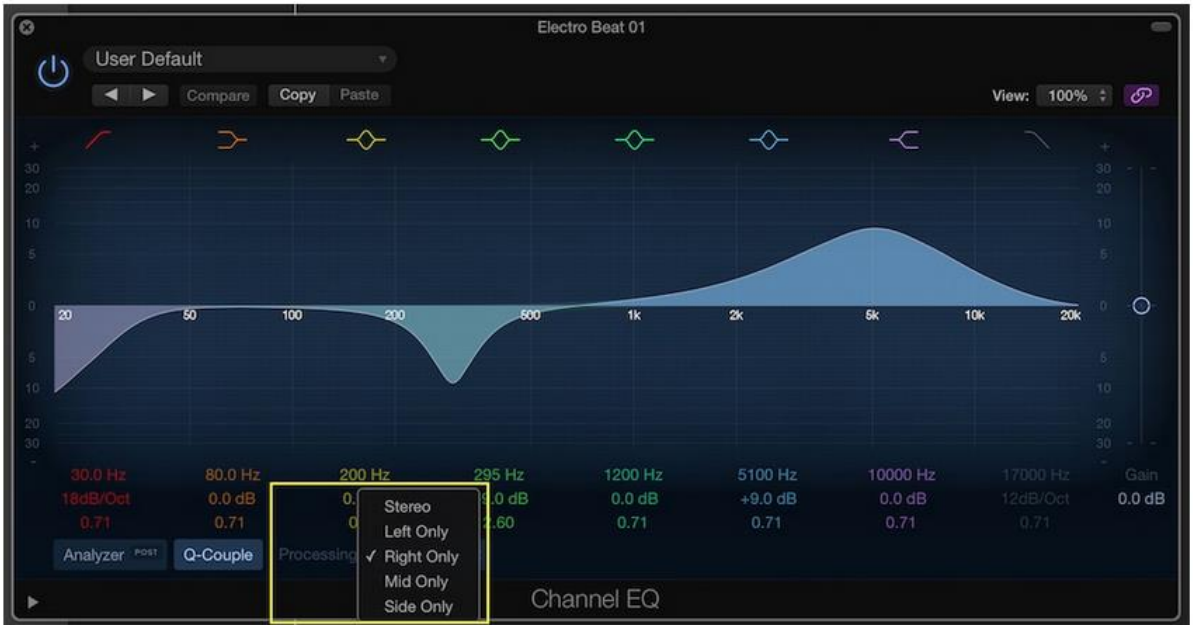


Figure 45: An example of an EQ curve created using the in-built para-graphic EQ in the *Logic Pro X* DAW. This image also nicely illustrates the 'bandwidth' parameter: you can see how the cut enacted around 350Hz has a smaller 'Q' than the boost at 5kHz which is very wide. Varying the bandwidth essentially enables the user to determine the extent to which a "cut" or "boost" at a defined centre frequency will affect the surrounding frequencies. (Image from https://www.pro-tools-expert.com/logic-pro-expert/logic-pro-blog/2016/03/28/logic-pro-x-channel-eq-power-user)



Figure 46: The 'rotary' interface of a Waves SSL G-Equalizer, modelled on the SSL *G Series EQ292* hardware EQ. Looking at each knob you can see how the user has not made any modifications to the lows (LF), performed a small cut to the low-mids (LMF) at 1.75kHz with a relatively small Q, a small boost to the high-mids (HMF) at 2kHz with an average-sized Q, and a small boost to the highs (HF) at 10kHz. The user has also added a high pass filter at 180Hz (on the left hand side) to remove most of the low end. (Image from waves.com product page: https://www.waves.com/plugins/ssl-g-equalizer)

EQ in Western popular music mixing: 'carving' and universal models of auditory perception

The discourse of realism in commercial pop vocal EQ has significant precedents in Western music mixing trends. Specifically, assumptions that the technique engages universal facets of human auditory perception are embodied in the notion of 'carving', discussed frequently across pedagogical mixing literature. To be clear, this conceptualization of EQ as somehow directly harnessing fundamental capacities of human auditory perception is a key idea underpinning the discourse of realism in vocal EQ, because it conflates an aesthetically motivated sound editing technique with an idea of the universal human experience of sonic reality.

As I explained in Chapter 2, the aesthetic discourse of vocal editing conceives of "the mix" as a bounded space within which certain elements must be positioned along three perceptual axes (vertical, horizontal, depth). According to my appraisal of relevant practice-based literature, music production and recording professionals consider EQ to be one of the most important and useful techniques for positioning sound sources in the vertical and depth domains especially. In terms of the vertical plane, EQ is seen as an indispensable tool for separating individual sound sources in a "mix" so that they can be heard distinctly—a technique that Huber & Runstein have termed 'carving' (Huber & Runstein 2018: 489). And, in terms of the depth or 'proximity' plane, EQ is used to either emphasize or attenuate high frequency content as a way of creating the perceptual impression of a sound source being further "backwards" or "forwards" within a mix (Hodgson 2010: 168; White 2009). These aesthetic principles guide the use of EQ to support vocal foregrounding in commercial pop.

More specifically, 'carving' involves using EQ to loosely confine individual sound sources in a mix to specific frequency ranges and to prevent 'masking' (otherwise known as frequency masking or spectral masking). 'Masking' is the idea that if two sounds with similar frequency spectra are occurring simultaneously, the louder tends to 'mask' the quieter one (Huber & Runstein 2018: 69), which, as a Sound on Sound article claims, can cause difficulties for 'retaining clarity and focus in (say) a lead vocal' (Robjohns 2017b). Mixing textbooks claim that carving remedies this by contributing to the perceptual impression that individual elements in the mix are distinct from one

another, helping to ‘balance’ the mix (Huber & Runstein 2018: 489; Oswinski 2013: Ch. 5). Other techniques such as panning and reverb are also intended to fulfil this function of ensuring that different mix elements occupy autonomous ‘zones’, but these techniques are more overtly spatial, operating within a horizontal, left-to-right conception of the mix space, whereas carving is intended to create an impression of textural separation relative to the other sounds in the mix (what mixing textbooks have called the ‘frequency domain’ (Izhaki 2017)). The embeddedness of ‘carving’ in Western music production is also evidenced by the popularity of instrument frequency chart graphics or posters which circulate widely online (one producer I observed had one of these charts fixed to his desktop background whilst working (2019G)) (Figure 47).

Most importantly, carving assumes that different sound sources have certain frequency areas which are more perceptually significant than others. For vocals, the assumption is that most of the perceptually important or fundamental sound material lies somewhere between approximately 50Hz to 2kHz, with excess harmonic information lying above this range (approx. 2kHz-10kHz) (see SOS Team 2012; Seydel 2016; Huber & Runstein 2018: 416-417). Anything underneath this range is considered unimportant and disposable, especially because it is likely to ‘conflict’ with lower frequency areas occupied by other elements in the mix (e.g. bass, drums, piano). My fieldwork suggested that it was common practice among Western popular music production professionals to ‘carve out’ the vocal’s autonomous space in the mix by cutting out most of the lowest frequencies in a vocal track (below 100Hz) and concentrating any further EQ modifications above this range. This practice was also deeply embedded in commercial pop vocal editing, evidenced by the fact that the first step each of my informants took when EQing a vocal was to perform a “low cut” (high-pass filter) using a para-graphic EQ. Most of the time they would perform this low cut instantaneously as soon as the para-graphic EQ interface was opened, without listening to the voice at all, both illustrating the extent to which “low cutting” was naturalized in the practice and implying that the malleable visual representation of vocal timbre in para-graphic EQ plug-ins encouraged its routine undertaking as a visual rather than an aural technique. Figure 48 shows a selection of para-graphic EQ curves taken from both my field data and internet materials, showing the ubiquity of the “low cut” (as

well as to introduce the reader to the similarities across the para-graphic EQ shapes I observed among commercial pop producer-songwriters). The pervasiveness of such an approach indicated that these ideas about universal auditory perception, derived from the established mixing concept of 'carving', were firmly and largely tacitly rooted in my informants' vocal EQing practices.

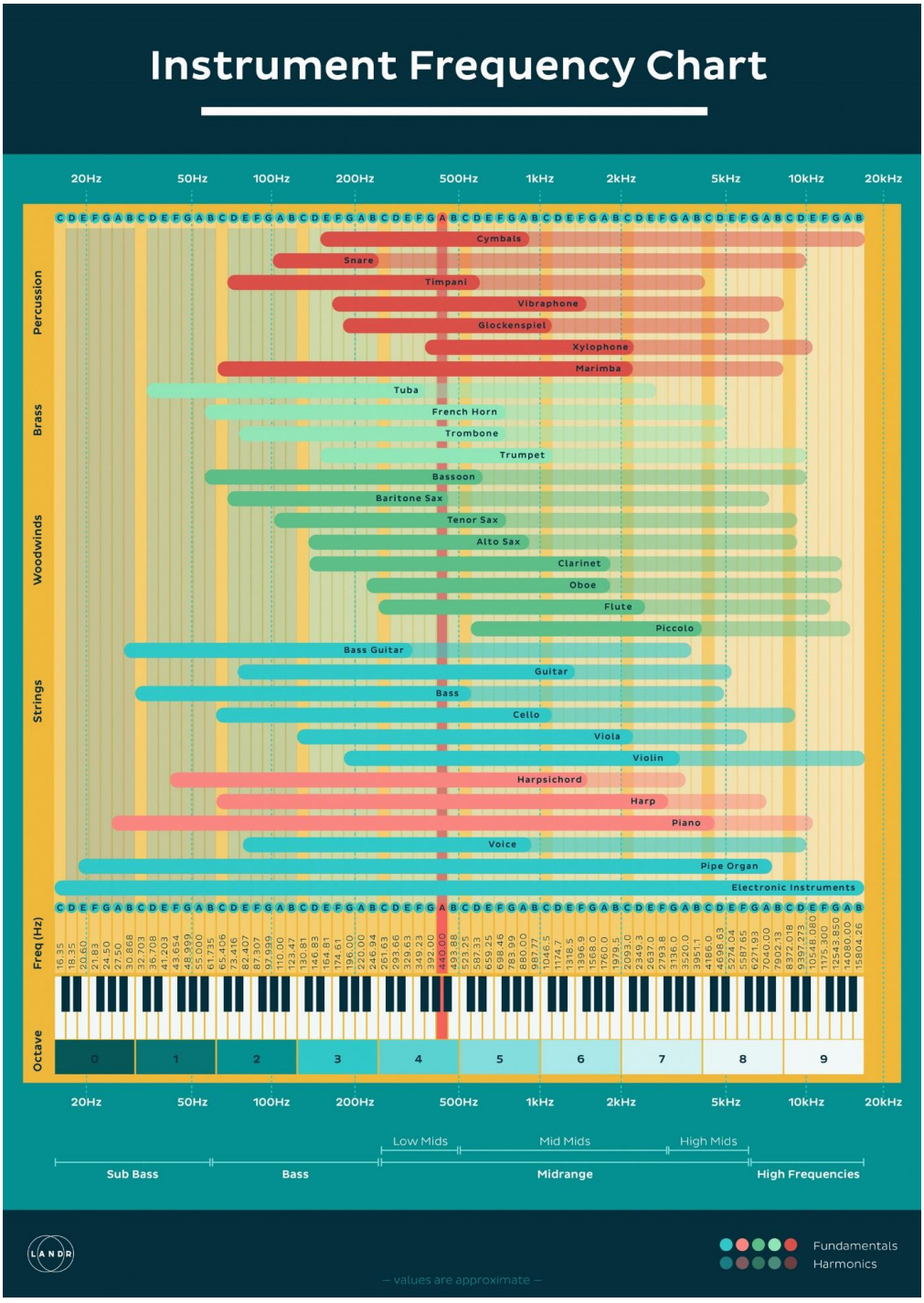


Figure 47: Instrument frequency chart from automated music mastering company LANDR (Seydel 2016). 'Voice' is third from the bottom.



Figure 48: A selection of first-round para-graphic EQ curves from producer-songwriters working in commercial pop. The first four curves are taken from my field data. The last four are taken from a selection of informal pedagogical videos available on YouTube, created by producers whose vocal editing styles were very similar to those of my informants. The "low cut" is highlighted on the first image and can be compared to Glen Campbell's representation of a high-pass wave filter in Figure 44. (The 2nd image performs a subtler modification, only slightly attenuating the lowest frequencies rather than cutting them out completely).

Paradoxically, in these broader mixing sources, notions of universal perceptual irrelevance which underpin 'carving' techniques also coexist with very clear statements that EQ ought to be used to heighten the aesthetic effect of vocal foregrounding. For example, music production textbooks and internet materials discuss how EQ is used to alter the perceptual impression of depth, adjusting how far "back" or "forward" a sound appears to be situated within a mix. This is supported by the idea that if a sound is both louder and contains more prominent high-frequency energy ("top-end"), it will sound more "up-front" or "closer" to the listener (White 2009). The prevalence of this idea in Western popular music mixing is evidenced by a further type of infographic which circulates widely in online music production resources, in which sound sources are represented by coloured boxes on a loosely drawn graph of frequency (y axis) against panning position (x axis) (Figure 49). This suggests that Western popular music mixing has an existing tendency to conflate notions of vocal realism with the aesthetics of vocal foregrounding, perhaps because this aesthetic trend has become so dominant as to be normalized. This conflation of realism with vocal foregrounding is a very important and established phenomenon which, as I will indicate in the next section, is even more embedded and naturalized in 21st-century commercial pop vocal editing.

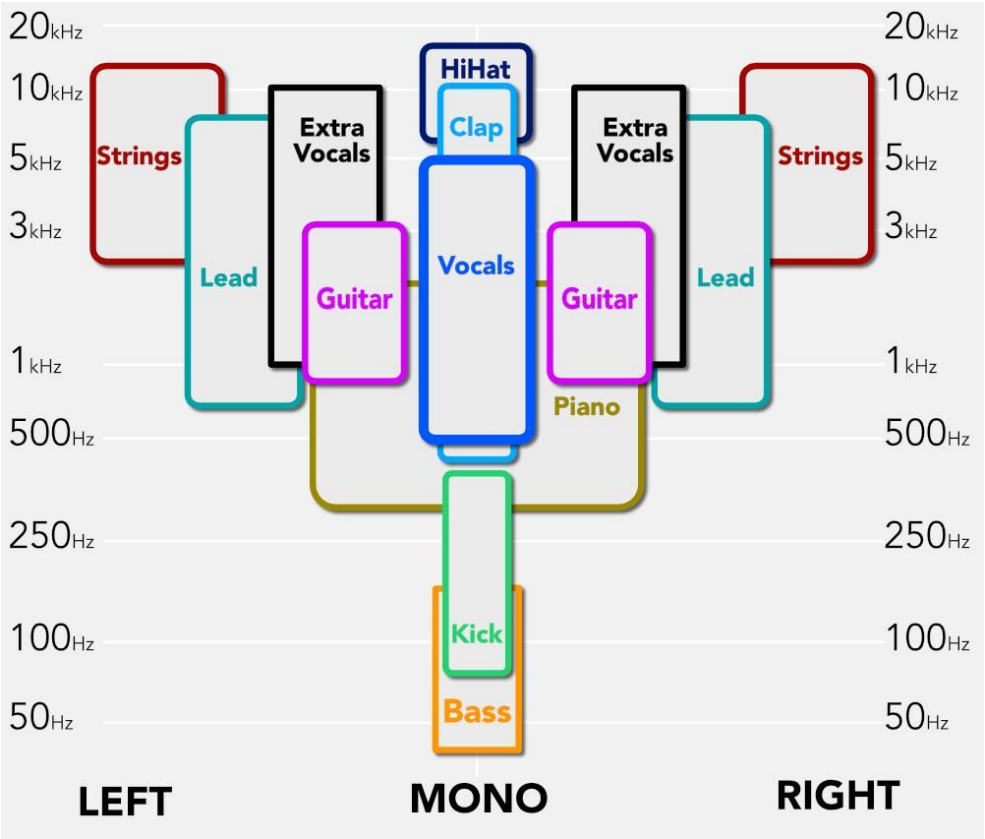


Figure 49: An example of a 'separation in the mix' infographic used frequently by informal music production pedagogy website masteringthemix.com. (Mastering the Mix 2018) The track represented here is 'Electricity' by Silk City and Dua Lipa (2018).

4.2. EQ in commercial pop vocal editing

Para-graphic EQ and assumptions about vocal realism

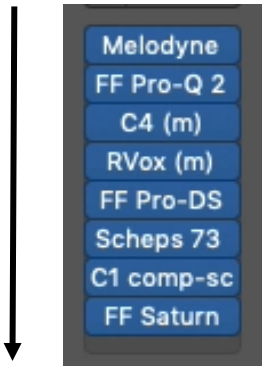
When using para-graphic EQs on the voice, my informants were guided by a very precise language which described what was or wasn't aesthetically desirable in a vocal frequency spectrum. This desirability was predominantly figured in terms of universalizing ideas about what was or wasn't perceptually "necessary" when hearing a singing voice in this genre. This section taxonomizes this vocabulary, and in the process of doing so it shows how my informants used para-graphic EQ "curves" to sculpt the voice into what they believed to be its most real or essential state by selectively emphasizing or attenuating clearly defined regions of its frequency spectrum. In fact, most of these techniques could also be interpreted as exaggerating vocal foregrounding, showing how the conflation of vocal realism with this aesthetic ideal had become naturalized in the practice. Furthermore, there

were several inconsistencies and contradictions within this discourse of realism applied to the various vocal EQ 'zones', and realist ideas were occasionally applied selectively and in varying degrees to the various EQ modifications undertaken in these 'zones'. In particular, they appeared to be much shakier in cases where the aims of vocal foregrounding were more explicit (e.g. "air"), further illustrating the culturally constructed quality of this realist discourse. Most significantly, this analysis shows how the practice of vocal EQ assumes a totalizing power over vocal timbre, which is strengthened by the conception of the practice as serving vocal realism: it splits up the voice into assumed archetypal frequency regions which are modified according to generalized, formulaic rules, further facilitated by the visualization of the vocal frequency spectrum provided by para-graphic EQ interfaces.

Briefly, although this chapter concentrates largely on para-graphic EQ modifications, it is worth clarifying that my informants frequently applied EQ in cumulative, iterative rounds, using both para-graphic and rotary EQ plug-ins to perform slightly different, subtle modifications to the frequency spectrum at different points in the "vocal chain". However, all my key informants started by applying an initial para-graphic EQ at an early stage in the vocal chain according to realist principles, both to 'carve' away at the low-end and to address any other perceived "problems" in the middle and upper ranges of the vocal frequency spectrum. In addition, this first para-graphic EQ was always the point at which the vocal sound was most transformed by modifications to the relative balance of different frequency bands. Any further EQ plug-ins added would either be making modifications which were thought to compensate for the effects of other plug-ins on the frequency spectrum (e.g. de-essing or compression) or to boost the high-frequency energy in the voice even further in a more recognizably aesthetic fashion. For example, Chris's vocal chain for Audio 2, represented in Figure 50 below, contained a first-round para-graphic EQ (Fabfilter *Pro-Q 2*) and a second-round hardware emulation (Scheps 73) after two steps of compression and "de-essing"[61] (2020H). He discussed how this para-graphic EQ was intended "for clarity of signal more than anything. Just kind of cutting out unnecessary stuff" (ibid.). Although the cumulative complexity of vocal chains is another key characteristic of this production culture which illustrates its compulsive

[61] This will be explained further in Chapter 5.

tendencies, I only focus on para-graphic EQs here due to their widespread use at this initial, fundamental stage in vocal chains, and the way such usage seemed to be seen as functional and in service of vocal realism.



Celemony Melodyne: Tuning
Fabfilter Pro-Q 2: first para-graphic EQ
Waves C4: Multiband compressor
Waves RVox: Compressor
Fabfilter Pro-DS: De-esser
Scheps 73: EQ, rotary, hardware emulation (also pre-amp)
C1 comp-sc: compressor/compander
Fabfilter Saturn: distortion/saturation plug-in

Figure 50: Chris' vocal chain (inserts only) for Audio Track 2 (2020H)

With this in mind, Figure 51 shows a 'Vocal EQ Cheat Sheet', taken from a very basic pedagogical blog intending to provide a starting point for novice producers learning how to create para-graphic vocal EQ curves (Fitzjohn 2020). This infographic generalizes the vocabulary I want to describe. To be explicit: this infographic is heavily generalized and not entirely representative of the way my informants applied para-graphic EQs—I know for a fact that some of my more professional or older interviewees would cringe at such a graphic for its amateurish attempts to codify the practice and appeal to "bedroom producers".[62] However it usefully demonstrates the core ideas and terms involved, as well as illustrating clearly how the process of creating a para-graphic EQ "curve" conceptually splits up the vocal frequency spectrum into different frequency areas which correspond to assumed universal sonic characteristics in a voice, which are either emphasized or attenuated. This infographic also demonstrated how vocal EQ was undertaken with a standardized, template para-graphic EQ curve in mind, further portraying the technique's association with a kind of totalizing or

[62] This term circulated widely among many of those I interviewed.

generalizing power over vocal timbre. Even though the blog containing this infographic also contained short asides suggesting that in spite of this template every voice will be different, and that producers ultimately ought to 'use their ears' when EQing (in a way which ought to account confusingly for the 'biological timbre of the vocalist's vocal folds' (ibid.: n.p.)), the strength of this totalizing approach to vocal EQ is indicated by the prevalence of similar methods in other online pedagogical sources,[63] and its operation in commercial pop vocal editing was evidenced by my informants' espousal of a very similar 'zoned' approach.

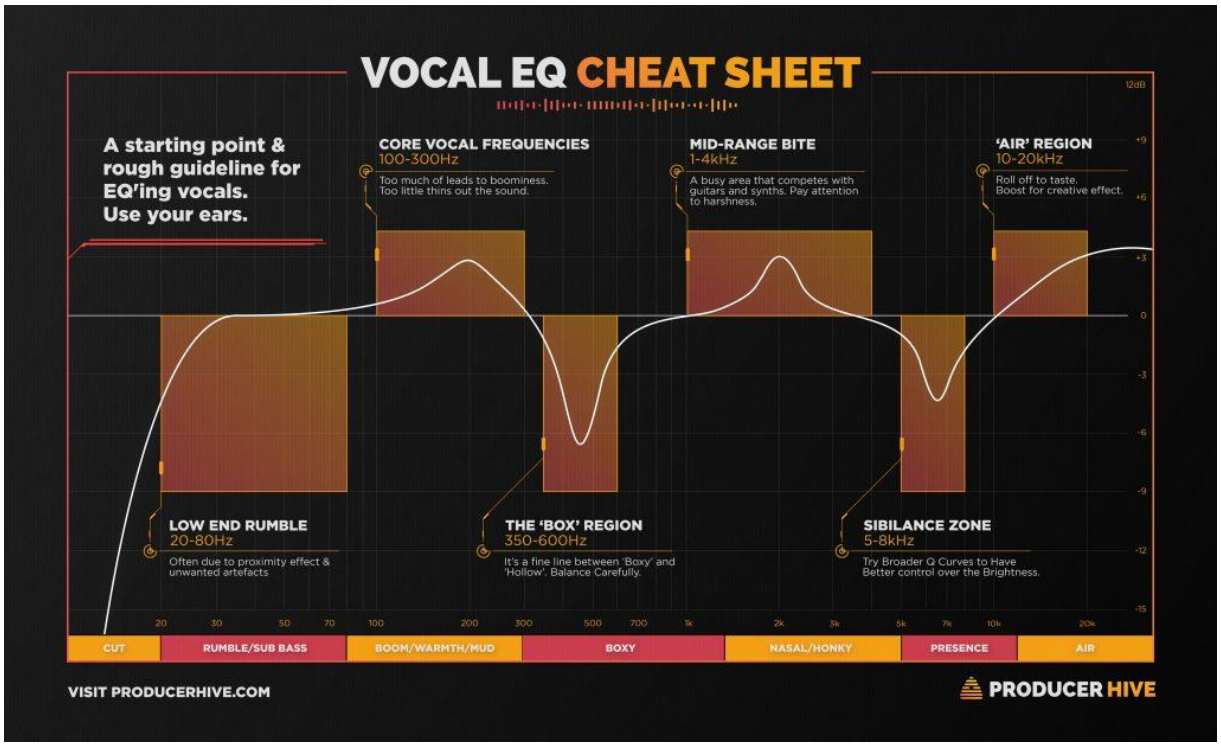


Figure 51: 'Vocal EQ Cheat Sheet' from ProducerHive (Fitzjohn 2020).

[63] For example, one blog claims that Vocal EQ entails focusing on 'five primary areas of modification' (Huff 2011) and another suggests similarly that there is a '6-step method' to professional vocal EQ involving similarly demarcated frequency regions (Jared H. 2021).

4.2.1. The “low cut”/“low end rumble”

As I have already mentioned, one of the most common EQ transformations made to the voice in this genre was the “low cut”, represented in this image as the “low end rumble”. The para-graphic EQ “curve” was used to create a high-pass filter, which “rolls off”[64] the lowest frequencies in the vocal—anything below around 80-100Hz. The quotes provided below illustrate some of my informants’ ideas about “low-cutting”, as well as demonstrating the extent to which it was a habitual, repeated practice in vocal EQing.

> [Q: Do you have a standard EQ curve that you use?] Yeah. So obviously a low cut. And then… (Tim, 2019CC).
>
> I’d say by habit I will cut up to 140Hz, because there is never anything happening that you want underneath that, in vocals. (Chris, 2020T).
>
> Straightaway, I’ll completely cut the lower sort of 80 or 100Hz, because that’s unnecessary stuff. You can always cut the low stuff out of the way because it’s just rumbly, muddy nonsense. (Robin, 2020I).
>
> The low cut I’ve got going on at 81Hz is just to remove any unnecessary low end. (Oliver, 2020S).

In short, my informants’ “low cutting” was guided by their belief that this area of the vocal frequency spectrum did not contain anything which significantly informed how the voice was perceived by a listener. Although only invoked implicitly, these ideas are based on assumptions about the most important spectral range for human auditory perception of the voice, regardless of whether such interpretations of perception match up with psychoacoustical research. One American producer who produces pedagogical YouTube videos elaborates on this reasoning, arguing that most of the “acoustic energy” in the singing voice is concentrated above a threshold of around 127Hz, so anything under this threshold will just be “rumble” or “room tone” generated during the recording process (Help Me Devvon 2020). The infographic terms such unnecessary features “unwanted artefacts” and “proximity effect”.[65]

[64] Another term which circulated widely across the internet materials I studied (e.g. Pro Audio Files 2020; Weiss 2017)

[65] To be clear, this refers to the specific technical term described in Chapter 2 (p. 118) rather than the desirable sense of proximity underpinning vocal foregrounding.

Similarly, two quotes included below (one from a YouTube tutor, and one from an informant) show how "low cutting" more accurately serves the specific aesthetic technique of 'carving' out a special space for the voice within the overall "mix". The autonomy or centrality of vocal sound within a mix space—ideas which underpin the concept of vocal foregrounding—are aesthetic ideals underpinning the logic of low cutting, even though producer-songwriters also linked the practice to these universalizing ideas about perceptual irrelevance. This again shows how the practice of commercial pop vocal EQ implicitly conflates notions of vocal realism with the aesthetics of vocal foregrounding.

> I just cut out all of the low end so there's no mic rumble or A/C rumble […] That way it's not taking up too much space in my mix. The more low end you have the more it'll start feeding into your final limiter and everything else just gets kind of crowded and cluttered and it gets a bit wonky. (Austin Hull, Make Pop Music 2019)

> As soon as you've got a vocal all kind of round the 100-200Hz range—where you get that little lump in the EQ—that's a nightmare. […] It's just all… the guitars are there, the keyboard's there, the vocal's there… all in that range, and it's really a pain in the ass to mix. (Robin, 2020I)

Further, notably missing from these quotes is the fact that the "low end rumble" which these producers consider perceptually irrelevant, and later invasive in the final stages of mixing, cannot be audibly detected in initial vocal recordings unless it is amplified. This anticipatory quality—a kind of pre-empting or acting on some problem that is not yet audible but will likely generate potential issues in further mixing—was a key characteristic of my informants' vocal editing practices that emerges throughout this chapter and Chapter 5. In the case of "low cutting" however, this belief in an imperceptible sonic defect was so strong and so fixed that most of my informants engaged in the technique habitually, without thinking, and never attempted to try and locate, amplify and hear the "low end rumble" which was understood as problematic. Again, this was facilitated by the interfaces of para-graphic EQ plug-ins which made the visualization and carrying-out of "low cutting" incredibly easy.

4.2.2. “Core vocal frequencies”/“body” in the lower range

The second area highlighted by the ProducerHive infographic is the “core vocal frequencies”, or what my informants more frequently termed the “body” or “fullness” of the vocal sound. Although the infographic suggests that this area can be boosted, most of my informants didn’t make any huge boosts in this area. Rather, they tried to keep it relatively intact or tried to be conservative with their boosts and cuts in the interests of retaining this “body” as far as possible. Another common sentiment expressed both by my informants and by YouTube pedagogues was that cutting away at this “body” too much would cause the vocal to sound “thin”, or would “weaken” the overall vocal sound:

> I think it might be too thin. […] where have I gone wrong… […] I’ll get a little bit of body back in there, I might be taking too much out. [He uses two EQs and a multiband compressor to reduce existing cuts in the 100-300Hz range] (Oliver, 2020Q).
>
> The very top part [of the vocal frequency spectrum] is important. You need to make sure that’s there, then you need all the body, the stuff down here [in the 100-300Hz range]. (Jason, 2019Z).
>
> You’re only going to be taking those frequencies out when it’s too much. And you want to leave everything else untouched. Otherwise it’s too much… that’s what makes it sound wrong, cause it just, it weakens the overall [vocal] sound, because you’ve basically killed some of the frequency spectrum. (Robin, 2020I).

Again, these ideas about “core vocal frequencies” and vocal “body” needing to be preserved showed how my informants were guided by ideas about which frequency ranges of the voice are the most important when perceived by an archetypal listener. The term is also significant because it makes a direct link between EQ modifications and the idea of a physical singing body producing a universal vocal sound. The metaphor of vocal “body” residing in a certain frequency range additionally becomes even more evocative in the context of the central role played by vocal timbre in erotic voice fetishism, as it appears to locate the voice’s physical essence not only in vocal timbre but in a very precise subsection of the archetypal vocal frequency spectrum. Robin’s quote, in particular, makes the corporeal flavour of the metaphor even more extreme, suggesting that the attenuation of these frequencies “kills” some of the vocal frequency spectrum.

4.2.3. “Boxiness” in the lower-mid ranges

My informants frequently encountered problems with so-called “boxiness” in the voice within the third region addressed by this infographic—the “box” region. Generally, ”boxy” seemed to refer to a subtle “telephone”-like sound in the voice, the source of which was, apparently, too much energy in these lower to middle frequency ranges: anywhere from 200-900Hz.[66] I also heard the term explained with reference to a sung or spoken tone produced when the hands were cupped over the mouth, associated with the idea of the voice sounding “covered” (Help Me Devvon 2020). This was inconsistent, however, as some informants and sources tended to talk about this “covered” kind of sound as if it were a completely different phenomenon to “boxiness”. As shown in the quotes below, one producer-songwriter also consistently equated the notion of “boxiness” with the widely acknowledged phenomenon of “proximity effect”[67] in audio recording, as well as with this notion of the vocal sounding as though it was sung into cupped hands, and I also witnessed the terms “muddy” or “cloudy” being used to describe similar ranges:

> [Using a preset called ‘de-mud’ to attenuate around 350Hz] I’m just using this to cut out that boxiness. (Chris, 2020H)
>
> The vocal feels messy, cloudy, like singing into a paper cup. (Help Me Devvon 2020)
>
> 263Hz was where I felt the proximity effect was most predominant. So I was just trying to get rid of some of that like… proximity. Some people use the term boxy. And then in the midrange at about 1156.1 ish I felt the vocal [makes cup with hands around mouth] ‘sounded a little bit like that’… so that just helps try and eliminate some of that telephone-y sort of sound and make it sound a bit warmer and a bit thicker I guess. (Oliver, 2020S)

The undesirability of “boxiness” therefore did not appear to be as strongly linked to vernacular assumptions about listener perception of voices as the previous terms, even though vocal realism still seemed to underpin the term more generally through its implication that the true vocal sound was being “covered” over or “muddied”, and EQing to remove “boxiness” therefore performed a restorative function. The term “boxy” also revealed more clearly how EQ was used to eliminate any

[66] Although the infographic pinpoints this region at around 350-600Hz, in my fieldwork the term cropped up in reference to points on a wider range from around 200-1000Hz as shown by the quotes in this section. Waves, for example, mark this region at between 250-900Hz (Waves, 2019).
[67] Refer to footnote 65.

intrusive vocal features which might disrupt the ideal aesthetic of vocal foregrounding, despite its widespread interpretation as a technique for preserving vocal realism. “Boxiness” was associated with the voice sounding distant and muffled, which directly impacted the sense of intense vocal proximity these producer-songwriters were trying to cultivate.

4.2.4. “Presence”/”harshness” in the mid-range

Moving into the upper parts of the vocal frequency spectrum, the following two regions—the “mid-range bite” and “sibilance zone”—revealed how the implicit aims of vocal foregrounding in EQ created undesirable sonic effects which producer-songwriters felt must be attenuated further using EQ. This demonstrated how the conflation of vocal realism with vocal foregrounding resulted in contradictory tensions within the practice.

Paradoxically, in the case of the “mid-range bite” region, the undesirability of such effects was justified with reference to their disruption of vocal realism, even though the process of exaggerating desirable characteristics already seemed to be understood as cultivating vocal realism. For example, my informants described how making boosts in this region could give a desirable sense of “presence” to a vocal, which directly related to EQ’s aims to cultivate vocal foregrounding. However, as the infographic warns, too much boosting in this range could also lead to undesirable “harshness”. Resulting from this tension, my informants made varying modifications to this range: sometimes subtle, wide, boosts in the name of enhancing “presence”, and occasionally narrow cuts to target particularly “harsh” frequency points. Referring to Figure 48, for example, you can see how image 5 makes a very slight, wide boost to this range, and images 8 and 9 make tiny, targeted cuts in the same range.

“Harshness”, was defined by informal pedagogical sources with reference to not only vernacular ideas about listener perception but also the physiology of the ear. The quotes below, for example, show harshness being defined as an amplified frequency zone that will, over continual listening, cause the listener “ear fatigue” or “listener fatigue” (see also Huber & Runstein, 2018: 418-317). My informants only indirectly referenced these ideas about pain and fatigue (e.g. see Chris’s

quote below), however they expressed unanimously that "harshness" in vocals was considered particularly undesirable and intrusive.

> This area between like 2-3kHz, you gotta be careful […] if you do boost that area you're going to get a harsh, pingy sound that's going to hurt a lot of people's ears at higher volumes. I don't like to boost in that area [because] when they get a little beltier it gets a little bit harsher. (SoundOracle 2018).
>
> [discussing the range around 2.7-3kHz] I try to take this frequency out early on in mixes so it doesn't cause too much pain in the ears. You will take your headphones off after listening over time if you hear this too much. That frequency range is… once it starts getting boosted it will cause pain and fatigue on your ear. (Help Me Devvon 2020).
>
> Even on a subconscious level if you've got like a niggling 1kHz frequency throughout a whole track, you might not even hear it but eventually its just going to be tiring to listen to. […] And to someone who doesn't know music it can almost go as far as 'I don't really like this song'. (Chris, 2020T)

These statements attempt to justify universalizing ideas about auditory perception by linking them to actual physical pain in the ears. Importantly, the notion of too much high-frequency energy causing auditory discomfort or "fatigue" is obviously not universal—it's not possible to define how much is 'too much'. Indeed, the notion of "ear fatigue" is a term rooted more in music production and audio than it is in psychoacoustic research or clinical diagnoses. The terms "ear fatigue" and "listener fatigue" circulate widely in audio engineering discourse in relation to prolonged mixing sessions (e.g. Waves 2018), however such terms do not appear to be scientifically recognized: the temporary hearing damage or loss resulting from exposure to loud sounds is usually discussed in terms of Threshold Shifts in psychoacoustics literature (see Ryan, et. al. 2017), and I am unaware of any scientific terms which address the notion of pain resulting from hearing too much high-frequency energy. This is therefore an extreme example of how commercial pop vocal EQ prioritizes ideas about how "real" voices sound: here, such ideas are figured in terms of how an archetypal listener can most comfortably hear a voice in a commercial pop track. These appeals to universal listener perception have little basis in scientific research and therefore further demonstrate the culturally contingent qualities of realist discourse in commercial pop vocal EQ.

Regarding the term "presence": thinking back to my discussion of this term in Chapter 2, it was used here to refer to a specific result of EQ modifications to this range ("mid-range bite"), even though the term was also used more broadly throughout the aesthetic discourse of vocal editing to

refer to a variety of aesthetically valued vocal traits. Although the infographic only pinpoints the location of this feature around 5-8kHz (at the bottom of the diagram), the term came up in reference to a much broader frequency range in my informants' discussions of EQ—anywhere from around 1kHz to 8kHz—which further illustrated the ambiguous pervasiveness of the term in the aesthetic discourse.

4.2.5. "Sibilance"

The fifth range, what the graphic terms the "sibilance zone", is where most of my informants pinpointed the high-frequency energy associated with sibilant consonants in the voice—"esses" (also including some other fricative consonants which fall under the rubric of sibilance e.g. "sh"/ [ʃ]). In a slight contrast to "body", "presence" and "harshness", the range associated with vocal sibilance was more consistently located by my informants as between 7-10kHz.

Across my fieldwork, "too much sibilance" was a universally undesirable vocal quality that my informants consistently endeavoured to attenuate when EQing the voice. Notably, however, the exaggeration of vocal sibilance is a product of pop vocal recording and editing's aiming towards vocal foregrounding: a commentator from Sound on Sound, for example, argues that amplified sibilance is an undesirable consequence of "modern" routine production decisions such as close-miking vocals with specific microphone types which tend to 'emphasise [esses] *unnaturally*' (Senior, 2009b, italics added), and, as I will show further in the following paragraphs, some informants discussed how EQ modifications made to the other parts of the curve in the name of boosting "presence" or "air" had the residual effect of amplifying sibilance. More generally, and as the quotes I have included below show, "esses" are ostracised for similar reasons to the mid-range "harshness" that I just described—which rely on the idea that too much "harshness" tires the ears. Indeed, sibilance was frequently described by my informants as "harsh"-sounding in the same way. The attenuation of sibilance also served the principles of 'carving': as Robin claimed in the quote below, sibilance has the potential to conflict with other mix elements containing higher-frequency energy (e.g. hi-hats). My informants often used para-graphic EQs to attenuate the sibilance range alongside using a dedicated "de-esser" plug-in at a later point in the vocal chain. This will be discussed further in Chapter 5. At this stage, however, the

use of dedicated "de-esser" plug-ins in this practice illustrated the extent to which sibilance—clearly a product of the desire for vocal foregrounding—was considered an invasive vocal problem that needed addressing by targeted EQ modifications.

> I wouldn't really recommend too much processing on the higher end of the frequency spectrum of the vocal, other than a de-esser just to take away some of the harsh silibances [sic] because they will clash with hi-hats. (Robin, 2020I).
>
> That plug-in's getting rid of some harshness in the top end, and it's the same with that de-esser… (Oliver, 2020S).
>
> [You should] boost the top end without it becoming incredibly harsh and sibilant. If you just lump top end on everything you're gonna get a load of sibilance and its gonna unbalance the mix. (Michael, 2020E).

4.2.6. "Air"

The last region on the graphic points to the uppermost ranges of the frequency spectrum—around 10kHz—which my informants associated with the addition of "air" to a voice. The high-frequency boost in this range was also described by my informants using other terms such as "sparkle", "brightness" or "lift". Although the infographic suggests a lower limit of 10kHz for this characteristic, I heard the term "air" applied to lots of different EQ boosts in the upper ranges of the frequency spectrum, from around 4kHz upwards, suggesting that the term may overlap somewhat with the notion of "presence". Most importantly, my informants linked the notion of "air" to ideas about how the voice was usually perceived in close proximity to a listener, illustrating even more explicitly how realism and the notion of vocal foregrounding seemed bound up together in the practice. For example, Jason described how the uppermost ranges of the vocal frequency spectrum were a subtle but essential detail differentiating between "close" and "far away" vocals:

> This very top part is important. The high frequency stuff above like 8-9k – that's all the detail that makes [the voice] sound close. If someone's far away that's the part that dies away quickly in real world situations. (Jason, 2019Z)

Further, and similar to the way "presence" was thought to lead to "harshness", "air" also existed in an interlocking relationship with "sibilance". Oliver, for example, described how his attempt to add extra "air" was likely to lead to sibilance problems during further mixing, because amplification of the higher-frequency energy associated with "air" might also amplify the unwanted sibilance.

> The boost at 6130Hz… that's probably just to try and add a little bit of extra air. But obviously… looking at that, that's probably about where the sibilance is hitting as well, so I'm probably giving myself more de-essing problems moving forward. (Oliver, 2020S)



Figure 52: Oliver's EQ curve, with the boost for "air" at 6130Hz circled.

My informants' discussion of "air" also revealed how they chose not to use EQ to cut away the highest part of the vocal frequency spectrum in the same way as the lowest part. This exposes an inconsistency in the logic of realism or perceptual irrelevance underpinning the practice of "low cutting". If, as one of my field sources claimed earlier, most of the "acoustic energy" in the human singing voice is concentrated in the middle ranges (around 100-300Hz specifically) meaning that the lowest ranges can be removed, then one wonders why such logic is not extended to this much higher range, which could also be cut to contribute even further to 'carving' techniques. But, although the ProducerHive graphic suggests that the "air" region can be "rolled off to taste", this was seldom the approach taken by any of my informants. Robin, for example, described how he had seen other

producers do this (what might be called a “high cut”), but believed that it was an amateur, ill-informed approach to para-graphic EQing, because it took away some of the most important and valuable features of the vocal sound (“air”, “sparkly stuff”):

> A lot of people will automatically do a high cut as well, at sort of above 20kHz, ‘cause they think that’s beyond the range of human hearing so it doesn’t matter. But actually, it kills a lot of the sound of the vocal. Because even me talking now, there are harmonics and subharmonics up that end of the scale which are actually affecting the resonance of my voice. And if you take that away, the sound’s different, and it just makes everything dull. It’s actually better to keep that sparkly stuff. (Robin, 2020I)

This inconsistency suggested that notions of realism were selectively used to describe EQ techniques in different areas of the vocal frequency spectrum, only when they did not conflict with the practice’s implicit aims of vocal foregrounding. In the case of vocal “air”, notions of the value of vocal “detail” and “closeness” were important, but such ideas contradicted the realist premise of the core vocal frequency ranges needed for perception which underpinned “low-cutting”, exposing that the language of vocal realism was a specific way of describing EQ in this culture.

4.2.7. Summary

To summarize, I have used this infographic to illustrate how my informants’ use of para-graphic EQ plug-ins was guided by a belief that vocal frequency spectra are divided up into specific ‘regions’ corresponding to specific desirable and undesirable vocal characteristics, conceived predominantly in terms of the cultivation of vocal realism. “Body”, “presence” and “air” were conceived as desirable realist characteristics that ought to be either preserved or emphasized, whereas “low-end rumble”, “boxiness”, “harshness” and “sibilance” were undesirable, intrusive and in need of attenuating. Among the desirable characteristics, “presence” and “air” most clearly referenced the exaggerated sense of vocal foregrounding that is so foundational to this aesthetic, despite the dominance of ideas about vocal realism, showing how these two ideas were blended in the practice. Further, the fact that the attempt to cultivate supposed vocal realism throws up new problems (“boxiness”, “harshness”, “sibilance”), which are themselves considered intrusive to vocal realism, suggests that the impression of vocal foregrounding has been naturalized in this genre to the extent that the techniques used to achieve it are conceived in realist terms. In addition, the fact that most of these characteristics are

defined with reference to informal ideas circulating among producers about how the human ear interprets the vocal frequency spectrum adds an additional quasi-scientific layer to this discourse—it naturalizes the aesthetic specificity of vocal foregrounding with reference to supposedly objective notions of how the voice is universally perceived.

On a side note, my informants seemed to have picked up their para-graphic EQ techniques uncritically both from online materials and university training. Although my informants suggested in follow-up interviews that their initial development of EQ techniques was a very intuitive, "reactive", "trial-and-error" process for them—which involved listening to different voices and trying to EQ them to "make them sound better"—it was clear that such tacit knowledge had been unconsciously absorbed from external sources:

> [My mentor at university] was big on the ol' (covers mouth with cupped hands) 'you don't want the vocal sounding like this' […]. And I think the rest was sort of like EQing any other instrument you know, you get the vocal up and do what you feel you need to do to make it sound better. Obviously, you know you've got to deal with some proximity, you've got to get rid of some low end, you know you want to add some brightness. But it's just, each vocal's different init, so it's just figuring out the way you want to do that. (Oliver, 2020S)
>
> I think its just from doing it, you know. I remember when I first started out I used to spend hours trying to work it out. To reach this curve. It's so dependent on the vocal, there's no real like hard and fast rules really, apart from I guess loosely those areas. I think it just comes from trial-and-error, and reacting to what you've got recorded. […] I think it's partially information that I've taken in that's like "500Hz is a bad frequency!" that has a naturally like, boxy, sound to it that has no place in a pop lead vocal. It just sounds… wrong. I couldn't really tell you why, it's just like that. It doesn't want to be there. (Chris, 2020T).
>
> So the initial placement is like habit, right. 140 needs to go, 250-300 is going to be problematic, 500 is going to be problematic… so I like pop those in, and then tweak according to what I'm hearing… (ibid.)

Oliver, for example, makes a direct link to one of his university lecturers who taught him to pay particular attention to the vocal sounding as though it was in cupped hands, and Chris's description implicitly links his EQ practices to a period of pre-university self-teaching (using internet materials, particularly YouTube videos, which we discussed later in the interview) (2020T). The speed with which my informants jumped to talking about the exact, clearly demarcated areas within para-graphic EQ curves also showed how such a conceptualization of the vocal frequency spectrum seemed ingrained rather than having arisen organically or reactively from listening as my informants claimed.

Clearly, the idea of EQ being a process of finding out "just what sounds good", innately, in the voice was influential in this discourse, even though my analysis of the codified regions of para-graphic EQ curves shows how they are mostly specific to the aesthetic of vocal foregrounding. This suggests that another effect of the pervasive realist discourse in this practice is that it makes producer-songwriters think that they have just happened upon these EQ techniques naturally, even though they clearly embody a specific approach that is taught pervasively throughout this culture.

The cultural specificity of realist discourse in commercial pop vocal EQ

Notably, the cultural specificity of realist discourse in commercial pop vocal EQ is illustrated by comparison with Bates' work on EQ practices in Anatolian folk music studio production (2010, 2016). Like my informants, Bates' studio engineers used EQ to exaggerate high-frequency energy, however this exaggeration was usually applied to both lead vocals and instrumental timbres for slightly different aesthetic ends. Bates' engineers in Istanbul call this desirable aesthetic *parlak* ("shine"), and this concept recalls the notions of "air", "brightness" and "presence" that my informants use. *Parlak* is viewed as an exaggeration of the perceived brightness already inherent in the timbre of Anatolian folk instruments (2010: 98). This discourse, where the aesthetic effect is perceived as an exaggeration of a sonic quality innate in the instruments themselves, is quite similar to the EQ discourse I have been discussing, where EQ is frequently conceived as a functional method of heightening or exaggerating the voice's inherent sonic qualities (even though these qualities often relate to the notion of imagined close proximity to a listener). As I have attempted to show, however, my informants working in commercial pop were not as forthcoming about this tendency to exaggerate specific qualities for aesthetic effect as the engineers in Istanbul studios. Bates discusses how Anatolian folk music engineers simply try to boost the high frequencies as much as possible to emphasize the *parlak* effect, an integral component of which is the phase distortion resulting from an extreme high-frequency boost in the 1.5-2.2kHz range (Bates 2010: 99-100). It also appears that in Anatolian folk music studio production the concept of *parlak* is more specifically linked to ideas about "brightness" and "shine", rather than being linked to notions of foregrounding within a mix space. Figure 53 below, which shows typical *parlak* EQ settings, suggests that this approach to

emphasizing high-frequency energy is also a lot more simplified and extreme, and doesn't have the intense, obsessive levels of complexity and dividing up of the vocal frequency spectrum into different regions that commercial pop vocal EQing has.

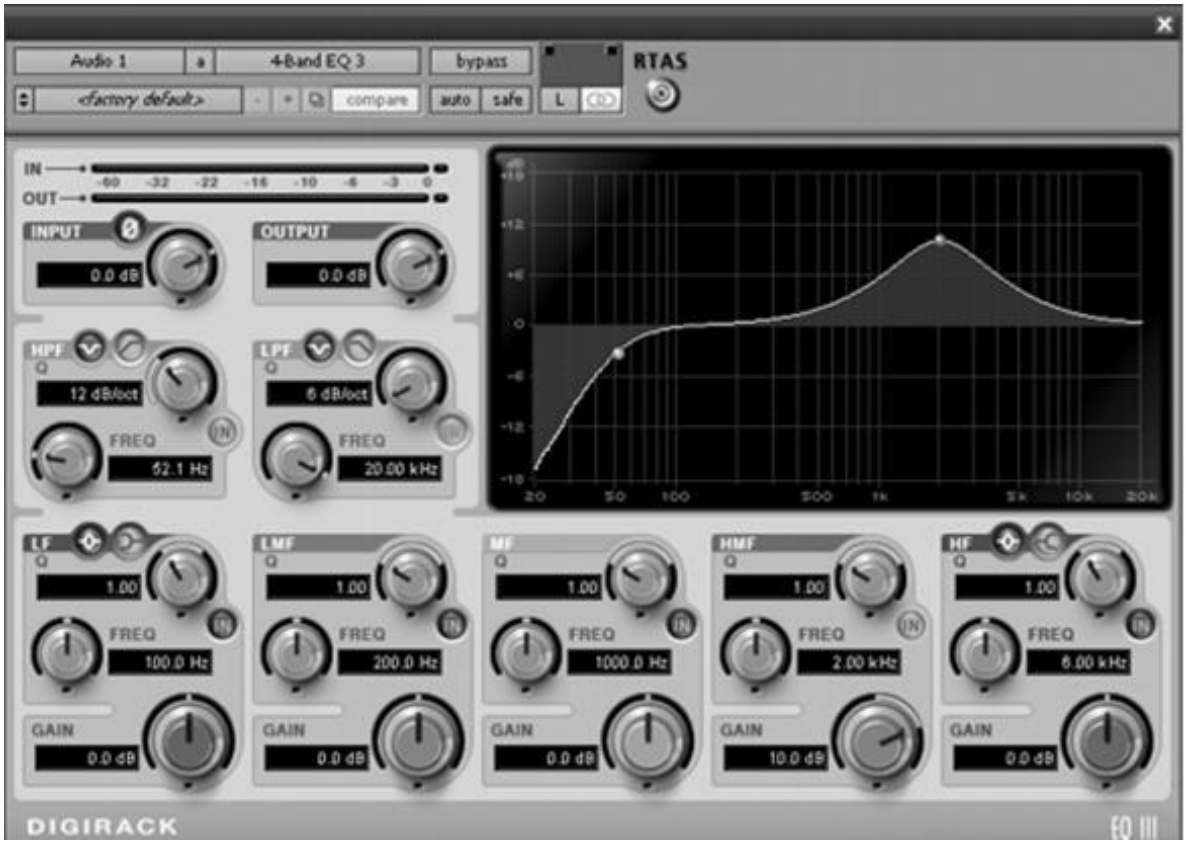


Figure 53: Typical 'Parlak' EQ settings, taken from Bates (2010: 99).

This brief comparison lends greater weight to my claim that the discourse of realism surrounding commercial pop vocal EQing relates to a key component of the erotic fetishism of the voice in this culture. As I have explained, this kind of fetishism is founded on the assumption that there are strong links between the voice—or more specifically vocal timbre—and essential human identity, physicality and uniqueness. Given that EQ provides a powerful way of modifying the over-arching timbral qualities of the voice, this technique therefore seems to give the user a direct power to manipulate the essential qualities of the voice which define this kind of voice fetishism. This interacts with the existing commodity fetishism of digital novelty para-graphic EQ plug-ins, leading to the technique's interpretation as a powerful, almost quasi-scientific way of preserving and enhancing these "real"

qualities. Bates' culture already has the commodity fetishism of the technology, but the instrumental context lacks erotic voice fetishism and therefore also lacks similar realist ideas.

4.3. Vocal realism, surgical metaphors and gender

"Subtractive" vocal EQing: vocal realism and surgical metaphors

Ideas about vocal realism also came to the fore in my informants' discussions of the distinction between "subtractive" and "additive" EQ (or "cutting" and "boosting)—a distinction which some of my informants used as a way of distinguishing professional from amateur vocal editing. Although the ProducerHive infographic gives the impression that para-graphic EQ practice consists of oscillations between cutting and boosting in different areas, my informants felt that a "subtractive" approach to para-graphic EQ—i.e. mostly cutting, and no boosting—was more professional, because it was more effective at preserving vocal realism. I use the terms "subtractive" and "additive" because they circulated widely among both my informants and online materials. Not insignificantly, this language also evokes the vocabulary of cosmetic surgery, and, as I will describe, my informants occasionally used the term "surgical EQ" to describe their "subtractive" para-graphic EQ modifications. The presence of this surgical language makes realist discourse in vocal EQ more extreme by creating a direct link between vocal EQ practices and medical interventions in the body. My analyses of this surgical language also suggest that the prevalent discourse of vocal realism in EQing may have subtle consequences for the gendered power dynamics between producer-songwriters and vocalists in vocal editing, relating to EQ's conception as a tool for diagnosing and repairing, or perfecting, the voice.

The following quotes are taken from interviews and observation where my informants Chris and Jason were discussing, in very similar ways, how they aimed to be as conservative as possible with additive EQ ("boosting"), instead endeavouring to prune the lower parts of the vocal frequency spectrum using subtractive EQ ("cutting"). Examples of Jason and Chris's para-graphic EQ curves, shown in Figure 54, illustrate these approaches, which involved, alongside a low cut, making several other small cuts in the lower to middle ranges of the vocal frequency spectrum.

> If you boost [the 3-5kHz range] it sounds horrible, but you need it in the mix to make it cut through. The way to get it isn't by boosting, the way is by distorting the lower part [of the frequency spectrum]. You want it to sound real, like a person. What people do with EQ is mess around trying to get it to sound ["present"], but it never sounds like that, because they're boosting all these horrible frequencies that you wouldn't hear in real life. (Jason, 2020Z)

> If I want the voice to be brighter it's more about cutting away the dark stuff. When you're boosting stuff… to me, anyway, it just ends up sounding a little bit harsh. Like they're making a frequency *more* that doesn't exist. Whereas when you're cutting, you're kind of removing something that does exist, if that makes sense. […] Cause all of those frequencies are there, you know like up to 20kHz. They still exist in the voice. And by lowering everything else, you're kind of bringing them out. (Chris, 2020F).

These statements show how both producer-songwriters viewed subtractive EQing as a more effective approach because they felt it didn't unnecessarily manipulate the most *real* aspects of the vocal frequency spectrum ("you want it to sound real, like a person"; "they're making a frequency *more* that doesn't exist").[68] Of course, the irony is that both subtractive *and* additive EQ obviously distort the vocal frequency spectrum in some way, so such statements show how the notion of a "real" voice is specific to this aesthetic context, which involves exaggerating the impression of vocal foregrounding. Here, it is only the high-frequency energy associated with proximity that indicates vocal realism.

Alongside this selective realism, these quotes also show how both Chris and Jason think that the aesthetically desirable exaggeration of high-frequency energy can be achieved *indirectly* according to the logic of carving, because pruning away at the lower ranges of the frequency spectrum will "make space" for the higher frequencies. A similar position is espoused by celebrity engineer Tony Maserati,[69] who, in the quote given below, praises the indirect effects of subtractive EQing on high-frequency content in mixes, arguing that cutting the lower parts of the frequency spectrum is equivalent to dramatic and aesthetically effective high-frequency boosts. The logic is explained in

[68] Chris and Jason's quotes also suggest that the idea that "subtractive" EQ "preserves" the "real" vocal frequency spectrum also functions as a way of distinguishing amateur from professional vocal editing. The two producer-songwriters attempt to differentiate their EQ practices from other less-informed producers more generally ("What people do with EQ is mess around…", "they're making a frequency *more* that doesn't exist"). I direct the reader to work by Thomas Porcello (2004) for a more in-depth study of the ways talk about timbre in recording studios reinforces professional boundaries among studio engineers.

[69] I am purposefully referencing Maserati's comments because he has a very public persona in the plug-in industry, specifically via his partnership with Waves (who have released a Tony Maserati branded plug-in series: https://www.waves.com/bundles/tony-maserati-signature-series). Because of this, it is likely that my informants had directly or indirectly engaged with his ideas. I elaborate further on the notion of 'celebrity engineers' and their role in plug-in marketing in the next chapter, particularly regarding rock engineer Chris Lord-Alge's position in discourse about vocal compression (pp. 262-266).



Figure 54: Chris and Jason experiment with surgical EQ (2020F, 2020N). Combinations of coloured dots and notches in the curves represent boosts and cuts. The dots are not always connected to the curve because FabFilter *Pro-Q 3* has a special dynamic EQ function which I will explain later in the chapter (pp. 232-235). (Note that Chris had FabFilter DS open at the same time which covers the lower right corner of the *Pro-Q* window.)

slightly more depth by Maserati, who suggests that by removing lower-end frequency content, "subtractive" EQ prevents low-frequency energy from "building-up" later in the mixing process (Waves 2020), and the implication is that such a build-up can negatively affect the desirable high-frequency "brightness" of a vocal. Although it is impossible to know whether such techniques actually figure in Maserati's own mixing practices,[70] the similarity to my informants' expressions is undeniable, and suggests that the notion of subtractive EQ being indirectly equivalent to higher-end boosting is an informal theory which circulates widely in this culture. More generally, such an assumption illustrates how the practice aims to hide the aesthetic effects of EQ as much as possible, not only under a language of vocal realism but also by using EQ techniques which attempt to achieve such effects indirectly.

> Subtractive EQ'ing will save you. A deep reduction at 500Hz is equal to a huge gain at 2kHz and above. I usually start [with] subtractive EQ on the tones I don't like or [which] make it harder to fit [the vocal] in the mix. (Tony Maserati, cited in White 2012).

Across my fieldwork, the term "surgical EQ" was occasionally used to refer generally to any kind of functional, instrumental use of para-graphic EQ plug-ins, usually for subtractive EQ procedures. The term was more commonly used in pedagogical materials online, however the techniques and ideas associated with the term were similarly invoked by my own informants. Broadly, the term encapsulates a kind of medicalized, diagnosis-and-repair approach to EQing the voice, where para-graphic EQ interfaces are used explicitly to identify "problematic" or "annoying" frequency bands or "resonances" that are subsequently smoothed out or attenuated using the EQ curve. YouTube pedagogue Matthew Weiss, for example, in the quote below, uses the term when trying to replicate vocal mixing techniques by a renowned pop producer Serban Ghenea[71] (Pro Audio Files 2020). This video is intended for amateur producers intending to replicate a quintessential pop vocal sound—represented by Ghenea's work—and Weiss is discussing how Ghenea tends to zone in carefully on precise frequency bands with a para-graphic EQ, performing "surgical cuts" to "clean up" the vocal.

[70] Similar sentiments were expressed by Maserati in real time on a live streamed mixing session that I observed in June 2020.
[71] I refer to this video because one of my informants mentioned Ghenea as a model vocal editing specialist whose "sound" they aspired to emulate in their own vocal editing (2020E).

This surgical language is not insignificant in the context of a working relationship where aesthetic and technological control over a vocalist is exercised by a producer-songwriter—a relationship which is also often gendered. The notion of "cleaning up" a voice makes it very clear that this "surgery" is aesthetic or cosmetic rather than medical, and the implication is that producer-songwriters have power to beautify the voice beyond that which the vocalist can achieve in performance.

> Basically there's [sic] these two tones that really kind of messed up the sound, and so pulling them out definitely cleaned [the vocal] up a lot, and I think that Serban is really thoughtful and careful about how he does his surgical cuts. (Pro Audio Files 2020)

One EQ procedure which was particularly emblematic of a "surgical" approach was the commonly termed "search and destroy" technique,[72] a method for isolating and attenuating undesirable frequency bands. This involved using a para-graphic EQ to temporarily make a very pronounced, tight boost on the EQ curve and slide it across the frequency spectrum, scanning to pinpoint a particularly disagreeable frequency band (or "resonance"), which would then be attenuated slightly in a very targeted way using a tiny, precise cut. Jason used this technique frequently, and I also observed it across multiple online sources (e.g. Pro Audio Files 2020, White 2012). The technique relies on the assumption that any particularly intrusive "resonances" in the vocal frequency spectrum cannot ordinarily be heard, but later editing processes (e.g. compression) might amplify such "problem frequencies" later on in the mixing process and so must be sought out via amplification to quell them in anticipation. Like the "low cut", this is another anticipatory tactic where problems in the voice are assumed to exist and sought out before they are actually detected audibly. Of course, most of the time these cuts are hardly noticeable, since the offending frequency has only been made audible by this searching process: yet there was this adamant sense across these sources that scanning for and "surgically" removing these hidden frequencies was vital for ensuring the overall quality of the voice. Although such a technique was not unanimously used by my informants, the mere existence of such a

[72] It's worth clarifying that this technique is not specific to vocals—one of my informants, for example, also described how he sometimes uses a similar approach to address difficult frequency resonances in drum parts (2020T). However, its frequent use on vocals is noteworthy specifically for its similarity to medical processes in the context of erotic voice fetishism and its assumptions about links between voices and bodies.

process is significant in the context of ideas about "subtractive" vocal EQ as a process in service of vocal realism, because it is reminiscent of medical intra-bodily scanning processes that reveal problems that cannot easily be observed externally.

> What some people do is kind of sweep through and listen for the bit that sounds the most harsh… [Jason proceeds to do an extended demonstration of the "search & destroy" technique, by making a very big boost, first pausing around 700Hz and cutting out something, then sweeping across between 1kHz – 5kHz and listening to locate any other "harsh" parts.] (Jason, 2020N)

To summarize, producer-songwriters uphold the "subtractive" approach as a more appropriate vocal EQing method because they think it is more sensitive to and preserving of the "real" voice. Related to this, discourse about "subtractive" EQing is saturated with metaphors of surgery: with the help of the para-graphic EQ interface, frequency bands which are considered problematic or intrusive are diagnosed, pinpointed with precision, and "surgically" attenuated. The quasi-scientific, medicalized flavour of this practice adds a new dimension to my argument that commercial pop vocal EQing is understood as technique which preserves vocal realism, even though it is geared towards cultivating a culturally specific aesthetic of vocal foregrounding. It suggests that such metaphors heighten the sense of control that producer-songwriters' have over the voice—and by extension, the vocalist—through the manipulation of vocal timbre; a sense of control and authority which also often implies a gendered power dynamic. Such metaphors also show with even greater richness how the erotic fetishism of the voice underpins vocal EQ practices: through its entanglement with notions of surgery, EQ, unlike the other techniques I describe, makes a strong link between the singing voice and an imagined singing body, to such an extent that vocal EQing interventions are conceptually aligned with ideas about physically cutting into such a body in order to transform it to suit the aesthetics of commercial pop.

"Presence", "harshness", and gendered interpretations of vocal realism

As I mentioned earlier, my informants found it difficult to achieve the desirable qualities of "presence" (along with "brightness" and "air") without letting the high-frequency boosting become too "harsh". I expand on this theme here as it seemed a particular cause for inconvenience in the vocal editing process. Specifically, during my fieldwork, discussions of the "harshness" issue frequently invoked essentialized differences between male and female vocals. This revealed how my informants'

conceptualization of EQ as a technique for refining the real or innate qualities of voices occasionally obscured universalizing assumptions about gender differences.

Links between the notion of "harshness" and gender gradually peaked my interest as I began to notice that discussions of "harshness"—both by my own informants and within pedagogical materials on YouTube—were often accompanied by statements regarding the role of the gender of the vocalist in determining the exact location of "harsh" resonances in the voice. This is demonstrated by several quotes from YouTube sources below. According to these producer-songwriters, "harshness" is something which is *especially* problematic with female singers:

> The issue that I've always had, *especially with female singers*, of why I don't like to boost in [the mid-range], is when they get a little beltier […] it gets a little bit harsher. (SoundOracle 2018, italics added)
>
> We've all wrestled with this in a mix: we want to get a vocal, *particularly a female vocal*, nice and bright and forward in the mix, and all the things we try end up making the vocal harsh and sibilant and nasty! (Ronan Chris Murphy 2015, italics added)
>
> [Speaking on a video titled "How to EQ Female Vocals That Won't Hurt Your Audience!"] Around 4kHz, I find—*especially on female vocals*—that's a really common spot to find issues with harshness. (Worship Sound Guy 2019, italics added)

This implies that, across Western contemporary popular music production, the diagnosis of "harshness" is disproportionately applied to female vocals, or perhaps deemed more intrusive or unpleasant in female voices. These producer-songwriters assume that "harsh" resonances can more easily be found in female voices, likely because they generalize that these vocals tend to rest in a higher pitch register and so boosting high frequencies—when there is more high-frequency energy present—is more likely to result in such unwanted resonances. However, this phenomenon is also significant given that sound studies research has suggested that there is a longer history of sound technologies rarely being designed with female singing voices in mind. Jonathan Sterne and Paula Lockheart have outlined aspects of this history in more detail, particularly pertaining to the ability of early microphone technology to capture the voices of operatic sopranos. In several written sources from the early 20th century, these voices were described as too abrasive, too harsh, and too much for the capabilities of sound recording technology at the time (Lockheart 2003: 368-373). Sterne (2003: 238), in particular, suggests that this problem was usually framed in terms of a deficiency in female

vocal delivery rather than as a problem resulting from the ways these technologies were usually designed first and foremost for (male) speech reproduction. Such a historical trend arguably extends to vocal EQing in commercial pop in which, due to the overarching ideas about vocal realism which define the practice, the problem is framed as an innate problem of female voices rather than an artefact of commercial pop's aiming for an aesthetic of vocal foregrounding, which emphasizes higher frequency areas. This suggests that prevalence of ideas about vocal realism in EQ has intensified gendered assumptions about the voice.

For example, one of my informants[73] discussed how he had learnt from his production training that there was an exact, repeated region of the frequency spectrum at which "harshness" could be found in a female vocal: around 2.8-3kHz. A very similar sentiment was expressed in a pedagogical video specifically designed for "mainstream" pop vocals on YouTube. Quotes from both instances are given below (the second was introduced briefly in the previous section about harshness). When I first came across statements like this, I was struck by their universalizing tone: they seemed to insist that *all* female vocals suffered from this same deficiency, which required attenuation using a specific precision cut. The informant even discussed how, under the instruction of one of his university mentors, he had started to use a unique EQ plug-in in his vocal chain specifically for addressing this issue.

> There's this little frequency – sort of like around 2.7, 3k… that sort of area, in a female vocal it's just… harsh (anonymous informant).

> Another thing that I try to look out for, especially with female vocals, is this area between like 2-3k. You gotta be careful with this area… some people swear by boosting this area but you have to be careful because if a vocalist is really powerful you can get a really pingy sound and it can kinda hurt the ears a little bit. If you do boost that area you're going to get a harsh, pingy sound that's going to hurt a lot of people's ears at higher volumes. (SoundOracle 2018)

Notably, when I asked the same informant what kind of approach he would take for a male vocal with a similar issue, he explained that he hadn't really done much work with male vocalists, and suggested he might try a "search and destroy" kind of approach to find any similar "harsh" resonances,

[73] Informants are fully anonymized in this section.

illustrating how stereotypes about the gendered voice seemed to be maintained uncritically even in the absence of any confirming evidence.



Figure 55: A producer makes a small cut specifically to address a "harsh" frequency in a female vocal, in a pedagogical video by SoundOracle.net (SoundOracle 2018). The frequency axis is not visible in the plug-in interface but the cut is at 2920.2Hz by -1.22 dB.

Strikingly, another of my informants used the same technique for cutting out a precise "harsh" frequency around 2.8-3kHz with the same, unique, plug-in that this first informant used to tame harshness in female vocals. When pressed about the gendered aspects of this technique, however, he argued "harshness" wasn't any more difficult or problematic in female vocals—and baulked at the possibility that other producer-songwriters might think this. He did say, however, that female and male voices generally had essential differences which ought to be considered during vocal editing:

> It is generally like a very different sound. Again, that's a massively sweeping comment, like some female vocals will sound the same as male vocals. But generally you know it's all like… it all feels as though things are shifted up in a female vocal, a softer female vocal. If I was to generalize (anonymous informant).

Such statements illustrate how the gendering of vocal "harshness" is founded on essentializing ideas about the links between vocal timbre and gender, an assumption that Eidsheim criticizes at length in her work (e.g. see Eidsheim 2019: 6-7). The implied act of generalization across all 'female' vocals also makes one wonder whether the same kind of binarized generalization of the difference between male and female vocals would be reproduced in the case of vocalists who do not conform to binary notions of gender. Such a generalization would surely invalidate such gender identities and gloss over their complication of the belief that the relationship between vocal timbre and identity is a stable one. The fact that the same, incredibly specific, technique was being used by producer-songwriters in completely different locales who had gone to completely different universities was testament to the power of the intersection between pedagogical materials on the internet and specialist university training in driving commercial pop vocal EQing practices, and their associated realist discourse which motivates these assumptions about gendered voices.

In summary, this section has provided examples of the way that the discourse of vocal realism which dominates vocal EQ fuels assumptions about the links between voices and gendered bodies or identities: a key concept which I have suggested is a fundamental component of erotic voice fetishism. More broadly, such realist assumptions intensify the often-gendered power dynamics at play in vocal editing, by turning vocal EQ into a process of diagnosing and repairing vocal "problems", and, by extension, enhancing the sense of control that producer-songwriters feel they have over the voice through the direct access to vocal timbre that para-graphic EQ plug-ins appear to provide.

4.4. "Intelligent" EQ plug-ins and vocal realism: Fabfilter's *Pro-Q 3*, oeksound's *Soothe* and SoundTheory's *Gullfoss*

The widespread usage of three key EQ plug-ins across my fieldwork illustrated how the plug-in industry had picked up on the currency of realist discourse in commercial pop vocal EQ, as such plug-ins attempted to rationalize strategic portions of the vocal EQ process that were frequently discussed in realist terms by my informants. This suggested that the plug-in industry had a significant role to play in sustaining the interrelationship between erotic voice fetishism and the commodity fetishism of

digital novelty plug-ins which fuelled the practice. More specifically, the niche functionalities offered by such plug-ins appeared to respond to producer-songwriters' core beliefs about the importance and intricacy of vocal timbre, encapsulating supposedly sensitive and efficient ways of controlling this timbre in the form of slick, easy-to-use, contemporary digital commodities. Appeals to the language of artificial intelligence (AI) in the design and marketing of these digital novelty plug-ins also strengthen the quasi-scientific dimensions of realist discourse in vocal EQ, relating to the technique's supposed sensitivity to universal human auditory perception. This further motivates perceptions of the technique as a particularly powerful and ostensibly non-aesthetic, realist way of manipulating the voice's most fundamental characteristics.

Primarily, and as the reader may have noticed from the screenshots included in this chapter so far, the plug-in *Pro-Q 3*, by FabFilter, was used by every single one of my informants as their go-to para-graphic EQ.[74] Beyond *Pro-Q 3*, however, there were three recently released plug-ins that my informants kept mentioning as additions to their vocal chain which were thought to complement *Pro-Q 3*, and which performed specific EQ-like functions. *Soothe* and *Gullfoss*, developed by plug-in companies oeksound and SoundTheory respectively, claim to provide "intelligent"[75] vocal "smoothing" functions: they allow the user to locate and attenuate "problematic" frequency bands in the voice, but in a way that is supposedly sensitive to the way that the vocal frequency spectrum changes with pitch and register, additionally sometimes offering ways to automate this process. FabFilter *Pro-Q 3* offers similar functionality, but this is embedded within its central para-graphic EQ interface, presented as an important and useful additional function. These plug-in companies attempt to condense vocal EQing techniques which are founded on ideas about vocal realism into a single plug-in (or function within a plug-in), capitalizing on the realist discourse which dominates

[74] If not *Pro-Q 3*, earlier iterations of the plug-in were sometimes used, for example when I first met Oliver he was using *Pro-Q 2*, however within 6 months he had purchased *Pro-Q 3* as he felt the upgrade was absolutely necessary.

[75] These brands do not make clear whether suggestions of intuition or intelligence refer specifically to the use of machine-learning algorithms. Plug-in designs are often heavily guarded and difficult to access, so the notion of intuition tends to refer to some kind of function the plug-in offers in which a conventional or commonly undertaken process can be achieved automatically—or if not immediately then very quickly and easily.

commercial pop vocal EQ by designing plug-ins which attempt to perform these specific kinds of EQ processes intuitively or automatically.

FabFilter's *Pro-Q 3*

Among my informants, *Pro-Q 3* was by far the most commonly used EQ plug-in for the first-round, subtractive EQ that I have described. It was a staple in all my key informants' vocal chains, along with those of multiple of my other informants working in commercial pop (2019G, 2019CC, 2020E). My informants described how *Pro-Q 3* was particularly popular for vocal EQing for two reasons: firstly, it had a very simple, easy-to-use interface which allowed them to work quickly and intuitively with the vocal frequency spectrum. More importantly, however, it provided some additional 'dynamic EQ' functionality that other para-graphic EQs did not offer. This 'dynamic EQ' function condenses an action which would usually be achieved via a combination of EQ, compression, multiband compression[76] and de-essing plug-ins, and was recognized by my informants as being more intuitive or sensitive to the way vocal frequency spectra naturally fluctuate over time.

An explanation of the development of the FabFilter brand contextualizes this 'dynamic EQ' feature further. Fabfilter is a software company that develops a wide variety of audio plug-ins for music production. Founded in 2002 by two Dutch software developers, the company initially started out by focusing on 'creative' plug-ins geared towards sound generation and experimentation with timbre (synths, delays etc.). According to an interview with one of their founders (Computer Music 2013), the success of their early plug-ins highlighted a demand in the pop production industry for simple, user-friendly plug-in interfaces: 'workhorse' plug-ins that sounded good, whilst also incorporating 'slick' and streamlined interface design. Subsequently, the company started to let the 'creative' plug-ins take a backseat and directed their resources towards cultivating their 'Pro' range of plug-ins, including the Fabfilter *Pro-Q* series (*Pro-Q 3* was released in 2018). Sound on Sound puts FabFilter's success down to their digital novelty status: or, more specifically, the way the company spearheaded the move away from skeuomorphic audio plug-in interfaces replicating pre-digital

[76] This technique will be explained in Chapter 5.

technology towards a wholly digital approach, 'embracing the new possibilities of software design' and creating 'beautiful' interfaces that perfectly balance 'ease of use and versatility' (Inglis 2017).[77] Even before *Pro-Q 3*, then, FabFilter had already tried to brand *Pro-Q* as an intuitive EQ plug-in which effectively streamlined the complexities of a frequency spectrum. The interface even has functions which suggest ways the user might want to manipulate their EQ curve, allow seamless manipulation of the 'Q' parameter with a single keyboard shortcut, and permit the user to work with the EQ curve in full screen with as little visual distractions as possible (Manchester Music 2019). Even though *Pro-Q 3* was not solely used on vocals, its brand image of functionality and instrumentality complemented producer-songwriters' ideas about vocal EQ preserving vocal realism.

Building on this brand image, FabFilter enhanced their third version of *Pro-Q* with the new 'dynamic EQ' function (FabFilter 2021). Put simply, this function allows the user to set a 'dynamic threshold' for any boost or cut they perform. This means that the selected frequency is only cut/boosted when it reaches a certain volume. Both my informants and multiple internet reviews of the plug-in claimed that this function was especially useful for vocals because it provided a more sophisticated way of preserving the innately unpredictable qualities of vocal timbre (2020F; 2019Z; Inglis 2019; Manchester Music 2019). For example, in order to negotiate the tension between wanting to obtain "presence" in the upper-mid range of the vocal whilst also mitigating "harshness", the user could set a dynamic threshold on their cut such that the "harsh" frequency was only attenuated when it became particularly noticeable (in higher or "beltier" portions of the vocal line), which would ensure that vocal "presence" was retained for the remainder of the track. *Pro-Q 3*'s dynamic EQ function therefore directly engages realist discourse in commercial pop vocal EQ because it makes producer-songwriters feel like they are EQing voices in a way which is more sensitive to their inherently fluctuating qualities. The assumption is that vocal timbre is inherently "dynamic" or complex, and that EQ modifications (while clearly contributing to vocal foregrounding) must preserve these innate

[77] This is very similar to the way Microsoft and Apple moved away from skeuomorphism to so called 'flat design' principles in the late 2000s, abandoning interfaces which mimic the look and feel of real-life objects and instead creating minimal, two-dimensional interfaces consisting of rounded elements and block colours (Turner 2014).

qualities of the voice as far as possible. It's also worth noting that dedicated dynamic EQ plug-ins are also available (e.g. Waves' *F6*[78]), but FabFilter's *Pro-Q* 3's unique selling point is the way the dynamic EQ functionality is built into the para-graphic EQ curve, represented by coloured regions whose outer limits indicate the dynamic thresholds. This act of constraining cuts and boosts within dynamic thresholds is even dramatized visually by the plug-in: when a user activates the dynamic EQ function, the central (yellow) EQ curve oscillates during playback, and the coloured regions restrict the width of these oscillations (Figure 56, for a video of the process see Manchester Music 2019).[79]



Figure 56: Jason uses Fabfilter *Pro-Q 3* to create a dynamic threshold at certain frequencies associated with 'harshness' and 'air' (green and blue respectively). (Image acquired through email correspondence)

FabFilter's introduction of the dynamic EQ feature was clearly responding to a perceived demand for this kind of realist function. My informant Robin, for example, described how he tried to EQ the voice in this kind of "dynamic" manner, even without or before FabFilter *Pro-Q 3* (2020I). Often, this would be achieved via a combination of both EQ and multiband compression[80] plug-ins, in an attempt

[78] https://www.waves.com/plugins/f6-floating-band-dynamic-eq (Accessed 11 Jun 2022).
[79] Here is a link for ease of access: https://www.youtube.com/watch?v=_lVUxIuR_nE (Accessed 11 Jun 2022).
[80] Multiband Compression is very similar to dynamic EQ, however they are subtly different. Put simply, the steps involved are reversed: where dynamic EQ adds a dynamic threshold to a selected centre frequency,

to create a series of EQ modifications that would only kick-in at certain points during a track, following the registral and pitch contours of the voice (ibid.). The plug-in therefore simultaneously responds to and perpetuates EQing the voice in ways which tap into ideas about preserving the innate dynamism of the vocal frequency spectrum. Not only are users able to use *Pro-Q 3* to 'zone in' on particular frequency regions using a very user-friendly, simple interface, but they are also able to add an extra temporal dimension to this process which appears to make their EQing even more sophisticated and effective at working with the 'real' qualities of the voice. Put another way, FabFilter have capitalized on commercial pop producer-songwriters' desire for EQ modifications to cultivate vocal realism, by developing a plug-in which increasingly claims to "intelligently" respond to the voice's innately unpredictable qualities.

oeksound's[81] *Soothe2*

Soothe and *Gullfoss* were more radical than *Pro-Q* in the ways their marketing languages directly tapped into pop producer-songwriters' ideas about universal auditory perception of the voice. *Soothe*, specifically, is marketed by the Finnish company oeksound as a 'dynamic resonance suppressor' which purports to automatically identify and attenuate 'problematic' frequencies (oeksound 2021). The plug-in has a para-graphic EQ-like interface which allows the user to adjust the automatically identified frequencies if they so desire (Figure 57). The plug-in aims to provide an easy fix for the "harshness" problems I have been discussing throughout this chapter, and is marketed as a neat labour-saving device. Indeed, the product webpage for *Soothe*'s most recent iteration (*Soothe2*) is subtitled with the catchphrase 'Soothe harshness. Save time' (ibid.). oeksound's advertising materials sell the plug-in's ability to achieve the desirable 'soothing' effect whilst also 'preserving the timbre of the original sound source' because—like *Pro-Q 3*—the attenuations which the plug-in applies are 'dynamic' meaning that they only kick in when 'needed' (ibid.). *Soothe* therefore directly taps into the discourse of vocal realism in EQ through its presentation as a plug-in which can intuitively 'read' and

multiband compression involves restricting the actions of a compressor plug-in—involving a user-defined dynamic threshold among other parameters—to a specific frequency band. The mechanisms of compression will be explained in more detail in chapter 5.

[81] The company's name purposefully lacks capitalization.

modify the innate timbral qualities of vocal frequency spectra. (Although the plug-in can be used on any sound source, it is primarily marketed for use on vocals specifically (ibid.).)



Figure 57: *Soothe2* interface (oeksound 2021)

Several of my informants spoke effusively about *Soothe* as a highly desirable vocal editing plug-in which had extraordinary abilities, illustrating how oeksound had successfully tapped into established ideas about vocal realism in EQ among commercial pop producer-songwriters. The first iteration of the plug-in was released in early 2018 and I first heard such sentiments in the summer of 2019, which illustrated the very fast uptake of the plug-in in the industry. Oliver, for example, described how in the latter stages of his vocal chain he focused on using two or three plug-ins to try and tame unwanted resonances, and wished he had enough money to buy *Soothe* because he believed it would make this process much easier and faster (2019BB). He seemed to believe strongly that this relatively new plug-in would become an indispensable part of his vocal chain. When I asked him where he heard about the plug-in, he suggested it was probably via word-of-mouth at university where "a few people were saying [this plug-in is] the best thing since sliced bread" (2020S). Similarly, Tim (2019CC) described how he always used *Soothe* at the end of his vocal chain because "it just does what it says it does—it

soothes any harshness out of the vocal." In my observation with Jason for Audio 1, *Soothe* was slotted in towards the end of the lead vocal chain (2020M). He simply opened the plug-in and automatically applied a preset which supposedly "soothed" any harshness in the voice conveniently, without him having to think about it too much.

The sheer popularity and quick uptake of this plug-in, alongside my informants' conceptualization of the plug-in as a functional time-saving device, illustrated just how naturalized the discourse of vocal realism had become in vocal EQing. *Soothe* condenses and speeds up the attenuation of a sound quality ("harshness") which has been emphasized, in the first place, by the desire to exaggerate vocal foregrounding (itself couched in terms of realism), however there was very little recognition of this in my informants' discussions of the plug-in. oeksound had successfully made these producer-songwriters feel like they *needed* to use this plug-in in their vocal chain to address this common issue, without considering how their desire to fulfil the aesthetic requisites of the commercial pop genre had created such a need in the first place. Further, although the development of this plug-in partly responds to the widespread practice of "soothing harshness" in the voice, the popularity of this plug-in is also likely to consolidate the technique. The name of the plug-in also has a domesticating quality as, to put it bluntly, (female) vocalists are being soothed or tamed by (male) producers. Most importantly, *Soothe* clearly both responds to and reinforces ideas about vocal realism in vocal EQing.

SoundTheory's *Gullfoss*

Gullfoss, also released in 2018 by British plug-in company SoundTheory, is very similar to *Soothe,* and was comparably used widely by my informants for 'intuitively' EQing vocals. Marketed as an 'intelligent equalizer' (SoundTheory 2021), the plug-in claims to automatically generate an EQ curve according to two main user defined parameters. The first is 'Recover', which Sound on Sound claims accentuates frequency content which the plug-in interprets as perceptually salient but not prominent enough, and 'Tame' which suppresses frequency content which is supposedly too prominent or invasive (e.g. "harshness") (White 2018) (Figure 58). More extreme than *Soothe* and *Pro-Q 3,* however, is SoundTheory's appeal to ideas about archetypal auditory perception in the marketing

materials for *Gullfoss*. Ostensibly, *Gullfoss* 'listens to a signal and decides how to prepare the audio so that your brain can get the most information out of it', as well as using a 'computational auditory perception model to understand which audible elements are competing for your attention.' (SoundTheory 2021). Of course, this 'computational auditory perception model' is heavily black-boxed and it is nearly impossible to find any concrete information about it online. SoundTheory are capitalizing on the way commercial pop producers understand EQ as process that is sympathetic to the universal auditory perception of the voice. Indeed, SoundTheory have worked hard to uphold this notion of creating audio plug-ins which somehow epitomise the true or innate qualities of sound at the centre of their brand, whose tagline reads 'REVEAL YOUR SOUND.' (SoundTheory 2021).



Figure 58: *Gullfoss* interface (from White, 2018). Recover and Tame parameters are shown at the top, and their respective effects represented on the curve.

Similar to *Soothe, Gullfoss'* popularity among my informants demonstrated how the uptake of the plug-in had been incredibly fast, and there was a general sense of its desirability among my informants; that owning and integrating this plug-in into their workflow was likely to transform their vocal editing process. In one interview, however, Chris went further by linking *Gullfoss'* capabilities to his informal understanding of how the voice is "naturally" perceived:

> I think just because of the way the human ear is, there are certain frequencies that are just naturally unpleasant to us. […] Have you ever heard of *Gullfoss*, before? That [plug-in is] just saying like: okay, here are the frequencies that are unpleasant to the human ear, and we'll just cut them out. I guess, naturally… when you hear someone speaking you don't naturally hear that frequency that's resonant, cause in the real world it just kind of gets lost, but for whatever reason through the recording process that seems to be accentuated, cause you know… science, I guess. (Chris, 2020T)

Here, Chris's understanding of auditory perception assumed an archetypal human ear which "naturally" finds some frequencies unpleasant, and he felt that *Gullfoss* attempted to return the voice to an ideal "real world" state without such "unpleasant" frequencies. Interestingly he also suggested that such "unpleasant frequencies" were "accentuated by the recording process", but in ways that were unknown to him. As I have shown, such frequencies are accentuated primarily by the way vocal EQing aims to foreground the voice, however this didn't seem to enter Chris's considerations of this plug-in, illustrating how the aesthetics of vocal foregrounding had been naturalized for him, re-coded as an inevitable consequence of the "recording process". Obviously, this is a tacit, vernacular interpretation, and Chris seemed to acknowledge that a full understanding was somewhat beyond his means ("it's science, I guess"). However, these discussions showed very clearly the extent to which universalizing ideas about vocal realism obscure the aesthetic aims of EQ, and plug-ins like *Soothe* and *Gullfoss* intentionally tap into these ideas.

In short, the apparently 'intuitive' features of *Pro-Q 3*, *Soothe* and *Gullfoss* show how the plug-in industry has identified and subsequently jumped on the way commercial pop producer-songwriters think that EQ should preserve vocal realism. Some of my informants also briefly highlighted some other less-widely used 'intelligent' EQs that performed similar functions—

specifically, SoundRadix's *Surfer EQ*[82] and sonible's *smartEQ*.[83] Jason, in particular, was interested in learning more about *SurferEQ* to see if it more intuitively EQed the voice in comparison to *Soothe* (2021D). This further illustrated the efficacy of the way these plug-in companies had tapped into the existing power of realist discourse in EQ. By extension, such plug-ins covertly increase the practice's intense focus on the power of vocal timbre, naturalizing it and presenting it in the form of alluring digital commodities which offer supposedly scientific ways of manipulating this timbre in order to make the voice sound as "real" as possible.

Conclusion

To summarize, this chapter has demonstrated how my informants conceptualized vocal EQ as a way of preserving and enhancing the singing voice's most "real" and universally perceptually salient qualities, even though EQ was clearly an aesthetic technique which aimed to exaggerate high-frequency energy in the voice in the service of vocal foregrounding. I began by historicizing the development of EQ and its use in Western music production, culminating in the widespread usage of digital para-graphic EQ plug-ins in commercial pop vocal editing. I also explained the role of EQ in the practice of 'carving' in Western music mixing, showing how, through its contribution to this approach, EQ plays a particularly important role in achieving the aesthetic of vocal foregrounding, while simultaneously being conflated with notions of universal auditory perception. Following this, I outlined how the vocabulary commercial pop producer-songwriters use to describe para-graphic EQ curves splits up a generalized version of the voice into particular regions which correspond to desirable and undesirable vocal qualities, with EQ modifications in these ranges being described in realist terms, including in relation to how the voice is universally perceived by the ear ("low-end rumble", "body", "boxiness", "presence", "harshness", "sibilance" and "air"). Thirdly, I demonstrated how realist discourse in vocal EQ subtly reflects and amplifies gendered power dynamics in vocal editing through its intersection with surgical metaphors. A "subtractive" approach is upheld as a more appropriate way of using para-graphic EQ plug-ins on vocals, because producer-songwriters think it is

[82] https://www.soundradix.com/products/surfer-eq/ (Accessed 11 Jun 2022).
[83] https://www.sonible.com/smarteq3/ (Accessed 11 Jun 2022).

more sensitive to the "real" voice—an approach which is also shot through with metaphors of cosmetic surgery. The dichotomy between "presence" and "harshness" illustrates how my informants' ideas about vocal realism in EQ obscured universalizing assumptions about vocal gender differences. Finally, I discussed how three recently-released "intelligent" EQ plug-ins—FabFilter *Pro-Q 3, Soothe* and *Gullfoss*—attempt to rationalize or condense realist approaches to EQing, demonstrating how plug-in companies are capitalizing on and reinforcing ideas about vocal realism in EQ.

This discourse of vocal realism is particularly prominent in relation to EQ and recedes into the background in the case of tuning, comping and compression. And as my comparison with Bates shows, the discourse of realism surrounding vocal EQ is specific to Western music production. A simple explanation for why this realist discourse is so prominent in commercial pop vocal EQ would be that the aesthetics of vocal foregrounding have become so naturalized and routine in this genre that one of the most important techniques producer-songwriters use to achieve this aim is now conceptualized as functional, objective and seemingly non-artistic. However, I do not think this adequately explains the strength of realist discourse in EQing, or the repeated allusions to universal listener perception of the voice which motivate such realism. The analyses I have provided in this chapter suggest that realism in vocal EQ is much more complex and culturally-driven, related to the technique's entanglement with specific characteristics of erotic voice fetishism and the commodity fetishism of digital novelty plug-ins: EQ is thought to provide powerful capacities to manipulate generalized characteristics of vocal timbre which, drawing on my theorization of erotic voice fetishism, is believed to be a defining characteristic of a voice's corresponding identity and physicality. This intersects in complex ways with the commodity fetishism of para-graphic EQ plug-ins, which appear to permit direct and flexible manipulation of objectified vocal frequency spectra. In combination with the existing idolization of vocal timbre, this gives EQ users the illusion that they have direct, unmediated and objective control over the voice itself. Very broadly, realist discourse in EQ therefore works to sustain the presumed power and importance of the voice in commercial pop vocal editing. It also arguably sustains the power asserted by producer-songwriters in the broader labour process of vocal editing, as they feel responsible for not just preserving but enhancing this

vocal realism in a way that (paradoxically) improves the vocalist's performance. It may also be the case that conceptualizations of vocal EQ as a realist technique maintain or even perhaps homogenize the aesthetic effects of vocal EQ in commercial pop, because they make this highly aesthetically specific technique seem natural, normal and inevitable.

Another speculative interpretation of the existence and function of realist discourse in vocal EQ would be that it works to preserve the foundations of erotic voice fetishism—specifically, the presumed deep links between vocal timbre and physicality or identity—in a situation where these foundations are increasingly disturbed because of the capacities provided by vocal EQ. If the timbral qualities of the voice can now so easily be manipulated using EQ, then what is left to tie them to the bodies and identities that produced them? This provides further reasoning for why discourse about EQ increasingly draws on notions of vocal realism even though there is a clear aesthetic specificity to what is being done with EQ, related to the aesthetics of vocal foregrounding in commercial pop.

My analysis also suggests that EQ's realist discourse maintains unhelpful assumptions about the gendered qualities of voices among producer-songwriters. The predominance of cosmetic surgery metaphors in EQing similarly suggests that realist discourse functions as a way of hiding or minimizing the aesthetic effects of EQ, making them seem normal. Just like cosmetic surgery's entanglement with the production and maintenance of gendered body standards (see Dull & West 1991: 64-67; Basalmo 1992: 219-214), this may have consequences for the ways female voices are perceived in this production culture more broadly. Put bluntly: female voices do not sound like this in 'real life', but the subtle effects of EQ and their conception in realist terms may make members of this culture increasingly think that they do. Metaphors of EQ as surgical intervention also embody notions of gendered control over the voice, again recalling the characteristics of erotic fetishism.

In terms of my arguments about the plug-in industry, this chapter also shows that there are several small, 'indie' plug-in companies who have developed specific EQ plug-ins that capitalize on the discourse of vocal realism. Indeed, my discussion of the three 'intuitive' EQ plug-ins represents a point at which the convergence between voice and technology fetishism that defines commercial pop

vocal editing can be observed most clearly. Moreover, the success of such indie plug-in developers in this arena suggests that producer-songwriters trust these smaller, less well-known organizations relatively more than established major music technology companies to fully understand and address their realist concerns in EQ. The most widely used tuning and compression plug-ins I discuss, in contrast, are produced by these bigger companies (Antares, Celemony, Universal Audio and Waves). oeksound and SoundTheory, in particular, centre their brand identities around these flagship "intelligent" EQ plug-ins (for example SoundTheory, so far, has only released *Gullfoss*), emphasizing the sense that a huge amount of time, effort and knowledge has gone into giving these plug-ins the ability to access and enhance the innate qualities of the voice. There is a sense that the smaller, developing nature of these companies—all start-ups who only produce 1 or 2 plug-ins—leads producer-songwriters to assume that a substantial amount of labour has gone into developing plug-ins which truly 'understand' and provide the ability to effectively bring out vocal realism.

As I will show in the next chapter, the scenario was slightly different with compression, as it seemed that digital compression technologies—particularly analogue emulations—were being more obviously fetishized as commodities in vocal editing. The combination of this analogue emulation commodity fetishism with a more general voice fetishism related to the perceived importance and volatility of vocal "dynamics" fuelled the iterative, compulsive nature of commercial pop vocal compression techniques. I consider the ramifications of this comparison between the different kinds of fetishisms at play in vocal EQ and compression more deeply in the conclusion to the thesis.

Chapter Five: The Commodity Fetishism of Analogue Emulations in Commercial Pop Vocal Compression

Compression is a technique for reducing the overall dynamic range of a sound signal (Huber & Runstein 2018: 421-422; Hodgson 2010: 81; Giannoulis et. al. 2012: 399). In commercial pop vocal editing, however, specific compression plug-ins are used to produce aesthetic effects beyond condensing dynamic range: they are intended to make vocals sound more "punchy", "smooth" or "warm"—to take my informants' terms—and to contribute to the overall sense of vocal foregrounding which I described in Chapter 2. This chapter claims that commercial pop producer-songwriters fetishize two digital plug-in emulations of analogue compressor units as commodities—the *1176* and *LA-2A*—compressors with reputations among music production professionals for their efficacy as vocal compressors. My informants considered the sonic effects of these compressors, emulated in digital form, to be important and aesthetically valuable but also mystical, elusive and difficult to comprehend. In this way, the same sort of commodity fetishism characteristic of analogue music technologies had been translated into a digital music production context, whereby both the inner workings of a technology and its sonic effects are veiled in mystery, in turn imbuing the form of the technology with a kind of magical power or intrigue (see Sterne 2019a, 2020; Théberge 1997). The chapter also shows how, when combined with the additional erotic fetishism of the voice in this music production culture relating to the unpredictability of vocal "dynamics", such commodity fetishism of digital compression technologies makes vocal compression techniques highly complex and iterative, involving multiple types of compression plug-ins in sequence. This benefits plug-in companies who both target and fuel both kinds of fetishism in the way they design and market compression plug-ins.

This chapter therefore builds on Sterne's (2019a, 2020) recent research into the development and commodity fetishism of analogue emulations in the contemporary music technology industries. Sterne uses the concept of commodity fetishism to interrogate the power of analogue emulations across these industries, which are conceived as digital musical instruments that musicians feel substantially contribute to their creative practice (Sterne 2019a). He also analyses how the digital modelling of these analogue units is achieved both sonically and visually, focusing in particular on the

important role of Universal Audio in this arena (Sterne 2020). I complement Sterne's approach by considering how the commodity fetishism of analogue compressor emulations in commercial pop vocal editing interacts with and is intensified by the fetishism of the voice in this culture, claiming that this interaction ultimately defines vocal compression's uniquely iterative character. This unrelenting iterability—in which one round of compression creates a problem or potential problem in the voice which must be solved by another, and so on—especially reveals how these interacting fetishisms drive vocal compression, as well as portraying the practice as more generally fetishistic or compulsive through its tendences to obsessively hone the sound of the voice using incessant subtle modifications.

With this in mind, the chapter will be in two main sections, preceded by a contextualization of compression discourse in Western music production. The preliminary discussion provides some history and context for the mystification of compression as a complex and elusive technique, which is linked to the way it affects both loudness and timbre (or "tone") in ways that are difficult to measure. I also explain how these ideas about compression translate into the aesthetic context of commercial pop vocal editing. Following this, I will interrogate the commodity fetishism of *1176* and *LA-2A* emulations in compression by: firstly, discussing the reasons behind commercial pop producer-songwriters' dominant usage of these plug-ins, which is rooted in the belief they lend elusive sonic 'signatures' or 'characters' to the vocal sound. This discussion also suggests that the plug-in industry is capitalizing on the commodity fetishism of such compression plug-ins through 'celebrity engineer' marketing tactics which reinforce ideas about the particular 'sound' these plug-ins offer. Such marketing rhetoric additionally promulgates an aggressive, hyper-masculine discourse surrounding the actions of compression which resonates with such sonic ideals, and arguably magnifies the gendered power dynamics of vocal editing. Secondly, I draw on closer analysis of moments from my field data to illustrate how the interactions between such commodity fetishism and the erotic kind of voice fetishism fuelled the complex, multi-stage compression processes my informants used. The character of this voice fetishism relates specifically to the notion of vocal "dynamics" discussed in Chapter 2. Throughout the chapter I refer to an excerpt of a track produced by my informant Oliver,

whose cumulative vocal compression techniques were particularly complex and iterative. Audio tracks 5, 6, and 7 contain the same portion of the track in different forms: Audio 5 contains the "raw" vocal for this track, without any compression or other plug-ins; Audio 6 contains the processed vocal by itself; and Audio 7 contains the processed vocal in the context of the overall track. These excerpts show how, even though the direct sonic effects of compression are difficult to perceive, compression has a recognizable aesthetic effect on the sound of the commercial pop voice, overall amplifying its energy and vitality throughout a track. These effects of compression combine with the other vocal editing technologies discussed to produce the commercial pop vocal aesthetic.

5.1. Dynamic range compression: definitions, history and context

Much of the magic and mystery surrounding compression can be linked to the difficulty of defining the technique. The term 'compression' in audio refers to the process of *dynamic range compression*. However, precise definitions tend to vary depending on whether they come from audio engineering or music production perspectives. Audio engineering sources describe the process as an act of reducing the overall dynamic range of a sound signal: for example, Giannoulis et. al. define it as 'the process of mapping the dynamic range of an audio signal to a smaller range, i.e., reducing the signal level of the higher peaks while leaving the quieter parts untreated.' (2012: 399). Music production textbooks, however, tend to augment such a definition with the aesthetic ends to which dynamic range compression is usually put, which accompany the reduction in dynamic range with a subsequent *increase* in overall level, so that the sound's overall perceived loudness[84] is magnified.[85] Sound engineer Alexander Case, for example, avoids defining the technique in any concise way, instead focusing primarily on the various aesthetic effects compression can be used to create in mixing, including to 'increase perceived loudness', 'improve intelligibility and articulation', 'smooth [out] performance' and to create 'ambience and artifacts' (2007: Ch. 6). Case's discussion is accompanied

[84] For this chapter it is especially important that a distinction is made between the terms 'level' and 'loudness': one is an objective measure of sound pressure (in dB) and the other is rooted in auditory perception (usually measured in phons). See Ballou (2008: 26) for fuller definitions of these terms.
[85] See Huber & Runstein (2018: 421-422), Case (2007: Ch. 6) and Hodgson (2010: 80-81) for examples of this kind of definition.

by a diagram of the process of peak limiting on a waveform, which is an extreme form of dynamic range compression (Figure 59). This figure should give the reader a general idea of the basic effect of dynamic range compression on a sound: the 'peaks' of the waveform are brought down in level, then the level of the entire waveform is increased. The total effect is that the dynamic range of the signal is condensed and its perceived loudness increases.

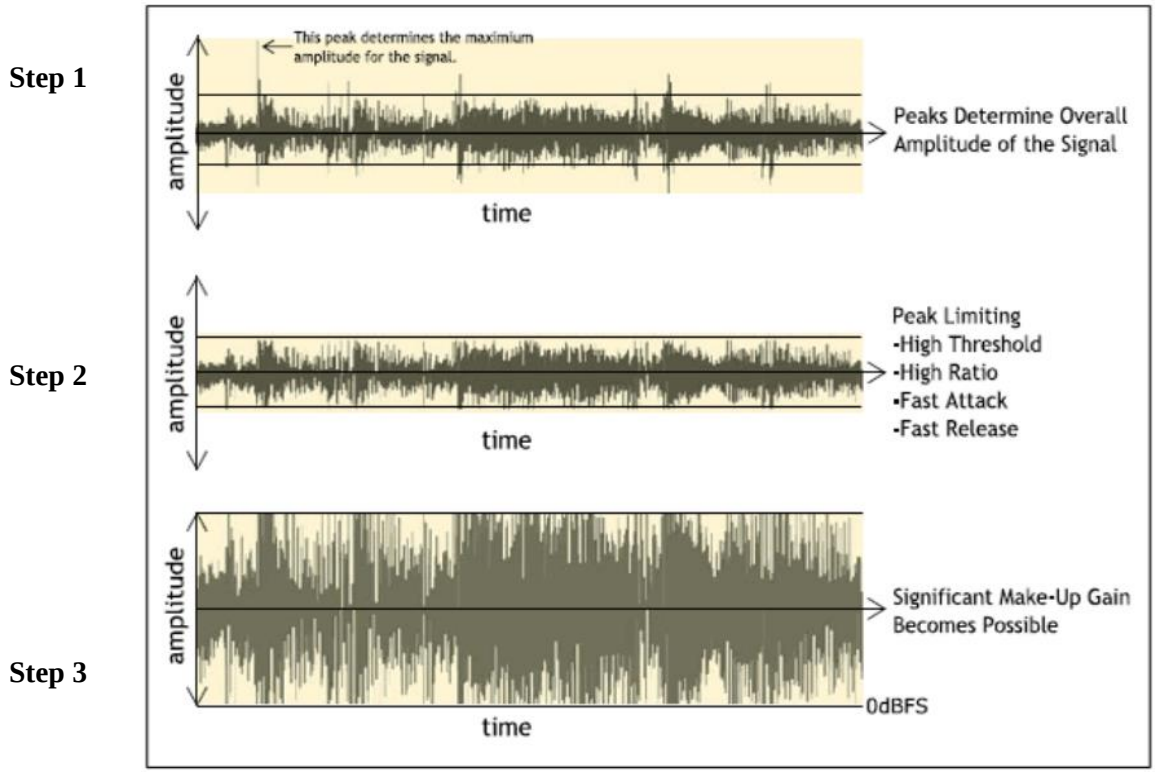


Figure 59: Diagram of the process of peak limiting, taken from Case (2007) (Ch. 6, Figure 6.3).

Notably, definitions of compression in music production and mixing textbooks are often accompanied or preceded by descriptions of the technique as mysterious, elusive and difficult to comprehend without years of training. Case, for example, describes compression as 'perhaps the most misunderstood' technique in mixing, and claims that truly mastering the technique is a 'fully learned, intellectual process' that is 'more difficult than learning to ride a bicycle (Case 2007: 161-162, cited again in Hodgson 2010: 79; see also Huber & Runstein 2018: 424; Izhaki 2017: Ch. 17). This

mystification of compression feeds my informants' commodity fetishism of analogue compressor emulations. The reason such attitudes towards compression exist in music production discourse more broadly is because its operations and effects are perceived as particularly difficult to trace audibly in comparison to other mixing techniques: more specifically, the technique does not just increase the overall perceived loudness of a sound, it also has subtle effects on "tone", and the effects on both loudness and "tone" overlap in complex, interrelated ways. As Case's descriptions imply, like tuning, compression is understood as more of a tone-shaping device: 'improving intelligibility and articulation' and 'smoothing out performance' do not just involve reductions in dynamic range. I will explain this complexity by briefly providing some historical background for compression as well as clarifying the technique more precisely.

Existing historical discussions of compression tend to conceive it as a *flattening* or *restricting* of the dynamic variability of a sound signal, often in the context of broader debates about the 'loudness war' in 20th-century Western recording practice. Sterne (2015), for example, situates dynamic range compression within a broader historical lineage of audio and data compression techniques which were designed to better 'fit' signals into the storage capacities of different media. Similar interpretations of compression as an act of containment underpin further interpretations of the technique, both academic and non-academic, which argue forcefully that it is an aesthetically reductive force in contemporary popular music (Sreedhar 2007; Smith 2008; Vickers 2010; Croghan et. al. 2012). Frequently referred to as the 'loudness war', numerous sound professionals argue that, over the past half-century or so, the belief that 'louder is better' in popular music recordings has led to an increase in the use of dynamic range compression to extreme levels (Vickers 2010; Devine 2013; Taylor 2017; Katz 2007). Such histories, however, focus on the over-arching, 'broad-band' use of compression applied to entire tracks in the mastering stage of popular music production, or to the output of entire radio stations, rather than the use of compression in popular music mixing for creative or aesthetic effects alongside the increased impression of loudness. The distinction between the terms 'macrodynamics' and 'microdynamics'—used in mixing discourse—is useful here: Roey Izhaki, for example, defines macrodynamics as an over-arching, longer-term sense of dynamic range ('variations

in level for events longer than a single note') and microdynamics as dynamic range or variability on a much smaller level ('level variations that happen within each note being played due to the nature of an instrument' e.g. 'the changing level between the verse and the chorus, or the level variations between snare hits, bass notes, or vocal phrases) (Izhaki 2017: Ch. 16). Sound engineer Emanuel Deruty argues against critics of the loudness war using this terminology, suggesting that these interpretations of compression as an aesthetically reductive force are too focused on macrodynamics: they do not acknowledge that the notion of 'dynamic range' operates on both micro and macro levels, and that even if macrodynamics have been 'flattened' somewhat in the past decades, microdynamic variation is still very important and prominent in current popular music tracks, as well as varying across genres (Deruty 2011). In commercial pop vocal editing, microdynamic variations are incredibly important to compression practices.

Explaining how the technique precisely affects the dynamics of a sound signal gives further context for the way contemporary Western music production perceives compression as a highly complex and elusive technique. Specifically, compression entails the automated attenuation of the peaks of a sound signal according to five main interrelated parameters that are set by the user (either using a hardware compressor unit or a digital compressor plug-in): *threshold, ratio, attack,* and *release*—followed by an increase in overall level using *make-up gain.*[86] The *threshold* determines the level at which the compressor is triggered into action: to quote Case, it forms 'a line separating too loud from not too loud' (2007: Ch. 6). When the level of a sound signal exceeds this threshold, the compressor will start compressing. The action of the compressor at this threshold is determined by the next three parameters. The *ratio* determines the amount or severity of the compression applied.[87] Figure 60 visualizes the effects of threshold and ratio on a sound signal. The audible effect of this process is that having a compressor set to a low threshold and high ratio will more noticeably create an effect of the sound being squashed into a narrow volume range, whereas higher thresholds and

[86] The following description of compression synthesizes information from textbooks by Case (2007: Ch. 6), Izhaki (2017: Ch. 16, Ch. 17) and Huber & Runstein (2018: Ch. 15).
[87] Limiting is essentially an extreme form of compression with a high ratio, meaning that the signal is nearly completely restricted from exceeding the threshold.

lower ratios have a more subtle effect. The *attack* and *release* parameters add a further degree of control by determining the speed with which the compressor starts and stops compressing the signal. Setting a fast attack time would mean that the level of the signal would decrease very quickly at the point the compressor kicks in, and setting a slow release time would mean that after the signal returned below the threshold the compressor would gradually return to an inert state and the level of the sound would gradually increase to the signal's original input level. For vocals, music production professionals claim that delicate manipulation of the attack and release parameters can very subtly manipulate the perceived energy of a vocalist's performance: setting a fast attack time, for example, will ensure that, after compression, the start of every vocal phrase will sound with the same loudness, contributing to a sense that the vocalist is exerting a consistent amount of energy throughout a track (Mayzes 2016; Huber & Runstein 2018: 426). However, these sources also claim that this would likely sound slightly *too* consistent (ibid.), and therefore unnatural, so they advocate setting a medium to slow attack time in order to generate a desirable amount of consistency without making a performance sound impossibly stable. Figure 61 is a simplified visual representation of what different attack and release settings might look like on the level envelope of a generic sound signal, taken from a mixing textbook. Modern compressors also permit the adjustment of *make-up gain* (sometimes just 'gain') which enables the user to turn up the overall level of the signal after all of this 'downward' compression has been applied. Figure 62 neatly illustrates the overall outcome this process of compression and make-up gain, which sound engineer Hugh Robjohns (2013) refers to as 'uplift compression': the overall loudness of the sound is increased to 'make up' for the 'squashing' that has been enacted by the compressor. Compressor interfaces tend to enable varying degrees of direct manipulation of each of these parameters. For vocals, when compression is applied, the overall effect is not only that the perceived loudness of the voice generally increases, but also that the energy of the vocalist's performance appears more consistent.

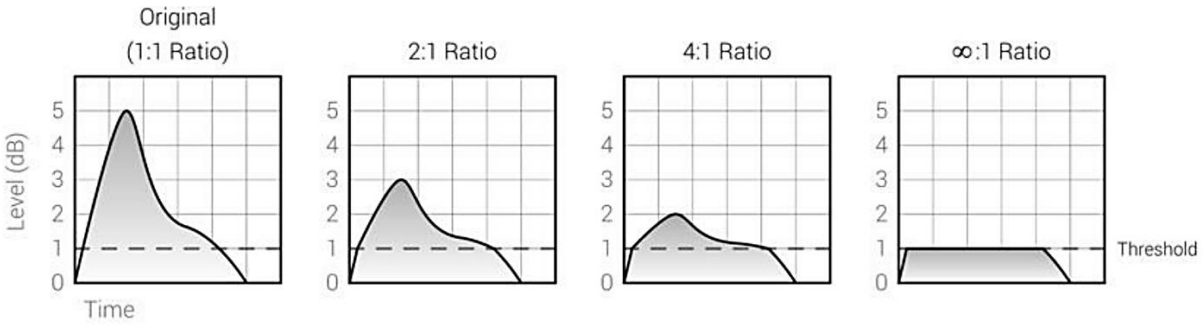


Figure 60: 'The effect of different ratios on the level envelope of a waveform' (from Izhaki 2017: Ch.17, Figure 17.8).

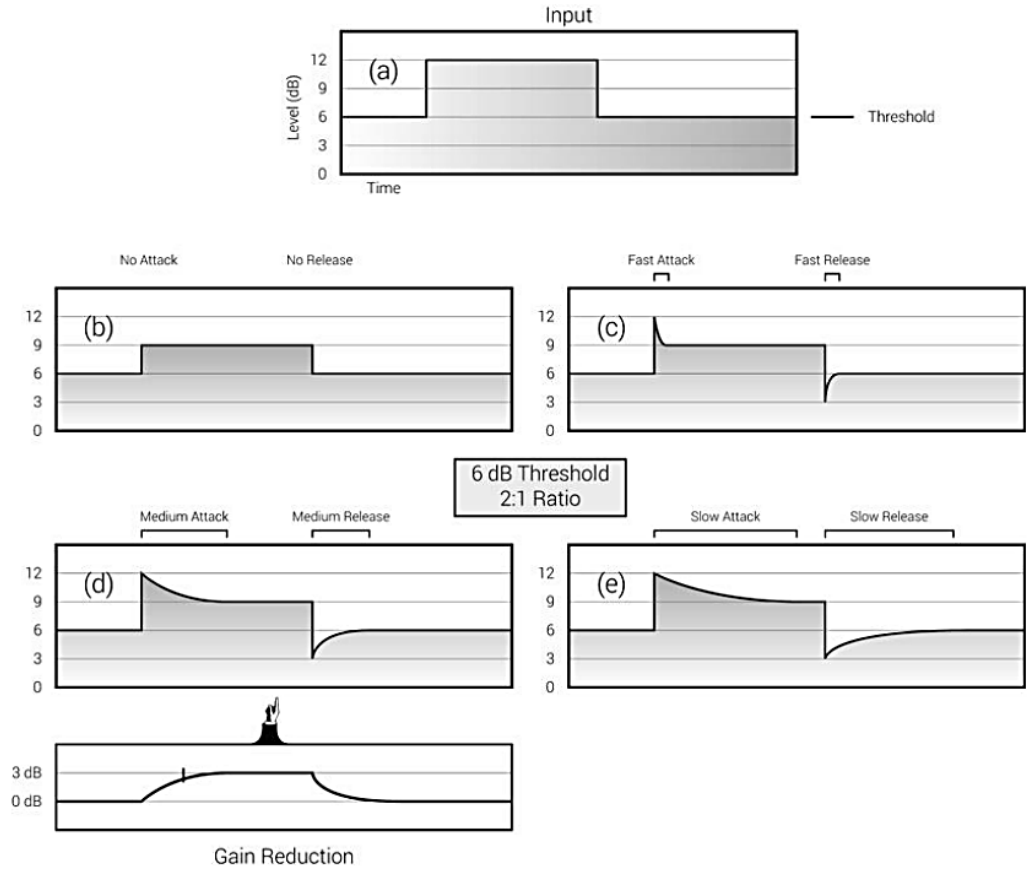


Figure 61: 'The effect of different attack and release settings on a waveform. (a) The original levels before compression. (b–e) The resultant levels after compression.' (From Izhaki 2017: Ch. 17, Figure 17. 11).

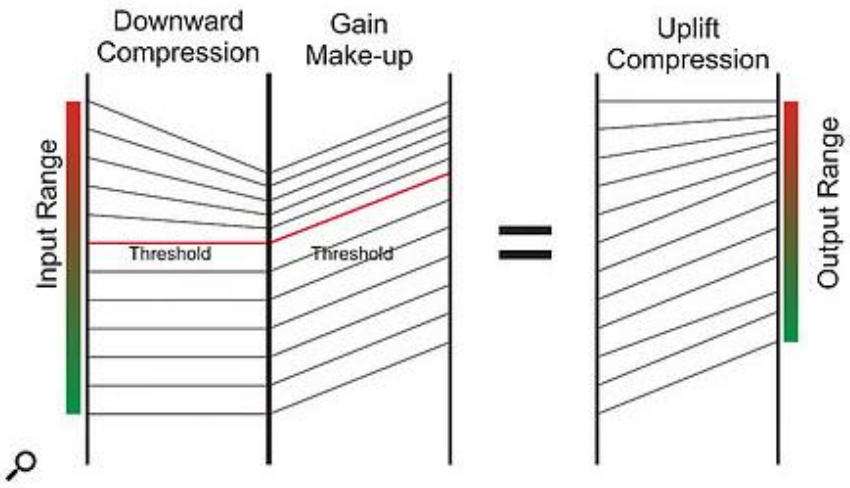


Figure 62: 'Uplift compression' represented diagrammatically by engineer Hugh Robjohns in a Sound on Sound article (Robjohns 2013).

Western music production professionals also conceive compression as a "tone"-shaping device, because the act of 'evening-out' the dynamic range of a sound will produce a corresponding 'evening-out' of the sound's spectral content as the most predominant frequencies will be emphasized and exaggerated. This view is expressed by engineer Andy Johns, quoted in a mixing handbook (Oswinski 2013: Ch. 9).[88] For vocals, this means that compression is believed to create further "tonal" enhancement by emphasizing existing frequency content in the voice or making it more consistent. It is this idea which underpins various claims in music production sources that certain compressors can add characteristics like, "warmth", "harmonic richness" or "clarity" to a voice (e.g. Krantzberg 2020; Senior 2009a)—and claims like this are often used to support the status of some of the 'legendary' analogue compressors that I discuss later in this chapter. Moreover, in commercial pop vocal editing, the interaction between the perceived complexity of vocal "dynamics" and "tone" (key features of the voice fetishism which drives the aesthetic discourse of vocal editing, as discussed in Chapter 2), and the widespread assumption across music production that compression also manipulates the

[88] Here is the full quote: 'I use compression because it's the only way that you can truly modify a sound. Whatever the most predominant frequency is, the more you compress it the more predominant that frequency will be. Suppose the predominant frequencies are 1k to 3 kHz. Put a compressor on it, and the bottom end goes away, the top end disappears, and you're left with "Ehhhhh." [Makes a nasal sound] So for me, compressors can modify the sound more than anything else.' Andy Johns, quoted in Oswinswki (2013: Ch. 9).

intersection between these parameters in complex ways, further intensifies the commodity fetishism of analogue compressor emulations.

Vocal compression in commercial pop vocal editing: contributing to vocal foregrounding by disciplining vocal "dynamics"

The use of compression in commercial pop vocal editing is not intended to completely flatten out or restrict the dynamic range of a voice. Rather, it is broadly intended to contribute to the sense of vocal foregrounding, which I described in Chapter 2. The following quotes from my field data and relevant internet materials illustrate some of the ways commercial pop producer-songwriters envisage the aesthetic effects of vocal compression in this way.

> I've used this compressor [the DBX-160] 'cause it has a very aggressive, punchy sound to it. It adds like, punch onto the beginning of the [backing vocal phrase]. It's just the nature of this compressor, it just sounds like that. (Chris, 2020H)

> That's the secret to vocals that are up-front—it's a very reduced dynamic range. It has to all be about the same volume, which is not realistic, but it's the way we like to hear vocals. It's the way to hear [vocals] on top of a dense mix. (Waves Audio 2017)

> Then there is compression. A lot of compression! The *1176* has an 8:1 ratio, so it is slamming the vocal, giving between 10 and 20 dB gain reduction. It also has a fast release, all because his voice is so powerful and dynamic. […] We compressed the vocal to bits, so things like breaths and tails get really loud, so it's important to fade them down. (Tingen 2019)

> He slams a compressor and then puts a Pultec after it, and that basically gives you this ridiculously compressed, warm, fat, vocal. And then you just blend that in, which just helps give it a bit more consistency and just like … the 'pop sound'" […] The way I would describe my vocals is… quite compressed, but natural sounding. (Oliver, 2019BB)

> The vocal sounds more tonally consistent with compression […] the settings I'm using are probably not appropriate for this record which is supposed to be sounding a bit more open and natural, but if this was a dance-pop type tune I don't think the settings I'm using are overly aggressive at all. (Pro Audio Files 2016)

> What I would normally do is put the compression on first, and like EQ into the compression, and try to make it sound smooth basically. […] You can almost not hear the difference in terms of compression, it just makes it sounds a bit more consistent. If you can really hear it, it's probably not right. [for example…] you can definitely hear how it brings out all the noise between the… uhh… the letters. All the weird little mouth sounds and stuff. I think they get super loud and it sounds kind of bad. So that's not what you want. (Jason, 2021A)

The most immediate thing which comes across from these quotes is the variety of ways compression is understood to contribute to the aesthetic aim of vocal foregrounding. Compression makes voices sound "aggressive", "punchy", "warm", "fat", "smooth" and "consistent", and helps them to "sit on

top of a dense mix". These aesthetic values are reminiscent of the categories Case puts forward to describe the aesthetic reasons behind compression: to repeat, compression may be used to 'increase perceived loudness', 'improve intelligibility and articulation', and to 'smooth [out] performance' (2007: Ch. 6). The problem is that these aesthetic goals often get in the way of one another: in the process of increasing the perceived loudness and improving the intelligibility and articulation of the vocal, the "smoothness" of the vocal is compromised because things like "breaths and tails" and "mouth sounds" "get really loud" and "sound kind of bad". Given the critical definition of compression I have just provided, one way of describing the aesthetics of vocal compression in commercial pop vocal editing would be that it is intended to even-out the voice's macrodynamics whilst simultaneously trying *not* to make the voice's microdynamics too consistent. Consistent loudness and intelligibility are important, but unwanted "mouth sounds" are brought up to the same level as the rest of the voice because of this increase in loudness. The dynamic range is condensed to increase the voice's foregrounded positioning within a mix, but the voice's "natural" sense of dynamic variability is ideally retained, and certain features of singing like breaths and consonant sounds are therefore made quieter using *further* compression techniques if the compression makes them too loud. In order to work towards these subtle aesthetic ideals, my informants usually compressed the voice in a series of stages, as I will describe in more detail in subsequent sections.

Further, however, compression techniques are also fuelled by the erotic type of voice fetishism, specifically in relation to the ideas about vocal "dynamics" I discussed in Chapter 2. Demonstrated by these quotes (especially quote 3) and extending throughout this chapter, producer-songwriters' discussions of compression are characterized by an assumption that the voice is an especially "dynamic" and volatile sound signal which demands special attention in comparison to other elements of mixes. In particular, the voice's "dynamics" are considered particularly unpredictable: its envelope characteristics change depending on the words sung, and its perceived loudness varies depending on the register and level of energy the singer is using. As I described in Chapter 2, these ideas about the voice both reflect and reinforce the erotic kind of voice fetishism underpinning the aesthetic discourse of commercial pop vocal editing, because they mark out the

voice as especially complex, aesthetically generative and ripe for modification in comparison to other elements of commercial popular music mixes. More importantly, even though compression partly aims to flatten out some of the dynamic qualities of voices to achieve vocal foregrounding, such fetishism of the voice also fuels a desire to retain as much of the dynamic volatility of the voice as possible in the face of this transformation, because it is perceived as an important, innate quality of voices. (This recalls the discourse of vocal realism I discussed in Chapter 4, however allusions to the "real" were nowhere near as prominent in relation to compression). These ideas about the voice combine with the commodity fetishism of analogue emulations to drive the continued usage of these emulations, as well as to intensify the multi-stage compression processes I discuss later in the chapter.

These quotes also demonstrate how descriptions of vocal compression are dominated by aggressive, violent language, incorporating terms which dramatize the mediation of a voice through a compressor (e.g. "slamming"). This language, and its use in a highly gendered context, becomes especially significant given this erotic fetishism of the voice, as compression seems to be a way of disciplining the voice's dynamic volatility. The interpretation of the voice as something volatile also seeped into the way some of my informants considered such sonic volatility to result from the vocalist's inability to control their dynamics. One anonymous producer-songwriter I observed, for example, discussed how his subtle use of a compressor was only for "safety" really, "just in case the singer goes really loud". The interaction between this technique and the fetishism of the voice therefore arguably strengthens producer-songwriters' conceptions of vocalists as particularly unpredictable beings, contributing to a more general devaluation of the vocalist's labour in vocal editing.

5.2. The commodity fetishism of *1176* and *LA-2A* emulations

Digital emulations of the Universal Audio *1176* Limiting Amplifier and the Teletronix *LA-2A* Leveling Amplifier compressor units—colloquially referred to by Western music production professionals as the *1176* and *LA-2A*—were used by most of my key informants for vocal

compression.[89] These compressors have a legendary status among older producers and engineers as well as various music production publications, and they are perceived as particularly effective when applied to vocals (Fuston 2020a, b; Senior 2009c; Universal Audio 2000, 2009; Interviews 2019C, 2019D, 2019F, 2019P, 2020E, 2020D). The popularity of these analogue emulations in commercial pop vocal compression—in comparison to other techniques in the vocal chain such as EQ and tuning in which analogue emulations were much less universally used—illustrates how the commodity fetishism of analogue emulations in compression complements the erotic fetishism of the voice in this culture. Specifically, the already established reputations of these compressors lead producer-songwriters to believe that their emulations lend elusive sonic 'signatures' or 'characters' to the sound of the voice in a digital context, and this combines with ideas that compression has powers to discipline the voice's dynamic unpredictability more generally to reinforce their widespread usage. In addition, all my informants used *1176* and *LA-2A* emulations produced by either Waves Audio and Universal Audio, suggesting that such companies played a big role in maintaining these interacting fetishisms in vocal compression.

To provide some context for the legendary status of these compressors, both the *1176* and *LA-2A* were first produced in the 1960s by Universal Audio and Teletronix respectively and have slightly different circuitry[90] and features. Universal Audio produced re-issues of these hardware units in the early 2000s and now produces some of the most widely used digital *1176* and *LA-2A* emulations. Importantly, and as Sterne suggests, the contemporary reputation of Universal Audio is built on the success of these two compressors (Sterne 2020: 167-168). Their publicity claims that the widespread initial success of the *1176* was due to its 'unique lightning-fast attack and release times', 'class A output stage', and flexibility of compression intensity from very transparent to very heavy using the 'all buttons in' mode (Fuston 2020a). The *1176* also has a unique interface which enables setting the threshold and make-up gain via input and output controls and provides four fixed ratio options (Figure

[89] Although these units were originally designated 'limiting amplifier' and 'levelling amplifier', in contemporary music production discourse these two units are usually just described as compressors.
[90] See Case (2007: ch.6) for a more detailed explanation of the differences between FET and Optical compressors.

5a). Audio engineer Austin Moore and others have also conducted a substantial amount of research exploring their belief that the *1176* has an identifiable 'aggressive' sound (Moore 2012, 2020; Moore, Till & Wakefield 2016; Moore & Wakefield 2017). According to Universal Audio, the *1176* is also renowned among older engineers for the 'brightness' and 'presence'[91] it can impart to a sound (Universal Audio 2000: 14-15).

Similarly, in the case of the *LA-2A*, Universal Audio suggests that this compressor's popularity is largely down to its 'slow, gentle and versatile multi-stage release time' which gives the compressor a 'unique, musical quality' (Fuston 2020b). Engineer Mike Senior, writing for Sound on Sound magazine, echoes this sentiment, arguing that 'the *LA2A*'s electro-optical feedback design works more slowly and in a non-linear fashion, which nonetheless produces very musical results, preserving the impression of performance dynamics despite quite extreme level management' (Senior 2009c). Setting aside the technicalities of these descriptions, the idea of the *LA-2A* somehow having a "musical" quality comes to the fore as a suggestive, hazily defined interpretation of the effects of this compressor. Further, and in contrast to the *1176*'s association with the term "aggressive", the *LA-2A* is frequently described by older audio engineers as a compressor which imparts an elusive "warmth" (as well as other terms like "fatness") to a vocal (Universal Audio 2000: 15). Much of the preliminary fieldwork I did with older engineers at the start of my ethnography corroborated this interpretation of the *LA-2A*'s sound (2019C, 2019D, 2019F, 2019P). The *LA-2A*'s interface is also particularly simple, which could partially explain its popularity. It provides two fixed ratios via the compress/limit switch (3:1, 100:1), and simplified threshold and make-up gain controls ('peak reduction' and 'gain') (Figure 64a).

The digital plug-in emulations of these compressors are aural and visual skeuomorphs: they attempt to model the actions of these compressors as closely as possible in their software algorithms (Universal Audio 2004; Sterne 2020), and they reproduce the visual character of the analogue hardware interfaces nearly identically.[92] The emulations produced by Waves even include options to

[91] Again, this illustrates the ubiquity of this term in the aesthetic discourse of vocal editing.
[92] See Sterne (2020: 163-166) for a discussion of visual and aural skeuomorphism in UA's analogue emulations.

add the ‘hum’ that would be produced by the power supply of the hardware unit via the ‘analog’ section at the bottom of the interface, which is testament to the allure of an elusive analogue ‘sound’ in digital production contexts (Figures 63c & 64c; Waves Audio 2019). Figures 63 and 64 show the original hardware interfaces of the *1176* (‘N’ model) and *LA-2A* alongside two of the equivalent software emulations by Universal Audio and Waves respectively.

(a)



(b)



(c)



Figure 63: Hardware interface (a) and plug-in interfaces by Universal Audio (b) and Waves (c) of the 1176 compressor. (Images from https://www.uaudio.com/hardware/1176ln.html, https://www.uaudio.com/uad-plugins/compressors-limiters/1176-collection.html, https://www.waves.com/plugins/cla-76-compressor-limiter#image)

(a)



(b)



(c)



Figure 64: Hardware interface (a) and plug-in interfaces by Universal Audio (b) and Waves (c) of the LA-2A compressor. (Images from https://www.uaudio.com/hardware/la-2a.html, https://www.uaudio.com/uad-plugins/compressors-limiters/teletronix-la-2a-collection.html, https://www.waves.com/plugins/cla-2a-compressor-limiter?#image)

The *1176* and *LA-2A* as vocal compressors: commodity fetishism and voice fetishism combined

In informal music production literature and marketing materials, various justifications are given for the status of the *1176* and *LA-2A* as legendary vocal compressors. These consistently allude vaguely

to the ‘sound’ or ‘character’ that the compressors impart to the voice, and such interpretations are touted by companies like Universal Audio which produce the compressors and their software emulations, in the form of supporting quotes on user manuals. For example, Universal Audio’s manual for the *LA-2A* includes a quote from engineer Mike Shipley:

> The *LA-2A* is not as versatile, but it also has a sound that I really like. On certain voices you can crank it heavily, to where you almost want to put a piece of tape over the meter because there's so much gain reduction that you don't want anyone else to see it! I'm not particularly into over-compression, but when you use it that way there's something about it that just sounds really great. […] If you have a singer with an intensely piercing voice I find that compressor […] incredibly useful. It makes things warmer, especially when you crank it […]. (Universal Audio 2020: 14)

This quote also illustrates how the compressors are perceived as having a desirable effect on the voice’s timbre or “tone”. Similarly, the manual also quotes engineer Jim Scott’s description of how the *1176* has ‘an equalizer kind of effect, adding a coloration that's bright and clear’ (ibid.: 15). Broadly, such statements illustrate the particular sound-oriented flavour of commodity fetishism that is common to these compressors: they are perceived as having identifiable, aesthetically desirable sonic effects on voices, and such effects are either described using vague, suggestive language or described as indescribable: ‘there’s just something about it that sounds really great.’

In commercial pop production, the use of the *1176* and *LA-2A* in sequence was perceived as a particularly useful strategy for dealing with the singing voice’s volatile dynamic range: by using the *1176* to reduce the dynamic “peaks” slightly, followed by using the *LA-2A* to do another round of compression to “smooth out” the voice in a way which sounded as subtle and “musical” as possible (Make Pop Music 2020). This was a common tactic for compressing commercial pop vocals among my informants (2020H, 2020Q). Pedagogical music production literature suggests that the efficacy of this technique lies in the fact that the fast attack and release times of the *1176* help to tame the highest peaks[93] of the vocal signal, then the *LA-2A* follows to give the vocal a “full, rounder sound” (Waves Audio 2019) that shouldn’t sound too obvious because the dynamic range has already been condensed significantly by the *1176* (ibid.; Make Pop Music 2020; Crane 2020.). Across my fieldwork, however,

[93] Among my informants, this term was used interchangeably with the term ‘transients’ to refer very generally to ‘the maximum instantaneous level of a signal.’ (Sound on Sound 2022).

there was no clear reason for precisely why the *1176* and *LA-2A* were considered such legendary compressors on vocals. Older sound engineers such as Universal Audio's founder Bill Putnam Jr. might argue that 'the heart and soul of the *LA-2A* is in the T4 cell' (Fuston 2020b), but most of my informants in commercial pop justified their usage of the *1176* and *LA-2A* by using words to the effect of 'they just sound good'. Oliver, for example, when discussing the two compressors, explained how he just "throws [them] on there—if it sounds great, [I] use it." (2020S). The fact that this sequential use of the *1176* followed by the *LA-2A* was not consistent across my secondary fieldwork materials either suggested that there wasn't necessarily any definite logic behind their sequential use on vocals, beyond the ideal of subtle-sounding compression. They were most used in this order, but they were also often inverted, or other compressors used in sequence: for example, one video in which a producer tried to deconstruct the vocal editing processes of celebrity mixer Serban Ghenea theorized that he inverted the two compressors in a way that produced a "distinct" "locked down" sound (Pro Audio Files 2020).[94]

All of this is to say that it is extremely likely that the legendary status of such compressors has been driven by their commodity fetishism, which is reproduced by these younger commercial pop producer-songwriters in a digital context. This is partially due to their tuition by older producers who grew up using these hardware compressors,[95] as well as their self-teaching using informal pedagogical materials on YouTube, but the marketing of *1176* and *LA-2A* emulations by plug-in companies like Universal Audio and Waves also effectively consolidates their legendary status, and their commodity fetishism, which relies on the assumption that they have elusive sonic signatures. Further, however, these widespread ideas about the power and efficacy of the *1176* and *LA-2A* when dealing with the voice's volatility resonate with voice fetishism, which combines with the existing commodity fetishism of the analogue units to strengthen the commodity fetishism of digital analogue compressor emulations in vocal editing.

[94] Again, I am referring to this video because Serban Ghenea was referenced by several of my informants as a reference for what good commercial pop vocal editing sounded like (e.g. 2020E).
[95] Oliver, for example, explained how he was taught about the effectiveness of the *1176* and *LA-2A* by a university lecturer. (2020R)

Waves' advertising partnership with Chris Lord-Alge and the masculine tone of *1176* and *LA-2A* commodity fetishism.

One of the most dominant plug-in companies in the market for compression plug-ins is Waves Audio, whose suite of analogue emulations called the 'CLA Classic Compressors'—which includes the *CLA-76* and *CLA-2A*—was used very frequently by my informants. The popularity of these plug-ins illustrated how Waves was playing an important role in reproducing the commodity fetishism of the *1176* and *LA-2A* in digital vocal editing contexts, as well as highlighting the masculine tone of such commodity fetishism.

The 'CLA' which prefaces the names of all these plug-ins refers to American rock engineer Chris Lord-Alge, supposedly renowned for his use of dynamic range compression across mixes (Tingen 2007). Lord-Alge partnered with Waves in 2009 to release this new set of compressor emulations (Waves Ltd. 2009), after Universal Audio had released a similar suite of emulations in the early 2000s (Universal Audio 2004). Figures 63c and 64c show these emulations. In my fieldwork, there appeared to be a strong, naturalized link between the Lord-Alge/Waves partnership and *1176/LA-2A* emulations. This suggested that this marketing partnership covertly increased the commodity fetishism of the emulations. Specifically, Lord-Alge has a reputation among music production professionals for powerful, exaggerated dynamic range compression more generally, nearly always using analogue units or emulations (Tingen 2007). Not only this, but Lord-Alge's reputation links to his work as an engineer on very successful pop-punk tracks during the 2000s (including My Chemical Romance, Green Day and Simple Plan), making a link between Lord-Alge's use of analogue compressor emulations to drastically and powerfully affect sound with a different kind of music more directly related to rock aesthetics. The way Lord-Alge talks about compression also emphasizes its felt, gestural sonic effects in a dramatic, flamboyant manner: in one interview, for example, he described his use of 'hard' compression on vocals with the *1176* as an addition of 'attitude' and 'personality', and compared the actions of compression more broadly to those of a gun:

> Compression can be dangerous in the wrong hands. It's like a gun; once you shoot a hole in something, you can't plug it. […] I would rather compress a vocal hard to get some personality out. It's like

> putting a snazzy jacket on a guy when he goes out at night—it adds attitude. And nothing adds attitude more than *1176*s. […] (Lord-Alge, quoted in Droney 2001)

Moreover, the marketing of these plug-ins has tapped into a particular 'macho' way of talking about compression which emphasizes its aggressive and violent effects. Such marketing materials include videos of Lord-Alge discussing the plug-ins, where he frequently extemporizes on their powerful sonic effects, alluding to their 'character', 'personality' or 'warmth' and their ability to 'absolutely slam' or 'spank' the vocal (Waves 2009; Waves Audio 2020a). The entire CLA marketing strategy is centred around these ambiguous semantic descriptors used to describe the effects of compression, which intensify ideas about the oblique 'sound' underpinning the commodity fetishism of these analogue emulations, as well as aligning such sonic effects with notions of violence and aggression.

Alongside the *1176* and *LA-2A* emulations, Waves produced a second suite of CLA plug-ins in the early 2010s—the 'CLA Signature Collection' (Inglis 2011)—which further exemplifies this 'sound-oriented' marketing. This included the plug-in *CLA Vocals* which provides the user with an entire vocal chain inside a single plug-in, intended to offer a quick and easy way to replicate the CLA vocal sound. Although not solely related to compression, this plug-in provides context for the marketing strategies Waves use which directly target the intermingling of erotic voice fetishism and the commodity fetishism of analogue emulations in compression. I observed several informants use this plug-in, often during the early stages of mixing to get a feel for what the vocal would sound like when fully processed (2020CC, 2020H, 2020L). The 'compression' section of the plug-in simply consists of a fader with three options at the top for the style of compression applied: 'push', 'spank' or 'wall' (Figure 65). As engineer Andrew Bourbon has argued, plug-ins which coalesce around 'semantic descriptors' like these represent an attempt by plug-in companies to innovate beyond producing analogue emulations, instead creating plug-ins which gesture vaguely at the 'sonic personality' or 'sonic impact' created by their application (Bourbon 2020: 219-220). Notably, even though this plug-in provides an entirely novel, digital way of manipulating sound that has no basis in any kind of audio hardware, the plug-in interface is designed as if it were an analogue emulation, incorporating sliders, buttons, meters and even screws that appear to hold the unit together. Lord-Alge

even claimed in a tutorial-cum-advertisement for the plug-in that simply applying the plug-in to a channel without adjusting any of the parameters would do something 'good' to the sound in question (Waves Audio 2013). This plug-in therefore reproduces the same kind of 'sound'-oriented commodity fetishism characteristic of the *1176* and *LA-2A* emulations, and contributes to the Waves/CLA partnership's maintenance of such commodity fetishism among commercial pop producer-songwriters.



Figure 65: *CLA Vocals* plug-in interface. (Image from https://www.waves.com/plugins/cla-vocals).

Furthermore, the 'macho' language surrounding these digital compression technologies is not insignificant given the gendered context of commercial pop vocal editing. The marketing of the CLA plug-ins takes this language to the extremes, recalling aggressive masculine violence in the context of the broader masculine coding of the commodity fetishism of technology across Western music

production (described in Chapter 1, pp. 93-95). Watching a few videos of Lord-Alge describing his process provides a feel for his unrelenting and somewhat unsettling machismo (Waves Audio 2013, 2020a, 2020b).[96] In a demonstration video for the CLA vocals plug-in, for example, Lord-Alge is filmed in a dimly-lit mixing booth wearing all black including a baseball cap and large metal dog tags (his signature outfit), and gesticulates at the CLA plug-in in action, extoling the plug-in's numerous virtues (Waves Audio 2013, statements to this effect include "the plug-in is completely designed how I mix" (5'35") and "I went with red [for the plug-in design] because when you see red, it's important" (1"35")[97]). My informants did not embody quite as extreme attitudes: indeed, one felt that the only reason these plug-ins were so popular among commercial pop producer-songwriters was that they were cheap and easy to use (2021D). But it is nonetheless significant that such plug-ins are in widespread usage given that the language surrounding vocal compression is consistently gendered in this way, and is undoubtedly intensified and promulgated by the Waves/CLA partnership. More targeted research would be needed to ascertain what kind of effects this violent discourse has, if any, on the gendered dynamics and interactions involved in studio recording and production.

It's worth pointing out that such marketing strategies did not appeal to some of my informants, however, who felt that the 'macho' language was cringey and overtly commercialised. Oliver, for example, felt that a recent Waves/CLA advert for a new reverb & delay plug-in suite called *CLA Epic* (released in November 2020), was a bit over-the-top—a bit "promoted"—and aimed at a certain "market" of music producer to which he did not feel he belonged (and which perhaps espoused a more extreme form of plug-in commodity fetishism than he did) (2020S). Despite Oliver's opinions, however, he occasionally incorporated the other CLA plug-ins into his own vocal editing processes, which suggested that the cheap and easy-to-use appeal of these plug-ins trumped any uncertainty regarding the language with which they were marketed, and was testament to the success of their marketing strategies.[98] This illustrated the efficacy of the 'sound'-oriented marketing rhetoric of the

[96] Links to these videos are provided in the bibliography.
[97] Here is a link to this moment for ease of access: https://youtu.be/qo1gZgycDGQ?t=95
[98] The product page for CLA vocals, for example, advertises how the plug-in can help the user to achieve 'CLA's radio-ready rock sound – instantly!' (Waves 2022b).

Waves/CLA partnership, which reinforces the commodity fetishism of these plug-ins. Figure 13, presented in Chapter 1 (p. 93), shows screenshots from the CLA advert in question, which should give the reader an idea of the kind of language and imagery used, which emphasizes an abstract notion of the 'sound' offered by the plug-in suite as well as presenting a suggestive representation of analogue audio hardware.

5.3. Iterative compression processes in commercial pop vocal editing

This section draws on closer analysis of my field data to demonstrate the variety of cumulative, iterative vocal compression processes used by my informants working in commercial pop. These processes gave the outward appearance of being incredibly complex and highly developed: my informants felt that compression *must* be done in stages, using a variety of different compressors, and compression plug-ins must be combined with both EQ and multiband compression/de-essing plug-ins throughout the vocal chain to make the voice sound as polished as possible. Despite this insistence on cumulative compression processes, my informants also tended not to be able to elaborate on the precise reasoning behind their operations. To me, this showed how such iterative processes were motivated by the interaction between the commodity fetishism of analogue emulations in compression and erotic voice fetishism. The *1176* and *LA-2A* supposedly created elusive yet desirable sonic effects in the voice, however there was also a sense that their usage should be balanced with the desire to retain the dynamic volatility of the voice, conceived as defining feature which ought to be preserved. This informed the gradual, cumulative application of compression plug-ins in commercial pop vocal chains, driven by the assumption that such gradual processes would ensure that the vocal compression sounded as subtle as possible, and satisfied both ideals. I will first explain two common multi-step compression processes I observed in my fieldwork: *serial compression* and *parallel compression*, both of which involved the use of multiple analogue emulations. Following this, I will also discuss the use of 'frequency-sensitive' compression techniques including multiband compression and "de-essing", which were conceived as ways of compensating for other undesirable effects caused by the combination of compression and EQ. Notably, frequency-sensitive compression was usually performed using digital novelty plug-ins, contrasting with the commodity fetishism of analogue

emulations in regular vocal compression, and the integration of these techniques within vocal compression suggests that iterative compression processes are further driven by interactions between the different kinds of voice and technology fetishism underpinning both compression and EQ.

Serial compression

One of the most common vocal compression techniques I observed was *serial compression* (or 'double-stacked' compression (Waves Audio 2017)), which refers to the use of two or more compressors in sequence on a vocal track. Although, as I have explained, both the *1176* and *LA-2A* supposedly have unique characteristics contributing to their commodity fetishism, in my fieldwork they were most frequently used in sequence, and their effects were thought to complement one another. My informants believed this helped to achieve a heavily compressed vocal sound whilst also keeping the effect subtle. 'Serial compression' was a term used more in informal music production pedagogy materials (e.g. Computer Music 2015; Pro Audio Files 2020; Krantzberg 2020; White 2012) rather than explicitly by my informants, but I am using it because it neatly explains the technique. This technique demonstrated, on the one hand, the ubiquitous usage of *1176* and *LA-2A* emulations in the practice, but it also illustrated how producer-songwriters' motivations to retain the voice's important "dynamic" qualities led them to use compression iteratively as an attempt to try and keep its desirable effects as subtle as possible. For example, in one session, Oliver used four steps of compression in his vocal chain, along with five rounds of multiband compression and de-essing to perform particular "smoothing" functions which responded to the effects generated by compression as well as EQ. Figure 66 shows Oliver's chain along with a description of each plug-in's function in the sequence, in Oliver's words. The main serial compression technique is represented by the addition of the Universal Audio *1176* plug-in at insert four, followed by the Universal Audio *LA-2A* three steps later in the chain. The other two steps of compression are much more subtle and happen at the end of the chain, accomplished by the Waves *RVox* and Universal Audio *SSL channel strip*, which Oliver interpreted as more "characterful" finishing or finalizing actions. Note that aside from the preamp and EQ towards the start, the entire chain consisted of compression and frequency-sensitive compression plug-ins.



Universal Audio *Neve 1073*: Preamp, to "work on proximity"
Fabfilter *Pro-MB*: **Multiband compressor**, to "work on proximity"
Fabfilter *Pro-Q 2*: EQ, to try and "richen it up a bit"
Universal Audio *1176*: **Compressor**, *usage unexplained*
C1: sidechain **multiband compressor**/compander, to "smooth a harsh frequency"
Fabfilter *Pro-DS*: **De-esser** set to "the standard frequencies on a female vocal"
Universal Audio LA-2A: **Compressor,** *usage unexplained*
Universal Audio Pultec: **EQ** for "presence", "air" and "low end boost"
Fabfilter *Pro DS*: **De-esser** again, "just to tame anything that's got boosted"
Fabfilter *Pro-MB*: **Multiband compressor**, "across the whole vocal to smoothen it off"
Waves *RVox*: **Compressor,** using a preset by a famous engineer called "good vocal"
Universal Audio *SSL channel strip*: **Compression** and **EQ**, "to give it that professional edge"

Figure 66: Oliver's vocal chain (inserts only) for Audio 3, with plug-in names and functions in Oliver's own words (2020Q). Oliver's descriptions are those he gave whilst he was producing the track, as opposed to during the later follow-up interview.

When I first observed Oliver using this vocal chain and talking about the order of the plug-ins, I assumed that he would have a precise reasoning behind these complex and repeated iterations of compression and multiband compression/de-essing, since they seemed to take up so much of his vocal chain. I also assumed that he would have a reason for using the *1176* followed by *LA-2A* serial compression technique, which I had read about online and observed repeatedly in YouTube videos. In fact, a follow-up interview with Oliver revealed how his use of serial compression techniques was a lot more intuitive rather than precise, mostly geared towards creating an overall 'compressed' vocal sound whilst also ensuring that no plug-in could be identifiably heard. Broadly, this showed how Oliver's serial compression technique was driven by the commodity fetishism of the analogue emulations and the interaction of such commodity fetishism with dominant ideas about the voice's innately "dynamic" qualities.

Nonetheless, Oliver did seem to have some suggestive ways of interpreting the 'sounds' of the *1176* and *LA-2A* on vocals, illustrated by his explanations of how these compressors worked on the bass guitar:

> I'll take it to bass for example. Like if it's like a funky sort of bassline I'd go for an *1176*, so translating that into vocals you wanna capture every word and things. And then for the *2A* I would use on a sustained bass, which is also good for the vocal for when they're doing longer notes and things like that. So I just like for a vocal to use both, to get the best of both worlds in there. […] it's just the

> different characteristics… each compressor's got a certain, like, use, I think. But I think, for vocals, the *1176* and the *2A* give a well-rounded sort of solid-sounding vocal. (Oliver, 2020S)

There was clearly a sense on Oliver's part that each compressor was different and imparted a different 'sound' to the vocal, but this was something which Oliver felt intuitively and it was not based on any precise reference to the functionality of the compressors' attack and release parameters in the way that Universal Audio describes. When I asked Oliver to explain the settings he used on his *1176* and *LA-2A*, he didn't have a reason—it was just that they sounded good to him at the time (ibid.). Also notable is the fact that, during observation, Oliver provided short explanations for his usage of all of the other compression and multiband compression plug-ins (e.g. "to give that professional edge" (2019BB)) but left the *1176* and the *LA-2A* without explanation, as if their efficacy as vocal compressors was taken for granted.

Oliver's honesty about his intuitive approach to serial compression was not universal across my fieldwork, however. Another commercial pop producer-songwriter who posts publicly on YouTube, for example, offered a slightly more descriptive reasoning behind his *1176/LA-2A* serial compression technique, explaining how he used the *1176* with a fast attack and release to "squash" the vocal and get more "consistency", especially in the chorus of the track he was working on where the voice was louder and the *1176* helped to "even these [transients] out" to avoid "massive volume jumps"—followed by the *LA-2A* which "provides the main squeeze on the vocal" (Make Pop Music 2020). But this reasoning was no more logical or precise than Oliver's intuition that "the *1176* and the *2A* give a well-rounded, solid-sounding vocal". Both producer-songwriters were simply using serial compression to give the vocal an overall 'compressed' sound whilst trying to maintain the subtlety of the effect by distributing the compression process between two compressors.

More generally, my observation of Oliver's serial compression process also illustrated the effects of the technique on the resulting vocal sound in Oliver's track. Audio files 5 and 6 show Oliver's lead vocal (along with a couple of harmony parts) before and after the chain in Figure 10 has been applied, and they should give the reader some idea of the sonic effects of multi-step compression processes like Oliver's (bear in mind that reverb has also been applied which gives the impression that

the voice is in a larger space). The difference is very subtle indeed, and there are few audible artefacts from the processing technologies Oliver has used, except perhaps small tremors in the voice that might be a result of the tuning process. Listening carefully, you can hear that the voice sounds a lot more "dynamically consistent" and "smooth" in the way that Oliver and others describe. For example, listen to the phrase 'like my dress you tore me' (0'13"-0'15"): in the first clip, the singer runs out of breath after the word 'you' and resumes the phrase at 'tore', and you can hear a noticeable jump in volume and energy for this word. In the second file, this transition sounds much smoother. This difference in "smoothness" and "consistency" is not something that can be represented easily using visual notation or spectrograms—perhaps providing some further context for the way compression has been mystified in music production more broadly. But these examples show how the multi-step compression processes Oliver is using *are* doing something to the vocal sound and contribute to creating the "well-rounded" and "solid-sounding" vocal he wants, even though he could not explain precisely what was happening. Similarly, the application of the serial compression technique by the YouTube producer-songwriter is audible as we hear him constructing his vocal chain in the video, even if it is subtle: comparing from timestamp 4'44" to 5'44",[99] it is possible to hear how the voice sounds a little louder and more "dynamically" consistent. These brief audio examples illustrate how compression does have a defining effect on voices in this genre, even though the actual effects of multi-stage compression processes are more-often-than-not difficult to distinguish from the effects of other compression techniques. More significantly, it shows how the interaction between voice and technology fetishism informs how the sound of commercial pop voices is produced.

Parallel compression

Often combined with serial compression techniques, some of my informants also used a technique called parallel compression on vocal tracks, which was similarly intended to both preserve a voice's dynamic variability as well as compressing it to increase the perception of overall loudness. My informants saw parallel compression as a useful technique which allowed them to try and get the

[99] Here's a link for ease of access: https://www.youtube.com/watch?v=N0lVM-J9JyQ&t=284s.

desirable effects of compression—that is, making the energy of a vocal appear more consistent and helping it remain foregrounded in a mix—without 'flattening out' the vocal sound in the process. Uniquely, parallel compression techniques therefore embodied both the commodity fetishism of analogue emulations and illustrated very clearly how fetishism of the voice's "dynamic" characteristics fuelled a specific compression technique using these compressors. Furthermore, the technique of parallel compression seemed more generally fetishized, having itself acquired a legendary status among pop producer-songwriters and being prized for the ambiguous yet productive sonic effect it has on pop vocals.

To provide some context for the technique: parallel compression, in the most basic sense, involves creating a copy of a vocal track, compressing it, and then blending it in with the uncompressed original vocal track (Izhaki 2017). A mixing textbook by Roey Izhaki sums up the aesthetic advantages of parallel compression succinctly: 'parallel compression lets us make sounds even bigger while retaining their dynamics' (2017: Ch. 17). This is because the uncompressed and compressed vocal sounds are sounding simultaneously: the parallel track is compressed very hard, so the dynamic range is condensed, but it is sounding at a much quieter level, so the overall effect is that the parallel track helps to amplify the quieter parts of the original track without pulling the peaks down too much (Izhaki 2017; Robjohns 2013). Pedagogical sources argue that this technique therefore helps to preserve the dynamics from the original vocal performance (ibid.). The technique is well-known across audio engineering and music production circles and has been around for at least 30 to 40 years: it is sometimes referred to as 'New York compression' due to its association with New York engineers who frequently used the technique (Izhaki 2017: 322).[100]

My informant Robin emphatically extolled the virtues of parallel compression to me in an interview, describing it as if it were the secret to creating great pop vocals:

> You kind of have your drier vocal there on one channel, but instead of slamming that through a compressor—you compress it a little bit—but then you run a parallel chain where you absolutely batter it, you hammer it, full blast with a compressor. And then you blend it in with the dry signal. So instead

[100] You can hear an example of quite subtle parallel compression on Coldplay's 'Violet Hill', mixed by engineer Michael Brauer who supposedly sent the lead vocal to five different compressors on a parallel track. (Tingen, 2008).

> of bringing the louder bits down, you're bringing the quieter bits up. Once you recognise the sound of it, you listen to everything else and you're like oh, that's what it is… it really flattens the signal […] Once you've heard it on a vocal you're like oh, that's the secret! (2020I)

Similarly, and to repeat an excerpt of a quote from earlier, Oliver explained how he frequently used parallel compression to get that "pop sound":

> It gives you this ridiculously compressed like warm, fat, vocal. And then you just blend that in, which just helps give it a bit more consistency and just like … the 'pop sound'… (2019BB)

Both producer-songwriters evidently believe in the sheer power of this technique, and its ability to define the sound of "pop" vocals, using similar metaphorical language which emphasizes the power and aggression of compression and its ability to create "warmth" and "fatness". These quotes are derived from initial interviews, however, and significantly, during observation both producers decided not to use parallel compression on their vocal tracks, despite having defined it as a key feature of their practice. Robin, for example, felt that he was doing "too much compression already" for parallel compression to be necessary, having used "quite a harsh compressor and a brick-wall limiter making it as punchy and flat as possible" in his vocal chain.

> The first thing that strikes me is that I didn't use parallel compression in this track. I think there's so much compression generally that it didn't need it. This highlights the fact there is no cookie cutter method to production! (2020I)

Oliver, in a follow-up interview, also explained how he chose not to opt for his usual parallel compression technique on the track in Audio 3 because the voice was too "soft"—in addition, he contrasted this track with some of his other productions in a slightly different style where the vocal needed to sound more "powerful", and therefore parallel compression would be more appropriate:

> The stuff I'm working on at the moment that's a bit like tech-house, or even piano-house stuff where the vocal's like a big powerhouse vocal, and you know it's like a forefront. That's where I'd look to parallel compress it there I think. Just to get this fat well-rounded sound that's present and consistent. Whereas here, I think with [her] voice being softer, I think… like it definitely added thickness, but not a powerful thickness, if that makes sense. I don't know how to explain that. It was sort of like soft and then… like, it was adding more softness underneath. Whereas if you have like this big powerful vocal and then you add that flavour back in it can sort of … [trails off] (2020T)

Oliver's assessment, in particular, suggested that parallel compression only worked on commercial pop vocals when the "tone" of the voice sounded a certain way—not so much gentle, soft and breathy,

but more powerful and sustained. The technique is therefore perhaps not as universally effective in creating the "pop sound" as the statements by Robin and Oliver initially suggested. Their descriptions of the technique, however, suggest that it is fetishized: it is thought to provide an elusive yet highly aesthetically effective vocal sound ("once you hear it… that's the secret!"), and yet the exact effects of the technique are very subtle and difficult to explain precisely. The rhetoric surrounding parallel compression is therefore an example of how the language surrounding vocal compression amps it up as something elusive, powerful and complicated, when in reality producer-songwriters' application of compression on vocals relies more on an intuitive sense of whether or not the compression has achieved the sound of vocal foregrounding which is so essential to the genre. Their discussions also suggest that the fetishism of this technique has been naturalized, and this fetishism is reproduced in discourse about parallel compression by my informants even though they did not use the technique in practice.

Jason and Chris's vocal editing also illustrated how the commodity fetishism of the *1176* and *LA-2A* emulations fuelled parallel compression techniques. Both Chris and Jason used Waves' CLA emulations to achieve parallel compression effects very quickly, which was evidence for the success of Waves' celebrity engineer marketing strategy in promoting such plug-ins. In every track I observed Jason produce, he compressed the vocal in parallel using the *CLA-2A*, including a quick round of EQ before and after for some "shaping" (2021B). Chris used the *CLA-76* in a similar way (2020H). Additionally, particularly notable about Chris and Jason's use of parallel compression was the way they seemed to take it for granted as a tried-and-tested, effective compression technique. The language they used to describe it was quite matter-of-fact, as if the technique was just assumed to be a defining feature of the sound. It was also not something they elaborated on naturally in video materials, often assuming that I would be well aware of the technique and its effectiveness—instead, it was something I had to ask them to expand upon in later follow-up interviews.

> It's just a parallel compression. [Chris opens *CLA-76* with a preset already loaded and fiddles a tiny bit with the settings, making the vocal sound slightly louder, and the mouth sounds start to become more prominent] So that's absolutely slammed, and I'll bring in a bit of that underneath the lead ... as you probably know. […] So going for overall quite a compressed vocal sound. (Chris, 2020H)

This analysis therefore suggests that the widespread commodity fetishism of these analogue compressor emulations, combined with voice fetishism, drives the use of complex multi-stage compression processes in commercial pop vocal editing. More specifically, parallel compression assumes that the accepted efficacy of either the *1176* or *LA-2A* on vocals, when applied in this covert manner, will become even more effective through combination with the "true" vocal signal. These informants feel that the technique is incredibly clever because it is a kind of 'having your cake and eating it too' approach—it allows for the desirable yet elusive sound of these analogue compressor emulations to sound at the same time as the voice's inherent dynamic volatility.

Interestingly it was also the younger producer-songwriters (Robin & Oliver) whose rhetoric about parallel compression more extremely extoled its virtues—the slightly older, more experienced producers seemed less effusive about the power of the technique, instead just using it without really talking about it very much. This illustrated the embeddedness of the idea of parallel compression being effective on vocals in this culture, since such ideas seemed to be touted uncritically by the younger producer-songwriters yet accepted as fact by the older ones who had likely been using the technique more frequently.

On a side note—this section has included more generalized quote-based data because the sonic effects of parallel compression are particularly difficult to analyse in this context. It is incredibly difficult for listeners untrained in music production to identify the effects of compression at all, let alone to distinguish between the effects of serial and parallel compression. In particular, more field data would be needed to ascertain whether the sonic effects of parallel compression on commercial pop vocals match up with the way producer-songwriters talk about it as simultaneously increasing loudness but also preserving vocal "dynamics". What this analysis does show, however, is that my informants tended to prioritize 'legendary', well-established compression techniques like parallel compression in their vocal editing because they had a strong belief that such techniques, applied using analogue emulations, would affect the vocal in a positive way, even if they only had an intuitive understanding of their effects.

Frequency-sensitive compression, digital novelty commodity fetishism and the interactions between compression and EQ

Alongside applying regular compression to the voice using analogue emulations, my informants also added "multiband compressors" and "de-essers" to their vocal chains, which were more likely to be digital novelties. These are 'frequency-sensitive' compression plug-ins (Jeffs et. al. 2005)—plug-ins which build on the baseline functionality of compression by enabling the user to target compression actions at specific frequency bands. Often, multiple iterations of such plug-ins would be used in an attempt to attenuate undesirable sounds in the voice that had been amplified through either EQ or compression (e.g. "harshness" or "sibilance"), or to pre-empt such sounds and nullify them before they could be heard. Most importantly, although they were considered forms of compression, these plug-ins shared more characteristics with para-graphic EQ plug-ins, illustrating how the entire discourse of compression shifted away from the commodity fetishism of analogue emulations when the notion of splitting up the voice into frequency areas became a primary concern. This section analyses some of the iterative compression processes used in commercial pop vocal editing which incorporated frequency-sensitive compression, showing how the iterability and compulsivity of these processes is even further heightened by the complex interactions between the different kinds of voice fetishism and technology fetishism relating to compression and EQ respectively, since frequency-sensitive compression plug-ins bring together characteristics of both techniques.

To define these processes more precisely: multi-band compression involves the splitting-up of an audio signal into two or more frequency bands that can be compressed independently and then recombined, enabling the user to apply different compression settings to each band (ibid.; Robjohns & White 2002). As one of my informants suggested, similar to serial compression and parallel compression, multi-band compression is particularly useful for vocals because it allows for more selective manipulation and handling of the voice's unpredictability: by targeting specific compression actions at specific frequency bands, the compression will respond to the way the voice's frequency spectrum fluctuates over time, only acting when specific frequencies reach a certain threshold (e.g. when the vocalist sings in a higher register), and therefore "leaving everything else untouched."

(2020H). As I explained in Chapter 4, it is a similar kind of functionality which makes Fabfilter *Pro-Q 3* so popular for commercial pop vocal EQing.[101] Furthermore, most of my informants defined "de-essing" as a kind of multi-band compression—usually enacted by a dedicated "de-esser" plug-in—that targeted the specific frequency band associated with "sibilance" (usually somewhere between 6-9kHz). As I explained in Chapter 4, "sibilance" in this culture is predominantly associated with 's' (s[102]) and 'sh' (ʃ) consonant sounds, but also encompasses other aspirated consonants like f (f) and t (t). These frequency-sensitive compression processes seemed to perform more corrective, technical roles, slightly distinct from regular vocal compression which seemed to be conceived as more generally enhancing the consistently foregrounded sound of a commercial pop vocal.

In addition, and perhaps unsurprisingly given my discussion of the brand in Chapter 4, plug-ins designed by FabFilter were commonly used for de-essing and multiband compression in my fieldwork, specifically by my informants Chris and Oliver (2020H; 2020Q). These included FabFilter's *Pro-DS* and *Pro-MB.* Given the reputation of FabFilter *Pro-Q*, the widespread use of these other FabFilter plug-ins suggested that the brand had acquired a broader reputation for work on the voice that was perceived as particularly detailed or functional, or that targeted specific frequency points within the vocal frequency spectrum. The fact that these kinds of plug-ins dominated frequency-sensitive vocal compression strengthens my claim that the different characteristics of EQ and compression techniques in commercial pop vocal editing are fuelled by different synergistic fetishisms of the voice and technology in each case: EQ practices, and the idealization of vocal timbre which underpins them, combine with the commodity fetishism of digital novelty plug-ins like *Pro-Q 3* (which give the user the illusion of great technical power over the voice) to produce realist conceptions of the technique. In contrast, compression mystifies the elusive sonic effects of analogue compressor emulations, and this interacts with the desire to remain sensitive to vocal "dynamics", fuelling complex iterative compression procedures including the iterative usage of frequency-sensitive compression plug-ins. Moreover, the use of frequency-sensitive compression plug-ins in commercial

[101] See Chapter 4, pp. 233-235.
[102] To clarify using the International Phonetic Alphabet.

pop vocal chains blends the commodity fetishism of analogue compressor emulations with the former kind of digital novelty fetishism related to EQ, because their modifications precisely target points on the frequency spectrum which are perceived to correspond to particularly undesirable characteristics of vocal timbre, rather than involving the application of over-arching enhancing gestures to the voice's dynamics using regular compression. Figures 67 and 68 show the interfaces of Fabfilter *Pro-DS* and Fabfilter *Pro-MB*.



Figure 67: Fabfilter *Pro DS* interface. This plug-in shows a horizontal rolling waveform when the track is played, highlighting the culprit "essy" moments that are being affected in green on the waveform. (Image from https://www.fabfilter.com/products/pro-ds-de-esser-plug-in)



Figure 68: Pro-MB interface. The different 'bands' of the frequency spectrum, set by the user, are represented by the vertical lines demarcating pink (low), blue (mid) and green (high) areas. (Image from https://www.fabfilter.com/products/pro-mb-multiband-compressor-plug-in)

The notion of "de-essing", although a form of multi-band compression, performs a similar function to the attenuation of sibilance by performing cuts at the upper end of a para-graphic EQ curve, which I discussed in Chapter 4 (pp. 214-215). Even though EQ was often used to attenuate sibilance earlier in the vocal chain, all my key informants incorporated a dedicated "de-esser" plug-in—sometimes applied more than once—as an integral part of their vocal chain. Reliable historical sources about de-essing are virtually non-existent, but some suggest that the desire for dedicated de-esser units has existed in music production since the 1970s: for example, one article in an audio magazine suggests that the earliest recorded standalone hardware de-esser unit was the Orban *516 EC Dynamic Sibilance Controller*, produced in the mid-70s (Szikla 2018). A later version of this hardware de-esser (*536A*) is visible in Figure 69, showing how the unit consists of a very simple interface which permits de-essing on two channels at user defined thresholds (SteamPoweredRadio 2020). Most importantly, the perceived need for de-essing results from commercial pop vocal editing's aims to foreground the

voice. Mike Senior from Sound on Sound, for example, describes vocal de-essing as a response to the problem of sibilance generated by various 'routine production decisions', including the 'preference for close-miking vocals', choice of mic, as well as 'high-frequency EQ boosts' and compression which 'controls sibilants less assertively than the rest of the vocal signal' (Senior 2009b).[103] Ironically, de-essing is an attempt to solve sibilance problems created by the interrelationship of EQ and compression by the addition of a further frequency-sensitive compression process, showing how de-essing—although it is a form of compression—begins to engage a similar kind of commodity fetishism of technology to that of digital novelty EQ plug-ins due to its focus on individual frequency components of the voice. De-essing therefore clearly embodies how the intensifying complexity of iterative compression processes is fuelled by the interactions between voice and technology fetishism, which result in back-and-forth, anticipatory and compensatory processes, using plug-ins with the same or related functionalities cumulatively in order to gradually shape the desired vocal sound.



Figure 69: The Orban *536A Dynamic Sibilance Controller* (from SteamPowered Radio.com, 2020).

It is also significant that during my fieldwork I was never able to discern the difference between a vocal before and after a de-esser had been applied. This is something which, again, is perhaps more easily discernible after the *entire* chain has applied: again, listening to the aspirated 't' of the word 'tore' on Audio tracks 5 and 6, the consonant sounds a lot less intrusive on the second track. Yet,

[103] Izhaki (2017: Ch. 17) explains this phenomenon in more detail in his description of de-essing.

listening on a smaller scale—for example as the first de-esser out of two was applied by Oliver (see Figure 66)—it was impossible to tell the difference. This suggests that de-essing is applied kind of intuitively in commercial pop vocal editing: it is interpreted as a necessity that will contribute toward the desirable result eventually, but it is difficult to track audibly in the moment. Indeed, one article in a prominent audio engineering magazine draws attention to this imperceptibility, suggesting that de-essing is only obvious when it is taken to the extremes, giving the impression that the singer has a lisp and 'making words like 'save' sound a bit more like 'shave'' (Szikla 2018: n.p.). According to sound engineer David Nicholas, 'with a de-esser, if you can hear it working, then it's too much.' (quoted in Szikla 2018).

If the combination between voice and technology fetishism in the case of regular vocal compression already contributes in part to the complex iteratability of the technique, then the interplay between the two pairs of fetishisms which operate in relation to compression and EQ respectively, blended together in the case of frequency-sensitive compression, fuels this iterability even further. The pedagogical video from Make Pop Music, mentioned earlier in this chapter in relation to serial compression, provides a neat example of such processes. The video starts with the producer-songwriter—Austin Hull—explaining the rationale behind each plug-in he is using in the vocal chain—represented in Figure 70. As discussed, Hull uses serial compression with a combination of *1176* and *LA-2A* emulations (*FG-1176* and *CLA-2A*), in an attempt to first "squash" the vocal and get more "consistency", followed by "providing the main squeeze on the vocal".



1. **Pitch-correction:** *Auto-Tune Pro*
2. Virtual Mix Rack including:
    a. **Tube microphone emulation:** *FG-800*
    b. **EQ:** *FG-N*
    c. **Compressor:** *FG-1176*
3. **Compressor:** *CLA-2A*
4. **De-esser**: Fabfilter *Pro-DS*
5. **De-esser**: Fabfilter *Pro-DS*
6. **EQ**: Fabfilter *Pro-Q3*
7. Virtual Mix Rack including:
    a. **EQ:** Slate *Custom Series EQ*

Figure 70: The vocal chain used by Austin Hull (Make Pop Music 2020), with each plug-in labelled in more detail. 'Virtual Mix Racks' are essentially skeuomorphic representations of hardware mixing practices—they allow the user to 'nest' a few plug-ins inside one insert on a chain. Here, each mix rack contains 3 more plug-ins.

After this, he applies two consecutive rounds of de-essing using FabFilter *Pro-DS* with different settings, claims that each round "attacks" a different kind of vocal sibilance rooted in different frequency bands of the voice, and which has resulted from the combined application of EQ and compression earlier in the chain. He suggests that the first de-esser is more subtle, employed just to reduce the prominence of "some of the 'esses' and 'ths' and stuff like that, that are peaking", and the second de-esser more targeted, "handling some of those more 'th', 'ess' kind of rolling sounds […] those really kind of pokey, piercing top-end frequencies". Indeed, this technique might be termed 'serial de-essing'—and Figure 71 below shows how each de-esser targets sibilance in a different way: the first encompasses a wider frequency range (around 5.6-13kHz) with a medium-high threshold, and the second seems to be targeting the sibilance in a slightly more intense way, with a narrower frequency range (8.7-13kHz) and a higher threshold. However, this sequential technique is so intricate and so subtle that the effects of its component stages can barely be heard—clearly, it embodies a more gestural way of applying de-essing in stages in order to achieve the desired effect subtly. This precise use of de-essing provides further evidence for the different kinds of voice and technology fetishism which underpin it, which combine ideas about the voice's innate frequency components and the supposed technical power offered by digital novelty plug-ins to adjust them with the notion of

compression as a felt, gestural sonic enhancement to the voice that is applied approximately in iterative stages. Ironically, it is also a response to the heightened ‘esses’ created by the use of serial compression with analogue emulations.



Figure 71: Settings for the two rounds of de-essing applied by Austin.

Following this, he attempts to compensate for the perceived attenuation of high-frequency vocal “air” which happens as a result of de-essing by “adding a little bit of the extra air back in” at the end of the chain. Using a final round of EQ (Slate *Custom Series Equalizer*), Austin adds a very subtle amplification to the highest frequencies in the voice, around 15kHz. Again, the exact effects of this

transformation—whether it has succeeded in compensating for something removed by the de-essing process—are impossible to detect in isolation. This idea that the process of removing high-frequency energy in the vocal that creates sibilance might also create a *further* problem in the voice as it adversely affects desirable high-frequency energy termed "air", shows how the fetishisms in the practice create complex feedback loops in vocal chains, where certain techniques and/or plug-ins create desirable effects but also undesirable side-effects that must be dealt with. This example illustrates a very long process to this effect where: The serial compression technique using analogue emulations, combined with EQ, creates a problem with sibilance which the de-esser solves; but then the de-esser creates another problem (removing desirable high frequency energy), which the final round of EQ is intended to solve.

This example therefore illustrates the highly complex, iterative nature of the process and the way such iterations are driven by the interactions between different kinds of technology fetishism and voice fetishism in the case of EQ, compression and frequency-sensitive compression (which blends the characteristics of both). The producer-songwriter is trying to prioritize what he believes are the most important (yet very complex) characteristics of the voice and to shape the vocal sound by balancing and accounting for all the parameters involved—here, it is an attempt to balance undesirable sibilance and desirable high frequency energy—but the process keeps endlessly throwing up additional problems which must be resolved. The different fetishisms underpinning analogue emulations and frequency-sensitive digital novelties in the practice bounce off one another and create reactive patterns which lead to more compression actions. Compression, at its most basic, is thought to create positive effects on the voice (vocal foregrounding), but it also leads producer-songwriters to notice the undesirable quality of sibilance, which is associated with the voice fetishism driving EQ, and so implies a technical modification using this kind of plug-in rather than a modification using an analogue emulation which is appreciated more for the elusive 'sound' it imparts to the voice. The intensifying nature of this process also works to the benefit of the plug-in industry, who are invested in creating and monetizing as many new plug-ins as possible, and—much like de-essing in the

1970s—producing innovative products by rationalizing specific vocal editing processes into individual plug-ins which then become indispensable parts of what is already a complex process.

Conclusion

In summary, this chapter has demonstrated how vocal compression in commercial pop is driven primarily by the commodity fetishism of digital emulations of legendary analogue compressors, specifically the *1176* and *LA-2A*. The sonic effects of these digital compression plug-ins on voices are construed as elusive and magical—a reputation which has been perpetuated by Universal Audio and Waves. Such commodity fetishism also interacts with the erotic fetishism of the voice in compression, embodied in producer-songwriters' conception of the voice's "dynamics" as particularly special and unpredictable—simultaneously in need of subtly taming using compression and a very important defining quality of the voice which must be preserved through applying compression subtly and gradually. The interaction between this voice and technology fetishism fuels my informants' iterative, multi-stage applications of compression in their vocal chains, including serial compression, parallel compression and frequency-sensitive compression (which additionally interacts with the slightly different kinds of fetishism related to vocal EQ). The last section of this chapter therefore functions as a culminating discussion of the various kinds of fetishism which run throughout chapters 4 and 5, interpreting frequency-sensitive compression as a nexus which brings these fetishisms of the voice and technology together.

Furthermore, my discussion of the fetishism of analogue compressor emulations—in particular, the clever marketing strategies employed by the Waves/CLA partnership—also further implies that plug-in companies have a stake in maintaining these fetishisms in vocal compression, as a way of ensuring that producers-songwriters keep buying their plug-ins. As the iterability and complexity of commercial pop vocal chains intensifies, more plug-ins are likely to be designed to address the demands for specific 'problem-solving' functions, which will further intensify the complex relationships between EQ, compression and frequency-sensitive compression and the fetishisms which define them. This is a tentative prediction for the future, but I also think its likeliness

is further evidenced by the fact that plug-ins which attempt to condense the functionality of the entire vocal editing process into one—like iZotope's *Nectar*—tend not to be popular in this production culture and are considered amateur. Building on Chapter 4, there are also further 'surgical' analogies to be made here, especially related to cosmetic surgery and the beauty industry, both of which rely on making consumers feel like complex and iterative routines involving multiple products are necessary to reach desired aesthetic standards, however more research would be needed to fully flesh out these comparisons. This chapter therefore further supports my argument that plug-in companies increasingly both respond to and perpetuate the fetishisms underlying vocal editing.

My discussions of the tracks by Oliver and Austin also show how multi-staged compression processes, in their totality, have an identifiable, audible effect on the sound of commercial pop voices. Even though compression sometimes seems shrouded in mystery, and often difficult to perceive in the context of the multiple layers of the vocal editing process, this chapter therefore also addresses the first main aim of my thesis by showing how compression alongside EQ, tuning and comping contributes to the sound of commercial pop vocals.

Finally, the 'macho' language which surrounds vocal compression—the extremes of which are perhaps represented by 'celebrity engineer' Chris Lord-Alge—also suggests that there are problematic gender ideologies at play in the discourse of compression and vocal editing more broadly, and that the commodity fetishism of compression plug-ins is intertwined with such ideologies. The notions of the voice's "dynamics" as being somehow both special and irritatingly unpredictable which underpin the fetishism of the voice in this context also recall gendered tropes of control and discipline, which are unsettling in the context of the broader gendering of roles across production and performance in this industry. Chapters 3 and 4 discussed such gendered tropes in the slightly different contexts of the visual objectification of the voice and metaphors of vocal surgery, and the conclusion which follows will attempt to tie together some of these threads relating to the theme of gender.

Conclusion

This thesis has shown how the culture of digital vocal editing among Western commercial pop producer-songwriters is driven by the interactions between fetishisms of both the voice and digital plug-in technologies. These fetishisms are additionally motivated by the activities of an expanding digital plug-in industry, and, cumulatively, the use of the technologies produced by such an industry informs the sound of commercial pop voices. These concluding pages draw the nuances of these themes together. First, I provide a summary of the ways these different fetishisms played out in the case of each of the thesis chapters, drawing out in the process some broader ideas related to the significance of such fetishisms. Secondly, I summarize how these digital editing techniques combine to create the commercial pop vocal sound. Following this, I also tie together the subtheme of gender, considering how the different interactions between voice and technology fetishism fuel the different ways gendered hierarchies covertly inform the uses of the digital technologies I describe. Lastly, I consider more fully the role of the plug-in industry in this landscape, claiming that it is a key agent which brings these multiple kinds of voice and technology fetishism into articulation, intensifying each by the other, and solidifying the vocal aesthetics of the genre in the process. After summarizing how this entanglement is structured, I provide some brief examples of recent plug-in releases and the response by producer-songwriters to their marketing to consider the future trajectories of this scenario.

Voice and technology fetishism combined

My contextualization of the production and songwriting industries in the introduction considered the increasing centrality of the voice and digital vocal editing in this economic and social organization, providing important background for my use of the concept of fetishism which emerged from my fieldwork. Subsequently, Chapter 1 outlined the voice and technology fetishisms which drive digital vocal editing, defining them both theoretically and ethnographically. The multiple theories of fetishism I delineated from anthropological literature through Marx and psychoanalytic theory through Freud provided two differing yet complementary uses of the term on which to build the

thesis' overarching theoretical frame. Commodity fetishism (extending to immaterial commodities) and erotic fetishism have similar conceptual foundations relating to the fetishized object concealing or standing in for something else: perceived as magical, indefinable, or indicating a larger, imagined, whole. More specifically: the concept of commodity fetishism embodies the form of a commodity standing in for or concealing the social relations of its production; erotic fetishism encompasses notions of bodily fragmentation, fixation and control; and acts of objectification are key to both.

Following this, I contextualized the two main kinds of voice fetishism pertaining to digital vocal editing, which map onto these two theoretical dimensions. The first is like erotic fetishism, and assumes that there are immutable links between the sound of the voice and a whole human identity or body from which it is fragmented, as well as entailing a belief that the most aesthetically successful singing voices contain a special, elusive quality (analogous to vocal timbre) which cannot accurately be described, and which can be linked directly to this imagined identity or body. The second is like commodity fetishism, as the songwriting and production industry's widespread trading in "topline" vocals encourages the voice to be appreciated solely as a sounding form divorced from the realities of labour performed by a vocalist. Notably, in my fieldwork, these two kinds of voice fetishism seemed to coexist, attesting to the power of the former despite the increasing presence of the latter.

Slightly differently, technology fetishism in digital vocal editing is a more explicit form of commodity fetishism related to producer-songwriters' proliferating acquisition of plug-ins. I outlined how this commodity fetishism differs across two main types of digital plug-in. Firstly, 'analogue emulations' are believed to lend desirable yet elusive sonic effects to the voice, a fetishism which is additionally promulgated by the way emulations of analogue hardware interfaces reproduce the same concealment and corresponding mystification of the inner workings of such technologies in the digital domain. Secondly, 'digital novelties' are believed to give the user significant technical power to manipulate the voice in ostensibly objective or even scientific ways, a commodity fetishism which is reinforced by the slick and minimal design of their interfaces which make the actual algorithmic construction of such plug-ins unimportant to users as well as creating a link between vocal editing and the contemporary technology industries more broadly. Voice and technology fetishism therefore each

have two main sub-fetishisms, and this theoretical structure drawn out from my fieldwork informed my discussion of the interrelationship between the use of digital vocal editing technologies and the vocal aesthetics of the commercial pop genre throughout the thesis.

Following this contextualization, Chapter 2 addressed the most fundamental aesthetic aim of vocal editing, as it was presented in my informants' aesthetic discourse. For the sake of simplicity, I called this 'vocal foregrounding', although my informants might call it the desire for "up-front and centre" vocals. More importantly, I claimed that such an aesthetic aim is fuelled specifically by erotic voice fetishism. The aim is to foreground the voice within the spatial environment of the "mix" to extreme proportions, in the name of reducing an imagined distance between the vocalist and the listener, and thus creating a sense of intimacy experienced by the listener. Like the erotic kind of voice fetishism, the voice is assumed to link to a whole human identity and body, and this assumption drives its connotation of physical intimacy. It also drives allusions to realism in the practice which figure the sonic exaggeration of intimacy in terms of how the voice is supposedly perceived in real-world situations of physical proximity. These ideas are naturalized in the aesthetic discourse of vocal editing, represented by terms such as "up-front", "close", and "present". Further, two more commonplace terms "tone" and "dynamics" have specific meanings in this aesthetic discourse, and their use applied to vocals reflects and reinforces this voice fetishism by encapsulating an idea of the voice being a more special, complex and unpredictable sonic entity in comparison to other mix elements.

Significantly, this aesthetic aim seemed to drive the practice of vocal editing first and foremost in my present-day fieldwork, and to precede any technology fetishism which interacted with it or intensified it. The consolidation of such an aesthetic aim may link in part back to the history of vocal recording and amplification technologies throughout the 20th century and the way they also engaged a combined fetishism of both the voice and technology. As I discussed briefly in Chapter 1, the initial reception of early sound reproduction technologies was loaded with emotive significance due to the technologies' perceived capacity to preserve the voices of the dead (Sterne 2003: 292; Stanyek & Piekut 2010: 16). The strength of this idea depends on the automatic—and arguably

culturally reinforced—link that is made between the voice and a human identity and body, as a part which implies a whole. Similarly, the favourable, erotically charged reception of crooning in the early-to-mid 20th century was energized by developing microphone technologies, and the ability they gave crooners to amplify soft singing (Frith 1996; Lockheart 2003). Like vocal foregrounding, such a vocal aesthetic depends on the connection that is made between a particular kind of vocal sound and an imagined singing body which resides in close proximity to a listener. In both cases, an existing fetishism of the voice is intensified by the development of new sound technologies. Digital vocal editing in commercial pop therefore represents a new era in which such interacting voice and technology fetishisms are exaggerated and deepened in complexity, in a way which I would suggest has audible effects on the voices heard in contemporary commercial pop tracks.

Moving on, Chapter 3 covered the ways tuning and comping technologies in vocal editing objectify the voice in visual form, fuelling the fetishism of the voice as a commodity divorced from the labour of vocal performance. Building on the theoretical discussions in Chapter 1, this chapter holds that objectification is a key process underpinning commodity fetishism, particularly in the case of immaterial commodities like the voice, the form of which is ephemeral and intangible. Tuning and comping technologies make the voice material, turning it into visualized waveforms and pitch-contours which can be sculpted at will by producer-songwriters. These visual objectifications of the voice carry over into the way producer-songwriters describe the voice and the performance of vocalists, using terms like “sliding” and “straightness”. Regarding technology fetishism, tuning plug-ins, specifically, are also ‘digital novelties’ which give the user a sense of great technical control over the voice through the ability to directly modify and re-construct visual representations of vocal pitch and timbre. Combined with the resulting capacities for technical control that they supposedly offer, such technologies and the ways they visually objectify the voice further inflect the often gendered collaborations between producer-songwriters and vocalists, intensifying the tendency for producer-songwriters to disregard or question the labour of vocal performance. Producer-songwriters’ responses to the vocalist’s necessary involvement in comping and tuning processes—perceived largely as an inconvenience—can also be interpreted through the lens of erotic voice fetishism, as the

voice is resolutely kept fragmented from its body and subject to an almost possessive kind of fixation and control. More generally, tuning and comping technologies therefore play a part in driving the broader commodity fetishism of topline vocals across the industry, in which the voice is valued and traded between parties solely as a sounding form divorced from the labour performed by a vocalist. This chapter should not be read in isolation, however, as this thesis has shown how such commodity fetishism of the voice sits within a broader vocal editing process in which the other kind of erotic voice fetishism is much more prominent, and in which the role of visual objectifications of the voice in driving such fetishism is different or recedes into the background altogether.

In Chapter 4, the erotic kind of voice fetishism comes to the fore, building on the theme of realism which was introduced in Chapter 2. Chapter 2 showed how the aesthetic discourse of vocal editing tends to describe vocal foregrounding in terms of the emulation of the sonic experience of "real-world" proximity, however EQ discourse makes these allusions to realism more extreme and totalizing, suggesting that the technique ought to make the voice sound "real, like a person" (2020Z). Chapter 4 analysed this unique discourse of vocal realism, linking it to the way the technique gives users the sense that they are directly controlling vocal timbre. As I explained in Chapter 1, a key feature of erotic voice fetishism is the belief that vocal timbre—or a similarly elusive notion of sonic quality—can be directly linked to or is indicative of the physicality and identity of a vocalist. This is a core concept which elevates the power of vocal timbre to define voices, and which, in the case of EQ, leads producer-songwriters to feel that they are manipulating the inner qualities of the voice itself (in a way that the other technologies do not provide to the same degree). Digital novelty 'para-graphic' EQ plug-ins are indispensable to this practice, and endow the user with a sense of specialized power to manipulate the voice represented purely as a graphical curve which can be split up into generalized frequency areas corresponding to particular vocal characteristics. In combination, these manifestations of voice and technology fetishism produce the widespread discourse of vocal realism surrounding EQ, in which the technique is spoken about as if it simply preserves and enhances what's already there in the voice, rather than contributing specifically to the aesthetic of vocal foregrounding by exaggerating high-frequency energy in the voice and carving out an isolated space for it in the mix more broadly.

Assumptions about vocal realism are repeatedly conflated in discourse about EQ with the exaggeration of vocal foregrounding. Indeed, this conflation can be interpreted more generally as a result of the naturalization of the aesthetics of vocal foregrounding: since this aesthetic trend is so embedded in commercial pop (and links back to a longer history of sound recording founded on the interaction between voice and technology fetishism), in EQ this specific Western pop aesthetic has become so normalized among practitioners that they think that the aesthetic use of EQ simply preserves and enhances the "real" voice.

Realism is therefore a pervasive theme across the discourse and practice of commercial pop vocal editing more broadly. The analyses in Chapters 2 and 4 point towards some of the reasons for the strength of realist discourses in this musical culture, as well as suggesting that such discourses have subtle effects on both the aesthetics of the genre and the social relations involved in the labour process of commercial pop production. As I have indicated, the prevalence of realist discourse in EQ may have resulted from a broader naturalization of the aesthetics of vocal foregrounding, implying that as realist conceptions of the technique become even more solidified in the practice, the aesthetic effects of EQ may become more homogenous or extreme across the commercial pop genre. Realist conceptualizations of EQ also obscure tacitly gendered assumptions about deficiencies in vocal timbre, couching them as vocal 'problems' in need of neutral, "surgical", diagnosis and repair. I would suggest that the pervasiveness of realism in vocal editing therefore covertly contributes to the maintenance of the gendered imbalance of agency in vocal editing more broadly. The predominance of realist discourse among my younger key informants—in comparison to the older generation of producers and engineers I spent time with at the start of my fieldwork—also suggests that the potency of realist discourse in vocal editing has been intensifying through the turn of the 21st century, and is particularly strong in the case of commercial pop. It may be the case that, since this younger generation of producer-songwriters have grown up hearing and editing voices that have already been subject to substantial technological mediation (e.g. through processing enacted by increasingly sophisticated microphone technology or the application of automated digital signal processing in recording sessions), they are more likely to assume that these vocal sounds are closer to what "real"

singing voices sound like. Regardless, this generational difference implies that the preservation and enhancement of vocal realism may become an even more routine mode of conceptualizing vocal editing among commercial pop producer-songwriters over the coming years.

In Chapter 5 I discussed how commercial pop vocal compression techniques are driven primarily by the commodity fetishism of digital emulations of legendary analogue vocal compressors (the *1176* and *LA-2A*), reproducing the existing commodity fetishism of such analogue units across Western music production in the digital domain. These digital commodities are fetishized for the elusive sonic effects they supposedly give to the voice, and this also links to the way compression more broadly is mystified among production practitioners as an inherently complex and elusive technique. Moreover, such fetishism of digital compression technologies interacts with the existing voice fetishism related to vocal dynamics, which conceives of the voice as a similarly elusive and unpredictable sonic entity: producer-songwriters believe that these characteristics of the voice need to be both preserved and controlled. The subtleties of these two fetishisms lead producer-songwriters to try and make their compression techniques serve multiple aims: they must control the dynamics of the voice in a way which serves the aesthetics of vocal foregrounding, but the true dynamics of the voice must also be preserved. These aims result in compulsive, iterative compression techniques such as serial compression and parallel compression, in which multiple rounds of compression are applied in stages because they are felt to be effective ways of balancing all these aims.

Remarkably, then, comparing the completely different ideas underpinning the practices of tuning/comping and EQ shows how different fetishisms of both the voice and technology seem to exist side by side within the practice of vocal editing, however there is an even more explicit interaction between these different fetishisms in relation to EQ and compression. Namely, my discussion of frequency-sensitive compression at the end of Chapter 5 showed how, in dialogue with iterative compression techniques, producer-songwriters maintain the same concerns about the "real" qualities of vocal timbre which motivate EQ, and this results in a constant switching of modes in their practice between the priorities of compression (ideas about the general sense of dynamic consistency it lends to the voice through the iterative application of analogue emulations), and the priorities of EQ

(the exaggeration and attenuation of particular vocal qualities which correspond to clearly-defined points on a frequency spectrum). The application of one technique consistently creates problems in the voice which need to be fixed through the application of the other, leading to iterative chains of not only compression but also frequency-sensitive compression and EQ. This phenomenon shows very clearly how the practice of digital vocal editing in commercial pop is driven by the dialogue between different kinds of voice and technology fetishism which are grouped in relation to each technology involved in the process. Moreover, since these alternating modifications often seem to cancel out or change earlier stages within the chain of EQ and compression—changes which seem to have been very carefully applied—this aspect of digital vocal editing acquires a compulsive, ritualistic quality, evoking the foundations of the concept of fetishism I described at the outset of the thesis.

The collective result of these interacting fetishisms within the practice of digital vocal editing is likely to be an over-arching fetishism of the commercial pop singing voice experienced by the listener, who may not be aware of the vast quantity of editing labour that has produced it. The process of digital vocal editing by a producer-songwriter involves the fragmentation of a voice into different sonic parameters and the manipulation of these parameters according to fetishisms of the voice and digital editing technologies. What's more, this vocal editing stage is nested within a much larger songwriting and production process, in which a voice may be recorded, re-recorded, edited and re-edited multiple times by different professionals, or the singer even changed entirely and the voice then subject to the same processes all over again. Given that the sonic results of vocal editing are so subtle, such processes are largely imperceptible to the listener untrained in music production, and the voice therefore seems to result effortlessly from the performance of the vocalist themselves. Slightly different from Marxist commodity fetishism, then, the fetishized commercial pop voice does not completely obscure the realities of vocal labour but rather creates an illusion that such labour is rooted in a single source—the vocalist—rather than the cumulative product of both vocal performance alongside the complex technological processes and agents involved in vocal editing. Notably, this explains why so much academic and journalistic attention has been devoted to interrogating vocal pitch-correction processes: since commercial pop vocals are conceived as directly linked to the

performance of vocalists themselves, the act of pitch-correction is seen as a cop-out or crutch which facilitates such vocal labour more easily, rather than the product of a very specific cultural and aesthetic moment driven arguably more by the developing aesthetic aims of music production rather than the performative labour of vocalists. Pointedly, a Google search conducted in April 2022 for 'with and without vocal editing' returns an abundance of video results instead showing vocals 'with and without autotune', rather than addressing the bigger picture I have painted here.[104] In contrast, this thesis has exposed how the sound of commercial pop vocals is produced cumulatively through a suite of multiple digital vocal editing technologies—a process of production driven by producer-songwriters and bolstered by the interactions between voice and technology fetishism.

The commercial pop vocal aesthetic

The result of these cumulative processes is a specific vocal sound, which I will now summarize in light of my analyses. First and foremost, at the end of Chapter 2 (pp. 139-140) I considered how the overall aesthetic aim of vocal foregrounding informs the sound of the voice (lead vocals, specifically) in the commercial pop genre very broadly. It drives its spatial centralization within the stereo field, its relative loudness across a mix, and its timbral and textural distinctness relative to other musical elements. These features remain consistent throughout a track. Most of the modifications performed by the constituent vocal editing technologies can be understood in terms of this overarching aesthetic ideal. For instance, comping processes (and related processes e.g. timing, manual gain automation) contribute to this aim by selectively amplifying and quietening individual consonants and vowels within words to make the voice's loudness more consistent, and by building up the voice's dominance in a track by layering vocal doubles and harmonies in a vocal arrangement. Moreover, alongside the correction of pitch, tuning is used as a vocal "tone"-shaping device which gives the voice a more stable sense of smoothness, for example by emphasizing the "sliding" characteristics of vocal pitch contours. Timing and other audio restoration techniques are also used to flatten out any imperfections

[104] https://www.google.com/search?q=with+and+without+vocal+editing&rlz=1C1CHBD_en-GBGB915GB915&oq=with+and+without+vocal+editing&aqs=chrome..69i57j0i546l5.2554j0j7&sourceid=chrome&ie=UTF-8 [Accessed: 22 April 2022].

in the way the voice is rhythmically aligned with the other components of a track, and to ensure that its timbre and diction are as smooth as possible. EQ directly contributes to the impression of vocal foregrounding by subtly emphasizing high-frequency energy in the voice to enhance the sense of proximity experienced by the listener, as well as by confining the voice to a specific frequency range to carve out an isolated and distinct space for it within the mix. Compression regulates the voice's overall loudness to make its "dynamics" more consistent, and therefore to increase the sense that it is the most central and dominant element of a track.

Without fully understanding the component processes of vocal editing, this sound is very difficult to describe, however certain ideas and key terms can gesture towards it. A brief account of my listening to Audio 1 with headphones provides a good example for such description.[105] Very generally, the voice is the leading musical element throughout the entirety of the track. Spatially, it occupies a clearly defined point as if right in the centre of the listener's head, at the middle point between two headphones, and it stays in this position throughout. In terms of volume, it is the loudest of all the component parts of the track, and the vocal line itself maintains a relatively consistent loudness throughout. This is the case between the different sections of the song but also across each word that is sung within individual vocal phrases. For example, the first line of the first verse (0'07"), starting 'your hands' sounds as though the vocalist is singing with little energy, even reticence, but it is the same volume as the chorus, which appears to be substantially higher in performative energy. Also within the second line of the first verse (0'14"), there is little contrast in loudness between the phrases 'your hands' and 'weakness', even though the voice appears to increase in energy on the 'ness' of weakness especially. The diction is also consistent throughout the song: each consonant of each word can be heard very clearly. In terms of pitch, the overall impression is that of the singer remaining consistently in-tune (relative to equal temperament) throughout the performance, however the pitch contours of the vocal melody are mostly continuous rather than discrete: by this I mean that there are core notes to the vocal melody but the voice curves and undulates around them, frequently

[105] The vocals for which, to remind the reader, were edited first by my key informant Jason and later by another producer before the track was released. See p. 41.

gliding between notes rather than jumping distinctly between them, or scooping up to notes rather than hitting them exactly. For example, the first line of the chorus 'I promise I'll give it, I'll give you all I got' (0'33"-0'46") traverses four core notes (G4, A4, B4, C5), however the pitch-contour of each word seems to pre-empt the next, containing a continuous glide towards the next note in the melody, which is fully attained by the time the voice reaches the next word. Timbrally, the voice has a brightness and intensity to it that at no point is obscured by the other elements in the track, which seem to support and surround it rather than compete with it. It also has a breathy quality to it, almost whispering, particularly in the verses: especially when listening with headphones, this creates an impression that the vocalist is singing right up close to my ears. Not all commercial pop tracks exaggerate this breathy quality—indeed, in this track, such a quality is less prominent in the song's choruses—however it can be read as an extreme manifestation of the vocal foregrounding which I have been arguing is a defining characteristic of this genre. Broadly, the sound of the voice on this track is emblematic of the commercial pop vocal aesthetic.

Gender and voice and technology fetishism

A running subsidiary theme throughout this thesis was the presence of covert gender ideologies in vocal editing processes, and the way such ideas about gender interlinked with voice and technology fetishisms. Within the context of a widespread gendering of roles across the industry between largely female vocalists and largely male producer-songwriters, several key themes came to the fore, including: the devaluation of a vocalist's performative labour, the female voice's greater obligation to satisfy ideas about vocal realism, and notions of controlling or disciplining the volatile voice through vocal editing. These ideas should also be situated within the more general erotic voice fetishism which, in Chapter 2, I argued was essential to the aesthetics of vocal foregrounding. Not insignificantly, the popular understanding of erotic fetishism tends to be gendered, as does Freud's conception of the term, involving a male fetishist who fetishizes a female trait or body part, additionally evoking gendered notions of objectification, fixation and control. The later chapters of the thesis touched on these themes in different ways.

Chapter 3, for example, addressed how the visual objectification of the voice in tuning and comping contributed to producer-songwriters' tendencies to disregard or override the vocalist's performance decisions, prioritizing their own aesthetic preferences in vocal editing instead, and sometimes even making edits that they thought were superfluous in an attempt to pacify the vocalist whose sincere creative feedback they felt was incorrect or unimportant. In all the situations I analysed in this chapter, the vocalists were female and the producer-songwriters male. There was clearly quite a definite hierarchy regarding aesthetic decision-making across this broader, fragmented creative process, a hierarchy which can be interpreted as a gendered imbalance of power given the wider gendered context of the industry. If, as I suggested earlier, listeners may tend to conflate the commercial pop vocal sound with the labour of vocal performance (i.e. mostly female labour) more so than vocal editing, the analyses in this chapter therefore characterize vocal editing as a tense process in which the vocalist wishes to take back control over the sound of their voice with this ultimate listener judgement in mind. Yet this remains difficult for vocalists to achieve given the producer-songwriters possession of the technical power to edit the voice and their at times deliberate policing of the boundaries of this power.

Chapter 4 showed how, within the context of EQ's realist discourse, negative assumptions about the term "harshness" were disproportionately applied to female vocals. I also explained how producer-songwriters' discussions of EQ were saturated by metaphors of surgery, implying that they had a significant amount of power to diagnose vocal complications and intervene accordingly using EQ. These findings show how the erotic voice fetishism underpinning ideas about vocal realism in EQ—involving the assumption that voices link inextricably to bodies, and that such voices can be fragmented into universal sonic characteristics—fuels notions of technical, medicalized control over the female voice. Not unlike cosmetic surgery, which relies on the fragmentation of the female body into idealized parts and the subjection of such a body to (usually male) clinical intervention according to certain naturalized Western beauty ideals (see Basalmo 1992), the diagnosis and repair of vocal "problems" using EQ seemed to be exaggerated in the case of female voices, and to be undertaken according to universalizing ideas about where such problems ("harshness") were conventionally

located in the female voice. As I have suggested, the discourse of realism in EQ works to sustain and arguably reinforce these tacitly gendered assumptions about vocal timbre.

Lastly, Chapter 5 painted a picture of vocal compression as an act of disciplining the dynamic volatility of the voice, even though there was also an attempt to simultaneously retain some of the voice's dynamic qualities which were thought to define the voice more fundamentally. Significantly, within this broad act of vocal disciplining, a violent, macho language was used to describe the application of compression plug-ins to the voice, including terms like "slamming", "spanking", "pushing" and references to aggression, especially in the case of Waves' partnership with Chris Lord-Alge. Such language is part of a broader masculine coding of the commodity fetishism of music technologies (See Annets 2015; Born & Devine 2015). In the gendered context of this industry, compression could therefore be read as a forceful domestication of the female voice.

Taken altogether, such themes show how longstanding gender hierarchies are furtively embedded in vocal editing processes, which likely maintain the control and devaluation of female creative labour in the production of commercial popular music more broadly. Such themes have also been elaborated by Provenzano who argues that vocal tuning increasingly functions as a mechanism of control over female voices in pop production scenarios, involving the anxious pursuit of what the producer imagines a voice 'should' sound like as well as, frequently, the non-consensual tuning of a vocalist in her absence (2019a: 296-316). This thesis has extended and deepened such claims about the gendered dynamics of commercial pop vocal recording and production by showing how similar ideas about gender influence all the cumulative digital processes that contribute to the commercial pop vocal aesthetic.

The plug-in industry drives voice and technology fetishism

Finally, this thesis has illustrated how the activities of the transatlantic plug-in industry both exploit and intensify the fetishisms and gendered ideologies of commercial pop vocal editing. Vocal editing's embeddedness within a wider labour process of songwriting and production where the voice occupies a central position of value drives the proliferation of vocal editing technologies, and this provides a

constantly evolving market which the plug-in industry intentionally exploits. It does this firstly by targeting the online spheres in which producer-songwriters acquire and share information about vocal editing, as well as by partnering with revered figures in production who advertise their plug-ins. More specifically, as shown most clearly in Chapter 4, it continually re-produces and extends the market for its products by developing new plug-ins which claim to respond to producer-songwriters' needs for more specific or isolated functionalities in vocal editing, and to speed up or refine such functions (e.g. *Soothe, Gullfoss)*. Even further, this thesis suggests that the specific details of the voice which such plug-ins target are borne out of the intensifying synergies between voice and technology fetishism, and drive a focus on, among other things: the detailed shapes of vocal pitch-contours; the intricate component parts of vocal frequency spectra; the complexity and unpredictability of vocal dynamics; the sonic character provided by special compressors; and the complex mutual interactions between EQ and compression throughout vocal chains. Across the thesis chapters I have addressed several key companies involved in such a pursuit—Antares, Celemony, Waves Audio, Universal Audio, FabFilter, oeksound and SoundTheory—all of whom have a role to play in furthering the complexity of vocal editing practices in this way for their own gain.

Although in 2022 there is a similarly constant proliferation of new plug-ins for tuning and vocal noise reduction (e.g. Antares' *AutoTune Unlimited* (2021) and Waves' *Clarity Vx* (2022)[106]), it is not coincidental that EQ is one of the main techniques where this intensified activity of the plug-in industry can be observed most clearly, whereas compression, in contrast, remained firmly centralized around the *1176* and *LA-2A* emulations. My discussions of the different kinds of fetishism relating to such techniques clarify this difference: the voice fetishism underpinning EQ and its intense focus on the "real" characteristics of vocal timbre parallels an increasing desire among producer-songwriters to cultivate these "real" qualities of voices more sensitively and scientifically. Plug-ins such as *Soothe* and *Gullfoss* both respond to and solidify such a desire, claiming to perform automatic, intuitive "smoothing" functions on the voice that are based on research into human perception. More broadly,

[106] Links: https://www.antarestech.com/product/auto-tune-unlimited/, https://www.waves.com/plugins/clarity-vx#clarity-vx-vocal-noise-reduction

such ideas resonate with the contemporary language of artificial intelligence (AI) in the technology industries, even though the marketing of such plug-ins does not clarify the extent to which they rely on machine learning, lending greater appeal to these digital novelties. In contrast, compression techniques are more firmly situated within a longer historical commodity fetishism of analogue compressor units which are perceived to be particularly effective on vocals. The plug-in industry is therefore likely to invest less resources in the development of new compression plug-ins given Waves and Universal Audio's monopoly over the analogue emulations, which producer-songwriters swear by.

In the case of EQ, new players in the plug-in market attempt to one-up competitors like *Soothe* by making vocal smoothing functions even more specific, as a way of generating new unique selling points for their products. For example, a 2021 plug-in release from company Mastering the Mix, *RESO* (Figure 72), claimed to exceed the functionalities provided by plug-ins like *Soothe* by not just 'identifying and resolving resonances' but 'dynamically fixing the problematic resonances without making your music sound thin' (Mastering The Mix 2021). The plug-in purports to automatically zone in on and attenuate problematic resonances in a frequency spectrum, like *Soothe*, but it incorporates further "dynamic" functionality into this automation, meaning that such "resonances" are identified individually and only attenuated when they reach a certain threshold. In an Instagram post publicising the software, a representative from Mastering the Mix directly responded with this statement to several comments from interested customers asking specifically how their new plug-in was different to the already popular *Soothe2* (Figure 73).



Figure 72: *RESO* plug-in interface. Note the similarity to the interface of *Soothe2* (shown in Figure 57, p. 236). (Mastering The Mix 2021).

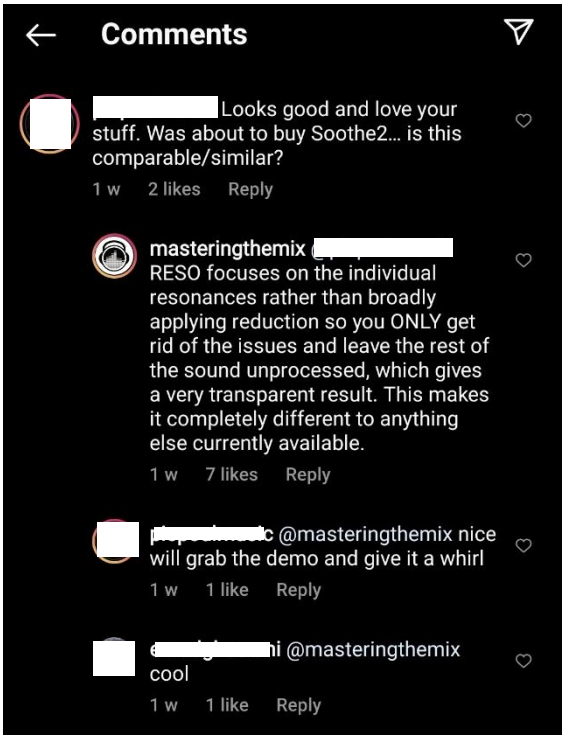


Figure 73: Comments on an Instagram post by @masteringthemix, November 2021. (https://www.instagram.com/p/CVxj1_8Nd9E/, accessed 23 April 2022).

Regardless of the actual algorithmic make-up of this plug-in, it shows how the level of automated technical control over the vocal frequency spectrum which such plug-ins claim to offer the user have reached an extreme level of intricacy and complexity—or claim to have done so, in the interest of acquiring new customers. Just as the unique selling point of FabFilter's *Pro-Q 3* was the ability it gave the user to add such "dynamic" functionality manually to individual EQ nodes, *RESO* attempts to surpass *Soothe* in its provision of the "dynamic" functionality in an automated context. Like *Pro-Q 3*, this plug-in therefore taps into ideas about the innate complexity of and fluctuations within vocal frequency spectra which underpin EQ, and which I have argued are constituted by voice fetishism, but it also homogenizes the aesthetic outcomes of such ideals through its automation of EQ processes. Although the success of this plug-in is yet to be determined, its aims and functions very clearly illustrate the role played by the plug-in industry in solidifying the aesthetics of commercial pop vocals through exploiting the foundations of voice and technology fetishism among producer-songwriters.

Moreover, in terms of innovations in the contemporary plug-in industry, it was these digital novelty EQ plug-ins which offered 'intuitive' ways of modifying very precise components of the voice that my key informants seemed most excited about, in contrast to another type of digital novelty plug-in which was also available. Represented by plug-ins like *Nectar* developed by iZotope (which I discussed briefly in Chapter 1, on pp. 80-81), as well as Waves' *CLA Vocals* (discussed in Chapter 5, on pp. 263-265), these plug-ins attempt to streamline the multiple stages of vocal editing into one convenient unit, including sections for tuning, EQ, and compression alongside other modules such as reverb and panning. Both kinds of digital novelty plug-in prompted critical reflections whenever they were raised among my field informants, however: as I mentioned previously, none of my informants incorporated *Nectar* into their vocal editing workflows, and one producer I interviewed described how he thought that the plug-in was aimed at "bedroom producers" (2019A) who were yet to learn music production fully and simply wanted to achieve passable pop vocals as fast as possible. Similarly, although some of my informants used *CLA Vocals* early on in their vocal editing processes as a way of creating a rough approximation of what the vocal would eventually sound like in the mix, those

with whom I did discuss the plug-in seemed to think of it as quite cheap and amateur, lacking any utility beyond this early stage "shortcut" process (2019CC, 2021D).

Although these 'all-rounder' plug-ins appear to fulfil a convenient market niche (iZotope, in particular, have recently received significant investment from a private equity firm who believe in their potential to 'democratize' musical production and to provide 'seamless' music making experiences (Terren 2021)), this thesis suggests that such plug-ins are unlikely to be as successful in commercial pop vocal editing as the 'intuitive', specific EQs, because they fail to sufficiently target the deeply-embedded fetishisms of the voice and technology which drive the practice, and which fuel producer-songwriters' use of complex chains of plug-ins in iteration including both analogue emulations and digital novelties. These commercial pop vocal editing professionals represent the higher end of the plug-in market, and as a result, plug-ins which appear to streamline the entire process of vocal editing do not appeal to them because not only do they require less skill to operate but they also seem to neglect core values or points of focus which producer-songwriters feel are vital to professional, high-end vocal editing. Neither *Nectar* nor *CLA Vocals* incorporate the complex visual objectifications of the voice which are essential to manual tuning processes, the technical control over representations of vocal timbre which drives EQ, or the unique sonic effects of revered analogue compressor emulations. Consequently, such plug-ins make things seem too easy for these producer-songwriters, who prefer to mature their vocal editing practices by increasing their complexity in a way which accounts for all these details. Another broadly fetishistic quality of the practice which these 'all-rounder' plug-ins fail to satisfy is simply commercial pop producer-songwriters' desire to modify and control increasingly specific components of the voice with ever greater detail and refinement, fragmenting of the voice itself into a whole host of isolated parts which themselves become fixations (e.g. "harshness"). Plug-ins like *Soothe, Gullfoss,* and *RESO,* in contrast, are more appealing because they only streamline a very precise part of solely vocal EQing, which relates more fully to the ways producer-songwriters conceptualize the technique as a way of preserving vocal realism.

It remains to be seen how the plug-in industry's driving of voice and technology fetishism in vocal editing will impact the aesthetics of commercial pop vocals over the coming years. What I have attempted to show with this thesis, however, is that this sound is founded on a synergy between two kinds of fetishism which have significant historical precedents in Western popular music cultures, and that a burgeoning sector of the music technology industries is able to engage such fetishisms and has a vested interest in promulgating them. Such fetishisms drive the usage of comping, tuning, EQ and compression in vocal editing, which cumulatively blend together to create the commercial pop vocal sound.

Reference List

Adorno, T. W. (1938). [2002]. 'On the Fetish-Character in Music and the Regression of Listening' In Leppert, R. [ed.] *Adorno. Essays on Music*. 288-312.

Akrich, M. (1992). 'The De-Scription of Technical Objects.' In Bijker, W. and Law, J. [eds.] *Shaping Technology/Building Society*. 1st edition. Cambridge, MA: The MIT Press. 205-224.

Alex Rome. (2021). 'The Secret to PRO Vocals (Layering)'. [online video]. Uploaded 18 Nov 2021. Available at: <https://www.youtube.com/watch?v=qhaWVi9tjgI&t=1s>. [Accessed 25 May 2022].

American Psychiatric Association. [ed.] (2013). 'Fetishistic Disorder, 302.81 (F65.0)'. *Diagnostic and Statistical Manual of Mental Disorders*. 5th Edition. American Psychiatric Publishing. p. 700.

Annets, A. (2015). 'Masculinity and Gear Fetishism in Audio Technology Community Discourse.' PhD Thesis. Cambridge, UK: Anglia Ruskin University. August 2015.

Antares Audio Technologies. (2018). 'Auto-Tune Pro® User Guide'. [product documentation]. Available online at: <https://www.antarestech.com/mediafiles/documentation_records/Auto-Tune_Pro_Manual.pdf>. [Accessed: 22 Jan 2021].

Appadurai, A. (1986). *The Social Life of Things: Commodities in Cultural Perspective.* Cambridge & New York: Cambridge University Press.

Apter, E. (1991). 'Fetishism in Theory: Marx, Freud, Baudrillard'. in Apter, E. *Feminizing the Fetish: Psychoanalysis and Narrative Obsession in Turn-of-the-Century France.* 1-14.

Atkinson, P. (1997). *The Clinical Experience: The construction and reconstruction of medical reality.* Aldershot & Vermont: Ashgate.

Auner, J. (2000). ''Sing it for Me': Posthuman Ventriloquism in recent Popular music'. *Journal of the Royal Musical Association.* 128 (1): 98-122.

Auner, J. (2000). 'Making Old Machines Speak: Images of Technology in Recent Music.' *ECHO: A Music-Centred Journal.* 2 (2). Available at: <http://www.echo.ucla.edu/Volume2-Issue2/auner/auner.pdf>. [Accessed: 8 Apr 2021].

Auto-Tune / Antares Audio Technologies. (2022). 'Made With Approach 1'. [online advertisement video]. Uploaded 16 Feb 2022. Available at: <https://www.youtube.com/watch?v=OhsLhHx5trM>. [Accessed: 25 May 2022].

Avid. (2018). 'Pro Tools® Reference Guide'. Version 2018.12. [product documentation]. Available online at: <http://resources.avid.com/SupportFiles/PT/Pro_Tools_Reference_Guide_2018.12.pdf>. [Accessed: 25 Jan 2021].

Baade, C. L. (2011). *Victory through Harmony: The BBC And Popular Music In World War II.* Oxford & New York: Oxford University Press.

Babich, N. (2020). 'Flat vs. Material vs. Skeuomorphic Design Examples.' [online article]. *Adobe XD Ideas.* 24 Nov 2020. Available at: < https://xd.adobe.com/ideas/principles/web-design/flat-vs-material-skeuomorphic-examples/>. [Accessed: 24 Feb 2022].

Ballou, G. (2008). [1987]. *Handbook for Sound Engineers.* 4th Edition. Oxford & Burlington, MA: Focal Press/Elsevier.

Barthes, R. (1972). [1977]. 'The Grain of the Voice.' In *Image, Music, Text*. Translated by Steven Heath. London: Fontana Press. 179-189.

Basalmo, A. (1992). 'On the Cutting Edge: Cosmetic Surgery and the Technological Production of the Gendered Body'. *Camera Obscura.* 10 (1 (28)): 206–237.

Bates, E. (2010). 'Mixing for Parlak and Bowing for a Buyuk Ses: The Aesthetics of Arranged Traditional Music in Turkey'. *Ethnomusicology*. 54 (1): 81-105.

Bates, E. (2013). 'Popular Music Studies and the Problems of Sound, Society and Method.' Journal of the IASPM. 3 (2): 15-32.

Bates, E. (2016). *Digital Tradition: Arrangement and Labor in Istanbul's Recording Studio Culture*. Oxford & New York: Oxford University Press.

Baugh, J. (2003). Linguistic Profiling, in Makoni, S. et al. [eds.] *Black Linguistics: Language, Society, and Politics in Africa and the Americas*. London & New York: Routledge. 155-63.
Bell, A., Hein, E. & Ratcliffe, J. (2015). 'Beyond Skeuomorphism: The Evolution Of Music Production Software User Interface Metaphors'. *Journal on the Art of Record Production*. 9. Available online at: <https://www.arpjournal.com/asarpwp/beyond-skeuomorphism-the-evolution-of-music-production-software-user-interface-metaphors-2/>. [Accessed: 25 Jan 2022].
Bell, S. A. (2015) 'From Voder to Vocaloid: A Media History of Voice Synthesis'. PhD Thesis. Salt Lake City, Utah: University of Utah, Department of Communications.
Bennett, J. (2011). 'Collaborative songwriting – the ontology of negotiated creativity in popular music studio practice'. Presented at the 6th Art of Record Production conference, Dec 2010, Leeds, UK. Available at: <https://joebennett.net/2011/07/07/collaborative-songwriting-academic-paper/>. [Accessed: 26 Jan 2021].
Blackburn, A. (2020). 'Voices That Matter: Authenticity, Identity, and Voice in the Musical Career of Lana Del Rey.' *Nota Bene: Canadian Undergraduate Journal of Musicology*. 13 (1): 84-114.
Bohn, D. A. & Rane Commercial Audio Products. (1982/87). 'Constant-Q Graphic Equalizers'. [online article]. Available at: <https://www.ranecommercial.com/legacy/note101.html> [Accessed 14 Nov 2020].
Bohn, D. A. (1988). 'Operator adjustable equalizers: An overview.' In *Proceedings of the Audio Engineering Society 6th International Conference on Sound Reinforcement*. Nashville, Tennessee, 5–8 May 1988: 369–381.
Bohne, B. A. & Harding, G. W. (1999). 'Noise & Its Effects on the Ear'. Noise-induced Hearing Loss. Dept. of Otolaryngology, Washington University School of Medicine, St. Louis, MO. Webpage archiving original article. Available at: <https://web.archive.org/web/20160701234404/http://oto2.wustl.edu/bbears/noise.htm>. [Accessed: 24 Sep 2020].
Born, G. & Devine, K. (2015). 'Music Technology, Gender, and Class: Digitization, Educational and Social Change in Britain'. *Twentieth-Century Music.* 12 (2): 135-172.
Born, G. (1995). *Rationalizing culture: IRCAM, Boulez and the Institutionalization of the Musical Avant-Garde*. London, Berkeley & Los Angeles: University of California Press.
Born, G. (1997). 'Modernist Discourse, Psychic Forms and Agency: Aesthetic Subjectivities at IRCAM'. *Cultural Anthropology.* 12 (4): 480-501.
Born, G. (2005). 'On Musical Mediation: Ontology, Technology and Creativity.' *Twentieth-Century Music*. 2 (1): 7-36.
Born, G. (2013). 'Introduction' in Born, G. [ed.] *Music, Sound and Space: Transformations of Public and Private Experience.* Cambridge: Cambridge University Press. 1–69.
Born, G. (2019). 'Workshop on Ethnographic Research Methods'. [Presentation]. 13 Jan 2019. Oxford: University of Oxford.
Bourbon, A. (2020). 'Plugging In'. in Hepworth-Sawyer, R., Hodgson, J., Paterson, J., & Toulson, R. [eds.] *Innovation in Music: Performance, Production, Technology and Business*. Abingdon, Oxon: Routledge.
Brackett, D. (2016). *Categorizing Sound: Genre and Twentieth-Century Popular Music.* Berkeley, CA: University of California Press.
Brøvig-Hanssen, R. & Danielsen, A. (2016). 'Autotuned Voices: Alienation and "Brokenhearted Androids."' in *Digital Signatures: The Impact of Digitization on Popular Music Sound*. Cambridge, MA: The MIT Press. 117-132.
Brøvig-Hanssen, R. (2018). 'Listening to or through technology: Opaque and transparent mediation in popular music'. In Bates, E. & Bennett, S. [eds.] In book: Critical Approaches to the Production of Music and Sound. London: Bloomsbury Academic. 195-210.
BSI (British Standards Institute). (2020). 'BS EN ISO 80000-8: 2020: Quantities and units. Acoustics.' Available at: <https://www.iso.org/standard/64978.html>. [Accessed: 28 Sep 2021].

Bushnell, B & Ferry, J. (2011). *From Downbeat to Vinyl: Bill Putnam's Legacy to the Recording Industry.* Morgan Hill, CA: Bookstand Publishing.
Cambridge Dictionary. (2022) 'gestalt' [webpage]. Available at: <https://dictionary.cambridge.org/dictionary/english/gestalt>. [Accessed: 22 Mar 2022].
Campbell, G. A. (1922). 'Physical theory of the electric-wave filter'. *The Bell System Technical Journal.* 1 (2): 1-32.
Case, A. (2007). *Sound FX: Unlocking the Creative Potential of Recording Studio Effects*. Oxford: Focal Press.
Cavarero, A. (2005). *For More Than One Voice: Toward a Philosophy of Vocal Expression*. Stanford, CA: Stanford University Press.
Celemony Software GmbH. (2018). 'Melodyne 4: User Manual'. [product documentation]. Available online at: <https://helpcenter.celemony.com/pdf/melodynestudio4_English.pdf>. [Accessed: 22 Jan 2018].
Celemony. (2020). 'What is Melodyne?'. [webpage]. *Celemony.com.* Available at: <https://www.celemony.com/en/melodyne/what-is-melodyne>. [Accessed: 22 Jan 2021].
Clarke, E. (2013). 'Music, Space and Subjectivity'. In Born, G. [ed.] *Music, Sound and Space: Transformations of Public and Private Experience*. Cambridge & New York: Cambridge University Press. 90-110.
Computer Music. (2013). 'Meet the Programmers: Fabfilter'. [online article]. *Musicradar.com.* 10 May 2013. Available at: <https://www.musicradar.com/news/tech/meet-the-programmers-fabfilter-574996>. [Accessed: 24 Sep 2020].
Computer Music. (2015). 'How to use serial compression'. [online article]. *Musicradar.com.* 16 October 2015. Available at: <https://www.musicradar.com/tuition/tech/how-to-use-serial-compression-628804>. [Accessed: 17 Dec 2020].
Connor, S. (2000). *Dumbstruck: A Cultural History of Ventriloquism*. Oxford & New York: Oxford University Press.
Connor, S. (2014). *Beyond words: sobs, hums, stutters and other vocalizations*. London: Reaktion Books.
Cousins, M. & Hussain, A. (1984). *Michel Foucault.* London: Palgrave Macmillan.
Crane, D. (2020). 'Chaining the 1176N and LA-2A Compressors for Maximum Control'. [online article]. *Universal Audio.* Available at: <https://www.uaudio.com/blog/chaining-1176ln-la2a-compressors/> [Accessed: 10 Dec 2020].
Croghan, N. B. H., Arehart, K. H. and Kates, J. M. (2012). 'Quality and loudness judgments for music subjected to compression limiting.' *Journal of the Acoustical Society of America*. 132 (2): 1177–1188.
D'Errico, M. (2016). 'Interface Aesthetics: Sound, Software, and the Ecology of Digital Audio Production.' PhD Thesis. Los Angeles, CA: University of California Los Angeles.
Danielsen, A. (2010). *Musical Rhythm in the Age of Digital Reproduction*. Abingdon, Oxon: Routledge.
Danielsen, A. (2019). 'Glitched and Warped: Transformations of Rhythm in the Age of the Digital Audio Workstation'. In Grimshaw-Aagaard, M., Walther-Hansen, M. and Knakkergaard, M. [eds.]. *The Oxford Handbook of Sound and Imagination, Volume 2.* Oxford & New York: Oxford University Press. 594-609.
De Brosses, C. (1760 [2011]). *Du Culte des dieux fétiches, ou parallèle de l'ancienne religion de l'Egypte avec la religion actuelle de Nigritie*. Charleston, CA: Nabu Press.
DeCordova, R. (2001). *Picture Personalities: The Emergence of the Star System in America.* Urbana, Ill.: University of Illinois Press.
Del Percio, A. & Duchêne, A. (2011). 'Commodification of Pride and Resistance to Profit: Language Practices as Terrain of Struggle in a Swiss Football Stadium.' in Duchêne, A. & Heller, M. [eds.] *Language in Late Capitalism: Pride and Profit.* New York & London: Routledge. 43-72.

Deruty, E. & Tardieu, D. (2014). ‘About Dynamic Processing in Mainstream Music’. *J. Audio. Eng. Soc.* 62 (1/2): 42-55.
Deruty, E. (2011). ‘‘Dynamic Range’ & The Loudness War’. [online article]. *Sound on Sound.* September 2011. Available at: <https://www.soundonsound.com/sound-advice/dynamic-range-loudness-war>. [Accessed: 8 Dec 2020].
Devine, K. (2013). ‘Imperfect Sound Forever: Loudness Wars, Listening Formations and the History of Sound Reproduction.’ *Journal of Popular Music*. 3 (2): 159-176.
Dibben, N. (1999). ‘Representations of Femininity in Popular Music.’ *Popular Music*. 18 (3): 331-355.
Dibben, N. (2006). ‘Subjectivity and the Construction of Emotion in the Music of Björk.’ *Music Analysis.* 25 (1/2): 171-197.
Dibben, N. (2009). ‘Vocal Performance and the Projection of Emotional Authenticity’, in Scott, D.B. [eds.] *The Ashgate Research Companion to Popular Musicology.* Farnham, Surrey: Ashgate. 317-333.
Dibben, N. (2013). ‘The Intimate Singing Voice: Auditory Spatial Perception and Emotion in Pop Recordings’. In Zakharine, D. & Meise, N. [eds.] *Electrified Voices: Medial, Socio-Historical and Cultural Aspects of Voice Transfer*. Göttingen, Germany: V&R Unipress.
Dickinson, K. (2001). ‘‘Believe’?: Vocoders, digitalised female identity and camp.’ *Popular Music.* 20(3): 333-347.
Dockwray, R. and Moore, A. F. (2008). ‘The establishment of the virtual performance space in rock’. *Twentieth-Century Music*. 5: 219-241.
Dockwray, R. and Moore, A. F. (2010). ‘Configuring the sound-box 1965-1972’. *Popular Music*. 29(2): 181-197.
Dolan, E. & Rehding, A. [eds]. (2020). *The Oxford Handbook of Timbre*. Oxford & New York: Oxford University Press.
Dolar, M. (2006). *A Voice and Nothing More.* Cambridge, MA: The MIT Press.
Downs, J. K. (2016). ‘Spatial imbrication, multitrack recording, and the problem of (hyper)reality.’ Unpublished M.St. dissertation. Oxford: University of Oxford. Copy acquired via correspondence with the author.
Downs, J. K. (2021). ‘Acoustic Territories of the Body: Headphone Listening, Embodied Space, and the Phenomenology of Sonic Homeliness.’ *Journal of Sonic Studies.* 21. Available at: <https://www.researchcatalogue.net/view/1260374>. [Accessed 15 Jun 2021].
Doyle, P. (2004). ‘From ‘My Blue Heaven’ to ‘Race with the Devil’: echo, reverb and (dis)ordered space in early popular music recording.’ *Popular Music.* 23 (1): 31-49.
Doyle, P. (2005). *Echo and Reverb: Fabricating Space in Popular Music Recording*. Middletown, CT: Wesleyan University Press.
Droney, M. (2001). ‘Chris Lord-Alge’. [online article]. *Mix Online.* Available at: <https://www.mixonline.com/recording/chris-lord-alge-372731>. [Accessed: 13 Jun 2022].
Dull, D. & West, C. (1991). ‘Accounting for Cosmetic Surgery: The Accomplishment of Gender’. *Social Problems.* 38 (1): 1991.
Dunsby, J. (2009). ‘Roland Barthes and the Grain of Panzéra’s Voice.’ *Journal of the Royal Musical Association*. 134(1): 113-132.
Eidsheim, N. (2008). ‘Voice as a Technology of Selfhood: Towards an Analysis of Racialized Timbre and Vocal Performance’. PhD diss., UC San Diego.
Eidsheim, N. (2009). ‘Synthesizing Race: Towards an Analysis of the performativity of Vocal Timbre’. *TRANS: Revista Transcultural de Música (Transcultural Music Review).* 13. Available online at: <https://www.sibetrans.com/trans/articulo/57/synthesizing-race-towards-an-analysis-of-the-performativity-of-vocal-timbre>. [Accessed: 7 May 2021].
Eidsheim, N. (2014). ‘The Micropolitics of Listening to Vocal Timbre’. *Postmodern Culture*. 24(3): 0-0.

Eidsheim, N. (2018). '"The Triumph of Jimmy Scott": A Voice Beyond Category'. in *The Relentless Pursuit of Tone: Timbre in Popular Music.* R. Fink, Z. Wallmark and M. Latour [eds]. Oxford & New York: Oxford University Press. 141-158.

Eidsheim, N. (2019). *The Race of Sound: Listening, Timbre, and Vocality in African American Music*. Durham, NC & London: Duke University Press.

Fabfilter. (2021). 'FabFilter Pro Q-3 – Equalizer Plug-in'. [product webpage]. *Fabfilter.com.* Available at: < https://www.fabfilter.com/products/pro-q-3-equalizer-plug-in>. [Accessed: 26 Mar 2021].

Feld, S. & Fox, A. (1994). 'Music and Language'. *Annual Review of Anthropology.* 23: 25-53.

Feld, S. (1981). 'Flow like a Waterfall: The Metaphors of Kaluli Musical Theory'. *Yearbook for Traditional Music.* 13: 22-47.

Feld, S. (1988). 'Aesthetics as iconicity of style, or 'Lift-up-over Sounding': Getting into the Kaluli Groove'. *Yearbook for Traditional Music.* 20: 74-113.

Feld, S., Fox, A., Porcello, T. & Samuels, D. (2004). 'Vocal Anthropology: From the Music of Language to the Language of Song'. In Duranti, A. [ed.] *A Companion To Linguistic Anthropology*. Malden, MA, Oxford & Victoria: Blackwell Publishing. 321-346.

Fink, R., Wallmark, Z. & Latour, M. (2018) 'Introduction' in *The Relentless Pursuit of Tone: Timbre in Popular Music.* Oxford & New York: Oxford University Press. 1-18.

Fitzjohn, S. (2020). 'The Vocal EQ Chart (Vocal Frequency Ranges + EQ Tips)'. [online article]. *Producerhive.com.* Available at: <https://producerhive.com/music-production-recording-tips/how-to-use-a-vocal-eq-chart/>. [Accessed: 25 Sep 2020].

Freud, S. (1899). [1997]. *The Interpretation of Dreams.* Translated by A. A. Brill. Ware, Herts: Wordsworth Editions.

Freud, S. (1910). [2001]. 'A Special Type of Choice of Object made by Men'. In Strachey, J. [trans.]. *The Standard Edition of the Complete Psychological Works of Sigmund Freud, Volume XI (1910): Five Lectures on Psycho-Analysis, Leonardo da Vinci and Other Works*. 163-176.

Freud, S. (1927). [1950]. 'Fetishism'. In Strachey, J. [trans.]. *Collected papers / Vol. 5, Miscellaneous papers, 1888-1938.* London : The Hogarth Press and the Institute of Psycho-analysis. 198-204.

Frith, S. & Zagorski-Thomas, S. (2012). *The Art of Record Production: An Introductory Reader for a New Academic Field*. Farnham, Surrey, UK: Ashgate.

Frith, S. (1986). 'Art versus technology: the strange case of popular music'. *Media, Culture and Society.* 8: 263-79.

Frith, S. (1996). *Performing Rites: On the Value of Popular Music.* Cambridge, MA: Harvard University Press.

Fuston, L. (2020a). 'UA's Classic 1176 Compressor – A History'. [online article]. *Universal Audio.* Available at: <https://www.uaudio.com/blog/analog-obsession-1176-history/>. [Accessed: 10 Dec 2020].

Fuston, L. (2020b). 'A History of the Teletronix LA-2A Leveling Amplifier'. [online article]. *Universal Audio.* Available at: <https://www.uaudio.com/blog/la-2a-analog-obsession/>. [Accessed: 10 Dec 2020].

Gell, A. (1998). *Art and Agency*. Oxford: Oxford University Press.

Giannoulis et. al. (2012). 'Digital Dynamic Range Compressor Design – A Tutorial and Analysis'. *J. Audio. Eng. Soc.* 60 (6): 399-408.

Gibson, D. (1997). *The Art of Mixing: A Visual Guide to Recording, Engineering and Production*. Vallejo, CA: Mix Books.

Green Light Sound. (2019). 'Andrew Scheps Vocal Mixing Trick | Get Your Vocals To Cut Through The Mix'. [online video]. Uploaded 25 Jul 2019. Available at: <https://www.youtube.com/watch?v=ysN5w7pD9nA&t=57s>. [Accessed: 31 May 2022].

Grossberg, L. (1992). *We Gotta Get Out of this Place: Popular Conservatism and Postmodern Culture*. New York & London: Routledge.

H. T. (2017). ‘What makes something a commodity?’ *The Economist*. [online article]. 3 Jan 2017. Available at: <https://www.economist.com/the-economist-explains/2017/01/03/what-makes-something-a-commodity>. [Accessed: 25 Feb 2022].
Hammersley, M. & Atkinson, P. (2007). *Ethnography: Principles in Practice*. 3rd edition. London & New York: Routledge.
Harding, P. (2020). *Pop Music Production: Manufactured Pop and Boybands of the 1990s*. London & New York: Routledge.
Harkness, N. (2014). *Songs of Seoul: An Ethnography of Voice and Voicing in Christian South Korea.* Berkeley, Los Angeles and London: University of California Press.
Hayles, N. K. (1999). *How We Became Posthuman*. Chicago: University of Chicago Press.
Hearn, A. (2017). ‘Commodification’ in Ouellette, L. and Gray, J. [eds.] *Keywords for Media Studies.* New York: New York University Press. 43-46.
Heidemann, K. (2016). ‘A System for Describing Vocal Timbre in Popular Song’. *Music Theory Online.* 22 (1). Available at: <http://www.mtosmt.org/issues/mto.16.22.1/mto.16.22.1.heidemann.html>. [Accessed: 17 May 2019].
Heller, E. J. (2012). *Why You Hear What You Hear: An Experiential Approach to Sound, Music and Psychoacoustics.* Princeton & Oxford: Princeton University Press.
Help Me Devvon. (2020). ‘Vocal EQ Techniques You Should Know.’ [online video]. Uploaded 18 May 2020. Available at: <https://www.youtube.com/watch?v=sZGKC5CAh7M&t=161s>. [Accessed: 24 Oct 2020].
Hepworth-Sawyer, R., Hodgson, J., Paterson, J. & Toulson, R. (2019). *Innovation in Music: Performance, Production, Technology and Business.* New York & Abingdon, Oxon: Routledge.
Hermes, K. (2020). ‘Optimizing Vocal Clarity in the Mix.’ in Hepworth-Sawyer, R., Hodgson, J., Paterson, J., & Toulson, R. [eds.] *Innovation in Music: Performance, Production, Technology and Business.* Abingdon, Oxon: Routledge. 191-210.
Hochschild, A. (1983). *The Managed Heart: Commercialization of Human Feeling.* Berkeley, Los Angeles & London: University of California Press.
Hodgson, J. (2010). *Understanding Records: A Field Guide to Recording Practice*. London & New York: Continuum.
Hodgson, T. (2021). ‘Spotify and the democratisation of music’. *Popular Music.* 40 (1): 1-17.
Horton, D. & Wohl, R. (1956). ‘Mass communication and para-social interaction: Observation on intimacy at a distance.’ *Psychiatry*. 19 (3): 215–229.
Houghton, M. (2016). ‘Sound Radix SurferEQ.’ [online product review]. *Sound on Sound.* January 2016. Available at: <https://www.soundonsound.com/reviews/sound-radix-surfereq>. [Accessed: 26 Mar 2021].
Huber, D. M. & Runstein, R. E. (2018). *Modern Recording Techniques.* 9th Edition. Abingdon, Oxon: Routledge.
Huff, C. (2011). ‘EQ Vocals: The Five Primary Areas of Modification’. [online blog post]. *Behindthemixer.com.* Available at: <https://www.behindthemixer.com/vocal-eq-settings/>. [Accessed: 10 Jun 2022].
Inglis, S. (2011). ‘Waves CLA Artist Signature Collection: Plug-in Suite’. [online review article]. *Sound on Sound.* February 2011. Available at: <https://www.soundonsound.com/reviews/waves-cla-artist-signature-collection>. [Accessed: 17 Dec 2020].
Inglis, S. (2014). ‘FabFilter Pro-Q 2: EQ Plug-in For Mac OS & Windows.’ [online review article]. *Sound on Sound.* December 2014. Available at: <https://www.soundonsound.com/reviews/fabfilter-pro-q-2>. [Accessed: 21 Feb 2022].
Inglis, S. (2017). ‘FabFilter Pro-R: Algorithmic Reverb Plug-in’. [online review article]. *Sound on Sound.* March 2017. Available at: <https://www.soundonsound.com/reviews/fabfilter-pro-r>. [Accessed: 24 Sep 2020].

Inglis, S. (2019). ‘FabFilter Pro-Q 3.’ [online review article]. *Sound on Sound.* April 2019. Available at: <https://www.soundonsound.com/reviews/fabfilter-pro-q-3>. [Accessed: 11 Jun 2022].
International Society for Presence Research. (2000). ‘The Concept of Presence: Explication Statement.’ [online article]. Available at: <https://ispr.info/about-presence-2/about-presence/>. [Accessed: 16 Jun 2021].
International Society for Presence Research. (2021). ‘About ISPR.’ [webpage]. *International Society for Presence Research.* Available at: <https://ispr.info/>. [Accessed: 16 Jun 2021].
Irigaray, L. (1974). *Speculum of the Other Woman.* Translated by Gillian C. Gill. Ithaca, NY: Cornell University Press.
Izhaki, R. (2017). [2008]. *Mixing Audio: Concepts, Practices and Tools*. 3rd Edition. Abingdon, Oxon: Routledge.
iZotope. (2018). ‘Removing plosives from a voice recording’. [online article]. *iZotope.com*. Available at*:* <https://www.izotope.com/en/learn/removing-plosives-from-a-voice-recording.html>. [Accessed: 4 February 2021].
iZotope. (2022). ‘Nectar 3 Plus Features.’ [product page]. *izotope.com.* Available at: < https://www.izotope.com/en/products/nectar/features.html>. [Accessed: 24 Feb 2022].
Jacob H. (2021). ‘How to EQ Vocals Professionally: The Easy Six Step Method’. [online blog post]. *Ledgernote.com.* Available at: <https://ledgernote.com/columns/mixing-mastering/how-to-eq-vocals/>. [Accessed: 10 Jun 2022].
Jared H. (2021). ‘EQ Cheat Sheet? Frequency Charts for Mixing Hurt More Than Help!’ [online blog post]. *Ledgernote.com.* Available at: <https://ledgernote.com/columns/mixing-mastering/eq-cheat-sheet/>. [Accessed: 22 Feb 2021].
Jeffs, R., Holden, S., Bohn, D. & Rane. (2005). ‘Dynamics Processors—Technology & Applications’. [online article]. *Rane Commercial Audio Products.* RaneNote 155, September 2005. Available at: < https://www.ranecommercial.com/legacy/note155.html>. [Accessed: 16 Dec 2020].
Katz, B. (2007). *Mastering Audio: The Art and the Science.* Waltham, MA: Focal Press.
Kealy, E. (1979). ‘From Craft to Art: The Case of Sound Mixers and Popular Music’. *Sociology of Work and Occupations*. 6 (1): 3-29.
Kockelman, P. (2008). ‘A semiotic ontology of the commodity.’ *Journal of Linguistic Anthropology.* 16 (1): 76-102.
Krantzberg. E. (2020). ‘Vocal Compression: Get Your Vocals Sounding Great!’ [online article]. *SonarWorks.* Jan 2020. Available at: <https://www.sonarworks.com/soundid-reference/blog/learn/vocal-compression-get-your-vocals-sounding-great/>. [Accessed: 8 Apr 2021].
Kristeva, J. (1980). *Powers of Horror: An Essay on Abjection.* New York: Columbia University Press.
Kromhout, M. J. (2017). ‘Noise resonance: Technological sound reproduction and the logic of filtering’. PhD Diss. Amsterdam: University of Amsterdam.
Kromhout, M. J. (2021). (forthcoming). *The Logic of Filtering*: *How Noise Shapes the Sound of Recorded Music.* New York: Oxford University Press.
Kvifte, T. (2010) ‘Composing a Performance: The Analogue Experience in the Age of Digital (Re)Production.’ In Danielsen, A. (ed). *Musical Rhythm in the Age of Digital Reproduction.* 213-229. Farnham, Surrey, England & Burlington, VT: Ashgate.
Lacasse, S. (2000). ‘Listen to my voice: the evocative power of vocal staging in rock music and other forms of expression’. PhD Thesis, University of Liverpool.
Lavengood, M. (2017). ‘A New Approach to the Analysis of Timbre’. PhD Thesis. New York: The City University of New York.
Lavengood, M. (2019). ‘What Makes it Sound ‘80s?: The Yamaha DX7 Electric Piano Sound’. *Journal of Popular Music Studies*. 31 (3): 73-94.
Ledsham, E. (2015). ‘Human After All: Auto-tune, Technology and Human Creativity’. [online article]. *Drowned in Sound*. Available at: <https://drownedinsound.com/in_depth/4149415-human-after-all--auto-tune-technology-and-human-creativity>. [Accessed: 20 Jan 2022].

Leyshon, A. (2001). ‘Time - space (and digital) compression: software formats, musical networks, and the reorganisation of the music industry’. *Environment and Planning.* 33: 49-77.
Lockheart, P. (2003). ‘A History of Early Microphone Singing, 1925-39: American Popular Singing at the Advent of Electronic Microphone Amplification’. *Popular Music and Society.* 26 (3): 367-385.
Lombard, M. & Ditton, T. (1997). ‘At the Heart of It All: The Concept of Presence.’ *Journal of Computer-Mediated Communication*. 3 (2). 0-0.
Lukàcs, G. (1923). [1967]. ‘Reification and the Consciousness of the Proletariat’. In Lukàcs, G. *History and Class Consciousness.* Translated by Rodney Livingstone. Merlin Press. [transcribed by Andy Berland for Marxists.org, available at: <https://www.marxists.org/archive/lukacs/works/history/hcc05.htm>. Accessed 25 Apr 2022].
MacGaffey, W. (1977). ‘Fetishism Revisited: Kongo Nkisi in Sociological Perspective.’ *Africa*. 47 (2): 172-84.
MacGaffey, W. (1990). ‘The Personhood of Ritual Objects: Kongo “Minkisi”’. *Etnofoor*. 3 (1): 45-61.
MacGaffey, W. (1994). ‘African Objects and the Idea of Fetish’. *Anthropology and Aesthetics.* 25: 123-131.
MacGaffey, W. (1998). ‘“Magic, or as we usually say, Art”: A framework for comparing European and African art’. in Schildkrout, E. & Keim, C. A. [eds.] *The Scramble for Art in Central Africa.* Cambridge: Cambridge University Press. 228-229.
Make Pop Music. (2019). ‘How To Mix Pop Vocals (Like Julia Michaels, Taylor Swift, & Chelsea Cutler) | Make Pop Music’ [online video]. Uploaded 5 Feb 2019. Available at: <https://www.youtube.com/watch?v=XJYH16OcZEk>. [Accessed: 15 April 2022].
Make Pop Music. (2020). ‘How To Mix Pop Vocals Like A Pro (FREE STEMS AND PRESETS) | Make Pop Music’. [online video]. Uploaded 5 Feb 2019. Available at: <https://www.youtube.com/watch?v=N0lVM-J9JyQ> [accessed: 10 Dec 2020].
Malawey, V. (2020). *A Blaze of Light in Every Word: Analyzing the Popular Singing Voice*. New York: Oxford University Press.
Manchester Music. (2019). ‘Fabfilter Pro Q3: Why is everyone in love with this thing?’. YouTube Video Review by Geoff Manch. Uploaded 14 Nov 2019. Available at: <https://www.youtube.com/watch?v=_lVUxIuR_nE>. [Accessed: 24 Sep 2020].
Marrington, M. (2017). ‘Mixing Metaphors: Aesthetics, Mediation and the Rhetoric of Sound Mixing.’ In Hepworth-Sawyer, R. and Hodgson, J. (eds.) *Mixing Music*. New York, Routledge. 199-215.
Marshall, MN. (1996). ‘The key informant technique’. *Family Practice*. 13: 92-97.
Marshall, O. (2017). ‘Tuning In-Situ: Articulations of Voice, Affect and Artefact in the Recording Studio’. PhD Thesis. Ithaca, NY: Cornell University.
Marx, K. (1847 [2008]). *The Poverty of Philosophy*. New York: Cosimo.
Marx, K. (1867 [2009]). *Das Kapital*. Washington DC: Regnery Publishing.
Mastering the Mix. (2018). ‘Decoding the Mix #7: All Star Collaboration.’ [online article]. *Masteringthemix.com.* Available at: <https://www.masteringthemix.com/blogs/learn/decoding-the-mix-7-all-star-collaboration>. [Accessed: 22 Feb 2021].
Mastering The Mix. (2021). ‘RESO: Dynamic Resonance Suppressor.’ [product webpage]. *Masteringthemix.com.* Available at: <https://www.masteringthemix.com/products/reso>. [Accessed: 11 Nov 2021].
Mayhew, E. (1999). ‘Women in Popular Music and the Construction of "Authenticity”.’ *Journal of Interdisciplinary Gender Studies.* 4 (1): 63-81.
Mayzes, R. (2016). ‘Vocal Compression: Learn How To Mix Like The Pros.’ [online article]. *Musicianonamission.com.* Available at: <https://www.musicianonamission.com/vocal-compression-how-to-compress-vocals/>. [Accessed: 8 Apr 2021].
McAdams, S. et. al. (2006). ‘A Meta-Analysis of Acoustic Correlates of Timbre Dimensions.’ *Journal of the Acoustical Society of America.* 120 (5): 3275–3276.

McIntyre, P. (2008). ‘Creativity and Cultural Production: Creativity and Cultural Production: A Study of Contemporary Western Popular Music Songwriting’. *Creativity Research Journal*. 20 (1): 40-52.
McMurray, P. (2017). ‘Once Upon Time: A Superficial History of Early Tape’. *Twentieth-Century Music*. 14(1): 25-48.
Meintjes, L. (2003). *Sound of Africa!: Making Music Zulu in a South African Studio*. Durham, NC: Duke University Press.
Messite, N (iZotope Contributor). (2018). ‘Vocal Mixing: 6 Popular Styles of Vox in Pop and Rock Mixes.’ [online article]. *iZotope.com.* 20 November 2018. Available at: <https://www.izotope.com/en/learn/vocal-mixing-6-popular-styles-of-vox-in-pop-and-rock-mixes.html>. [Accessed: 15 Jun 2021].
Michael White. (2014). ‘Mixing With Mike Mixing Tip: How to Pull A Vocal Forward in a Mix’. [online video]. Uploaded 30 Oct 2014. Available at: <https://www.youtube.com/watch?v=SKdDOmOEAX4>. [Accessed: 31 May 2022].
Middleton, R. (1990). *Studying Popular Music.* Milton Keynes & Philadelphia: Open University Press.
Mills, M. (2020). ‘”Everything is a Filter?”: A Social History of the Electrical Filter.’ Keynote Presentation at *Sound Instruments and Sonic Cultures: An Interdisciplinary Conference.* The National Science and Media Museum: Bradford, UK. [Conference hosted virtually during Covid-19 pandemic]. 18 Dec 2020. Currently available at: <https://www.crowdcast.io/e/keynote-mara-mills>. [Accessed: 18 Feb 2020].
Milner, G. (2009). *Perfecting Sound Forever: The Story of Recorded Music*. London: Granta.
Moore, A. F. (1997). *The Beatles: Sgt. Pepper’s Lonely Hearts Club Band.* Cambridge & New York: Cambridge University Press.
Moore, A. F. (2000). ‘Authenticity as Authentication.’ *Popular Music*. 21 (2): 209-223.
Moore, A. F. (2005). ‘The Persona-Environment Relation in Recorded Song’. *Journal of the Society for Music Theory*. 11(4). [online]. Available at: <http://www.mtosmt.org/issues/mto.05.11.4/mto.05.11.4.moore.html>. [Accessed: 22 Apr 2017].
Moore, A. F. (2009). ‘Interpretation: So What?’. in Scott, D. B. [ed.] *The Ashgate Resarch Companion to Popular Musicology.* Farnham, Surrey: Ashgate, 411-412.
Moore, A. F. (2010). ‘Where is Here? An Issue of Deictic Projection in Recorded Song’. *Journal of the Royal Musical Association*. 135(1): 145-182.
Moore, A. F. (2012). ‘Beyond a Musicology of Production’. in Frith, S. and ZagorskiThomas, S. [eds.] *The Art of Record Production: An Introductory Reader for a New Academic Field.* Farnham, Surrey: Ashgate. 99-111.
Moore, A. F., Dockwray, R. and Schmidt, P. (2009). ‘A Hermeneutics of Spatialization for Recorded Song’. *Twentieth-Century Music*. 6: 83-114.
Moore, Austin. & Wakefield, J. P. (2017). ‘An Investigation into the Relationship Between the Subjective Descriptor Aggressive and the Universal Audio 1176 FET Compressor’. Paper presented at: *142nd International AES Convention.* 20-23rd May, Berlin, Germany.
Moore, Austin. (2012). ‘All Buttons In: An investigation into the use of the 1176 FET compressor in popular music production’. *Journal on the Art of Record Production.* 6. Available at: <http://eprints.hud.ac.uk/id/eprint/27391/>. [Accessed: 14 Dec 2020].
Moore, Austin. (2020). ‘Dynamic Range Compression and the Semantic Descriptor Aggressive’. *Appl. Sci.* 10 (2350).
Moore, Austin., Till, R. & Wakefield, J. P. (2016). ‘An Investigation into the Sonic Signature of Three Classic Dynamic Range Compressors’. Paper presented at: *140th International AES Convention.* 4-7th June 2016, Paris, France.
Moore, B. C. J. (2013 [1977]). *An Introduction to the Psychology of Hearing.* Leiden, Netherlands: Brill.

Moylan, W. (2012). ‘Considering Space in Recorded Music’. in Frith, S. and Zagorski-Thomas, S. [eds.] *The Art of Record Production: An Introductory Reader for a New Academic Field.* Farnham, Surrey: Ashgate. 163-188.
Moylan, W. (2014). [2002]. *Understanding and Crafting the Mix: The Art of Recording*. 3rd edition. Oxford, UK: Focal Press.
Müller, M. (2015). *Fundamentals of Music Processing: Audio, Analysis, Algorithms, Applications. New York*: Springer.
Mulvey, L. (1975). ‘Visual Pleasure and Narrative Cinema.’ *Screen*. 16 (3): 6-18.
Nagle, P. (2015). ‘Izotope Iris 2: Software Synthesizer’. [online article]. *Sound on Sound.* March 2015. Available at: <https://www.soundonsound.com/reviews/izotope-iris-2>. [Accessed: 25 Jan 2021].
Nettl, B. (1956). *Music in Primitive Culture.* Cambridge, MA: Harvard University Press.
oeksound. (2021). ‘soothe2: a dynamic resonance suppressor.’ [product page]. Available at: <https://oeksound.com/plugins/soothe2/>. [Accessed: 26 Mar 2021].
Oswinski, B. (2013) [1999)]. *The Mixing Engineer’s Handbook.* 3rd edition. Burbank, CA: Bobby Oswinski Media Group.
Ouzounian, G. (2021). *Stereophonica: Sound and Space in Science, Technology, and the Arts.* Cambridge, MA: The MIT Press.
Papadaki, E. (2019). ‘Feminist Perspectives on Objectification’. [online article]. *Stanford Encyclopaedia of Philosophy*. Available at: <https://plato.stanford.edu/entries/feminism-objectification/>. [Accessed: 25 Apr 2022].
Papenburg, J. (2020). ‘Enhanced Sound. Filter der Musikproduktion und des Musikhörens’. In: *Navigationen - Zeitschrift für Medien- und Kulturwissenschaften*. Filter(n) – Geschichte Ästhetik Praktiken. 20 (2): 115– 132.
Parys, B. (2020). ‘The ‘T-Pain Effect’ Is About Way More Than Auto-Tune’. [online article]. *BerkleeNow.* (Berklee College of Music Magazine). Available at: <https://www.berklee.edu/news/berklee-now/t-pain-effect-about-much-more-auto-tune>. [Accessed: 21 Jan 2021].
Peres, A. (2016). ‘The Sonic Dimension as Dramatic Driver in 21st-century Pop Music’. PhD Thesis. Ann Arbor, MI: University of Michigan.
Peres, A. (2018). ‘Camilla Cabello, Charlie Puth and the Vocal Range Wars’. [blog post]. *Top 40 Theory.* Available at: <https://www.top40theory.com/blog/camila-cabello-charlie-puth-and-the-vocal-range-wars>. [Accessed: 25 January 2022].
Perlman, M. (2004) ‘Golden Ears and Meter Readers: The contest for epistemic authority in audiophilia’. *Social Studies of Science.* 34 (5): 783-807.
Piatier, A. (1987/88). ‘Transectorial Innovations and the Transformation of Firms.’ *The Information Society* 5 (4): 205-31.
Pietz, W. (1985). ‘The Problem of the Fetish, I’. *RES: Anthropology and Aesthetics*. 9 (9): 5–17.
Pietz, W. (1987). ‘The Problem of the Fetish, II: The Origin of the Fetish’. *RES: Anthropology and Aesthetics*. 13: 23-45.
Pietz, W. (1988). ‘The Problem of the Fetish, IIIa: Bosman's Guinea and the Enlightenment Theory of Fetishism’. *RES: Anthropology and Aesthetics.* 16: 105-124.
Pietz, W. (2003). ‘Fetish’ in Nelson, R. S. and Shiff, R. [eds.] *Critical Terms for Art History.* 2nd edition. London & Chicago: University of Chicago Press. 307-317.
Pool, R. (1990). ‘Fetishism Deconstructed.’ *Etnofoor.* 3 (1): 114-127.
Porcello, T. (2004). ‘Speaking of Sound: Language and the Professionalization of Sound Recording Engineers’. *Social Studies of Science.* 34 (5): 733-758.
Prior, N. (2017). ‘On Vocal Assemblages: From Edison to Miku’. *Contemporary Music Review*. 37 (5-6): 488-506.
Prior, N. (2018). *Popular Music, Digital Technology and Society.* Thousand Oaks, CA & London: Sage.

Pro Audio Files. (2016). ‘Why Use Multiband Compression?’ [online video]. Uploaded 26 May 2016. Available at: <https://www.youtube.com/watch?v=fGDFL-oVocM> [Accessed and transcribed: 20 Nov 2020, video has since been deleted]

Pro Audio Files. (2020). ‘Getting That Serban Ghenea Vocal Sound’. [online video]. Uploaded 26 April 2020. Available at: <https://www.youtube.com/watch?v=UtVw_4c5Zo4>. [Accessed and transcribed: 15 Dec 2020, video has since been deleted].

Provenzano, C. (2018). ‘Auto-Tune, Labor and the Pop-Music Voice’. In Fink, R., Latour, M. & Wallmark, Z. [eds]. *The Relentless Pursuit of Tone: Timbre in Popular Music*. Oxford & New York: Oxford University Press. 159-182.

Provenzano, C. (2019a). ‘Emotional Signals: Digital Tuning Software and the Meanings of Pop Music Voices’. PhD Thesis. New York: New York University. May 2019.

Provenzano, C. (2019b). ‘Making Voices: The Gendering of Pitch Correction and The Auto-Tune Effect in Contemporary Pop Music.’ *Journal of Popular Music Studies.* 31 (2): 63–84.

Ragona, M. (2013). ‘Doping the Voice’. in Vernallis, C., Herzog, A. and Richardson, J. [eds.] *The Oxford Handbook of Sound and Image in Digital Media.* Oxford & New York: Oxford University Press. 154-170.

Recording Revolution. (2016). ‘Mixing Vocals to Sit Upfront’. [blog post]. *Recordingrevolution.com.* Available at: <https://www.recordingrevolution.com/mixing-vocals-to-sit-upfront/>. [Accessed: 23 Oct 2020].

Reiss, J. D. & McPherson, A. (2015). ‘Filter effects’ (Chapter 4). In *Audio Effects: Theory, Implementation and Application*. CRC Press: Boca Raton, FL, USA. 89–124.

Ritchie, J., Lewis, J., McNaughton Nicholls, C. & Ormston, R. (2014). *Qualitative Research Practice: A Guide for Social Science Students and Researchers.* 2nd Edition. Los Angeles, London, New Delhi & Singapore: Sage.

Robinson, P. (2017). ‘‘Whisperpop’: why stars are choosing breathy intensity over vocal paint-stripping’ [online article]. *The Guardian.* 11 Nov 2017. Available at: <https://www.theguardian.com/music/2017/nov/11/whisperpop-why-stars-choosing-breathy-intensity-over-vocal-paint-stripping>. [Accessed: 10 Nov 2021].

Robjohns, H. (2001). ‘UNIVERSAL APPEAL: Universal Audio 1176LN Limiting Amplifier.’ [online article]. *Sound on Sound.* June 2001. Available at: <https://www.soundonsound.com/reviews/universal-appeal>. [accessed: 18 Feb 2022].

Robjohns, H. (2012). ‘Universal Audio Apollo: Firewire Audio Interface.’ [online article]. *Sound on Sound.* June 2012. Available at: <https://www.soundonsound.com/reviews/universal-audio-apollo>. [Accessed: 18 Feb 2022].

Robjohns, S. (2013). ‘Parallel Compression: The Real Benefits’. [online article]. *Sound on Sound.* February 2013. Available at: <https://www.soundonsound.com/techniques/parallel-compression>. [Accessed: 8 Dec 2020].

Robjohns, H. (2017a). ‘Izotope RX6: Restoration Software [MacOSX/Windows]’. [online article]. *Sound on Sound.* Available at: <https://www.soundonsound.com/reviews/izotope-rx6>. [Accessed: 25 Jan 2021].

Robjohns, H. (2017b). ‘Q: Can static EQ tackle masking effectively?’ [online article]. *Sound on Sound.* Available at: <https://www.soundonsound.com/sound-advice/q-can-static-eq-tackle-masking-effectively>. [Accessed: 22 Feb 2021].

Robjohns, S. & White, P. (2002). ‘Multi-band Compression: Practical Techniques’. [online article]. *Sound on Sound.* August 2002. Available at: <https://www.soundonsound.com/techniques/multi-band-compression-tips>. [Accessed: 16 Dec 2020].

Rodgers, T. (2010). *Pink Noises: Women on Electronic Music and Sound.* Durham, NC & London: Duke University Press.

Rogerson, B. (2021). ‘Gearslutz announces that it will be changing its name, following an online campaign’ [online article]. *Musicradar.com.* 20 Jan 2021. Available at:

<https://www.musicradar.com/news/gearslutz-announces-that-it-will-be-changing-its-name-following-an-online-campaign>. [Accessed: 24 Feb 2022].
Ryan, A. F. et. al. (2016). 'Temporary and Permanent Noise-induced Threshold Shifts: A Review of Basic and Clinical Observations.' *Otology & neurotology* 37(8): 271-5.
Savage, S. (2011). *Bytes and Backbeats: Repurposing Music in the Digital Age.* Ann Arbor: University of Michican Press.
Schmidt-Horning, S. (2013). *Chasing Sound: Technology, Culture, and the Art of Studio Recording from Edison to the LP. Baltimore*, Maryland: Johns Hopkins University Press.
Seabrook, J. (2015). *The Song Machine: Inside the Hit Factory.* New York: Penguin Random House.
Senior, M. (2009a). 'Advanced Vocal Compression'. [online article]. *Sound on Sound.* April 2009. Available at: <https://www.soundonsound.com/techniques/advanced-vocal-compression>. [Accessed: 8 Apr 2021].
Senior, M. (2009b). 'Techniques for Vocal De-essing'. [online article]. *Sound on Sound.* May 2009. Available at: <https://www.soundonsound.com/techniques/techniques-vocal-de-essing>. [Accessed: 16 Dec 2020].
Senior, M. (2009c). 'Classic Compressors: Choosing The Right Compressor For The Job'. [online article]. *Sound on Sound.* September 2009. Available at: <https://www.soundonsound.com/techniques/classic-compressors>. [Accessed: 10 Dec 2020].
Senior, M. (2011). 'Comping: Reaper Tips & Techniques'. [online article]. *Sound on Sound.* April 2011. Available at: <https://www.soundonsound.com/techniques/comping> . [Accessed: 20 Jan 2021].
Senior, M. (2019) [2011]. *Mixing Secrets for the Small Studio.* 2nd edition. Milton: Routledge.
Seydel, R. (2016). 'EQ Cheat Sheet: How to use an instrument frequency chart.' [blog post]. *LANDR Blog.* Available at: <https://blog.landr.com/eq-cheat-sheet/>. [Accessed: 22 Feb 2021].
Shankar, S. & Cavanaugh, J. R. (2012). 'Language and Materiality in Global Capitalism.' *Annual Review of Anthropology.* 41: 355–69.
Sharp, L. (2000). 'The Commodification of the Body and Its Parts'. *Annual Review of Anthropology.* 29: 287-328.
Smith, E. (2008). 'Even heavy-metal fans complain that today's music is too loud!!!' *Wall Street Journal*. 9–26, September 2008. Available at: <https://www.wsj.com/articles/SB122228767729272339>. [Accessed: 8 Apr 2021].
Smith, J. O. (2007) [First published 1985]. 'What is a Filter?' in *Introduction to Digital Filters with Audio Applications.* Centre for Computer Research in Music Acoustics (CCRMA), Stanford, CA. Available at: <https://ccrma.stanford.edu/~jos/filters/What_Filter.html>. [Accessed: 18 Feb 2021].
Sonible. (2022). 'smart: EQ 3' [product webpage]. *Sonible.com.* Available at: <https://www.sonible.com/smarteq3/>. [Accessed: 23 Feb 2022].
SOS Team. (2012). 'SOS Audio Frequency Chart: Fundamental Frequency Ranges of Instruments.' [online resource]. December 2012. Available at: <https://www.soundonsound.com/sound-advice/sos-audio-frequency-chart>. [Accessed: 12 May 2022].
Sound on Sound. (2008). 'iZotope ANR-B'. [online article]. 17 April 2008. Available at: <https://www.soundonsound.com/news/izotope-anr-b>. [Accessed: 5 Feb 2021].
Sound on Sound. (2009). 'Q. What are filters and what do they do?'. [online article]. July 2009. Available at: <https://www.soundonsound.com/sound-advice/q-what-are-filters-and-what-do-they-do>. [Accessed: 18 Feb 2021].
Sound on Sound. (2022). 'Glossary'. [webpage]. Available at: <https://www.soundonsound.com/glossary> [Accessed 15 Jun 2022].
SoundOracle. (2018). 'Tutorial: How To Properly EQ The Modern Voice (The Sound of Mainstream Vocals) ▏SoundOracle.net'. [online video]. Uploaded 13 Aug 2018. Available at: < https://www.youtube.com/watch?v=rkljNdhjikk&t=1180s>. [Accessed: 24 Oct 2020].
SoundRadix. (2021). 'SurferEQ: The Pitch-Tracking Equalizer'. [product webpage]. Available at: < https://www.soundradix.com/products/surfer-eq/>. [Accessed: 26 Mar 2021].

SoundTheory. (2021). ‘Gullfoss – Reveal Your Sound.’ [website homepage]. Available at: < https://www.soundtheory.com/home>. [Accessed: 26 Mar 2021].
Sreedhar, S. (2007). ‘The Future of Music’. [online article]. *IEEE Spectrum.* August 2007. Available at: <https://spectrum.ieee.org/computing/software/the-future-of-music>. [Accessed: 8 Dec 2020].
Stanyek, J. & Piekut, B. (2010). ‘Deadness: technologies of the intermundane.’ *The Drama Review.* 54 (1): 14-38.
SteamPoweredRadio.com. (2020). ‘The Orban 536A Dynamic Sibilance Controller’. [Product Brochure]. Date Unknown, c. 1980. Available at: <https://www.steampoweredradio.com/pdf/orban/orban%20536%20dynamic%20sibilance%20controller/Orban%20536A%20Dynamic%20Sibilance%20Controller%20Brochure.pdf> [Accessed: 16 Dec 2020].
Steingo, G. (2018). ‘Actors and Accidents in South African Electronic Music: An Essay on Multiple Ontologies.’ *Contemporary Music Review*. 37(5-6): 554-574.
Sterne, J. & Rodgers, T. (2011). ‘The Poetics of Signal Processing.’ *Differences.* 22 (2-3): 31-53.
Sterne, J. (2003). *The Audible Past: The Cultural Origins of Sound Reproduction.* Durham, NC & London: Duke University Press.
Sterne, J. (2012). *Mp3: The Meaning of a Format.* Durham, NC: Duke University Press.
Sterne, J. (2015). ‘Compression: A Loose History’. In Parks, L. and Staroskielsi, N. [eds]. *Signal Traffic: Critical Studies of Media Infrastructures.* Urbana, Chicago and Springfield: University of Illinois Press. 31-48.
Sterne, J. (2016). ‘Analog.’ In Peters, B. [ed.] *Digital Keywords: A Vocabulary of Information Society and Culture*. Princeton, NJ: Princeton University Press. 31-44.
Sterne, J. (2019a). ‘Spectral Objects: On the Fetish Character of Music Technologies.’ In Steinrager, J. A. and Crow, R. [eds.] *Sound Objects.* Durham, NC: Duke University Press. 94-109.
Sterne, J. (2019b). ‘Plug-in | software’. Encyclopaedia Britannica. [webpage]. Available at: <https://www.britannica.com/technology/plug-in>. [Accessed: 20 Jan 2022].
Sterne, J. (2020). ‘The Software Passes the Test When the User Fails It: Constructing Digital Models of Analog Signal Processors.’ In Tcakzyk, V., Mills, M. & Hui. A. [eds.] *Testing Hearing: The Making of Modern Aurality*. Oxford & New York: Oxford University Press. 159-185.
Sterne, J. (forthcoming). ‘On Auto-Tune, Or Voice and the Politics of Frequency’. Unpublished chapter from Sterne, J. & Mills, M. [eds]. *Tuning Time: Histories of Sound and Speed.* [pre-publication copy acquired via correspondence with the author].
Stillitoe, S. (1999). ‘Recording Cher’s “Believe”’. [online article]. *Sound on Sound.* February 1999. Available at: <https://www.soundonsound.com/techniques/recording-cher-believe>. [Accessed: 21 Jan 2021].
Synchro Arts. (2013). ‘VocAlign PRO 4 AU User Manual’. Version 2.0.4. Available online at: <http://www.synchroarts.com/downloads/VocALign%20Pro%204%20AU%20Manual%20v2_0_4.pdf>. [Accessed: 5 Feb 2021].
Szikla, A. (2018). ‘View From The Bench: DE-Esser’. [online article]. 29 March 2018. *Audio Technology Magazine.* Available at: <https://www.audiotechnology.com/tutorials/view-from-the-bench-de-esser>. [Accessed: 16 Dec 2020].
Tagg, P. (1979). *Kojak: 50 Seconds of Television Music*. Ph.D. Dissertation, Gøteborg University.
Tagg, P. (2019). *Fernando the Flute: Analysis of musical meaning in an Abba mega-hit*. 4th Edition [ongoing]. New York & Huddersfield: Mass Media Music Scholars’ Press. Available at: <https://www.tagg.org/bookxtrax/Fernando/FRNongoing.pdf>. [Accessed: 25 Sep 2020].
Taussig, M. (1980). ‘Reification and the Consciousness of the Patient’. *Soc. Sci. Med.* 14B: 3-13.
Taylor, R. W. (2017). ‘Hyper-Compression In Music Production: Agency, Structure And The Myth That Louder Is Better’. *Journal on the Art of Record Production*. Issue 11, March 2017. Available at: <https://www.arpjournal.com/asarpwp/hyper-compression-in-music-production-agency-structure-and-the-myth-that-louder-is-better/>. [Accessed: 14 Dec 2020].

Taylor, R. W. (2018). ‘Understanding The Use Of Hyper-Compression In Music Production: A Systems-Based Approach To Examining Innovative Change In The Field Of Music Production’. PhD Thesis. The University of Newcastle Australia, Faculty of Science, Information and Technology. July 2018.

Terren, M. (2021). ‘Plug-in Capitalism’. [online article]. *Bellona Magazine.* November 2021. Available at: <https://bellonamag.com/plug-in-capitalism>. [Accessed: 28 Apr 2022].

Théberge, P. (1989). ‘The ‘Sound’ of Music: Technological Rationalization and the Production of Popular Music.’ *New Formations.* 8: 99-111.

Théberge, P. (1997). *Any Sound You Can Imagine: Making Music/Consuming Technology*. Hanover, NH: Wesleyan/University Press of New England.

Théberge, P. (2004). ‘The Network Studio: Historical and Technological Paths to a New Ideal in Music Making.’ *Social Studies of Science*. Special Double Issue: ‘Technology and Music’. 34 (5): 759–781.

Théberge, P. (2012). ‘The End of the World as We Know It: The Changing Role of the Studio in the Age of the Internet.’ in Frith, S. & Zagorski-Thomas, S. [eds.] *The Art of Record Production: An Introductory Reader for a New Academic Field*. Farnham, Surrey, UK: Ashgate. 77–90.

Théberge, P. (2013). ‘Visualizing Music and Sound in the Digital Domain’. Paper presented at “Music | Digitisation | Mediation: Towards Interdisciplinary Music Studies”, a conference organized by the MusDig research group. University of Oxford: 11-13 July 2013.

Théberge, P., Devine, K. And Everrett, T. (eds.) (2015) ‘Introduction.’ in *Living Stereo: Histories and Cultures of Multichannel Sound.* London and New York: Bloomsbury. 1-36.

Thiemel, M. (2001). ‘Dynamics.’ In *Grove Music Online.* [updated 2013]. Oxford University Press. Available at: <https://www.oxfordmusiconline.com/grovemusic/view/10.1093/gmo/9781561592630.001.0001/omo-9781561592630-e-0000008458>. (Subscriber only access). [Accessed: 27 Sep 2021].

Thompson, A. & Taylor, B. N. (2008 [1995]). *Guide for the Use of the International System of Units (SI). 2008 edition*. Gaithersburg, MD: National Institute of Standards and Technology.

Thompson, P. & Harding, P. (2019). ‘Collective Creativity: A ‘Service’ Model of Commercial Pop Music Production at PWL in the 1980s’. in Hepworth-Sawyer, R., Hodgson, J., Paterson, J. & Toulson, R. [eds.] *Innovation in Music: Performance, Production, Technology and Business.* New York & Abingdon, Oxon: Routledge. 143-159.

Thornton, M. (2005). ‘Better Vocal Sounds in Pro Tools’. [online article]. *Sound on Sound.* April 2005. Available at: <https://www.soundonsound.com/techniques/better-vocal-sounds-pro-tools>. [Accessed: 20 Jan 2021].

Thornton, M. (2008). ‘iZotope RX: Noise Reduction Software [Mac/PC]’. [online article]. *Sound on Sound.* Available at: <https://www.soundonsound.com/reviews/izotope-rx>. [Accessed: 25 Jan 2021].

Thornton, M. (2009). ‘Track Comping: Digidesign (Avid) Pro Tools Tips & Techniques’. [online article]. *Sound on Sound.* April 2009. Available at: <https://www.soundonsound.com/techniques/track-comping> [Accessed: 20 Jan 2021].

Tingen, P. (2007). ‘Secrets of the Mix Engineers: Chris Lord-Alge’. [online article]. *Sound on Sound.* May 2007. Available at: <https://www.soundonsound.com/techniques/secrets-mix-engineers-chris-lord-alge>. [Accessed: 18 Dec 2020].

Tingen, P. (2008). ‘Secrets of the Mix Engineers: Michael Brauer’. [online article]. *Sound on Sound.* November 2008. Available at: <https://www.soundonsound.com/techniques/secrets-mix-engineers-michael-brauer>. [Accessed: 18 Dec 2020].

Tingen, P. (2019). ‘Inside Track: Lewis Capaldi's 'Someone You Loved' - Secrets Of The Mix Engineers: TMS LDN’. [online article]. *Sound on Sound.* May 2019. Available at: <https://www.soundonsound.com/techniques/inside-track-lewis-capaldis-someone-you-loved>. [Accessed: 30 Sep 2021].

Toft, R. (2010). ‘Hits and misses: crafting a pop single for the top-40 market in the 1960s.’ *Popular Music.* 29 (2): 267–281.
Toulson, R. (2008). ‘Can We Fix It? – The Consequences Of ‘Fixing It In The Mix’ With Common Equalisation Techniques Are Scientifically Evaluated.’ *Journal on the Art of Record Production.* 3. Available online at: <https://www.arpjournal.com/asarpwp/can-we-fix-it-%E2%80%93-the-consequences-of-%E2%80%98fixing-it-in-the-mix%E2%80%99-with-common-equalisation-techniques-are-scientifically-evaluated/>. [Accessed: 19 Feb 2021].
Tremblay, M. (1957). ‘The Key Informant Technique: A Nonethnographic Application’. *American Anthropologist*. 59: 688-701.
Tsing, A. (2015). *The Mushroom at the End of the World: On the Possibility of Life in Capitalist Ruins.* Princeton & Oxford: Princeton University Press.
Turino, T. (1999). ‘Signs of Imagination, Identity, and Experience: A Peircian Semiotic Theory for Music.’ *Ethnomusicology.* 43 (2): 221-255.
Turner, A. L. (2014). ‘The history of flat design: How efficiency and minimalism turned the world flat’. [online article]. *TNW (The Next Web).* 19 Mar 2014. Available at: <https://thenextweb.com/dd/2014/03/19/history-flat-design-efficiency-minimalism-made-digital-world-flat/>. [Accessed: 24 Sep 2020].
TVHomeStudio. (2012). ‘The Art of Mixing (A Arte do Mixagem) – David Gibson’. [online video]. Uploaded 26 May 2012. Available at: <https://www.youtube.com/watch?v=TEjOdqZFvhY> [Accessed: 30 Sep 2021].
Universal Audio. (2000). ‘Model LA-2A Leveling Amplifier’. [user guide]. Available at: <https://media.uaudio.com/assetlibrary/l/a/la-2a_manual.pdf>. [Accessed: 17 Dec 2020].
Universal Audio. (2004). ‘2192 AND 2-1176 NOMINATED FOR 2004 TEC AWARDS, LA-2A INDUCTED INTO TECNOLOGY HALL OF FAME’. [press release]. 15 July 2004. Available at: <https://www.uaudio.com/press/releases/2004/0715_TEC-nominated.html> [Accessed: 11 Dec 2020].
Universal Audio. (2009). ‘Model 1176LN: Solid State Limiting Amplifier’. [user guide]. Available at: <https://media.uaudio.com/assetlibrary/1/1/1176ln_manual.pdf>. [Accessed: 17 Dec 2020].
Universal Audio. (2022). ‘About Universal Audio.’ [webpage]. *Uaudio.com.* Available at: <https://www.uaudio.com/about/>. [Accessed: 18 Feb 2022].
Välimäki, V. & Reiss, J. D. (2016). ‘All About Audio Equalization: Solutions and Frontiers’. *Applied Sciences*. 6 (29).
Vickers, E. (2010). ‘The loudness war: Background, speculation and recommendations’. *AES 129th Convention, San Francisco, CA.* 4-7.
Walden, D. (2019). ‘Pitch vs Timbre.’ In Dolan, E. I. & Rehding, A. [eds.] *The Oxford Handbook of Timbre.* Oxford & New York: Oxford University Press.
Waves. (2009). ‘Chris Lord-Alge on the CLA Classic Compressors’. [online video]. *Waves.* Available at: <https://www.waves.com/chris-lord-alge-on-the-cla-classic-compressors>. [Accessed: 11 Dec 2020].
Waves. (2018). ‘How to Avoid Ear Fatigue While Mixing’. [online article]. *Waves.com.* 15 Feburary 2018. Available at: <https://www.waves.com/how-to-avoid-ear-fatigue-while-mixing>. [Accessed: 23 Oct 2020].
Waves. (2019). ‘How to Fix Boxy Vocals’. [online article]. *Waves.com.* 25 Dec 2019. Available at: < https://www.waves.com/how-to-fix-boxy-vocals> [Accessed: 23 Sep 2020].
Waves. (2020). ‘Vocal Production and Mixing With Tony Maserati.’ [online video of previous live-streamed “open mixing session”]. Uploaded 10 June 2020. Available at: <https://www.waves.com/open-mix-session-tony-maserati>. [Accessed: 22 Mar 2021].
Waves. (2022a). ‘About Us’. [webpage]. *Waves.com.* Available at: <https://www.waves.com/about-us>. [Accessed: 18 Feb 2022].

Waves. (2022b). ‘CLA Vocals’. [product webpage]. *Waves.com.* Available at: <https://www.waves.com/plugins/cla-vocals?#one-stop-shop-mixing-vocals>. [Accessed: 15 Jun 2022].
Waves Audio. (2013). ‘Waves Event with Chris Lord-Alge – Part 3: CLA Vocal Plugin’. [online video]. Uploaded 9 Jun 2013. Available at: <https://www.youtube.com/watch?v=qo1gZgycDGQ>. [Accessed: 17 Dec 2020].
Waves Audio. (2016). ‘How to Help Vocals Stand Out AND Sit in the Mix using Two Compressors’. [online video]. Uploaded 22 Mar 2016. Available at: <https://www.youtube.com/watch?v=LotWDQWNviQ>. [Accessed: 31 May 2022].
Waves Audio. (2017). ‘Vocal Mixing Tip – Double Stacked Compression’. [online video]. Uploaded 26 Mar 2017. Available at: <https://www.youtube.com/watch?v=P8BFr-sspMU&t=92s>. [Accessed: 17 Dec 2020].
Waves Audio. (2019). ‘Vocal Compression Shootout: CLA Classics Compared’. [online video]. Uploaded 4 Oct 2019. Available at: <https://www.youtube.com/watch?v=vTs05OQzUew> [Accessed: 10 Dec 2020].
Waves Audio. (2020a). ‘Open Mix Session with Chris Lord-Alge’. [online video]. Uploaded 29 Apr 2020. Available at: <https://www.youtube.com/watch?v=CXNkqh_PXIg>. [Accessed: 11 Dec 2020].
Waves Audio. (2020b). ‘NEW: CLA Epic—Complete Reverbs & Delays by Chris Lord-Alge’. [online video]. Uploaded 30 Nov 2020. Available at: <https://www.youtube.com/watch?v=5VsTFRHO04c> [Accessed 17 Dec 2020].
Waves Ltd. (2009). ‘WAVES PRESENTS CLA CLASSIC COMPRESSORS’. [press release]. 15 July 2009. Available at: <http://web.archive.org/web/20100106082053/http:/www.waves.com/objects/pdf/press/CLA_CLASSI_COMPRESSORS_July_2009.pdf>. [Accessed: 11 Dec 2020].
Weheliye, A. G. (2002). ‘“Feenin”: Posthuman voices in contemporary black popular music’. *Social Text* 71, 20(2): 22-47.
Weidman, A. (2021). *Brought to Life by the Voice: Playback Singing and Cultural Politics in South India.* Berkeley, CA: University of California Press.
Weiss, M. (2017). ‘How to EQ Vocals (+ 10 Problems to Look Out For)’. [online article]. *Pro Audio Files.* Available at: <https://theproaudiofiles.com/vocal-eq/>. [Accessed: 15 Apr 2022].
White, P. (2005). ‘Pop Shields: Why You Need Them’. [online article]. *Sound on Sound.* May 2005. Available at: <https://www.soundonsound.com/sound-advice/pop-shields-why-you-need-them>. [Accessed: 20 Jan 2021].
White, P. (2006). ‘Vocal Recording & Production Masterclass: Techniques & Tips’ [online article]. *Sound on Sound.* June 2006. Available at: <https://www.soundonsound.com/techniques/vocal-recording-production-masterclass>. [Accessed: 24 Jan 2022].
White, P. (2009). ‘Creating A Sense Of Depth In Your Mix.’ [online article]. *Sound on Sound.* February 2009. Available at: <https://www.soundonsound.com/techniques/creating-sense-depth-your-mix>. [Accessed: 22 Feb 2021].
White, P. (2012). ‘Vocal Production: Contemporary Mixing and Processing Techniques.’ [web article]. *Sound on Sound.* February 2012. Available at: <https://www.soundonsound.com/techniques/vocal-production>. [Accessed: 15 Jun 2021].
White, P. (2018). ‘Soundtheory Gullfoss: Automated EQ Plug-in for Mac OS’. [online article]. *Sound on Sound.* April 2018. Available at: <https://www.soundonsound.com/reviews/soundtheory-gullfoss>. [Accessed: 15 Jun 2022].
Whyte, W. F. (1943). *Street Corner Society: The Social Structure of an Italian Slum*. Chicago, IL: University of Chicago Press.
Wilson, K. A. (2020). ‘nkisi: west-central African lore’. [online article]. *Encyclopedia Britannica.* 7 Oct 2020. Available at: <https://www.britannica.com/art/nkisi>. [Accessed: 24 May 2022].

Wolfe, P. (2021). *Women in the Studio: Creativity, Control and Gender in Popular Music Sound Production*. Abingdon, UK: Routledge.
Young, M. (2015). *Singing the Body Electric: The Human Voice and Sound Technology*. Farnham, Surrey, UK: Ashgate.
Zagorski-Thomas, S. (2007). 'The musicology of record production.' *Twentieth-century music.* 4 (2): 189-207.
Zagorski-Thomas, S. (2010). 'Real and Unreal Performances: The Interaction of Recording Technology and Rock Drum Kit Performance.' In Danielsen, A. [ed.] *Musical Rhythm in the Age of Digital Reproduction.* Farnham, UK: Ashgate. 195-212.
Zagorski-Thomas, S. (2014). *The Musicology of Record Production*. Cambridge & New York: Cambridge University Press.
Zak, A. (2001). *The Poetics of Rock: Cutting Tracks, Making Records.* Berkeley & Los Angeles: University of California Press.
Zorn, D. (2020). 'Mixing Vocals: Sculpting the High End.' [online article]. *Waves.com.* 1 January 2020. Available at: <https://www.waves.com/mixing-vocals-sculpting-the-high-end>. [Accessed: 15 Jun 2021].
Zuboff, S. (2019). *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. London: Profile Books.

Appendix 1: Fieldwork Sources

NOTE: All names are pseudonyms. Sources are arranged chronologically, with capital letters denoting different sources from the same individual in date order. Sources come in four main categories: informal meeting, informal telephone conversation, semi-structured interview, and observation. These categories reflect my method of acquiring field informants (starting with initial informal conversations and culminating in more prolonged periods of observation or more in-depth interviews), as well as the tenor of my conversations with specific informants (sometimes, informal meetings turned out to be more like semi-structured interviews, depending on the individual in question). Informal meetings or conversations were usually more unstructured and sometimes weren't recorded to preserve the informality of the situation, whereas semi-structured interviews were always recorded. All fieldwork became remote rather than in-person following the start of the Covid-19 pandemic in March 2020. The chronology of the list reflects my gradual acquisition of key informants working in commercial pop. Towards the end, the focus on these key informants solidifies, including repeated observation and follow-up interviews.

Each entry includes four key pieces of information: the informant(s') role(s) (sound engineer, music producer, vocalist); professional status (amateur, aspiring student, professional); genre, in square brackets (e.g. commercial, trap, country, only if applicable as some engineers in particular consider themselves all-rounders, and often including the informant's own labelling in " "); and location (redacted if felt to compromise anonymity).

YEARX. [Category of source] with [pseudonym], role [genre, if applicable]. [Location], [Precise date].

2018A. Informal meeting with Ian, sound engineer based in a UK University Music Department. October 2018.

2018B. Informal meeting with Peter, sound engineer based in a UK University Music Department. November 2018.

2018C. Informal telephone conversation with Luke, sound engineer, based at a studio in the south-east of England [Indie-pop]. November 2018.

2018D. Observation of a full-day vocal recording session with sound engineer Anthony and Vocalist Leonardo. December 2018, London, UK.

2018E. Informal interview with vocalist Sarah ["Hyper-pop", "Bubblegum pop"]. December 2018, London, UK.

2018F. Observation of a vocal recording session with Luke, sound engineer [Indie-pop]. December 2018, London, UK.

2018G. Informal telephone conversation with Joe, sound engineer and music producer based at a studio complex in the south-east of England. December 2018.

2019A. Informal telephone conversation with Elliott, music producer [Dub]. January 2019.

2019B. Informal telephone conversation with George, sound engineer based in a UK University Music Department. January 2019.

2019C. Observation of vocal mixing session with sound engineer Paul. January 2019.

2019D. Semi-structured telephone interview with sound engineer Andrew. January 2019.

2019E. Informal meeting with pop music university lecturer Dominic. January 2019.

2019F. Observation of vocal recording workshop with Martin & Liam, sound engineer and producer, and university students. January 2019.

2019G. Observation of a full-day vocal recording session with Magnus and Bella, aspiring producer and vocalist [Commercial pop, “Pop/R&B”]. February 2019.

2019H. Informal meeting with Victor and Naomi, a pop vocalist and his agent [Pop]. February 2019, London, UK.

2019I. Informal meeting with Brian, university lecturer in popular music. February 2019, London, UK.

2019J. Informal meeting with Philip, music producer and university lecturer. February 2019.

2019K. Informal meeting with Robert, sound engineer. February 2019.

2019L. Semi-structured interview with Joseph, music producer [Pop & K-Pop]. February 2019, London, UK.

2019M. Semi-structured interview with Fabio, aspiring student music producer [Indie-pop]. March 2019, London, UK.

2019N. Informal meeting with Emanuel, music producer [“Electronic pop”]. March 2019, London, UK.

2019O. Informal video call with Kieran, US-based pop music academic and amateur music producer [Commercial pop]. March 2019.

2019P. Phone interview with Graham, US-based sound engineer and music producer [Country]. March 2019.

2019Q. Observation of a vocal recording session with Hassan, Fabio (from 2019M) and Adi, two aspiring student music producers and one rap vocalist [Trap]. March 2019.

2019R. Observation of a vocal recording and songwriting session with Hassan, aspiring student music producer [Trap, “Ambient”]. April 2019.

2019S. Collaborative songwriting and vocal recording session between the author and Emily, pop vocalist [Alternative/Jazz]. Oxford, UK, August 2019.

2019T. Informal interview with Emily [Alternative/Jazz]. Oxford, UK, August 2019.

2019U. Informal meeting with Diana and John, vocal editing specialist and sound engineer [Commercial pop, Pop]. London, UK, August 2019.

2019V. Observation of a vocal recording and songwriting session with Hassan, aspiring student music producer [Trap, “Ambient”]. London, UK, September 2019.

2019W. Observation of a vocal recording session with Emily and several other unnamed university students [Alternative/Jazz]. London, UK, October 2019.

2019X. Informal meeting with Jason, Commercial vocal producer, songwriter and topliner [Commercial pop]. London, UK, October 2019.

2019Y. Observation of a songwriting and vocal recording session with Jason and vocalist Martina [Commercial pop]. London, UK, October 2019.

2019Z. Observation of a vocal mixing session with Jason [Commercial pop]. London, UK, November 2019.

2019AA. Observation of a live performance by Emily [Alternative/Jazz]. London, UK, November 2019.

2019BB. Informal meeting with Oliver, aspiring student music producer [Commercial pop, "Commercial house"]. November 2019.

2019CC. Semi-structured interview with Tim, music producer [Pop, Indie-pop and Worship]. November 2019.

2019DD. Semi-structured interview and observation with Ben, aspiring student music producer [Pop, Neo-soul]. London, UK, November 2019.

2019EE. Informal meeting with Daniel, amateur music producer and lecturer based in a UK university music department. [Electronic]. London, UK, November 2019.

2020A. Semi-structured interview and observation with Harriet, singer-songwriter and producer. [Indie/alternative pop]. Surrey, UK, February 2020.

2020B. Semi-structured video interview with vocalist/artist Will [Pop, Indie-pop, singer-songwriter]. February 2020.

2020C. Informal meeting and observation with Penny, session vocalist and producer [Pop, Country]. Surrey, UK, February 2020.

2020D. Semi-structured interview with Fred, music producer and sound engineer [Pop, Rock, 80s]. London, UK, March 2020.

2020E. Informal telephone conversation with Michael, music producer and member of songwriting team [Commercial pop, Pop]. London, UK, March 2020.

2020F. Semi-structured video interview with Chris, music producer and university teacher [Commercial pop, alternative pop]. March 2020.

2020G. Semi-structured video interview with Ralph, US-based music producer, songwriter and multi-instrumentalist. [Commercial pop, Pop, Latin, 80s]. March 2020.

2020H. Remote observation of vocal editing with Chris. April 2020.

2020I. Semi-structured video interview with Robin, music producer. [Commercial pop, electronic/house]. April 2020.

2020J. Remote observation of vocal editing with Jason, Track 1. London, UK, May 2020.

2020K. Informal telephone conversation with Ryan, vocal tuning specialist [Commercial pop, Pop]. London, UK, June 2020.

2020L. Remote observation of vocal editing with Jason, Track 2. London, UK, June 2020.

2020M. Remote observation of vocal editing with Jason, Track 3. London, UK, June 2020.

2020N. Remote observation of vocal editing with Jason, Track 4. London, UK. June 2020.

2020O. Remote observation of vocal editing with Jason, Track 5. London, UK. June 2020.

2020P. Remote observation of vocal editing with Jason, Track 6. London, UK. June 2020.

2020Q. Remote observation of vocal editing with Oliver. August 2020.

2020R. Remote observation of vocal editing with Robin. August 2020.

2020S. Semi-structured follow-up interview with Oliver. December 2020.

2020T. Semi-structured follow-up interview with Chris. December 2020.

2021A. Remote observation of vocal editing with Jason, Track 7. January 2021.

2021B. Remote observation of vocal editing with Jason, Track 8. January 2021.

2021C. Remote observation of vocal editing with Jason, Track 9. January 2021.

2021D. Semi-structured follow-up interview with Jason, January 2021.