Cortical codes for pitch and its role in selective listening

Veronica Tarka

Wolfson College, University of Oxford

A thesis submitted for the degree of Doctor of Philosophy

Hilary, 2026

Cortical codes for pitch and its role in selective listening

Veronica Tarka
Wolfson College, University of Oxford
A thesis submitted for the degree of Doctor of Philosophy
Hilary, 2026

Pitch is a fundamental feature of sound that underlies our perception of speech and music, and enables us to segregate sound sources in complex acoustic environments. Despite this essential role, it is unclear how pitch is represented in the brain. Single neurons that selectively encode pitch have been identified in marmosets, but studies in other species, including humans, have instead identified distributed pitch codes that are often specific to acoustic structure. In this thesis, I addressed these conflicting findings through electrophysiology recordings in ferret auditory cortex. I found that a subset of neurons represented pitch based on harmonic spacing, while another subset derived pitch from temporal periodicity. A third subset invariantly represented pitch across both cue classes, the first evidence of such neurons outside of marmosets. Each cue class evoked a different population code for pitch, and population responses at sound onset and offset were equally and independently informative about pitch. I found that cortical representations of the pitch of complex sounds did not systematically relate to pure tone frequency encoding, either in single neurons or population codes. In a novel two-tone streaming paradigm, I showed that ferrets can use the pitch of both pure tones and complex sounds to segregate competing auditory streams, the first demonstration of pitch as a cue for selective listening outside of humans. Pitch perception and selective listening often worsen with ageing, even when audiograms reveal only mild hearing impairment, and improved clinical interventions will depend on a better understanding of the neural mechanisms underlying these fundamental aspects of hearing. The cortical codes for pitch I have described in ferrets unites diverging findings across several species, and the novel auditory streaming paradigm I designed opens the door for future studies to use invasive techniques only possible in animals to identify the neural mechanisms underlying selective listening based on a variety of acoustic cues.

Acknowledgements

A doctorate is a long and varied journey, and I've been lucky to have Dr Kerry Walker's and Prof Andy King's guidance and support, not just for the science but also for the scientist, throughout. My time at Oxford has been a genuine joy, and I thank both of my supervisors and the entire Auditory Neuroscience Group for the mentorship and collaborations that have inspired and supported my growth over the past three years.

I do not take for granted that I am still a curious and passionate scientist at this juncture which is so liable to burn-out, and for this I largely credit Kerry's own infectious curiosity, level-headedness, and wisdom which have kept my energy focused on the truly important matters. I am also grateful to the University of Oxford, the Clarendon Fund, Wolfson College, and the Department of Physiology, Anatomy, and Genetics for their generous support which made this degree possible.

Particular gratitude is due to Ben Willmore who has been eternally generous with his time, and never fails to give invaluable guidance when I feel utterly stuck. Sebastian Klavinskis-Whiting and Lorenzo Mazzaschi, my fellow DPhil students, have been a source of sanity, perspective, and much laughter since the beginning.

My time at Oxford has been as much an education in music as neuroscience, and I thank Christian and Martin not only as my constant friends but as major pillars of my life as a musician. Vanda and Leo have proved that the line between friend and family is in fact blurry at best, and I doubtless would've failed the transatlantic transplant without their companionship. The generous love and support of Kim, Darryl, and Adam, has carried me through many low points and encouraged me to aim higher than I would've thought possible. Finally, credit is due to those 'companions' Bach, Mahler, and Ravel whose music filled many days of solitary analysis, reading, and writing.

Contents

Acknowledgements 3
Contents 4
Abbreviations 8
1. General Introduction 9
1.1. Overview 9
1.2. What is pitch? 9
1.3 How is pitch encoded throughout auditory stations? 11
1.4 How is pitch encoded at the cortical level? 14
1.5. How is pitch relevant for selective listening? 14
1.6. Aims 16
2. Single neuron codes for pitch 18
2.1. Rationale 18
2.2. Contributions 18
2.3. Introduction 18
2.4. Methods 20
A. Experimental model and study participant details 20
B. Surgical Procedure 20
C. Sound presentation 21
D. Sound stimuli 21
E. Spike sorting 23
F. Isolating F0-sensitive neurons 23
G. Assessing tuning similarity 24
2.5. Results 24
A. Harmonicity neurons derive F0 from resolved harmonics 25
B. Temporal neurons derive F0 from unresolved harmonics 28

C. Temporal neurons are sensitive to phase manipulations 31
D. Pitch neurons encode F0 using both harmonic and temporal cues 33
E. Sensitivity to pitch salience 36
F. F0-tuning of pitch-selective neurons is not explained by pure tone frequency tuning or cochlear distortion artefacts 39
G. Anatomical locations of pitch-selective neurons 43
2.6. Discussion 46
A. Dual pitch processing in auditory cortex 49
B. Pitch-selective neurons in auditory cortex 50
C. Conclusions and future directions 51
3. Population codes for pitch 52
3.1. Rationale 52
3.2. Contributions 52
3.3. Introduction 52
3.4. Methods 54
A. Animals 54
B. Surgical procedure 54
C. Sound stimuli 55
D. Spike sorting 57
E. Isolating sound-responsive neurons 57
F. Decoding F0 57
G. Principal component analysis 59
H. Single neuron information 60
3.5. Results 60
A. Population F0 codes generalize well within, but not across, classes of pitch cues 62
B. Population F0 codes are composed of non-redundant onset and offset responses 65

C. Different neural populations encode different pitch cue classes 68
3.6. Discussion 72
A. Dual pitch processing in auditory cortex 72
B. F0-encoding in single neurons 73
C. Sound offset responses 74
D. Summary 75
4. Pitch as a selective listening cue 76
4.1. Rationale 76
4.2. Contributions 76
4.3. Introduction 76
4.4. Methods 78
A. Training apparatus 78
B. Stimuli 78
C. Behavioural paradigm 81
D. Data analysis 85
4.5. Results 85
A. Pure tones 88
B. Complex tones with fixed harmonic number 90
C. Complex tones with fixed bandwidth 92
D. Complex tones with jittered harmonic frequencies 93
4.6. Discussion 95
5. General discussion 98
5.1. Overview 98
5.2. Dual-encoding mechanisms for pitch 98
5.3. Anatomical distribution of pitch-selective neurons 99
5.4. Pure tone pitch 100
5.5. Pitch and selective listening 101
Appendix 103

Spike sorting quality metrics 103
Harmonicity neurons 104
Temporal neurons 105
References 108

Abbreviations

α	significance level
A1	primary auditory cortex
AAF	anterior auditory field
ANOVA	analysis of variance
C	Celsius
CC	cross correlation
CV	coefficient of variation (σ/ μ)
corr.	correlation
CT	click train
dB	decibel
F0	fundamental frequency
FA	false alarm
fMRI	functional magnetic resonance imaging
FRA	frequency response area
HSD	honestly significant difference
Hz	hertz (cycles per second)
i.m.	intramuscular
ISI	inter-spike interval
μ	mean
MF	missing fundamental harmonic tone
PC(A)	principal component (analysis)
R	rostral field
RMSE	root mean square error
σ	standard deviation
SAM	sinusoidally amplitude modulated
SEM	standard error of the mean
s.c.	subcutaneous
spks	spikes
SPL_A	A-weighted sound pressure level

1. General Introduction

1.1. Overview

The acoustic world is busy: we are often surrounded by a variety of sounds from the hum of the fridge, to chatter in the corridor, to the siren of a passing police car. How does the brain make sense of these competing noises? How can we know which particular sounds came from our fridge or our friend when it all reaches the ear as a single, jumbled soundwave? While the basic mechanics of the auditory nervous system have been well mapped, the ways in which these mechanics are employed to listen to natural sounds in natural contexts remains a major question in the field.

Pitch is a fundamental attribute of natural sounds, defined as the tonal (i.e. 'high' or 'low') quality of a sound (ANSI-S1.1, 1994). It underlies our perception of musical melody and speech prosody, and underpins our ability to segregate competing sounds into separate sources (Assmann and Summerfield, 1987; Bird and Darwin, 1998; de Cheveigné, 1997; McPherson et al., 2022). Despite the essential role of pitch in hearing, it remains unclear how pitch is extracted from acoustic structure, represented in the brain, and utilized in selective listening.

1.2. What is pitch?

Sound is often conceptualized as a sine wave oscillating at a given frequency, but natural sounds in fact contain many frequencies, with waveforms that resemble dense squiggles more than perfect oscillations. Many sounds, such as vowels in speech and notes in music, are periodic sounds composed of multiple frequencies called *harmonics*, where each harmonic is an integer multiple of the fundamental frequency of that sound. The pitch of such a sound is perceived as a single note at the fundamental frequency (F0). This is true even when there is no physical energy at the F0 (Houtsma and Goldstein, 1972). In other words, the perception of pitch arises from the physical relationship amongst a sound's harmonics and does not correspond simply to the lowest frequency present in a sound. Many animals, like humans, perceive pitch as demonstrated by their ability to discriminate and order the F0 of sounds (Brosch et al., 2004; Klinge and Klump, 2009; Osmanski et al., 2013; Walker et al., 2019, 2009).

Pitch can be conceptualized in two domains: spectral and temporal (Figure 1.1). Because harmonics are integer multiples of F0, the frequency spectrum (i.e. Fourier transformation) of a harmonic sound shows regularly spaced peaks of energy corresponding to the harmonic frequencies (Figure 1.1a). The pitch of a sound can therefore be derived by determining the greatest common divisor of these harmonic frequencies. For example, a sound with harmonics at 100, 200, 300, and 400 Hz has a fundamental frequency of 100 Hz, as this is the greatest common divisor across these numbers. Similarly, a sound with harmonics of 500, 750, and 1250 Hz has a pitch of 250 Hz (the greatest common divisor of this set), despite the fact that one harmonic in the sequence is absent (250 Hz * 4).

When we examine a harmonic sound in the temporal domain (i.e. its waveform), we find that it is composed of a series of short, repeating segments (Figure 1.1b). A sound's *periodicity* is the temporal regularity of these segments and the rate at which they repeat. For harmonic sounds, the repetition rate corresponds to the F0 of the sound. This arises by definition from the common ratio between the harmonic frequencies, whereby each component frequency's waveform completes a single cycle at different rates, but the entire pattern of the combined frequencies' oscillations repeats after 1/F0 seconds. Pitch can be equivalently derived from harmonic spectra and temporal periodicity, and the auditory system indeed utilizes both aspects of pitch when representing F0.

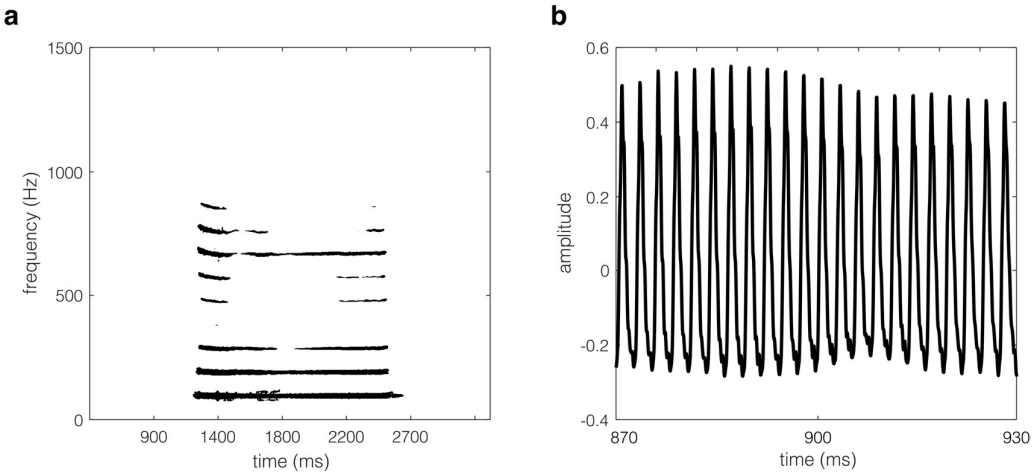

Figure 1.1. Harmonic structure and waveform of the vowel /i/. a) Harmonic structure for the vowel /i/, where black bands show frequencies with power greater than -70 dB/Hz in the sound. Pitch-evoking sounds such as vowels contain harmonic frequencies (black horizontal bands in plot) which are integer multiples of the fundamental frequency (F0). The perceived pitch of the sound corresponds to F0. **b**) The waveform for a 60 ms excerpt from the /i/ vowel. Harmonic sounds are "periodic" because their waveform is formed of short segments that repeat at regular intervals, and this interval corresponds to 1/F0.

1.3 How is pitch encoded throughout auditory stations?

The ways in which the auditory system extracts and represents a sound's harmonic structure has been examined at several auditory stations across the hierarchy. After transformation and amplification by the outer and middle ear structures, sound reaches the cochlea in the form of mechanical displacement of the fluid surrounding its interior duct, where the mechanotransducing hair cells are located. The basilar membrane that divides the fluid-filled chamber from the cochlear duct is displaced by the fluid's motion, and this displacement is maximal at a given location along the membrane for a specific, narrow frequency range. In this way, the basilar membrane produces a topographic map of sound frequency (hereafter referred to as a "tonotopic map"), which is transduced from mechanical energy into graded membrane potentials by hair cells that further refine the tonotopic gradient and subsequently synapse onto auditory nerve fibres.

Each auditory nerve fibre is most sensitive to a specific frequency, known as its characteristic frequency. The bandwidth of the fibres' frequency tuning is constant as a proportion of the characteristic frequency, meaning the range of frequencies

that fall within a fibre's receptive field is larger for fibres with higher characteristic frequencies. This necessitates two strategies for representing a sound's F0 (Figure 1.2). Lower order (i.e. low frequency) harmonics fall within receptive fields that are sufficiently narrow to spatially isolate each harmonic frequency. Such frequencies are termed *resolved harmonics* because they excite spatially separated nerve fibres and therefore produce a place-rate code for F0. Higher order (i.e. high frequency) harmonics fall within broad receptive fields that capture multiple harmonics, and are therefore termed *unresolved harmonics*. Because the combined information of spike rate and place is insufficient to determine the individual unresolved frequency components in a sound, F0 is instead encoded in the spike times of these nerve fibres, which are phase-locked to the sound's waveform.

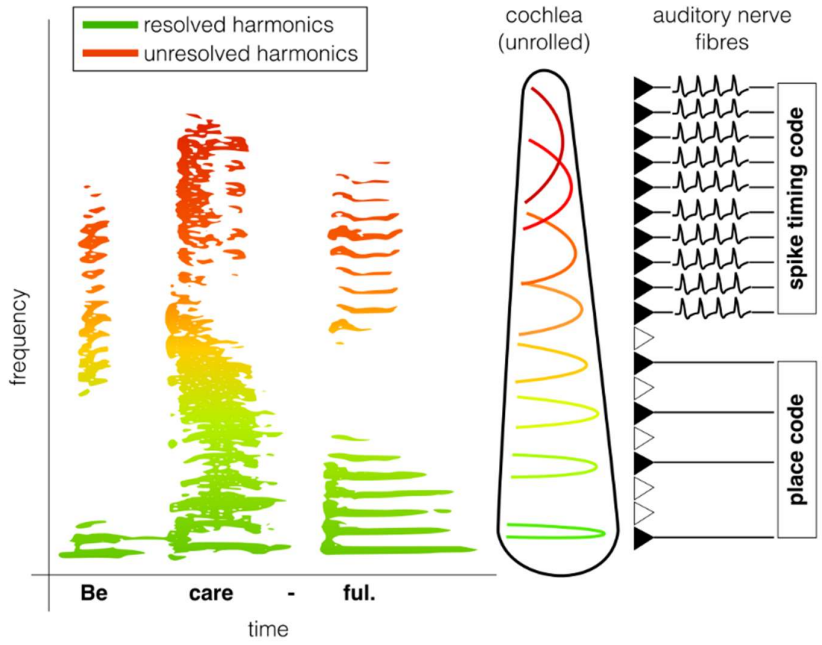

Figure 1.2. Resolved and unresolved harmonics are encoded separately in the periphery. Harmonic sounds, such as the vowels in the sentence shown, contain both low-numbered resolved harmonics (green), and high-numbered unresolved harmonics (red). Resolved harmonics fall within the narrow receptive fields of low-frequency auditory nerve fibres, producing a place code for F0 across well-separated regions along the cochlea. Because receptive fields in the auditory nerve are broader for higher frequencies, multiple unresolved harmonics may excite the same nerve fibres, making a place code for F0 uninformative. F0 is instead encoded through the spike timings of these nerve fibres, which are phase locked to F0. Adapted from Ward, Tarka et al., 2026 (Ward et al., 2026).

Auditory nerve fibres synapse onto the cochlear nucleus in the brainstem, which has a dorsal and ventral division that are each tonotopically organized. Neurons in the ventral cochlear nucleus are generally grouped into two main categories: primary-like and chopper. Primary-like neurons have response properties mostly adopted from the auditory nerve fibres that synapse onto them. They produce graded responses that align with the temporal structure of the waveform (i.e. large peaks in amplitude produce the largest firing rates). When a primary-like neuron has a characteristic frequency close to a harmonic frequency, it phase-locks to this single harmonic rather than the overall waveform (Kim and Leonard, 1988). In fact, these neurons may be able to phase-lock to even lower frequencies than auditory nerve fibres (Joris et al., 1994). Primary-like neurons can therefore encode either individual harmonics or the fine temporal structure of the sound as a whole depending on their characteristic frequency. Chopper neurons respond almost exclusively to the strongest peaks in the waveform, and the rate of these peaks corresponds by definition to the F0. The dual-encoding approach of the auditory nerve is therefore retained in the cochlear nucleus as F0 is represented as a rate-place code across the pattern of primary-like neurons across the tonotopic map, as well as a spike-timing code in chopper neurons.

The inferior colliculus is the next obligatory station for auditory afferents, and receives both ipsi- and contralateral inputs. It retains the tonotopic organization of lower auditory stations (Merzenich et al., 1975), and the acuity of this frequency map is sharpened at this stage by lateral inhibition, creating a finer representation of spectral peaks (i.e. rate-place F0 code) (Walker et al., 2011a). Neurons in the inferior colliculus show selectivity to amplitude modulation rates, which could be the earliest instance of periodicity extraction in the auditory hierarchy (Krishna and Semple, 2000; Langner et al., 2002; Langner and Schreiner, 1988; Rees and Palmer, 1988; Schreiner and Langner, 1988). A gradient organization of modulation frequency selectivity has been proposed that runs orthogonally to the tonotopic gradient, but the response characteristics of this neural population fail to align with pitch perception in key ways, such as acuity and frequency range, calling into question its relevance for extracting and encoding F0 specifically (Winter, 2005).

1.4 How is pitch encoded at the cortical level?

While subcortical auditory stations encode the acoustic features underlying pitch, lesion studies have shown that auditory cortex is critical for pitch perception (Stewart et al., 2006; Tramo et al., 2002; Whitfield, 1980; Zatorre, 1988). Pitch-selective neurons have been identified in marmoset auditory cortex which represent F0 irrespective of the harmonic structure (*timbre*) of the sound (Bendor et al., 2012; Bendor and Wang, 2010, 2005); however, a larger corpus of work across many species has instead found evidence of distributed F0 representations across auditory cortex (Allen et al., 2017; Berger et al., 2023; Bizley et al., 2010; Fishman et al., 2014; Gander et al., 2019; Griffiths et al., 2010; Norman-Haignere et al., 2019; Steinschneider et al., 1998). Neurons that are informative about F0 in these distributed representations are also generally sensitive to the sound's timbre (Bizley et al., 2009; Bizley and Walker, 2010; Walker et al., 2011b) and therefore do not invariantly represent F0. Given the diversity of findings across species, it remains unclear whether pitch is primarily encoded in single neurons or across cortical populations.

The dual-encoding strategy evident in subcortical stations appears to be preserved in auditory cortex. Neurons in macaque primary auditory cortex represent the spectral peaks of resolved harmonics as a place code, but phase-lock to F0 when there are limited resolved harmonics (Fishman et al., 2013). The integration of place code F0 representations may in fact first occur independently from that of spike time representations. Subsets of single neurons in marmosets are specialized to process the spectral F0 cues present in resolved harmonics, while other neurons represent F0 derived from the temporal periodicity cues in unresolved harmonics (Bendor et al., 2012).

1.5. How is pitch relevant for selective listening?

Pitch is a powerful acoustic cue for selective listening, which can be conceptualized as a two-stage process: feature binding and attentional enhancement. In the former, acoustic features like spatial location, harmonic structure, and temporal synchrony are extracted and grouped into perceptual objects by ascending auditory stations. In the latter, representations of target auditory objects are enhanced by top-down

attentional modulation. These stages are experimentally challenging to disentangle, but a variety of studies have indicated a key role played by pitch in feature binding and auditory attention in both humans and animal models through a range of experimental paradigms.

Harmonic regularity (i.e. the relationship of all harmonic frequencies as integer multiples of F0) is a fundamental cue for auditory feature binding. When a single harmonic frequency is adjusted upwards or downwards by even 1%, this frequency is perceived by both humans and animals as a separate auditory object from that of the remaining, spectrally regular harmonics (Figure 1.3b; Hartmann et al., 1990; Moore et al., 1986, 1985; Roberts and Brunstrom, 1998). Pitch also helps segregate overlapping sounds. Humans are better at identifying vowels presented simultaneously when there is a greater pitch separation between them (Figure 1.3a; Assmann and Summerfield, 1994, 1990; Culling and Darwin, 1993; Zwicker, 1984). When tasked with attending to speech from a female speaker against simultaneous distractor speech, people more easily track the target voice when the distractor speaker is male, specifically due to the larger pitch differences between this voice pairing (Figure 1.3c; Brokx and Nooteboom, 1982; Brungart et al., 2001).

ABA- streaming (or two-tone streaming) is one of the most widely employed paradigms for studying sound segregation and selective listening as it allows parametric control over specific acoustic features. In this design, a stream of alternating low (A) and high (B) pitch tones are presented in an ABA- pattern, where the hyphen indicates a short silent interval between tone triplets. When A and B tones are close in pitch, listeners perceive a single auditory object with a "galloping" rhythm, while large frequency differences produce a percept of two segregated A and B streams (Bregman, 1990). The majority of studies that have employed this paradigm have used pure tones, but this effect holds true for sounds with multiple harmonics (Bregman et al., 1990; Singh, 1987). Two-tone streaming has proven difficult to translate to animal models as it traditionally relies on subjective reporting of the perception of one or two streams, but frequency-dependent streaming behaviour has been demonstrated in macaques (Izumi, 2002) and ferrets (Ma et al., 2010). Critically, however, these studies only presented pure tones, leaving the question of how the brain groups and segregates complex, naturalistic sounds practically unexplored in animal models.

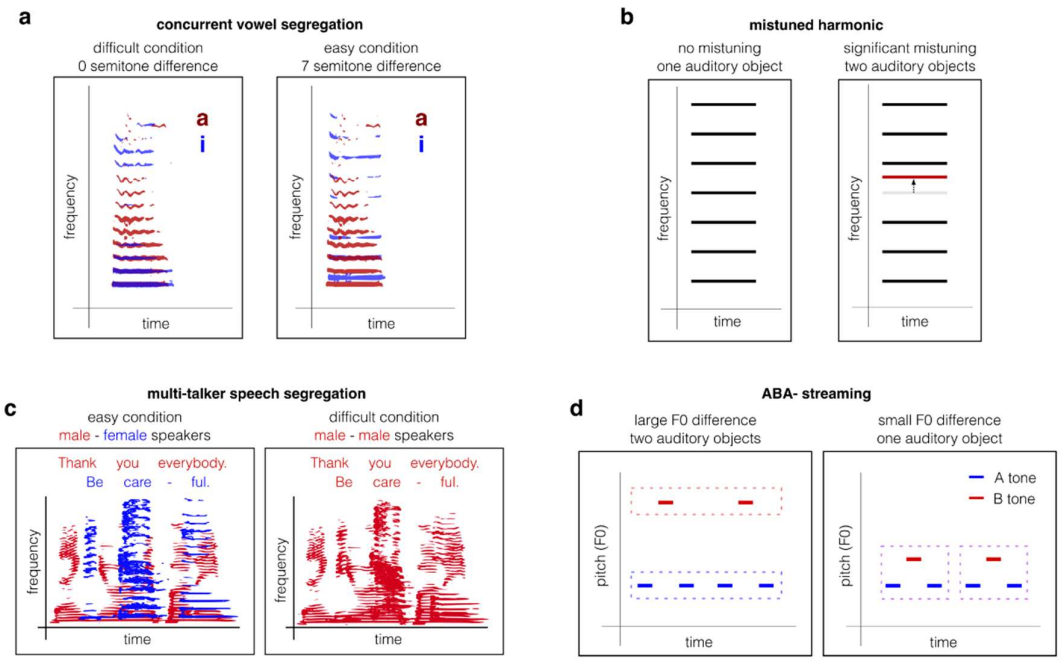

Figure 1.3. Common stimulus paradigms for studying pitch in selective listening. **a**) Concurrent vowel segregation tasks listeners with identifying two vowels presented simultaneously. This is easier with larger pitch difference between the vowels. **b**) Pitch-evoking sounds have regularly spaced harmonic frequencies. When a harmonic is adjusted (mistuned) to be lower or higher, it "pops out" as a separate auditory object. **c**) In multi-talker speech paradigms, listeners are tasked with reporting speech from a target speaker. This is easier when male and female speakers are paired together, than when two male or two female speakers are paired. **d**) ABA- streaming presents low (A) and high (B) pitched tones in an interleaved sequence, which is perceived as two independent sounds when the pitch separation is large, but as a unified object with a "galloping" rhythm when the pitch separation is small. Adapted from Ward, Tarka et al., 2026 (Ward et al., 2026).

1.6. Aims

This thesis aims to expand our understanding of the cortical representations of pitch and how pitch is utilized in selective listening. Decades of research on the auditory system have used pure tones to effectively characterize and isolate response properties of the auditory hierarchy, but the neural mechanisms underlying complex listening behaviours with naturalistic sounds remain poorly understood, particularly in animal models. Pitch offers a parameter for study that can be controlled and isolated with precision rivalling pure tone frequency, while also defining a core feature of the most salient natural sounds in our environment (e.g. speech, music). This unique position has motivated research into pitch coding for many decades, but

the conflicting evidence resulting from this corpus (see Chapter 1.4) has stalled many authors' interest in the topic. Here, I aim to address the crux of the conflicting evidence in pitch coding through electrophysiology recordings in ferrets, and I utilize a novel behavioural paradigm to explore the role of pitch in sound segregation and selective listening.

- **Chapter 2: Single neuron codes for pitch.** The first experimental chapter addresses the hypothesis that the ferret auditory cortex contains single neurons that invariantly represent F0. I analyse electrophysiology recordings of single neurons in anaesthetized ferrets while a battery of pitch-evoking sounds is presented. I find evidence for F0-selective neurons, as well as two subtypes of neurons with F0-encoding that mirrors the dual-encoding strategy of lower auditory stations.
- **Chapter 3: Population codes for pitch.** The second experimental chapter explores population codes as substrates of perception. I decode F0 from the responses of the neural population and compare F0 codes evoked by a variety of sounds that vary in acoustic structure. I find that population codes for F0 vary depending on the harmonics present in the sound, expanding evidence of a dual-encoding strategy for F0 in auditory cortex.
- **Chapter 4: Pitch as a selective listening cue.** The final experimental chapter describes a novel behavioural paradigm to measure selective listening in which ferrets are trained to discriminate between target and distractor oddballs within a two-tone streaming sequence. I find that greater pitch separations between the competing auditory streams help the animals segregate the sounds and attend to the target stream. I also find that the ferrets were more able to segregate the streams when they were composed of complex tones than pure tones, suggesting naturalistic stimuli may be critical to studying complex listening behaviours in animal models.

2. Single neuron codes for pitch

2.1. Rationale

Attempts to identify the neural mechanisms underlying cortical pitch encoding have variously reported either anatomically distributed population codes or clustered single neuron codes for F0 depending on the species and stimuli employed. While the seminal findings of Bendor and Wang, 2005 (Bendor and Wang, 2005) of single neurons that selectively encode F0 have not been replicated in non-primates, an effort to closely replicate their methodology has never been attempted. Here, we employ a battery of stimuli that includes the exact stimuli used to identify and characterize pitch-selective neurons in Bendor and Wang, 2005 (Bendor and Wang, 2005), while recording from single neurons in ferret primary auditory cortex.

2.2. Contributions

Kerry Walker and Quentin Gaucher conceptualized the experiment, designed the stimuli, performed the craniotomies, recorded the electrophysiology, and pre-processed the data including spike sorting. Veronica Tarka wrote software for all analysis after pre-processing and produced all figures.

2.3. Introduction

Sounds in nature are typically composed of many frequencies, yet we often perceive a natural sound to have a single tonal quality along a low-to-high scale, known as its pitch. Pitch perception supports recognition of musical melody, prosody in speech, and the segregation of sound sources in a complex acoustic environment, such as following a conversation in a noisy restaurant (Assmann and Summerfield, 1987; Bird and Darwin, 1998; de Cheveigné, 1997; McPherson et al., 2022). Despite its essential role in hearing, it remains unclear how pitch is extracted from acoustics and represented in the brain (Wang and Walker, 2012).

Many sensory qualities are not represented by a labeled line code in the periphery. For example, color is a percept synthesized by the visual system that is not directly dictated by a single wavelength of light. Similarly, the perceived pitch is not simply the range of frequencies spanned by a sound. In the spectral domain, when all frequency components are integer multiples (i.e. harmonics) of a common

fundamental frequency (F0), a pitch is perceived at the F0. This is true even for "missing fundamental" sounds, which contain no physical energy at F0 itself (Houtsma and Goldstein, 1972). The frequency-tuned auditory nerve fibers of the cochlea can represent this spacing of harmonics as a population place code, but only for lower "resolved" harmonics (Moore and Gockel, 2011). In the temporal domain, harmonic sounds are periodic, with a waveform repeating at 1/F0. Auditory nerve fibers tuned to higher, unresolved harmonics represent F0 in their spike timing by phase-locking to the sound envelope (Cariani and Delgutte, 1996). Psychophysical studies in humans and animals have shown that pitch can be perceived when sounds contain exclusively harmonic or temporal cues (Shackleton and Carlyon, 1994; Song et al., 2016; Sumner et al., 2018; Walker et al., 2019). Therefore, the auditory system may use dual processing to identify F0.

Neurological studies (Stewart et al., 2006; Tramo et al., 2002; Zatorre, 1988), lesion experiments in cats (Whitfield, 1980), and neuroimaging in human listeners (Warrier and Zatorre, 2004; Zatorre, 1988) point to a key role of auditory cortex in pitch processing. Microelectrode studies across a range of mammalian species have described auditory cortical neurons sensitive to F0 (Bizley et al., 2009; Fishman et al., 2013; Griffiths and Hall, 2012), but only in marmosets have specialized "pitch neurons" been identified that extract F0 invariantly across sound types, including missing fundamentals (Bendor and Wang, 2005). These neurons were clustered at the low-frequency border of primary and secondary auditory cortex, raising the possibility of a specialized pitch area in the brain.

To investigate whether invariant representations of pitch exist in non-primate auditory cortex, we recorded the F0 tuning of individual neurons in ferret primary auditory cortex while presenting sounds that varied in their harmonic structure, phase, and temporal regularity. We found a population of pitch-selective neurons in auditory cortex, with some of these neurons integrating the place-coded harmonic cues ("harmonicity neurons") and others tuned to periodicity ("temporal neurons"). A subset of "pitch" neurons encoded F0 invariantly across both cue types, reflecting the unified perception of the pitch evoked by these different sounds. Importantly, the pure tone frequency tuning of these pitch-selective neurons was not predictive of their F0-tuning, suggesting that cortical pitch representations necessarily integrate over multiple frequency components of a sound. These findings demonstrate that pitch-selective neurons are not unique to primates.

2.4. Methods

A. Experimental model and study participant details

Two female and two male adult ferrets (age: $\mu = 29.3$ weeks, $\sigma = 18.4$ weeks; female weight: 750-800 g, male weight: 1220-1800 g; *Mustela putorius furo*; Marshall BioResources, UK) were used in this study. Animals were housed with littermate of the same sex, with a normal nictemeral cycle (winter: 8h light/16h dark; summer: 14h light/10h dark) and unrestricted access to food and water. Animals were pathogen free. The animal procedures in this study were approved by the local ethical review committee of the University of Oxford and were performed under UK Home Office license.

B. Surgical Procedure

We performed terminal electrophysiological recordings on each ferret. The animals were put under general anaesthesia via an intramuscular injection of ketamine (Vetalar; 5 mg/kg) and medetomidine (Domitor; 0.02 mg/kg). Anaesthesia was maintained for the duration of the surgery and electrophysiology recording with a continuous intravenous infusion of ketamine and medetomidine in Hartmann's solution with 3.5% glucose and dexamethasone (0.5 mg/ml/hr) via intravenous cannula. The ferrets were intubated and artificially ventilated with medical O_2. Blood oxygenation, electrocardiogram, respiratory rate, and end-tidal CO_2 were continuously monitored throughout the recording session, and a body temperature of 36-38°C was maintained with a heating pad and warm air (3M Bair Hugger). To prevent corneal desiccation, eye ointment (Maxitrol; Alcon, UK) was applied throughout the procedure. Every 6 hours, the animals received atropine (Atrocare; 0.06 mg/kg i.m.), or when bradycardia or arrhythmia was observed.

Animals were placed in a custom stereotaxic frame and secured in place with a mouthpiece and ear bars. The scalp was then shaved, cleaned (ChloraPrep; 2% chlorhexidine gluconate), and injected with bupivacaine (Marcain, <1 mg/kg s.c.). The scalp was then incised, and the temporal muscles retracted. Using dental cement (SuperBond; C&B, UK) and a stainless steel bone screw (Veterinary Instrumentation, UK), a steel holding bar was attached to the skull. A craniotomy (10 mm diameter) was carried out over the right auditory cortex, and the exposed dura was removed. Anatomical location was confirmed from stereotaxic coordinates (11 mm ventral to

the midline and 8 mm anterior to Lambda) and visual identification of the ectosylvian gyrus. The surface of the brain was covered with a solution containing 1.25% agarose in 0.9% NaCl. Regularly throughout recording, silicone oil was applied to the craniotomy.

Immediately prior to recording, the ear bars were removed, and the animal with the frame was transferred to an electrically isolated anechoic chamber. An Ag/AgCl reference wire was inserted between the skull and the dura at the edge of the craniotomy. A microelectrode probe (Neuropixels Phase 3; Jun et al., 2017) was inserted through the entire depth of auditory cortex, orthogonally to the brain surface. The cortical area probed by each penetration was identified based on its location relative to the ectosylvian gyrus, the local field potential response latencies, and the frequency response area (FRA) shapes of cortical units recorded. Across the 18 cortical penetrations that yielded sound-responsive units, 10 were in low frequency A1, 3 in high frequency A1, 2 in low frequency AAF, and 3 in high frequency AAF. After a complete presentation of the stimulus set, the probe was removed and reinserted in a new location within auditory cortex. Data were acquired using the SpikeGLX software (https://github.com/billkarsh/SpikeGLX; Karsh, 2017) and custom MATLAB scripts (Mathworks) at a 30 kHz sampling rate.

C. Sound presentation

Stimuli were presented binaurally via earphones (Panasonic RP-HV094E-K) driven by a System 3 RP2.1 multiprocessor and headphone amplifier (Tucker-David Technologies). The earphones were closed-field calibrated using a GRAS amplifier and 1/8 inch condenser microphone placed at the end of a model ferret ear canal. Inverse filters were designed to ensure that the devices produced flat (less than ±3 dB) outputs across 200 Hz – 30 kHz. During the experiment, the earphones were coupled with otoscope speculae which were inserted into each ear canal and sealed in place with Otoform (Dreve Otoplastik). Sounds were presented at 48,828 Hz sampling rate.

D. Sound stimuli

The main stimulus set consisted of 13 sound types: pure tones and 12 different complex sounds, each presented over a range of 17 F0s (200-4000 Hz in 0.5 octave

increments). All sounds were generated with custom Matlab scripts, and bandpass filtered from 200 Hz – 30 kHz. Each sound was presented for either 200 ms (3 ferrets) or 300 ms (1 ferret) at 70 dB SPL_A. Each unique combination of sound type and F0 (e.g. resolved harmonics at 1000 Hz) was presented 13 times, in pseudo random order across type and F0, with an interstimulus silent interval of 600 ms (3 ferrets) or 400 ms (1 ferret), and linear onset and offset ramps of 25 ms. All frequency components were presented in phase unless otherwise stated.

There were three broad categories of acoustic manipulations in our complex sounds: temporal regularity, harmonic components, and phase. In the temporal regularity category, we presented click trains that were composed of biphasic pulses at a rate of F0. The 5 different types of "click train" stimuli were composed with different levels of temporal jitter between clicks (0, 5, 10, 20, and 40%). In the 0% jitter condition, the clicks were perfectly periodic, occurring with a spacing of 1/F0. In the jittered condition, each click from a regular click train was randomly shifted in time. The maximal amount of time shift was defined as a percentage of the inter-click interval (5, 10, 20, or 40%), centered on the timing of that click in a regular click train (Bendor and Wang, 2005; Pollack, 1968). Therefore, stimuli with an increased jitter have less temporally regular waveforms as well as less defined harmonic peaks in the spectral domain, and these are perceived by humans (Pollack, 1968) and ferrets (unpublished data from our lab) to have a weaker pitch salience.

The second category, harmonic component manipulations, comprised harmonic sounds designed to include either: 1) energy at all harmonics within our passband except at F0 (200 Hz – 30 kHz; "missing F0"); 2) only harmonics that would be resolvable on the ferret cochlea; or 3) only harmonics that are unresolved for ferrets. The resolved frequencies for each F0 were determined using a model of ferret cochlear filters based on ferret auditory nerve recordings (Sumner and Palmer, 2012; Walker et al., 2019). In addition, resolved and unresolved harmonic stimuli were presented with and without a pink noise masker ("low mask" and "high mask") to mask cochlear distortion artefacts. The masker was presented at a sound level that was 5 dB below the harmonic complex tone. In this condition, the noise masker would be 10 to 15 dB above the level of the expected individual tonal components at F0 (Pressnitzer and Patterson, 2001).

For the phase manipulations, the third category, we presented unresolved harmonic stimuli with either randomized phases across all harmonics ("rand. phase"), or with neighboring harmonics in alternating opposite phases (sine and cosine phase, "alt. phase"). These stimuli were generated by adding sine waves at the frequency of the desired harmonics for each combination of F0 and harmonic content. These stimuli did not include energy at F0.

In addition, a fuller set of pure tone stimuli was presented to assess frequency tuning in each penetration, so that we could localize each of our microelectrode penetrations on the tonotopic map. Broadband gaussian noise bursts of 100 ms were used as a search stimulus to identify sound-responsive neurons at the beginning of each electrode penetration, followed by presentation of pure tones for frequency tuning classification, and then the full set of pitch stimuli described above. Pure tones were presented at one of 15 different frequencies (500 Hz – 40 kHz in 0.45 octave increments) at 40, 50, 60, 70, or 80 dB. Tones lasted 100 ms followed by 500 ms of silence. Each unique combination of tone frequency and intensity was presented 20 times to capture the frequency response area of the electrode shank. FRAs were recorded for each penetration to identify the tonotopic gradients across AAF and A1.

E. Spike sorting

The recorded signal was processed offline by first digitally high-pass filtering at 150 Hz. Common average referencing was performed to remove noise across electrode channels (Ludwig et al., 2009). Spiking activity was then detected and clustered using Kilosort2 software (https://github.com/MouseLand/Kilosort2; Pachitariu et al., 2024, 2016). Responses from single units were manually curated using Phy (https://github.com/cortex-lab/phy; Rossant et al., 2016). Units showing stereotypical spike shapes with low variance, and a clear refractory period in their autocorrelation spike histogram were classified as single units. Only well isolated single units were included in subsequent analyses. Spike-sorting quality metrics may be found in the Appendix at the end of this thesis.

F. Isolating F0-sensitive neurons

Each neuron was first assessed for sound responsivity using a paired-sample t-test (alpha = 0.05), which compared spike rates during 150 ms of silence immediately

prior to sound onset to rates during the 150 ms after sound onset across all presentations of our sound stimuli. Only neurons found to be significantly responsive to sound were included in subsequent analyses. We assessed neural sensitivity to F0s in response to each of the 13 sound types using a one-way ANOVA (alpha = 0.05), with spike rates during the first 100 ms after sound onset as the dependent variable and F0 as the grouping variable. Our previous work has shown that auditory cortical responses in anaesthetized ferrets contain most information about sound F0 in this onset time window (Walker et al., 2011b). All ANOVA results were corrected for multiple comparisons across neurons using the False Discovery Rate. The best F0 for a given neuron was taken as the F0 eliciting the largest trial-averaged spike rate for a given stimulus type (e.g. click trains).

G. Assessing tuning similarity

We assessed the similarity of a neuron's F0-tuning across pairs of sound types using a bootstrapped sampling procedure. First, F0 tuning curves were constructed for a given neuron and for each stimulus type by averaging the spike rate in the first 100 ms after sound onset across trials. Next, we calculated the correlation of F0-tuning curves for every possible pair of sound types for that neuron. These pairwise correlations were averaged across stimulus pairs to create a single metric of how consistent the neuron's F0-tuning was across different sound types. To test whether this average correlation of F0-tuning was greater than would be expected by chance, we created a null correlation distribution from "shuffled" tuning curves, in which the neuron's average responses were randomly re-assigned across F0s. We repeated this procedure for 10,000 shuffled versions of each neuron's dataset. If the pairwise correlation of the unshuffled tuning curves was greater than the 95th percentile of the null distribution, the neuron was considered to have consistent F0-tuning across the different sound types.

2.5. Results

We examined the hypothesis that pitch is represented in individual neurons in the primary auditory cortex of ferrets. We presented anaesthetized ferrets with 13 stimulus types that varied in harmonic components, phase, and the presence of a pink noise masker. Each stimulus type was presented across 17 fundamental frequencies (1/4 octave increments across 250 – 4000 Hz). We recorded responses

from 1266 well-isolated single neurons across A1 (n=1077) and AAF (n=189) in 4 ferrets. We found that 872 (81%) of these neurons were sound responsive (paired t-test, $p<0.05$), and 165 (19%) sound-responsive neurons were also sensitive to the F0 of periodic click trains (1-way ANOVA, $p<0.05$). As click trains contain both resolved harmonics and temporal envelope cues, these "F0-sensitive neurons" could use either of these two F0 cues.

A. Harmonicity neurons derive F0 from resolved harmonics

We first aimed to test whether a subset of "harmonicity neurons" in auditory cortex may extract pitch exclusively from the resolved harmonic place code. Such neurons should show similar F0-tuning across all sounds that contain resolved harmonics, but would be either unresponsive or untuned to the F0 of sounds containing only unresolved harmonics (Figure 2.1A).

We identified putative harmonicity neurons as those that were F0-sensitive to click trains (1-way ANOVA, $p<0.05$) and tone complexes that contained only resolved harmonics (1-way ANOVA, $p<0.05$), but were not F0-sensitive to sounds containing only higher, unresolved harmonics (1-way ANOVA, $p>0.05$). These criteria identified 49 putative harmonicity neurons (Figure 2.1B and 2.2).

Note that this initial analysis identified neural responses that were modulated by the F0 of click trains and resolved harmonics, but it did not require that the neuron have similar F0-tuning curves for these 2 stimuli. A harmonicity neuron should give a reliable pitch read-out across all sounds containing resolved harmonics. Therefore, we next tested if these putative harmonicity neurons showed consistent F0-tuning across sounds with different timbres. The 3 example neurons shown in Figure 2.1 maintained their F0 tuning when we removed the energy at F0 (missing F0 sounds; Figure 2.1C). Similarly, when we introduced temporal and spectral noise to the original click trains by modestly (5%) jittering the timing of their individual clicks, these neurons were still tuned to the fundamental (CT 5% jitter; Figure 2.1C). This timbre-invariant F0 tuning is visualized as vertical bands of high spike rates at similar F0s across all 4 resolved harmonic stimuli in Figure 2.1C, although there is some variation in the width and peak of F0-tuning.

We used a bootstrapping approach to test if the timbre-invariant F0 tuning of these neurons was statistically significant. For each putative harmonicity neuron, we

calculated the average correlation between F0-tuning curves of all possible pairs of the 4 resolved harmonic sounds. We created a null distribution of these average correlations from F0-randomized versions of the same neuron's responses. If the true F0-tuning curve correlation was larger than the 95th percentile of the null correlation distribution, we concluded that the neuron showed significant generalization of pitch tuning across resolved harmonic stimuli (Figure 2.1D). Of the 49 putative harmonicity neurons tested, 55% (n=27) showed significantly correlated F0-tuning across the resolved harmonic stimuli ($p<0.05$), and these were classified as harmonicity neurons.

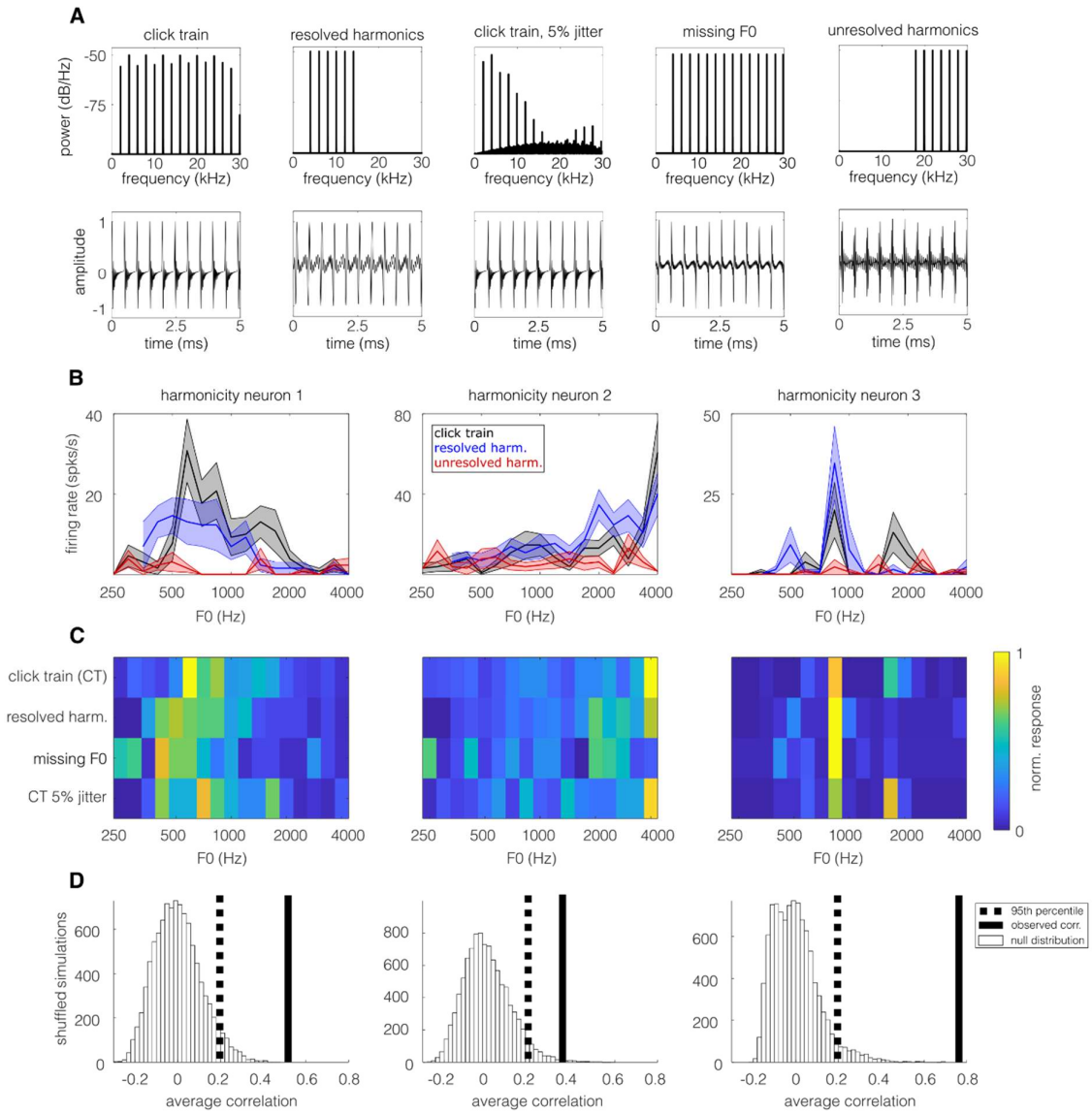

Figure 2.1. Harmonicity neurons derive F0 from resolved harmonics. A) Spectra (top row) and waveforms (bottom row) of the 5 different sound types used to classify harmonicity neurons. Each of these examples has an F0 of 2000 Hz. **B**) Each of the 3 panels shows the trial-averaged spike rate response of an example harmonicity neuron, as a function of F0. Shaded areas show the standard error of the mean (SEM). Harmonicity neurons derive F0 from resolved harmonics and therefore are sensitive to the F0 of click trains (black line) and resolved harmonics (blue line), but not to unresolved harmonics (red line). **C**) The color scale in each plot indicates the normalized trial-averaged spiking responses to 4 types of stimuli that contain resolved harmonics. **D**) In each plot, the solid black line indicates the average correlation between F0 tuning curves across all possible pairs of stimuli shown in (**C**). The histogram (open bars) shows a null distribution of the average F0-tuning correlation calculated when the tuning curves in (**C**) were shuffled randomly across F0. The dashed black line shows the 95th percentile of this null distribution. Results for the same 3 example harmonicity neurons are shown in (**B**), (**C**), and (**D**).

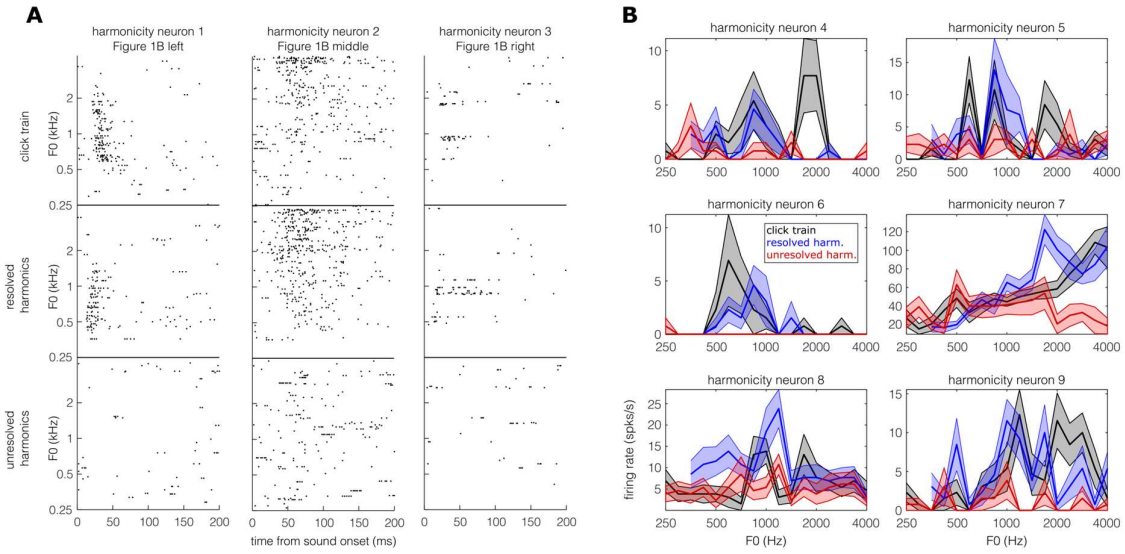

Figure 2.2. Raster plots of Figure 2.1 harmonicity neurons and additional examples.
A) Raster plots for the 3 example harmonicity neurons depicted in Figure 2.1, shown for click trains (top), resolved harmonics (middle), and unresolved harmonics (bottom) in the first 200 ms after sound onset. Each black dot represents a spike. Responses are plotted according to the F0 of the acoustic stimulus. **B**) F0-tuning for 6 additional harmonicity neurons, shown as described in Figure 2.1B.

B. Temporal neurons derive F0 from unresolved harmonics

We next examined whether a separate group of "temporal neurons" in auditory cortex could extract F0 exclusively from unresolved harmonics, possibly by integrating the spike timings of inputs that are phase-locked to the sound's temporal periodicity. Unlike harmonicity neurons, temporal neurons would show similar F0-tuning across all sounds that contain unresolved harmonics, regardless of whether resolved harmonics were also present. In principle, temporal periodicity could also be derived from the phase-locked activity of neurons responding to resolved harmonics, so temporal neurons may be either unresponsive, untuned, or tuned to the F0 of sounds containing only resolved harmonics.

Putative temporal neurons were identified as those that were sensitive to the F0 of click trains containing both resolved and unresolved harmonics (1-way ANOVA, $p<0.05$), as well as tone complexes containing only high, unresolved harmonics (1-way ANOVA, $p<0.05$; Figure 2.3A). In total, 41 neurons showed this temporal-based F0 sensitivity, with examples shown in Figure 2.3B and 2.4.

As with harmonicity neurons, these putative temporal neurons should provide a reliable read-out of F0 across stimuli with different spectra in order to support pitch

perception. We tested if these putative temporal neurons showed consistent F0-tuning across different sounds that contained unresolved harmonics (Figure 2.3A). These comprised the click trains and unresolved harmonic stimuli already described, as well as missing fundamental sounds and click trains with 5% jitter. A visual inspection of the 3 putative temporal neurons in Figure 2.3C showed a striking similarity in F0-tuning across these 4 sound types. We applied the same bootstrapping procedure described above for harmonicity neurons, and found that over 85% (35/41) of these putative temporal neurons showed statistically similar F0-tuning across the 4 stimuli ($p<0.05$; Figure 2.3D), and were classified henceforth as temporal neurons (n=35).

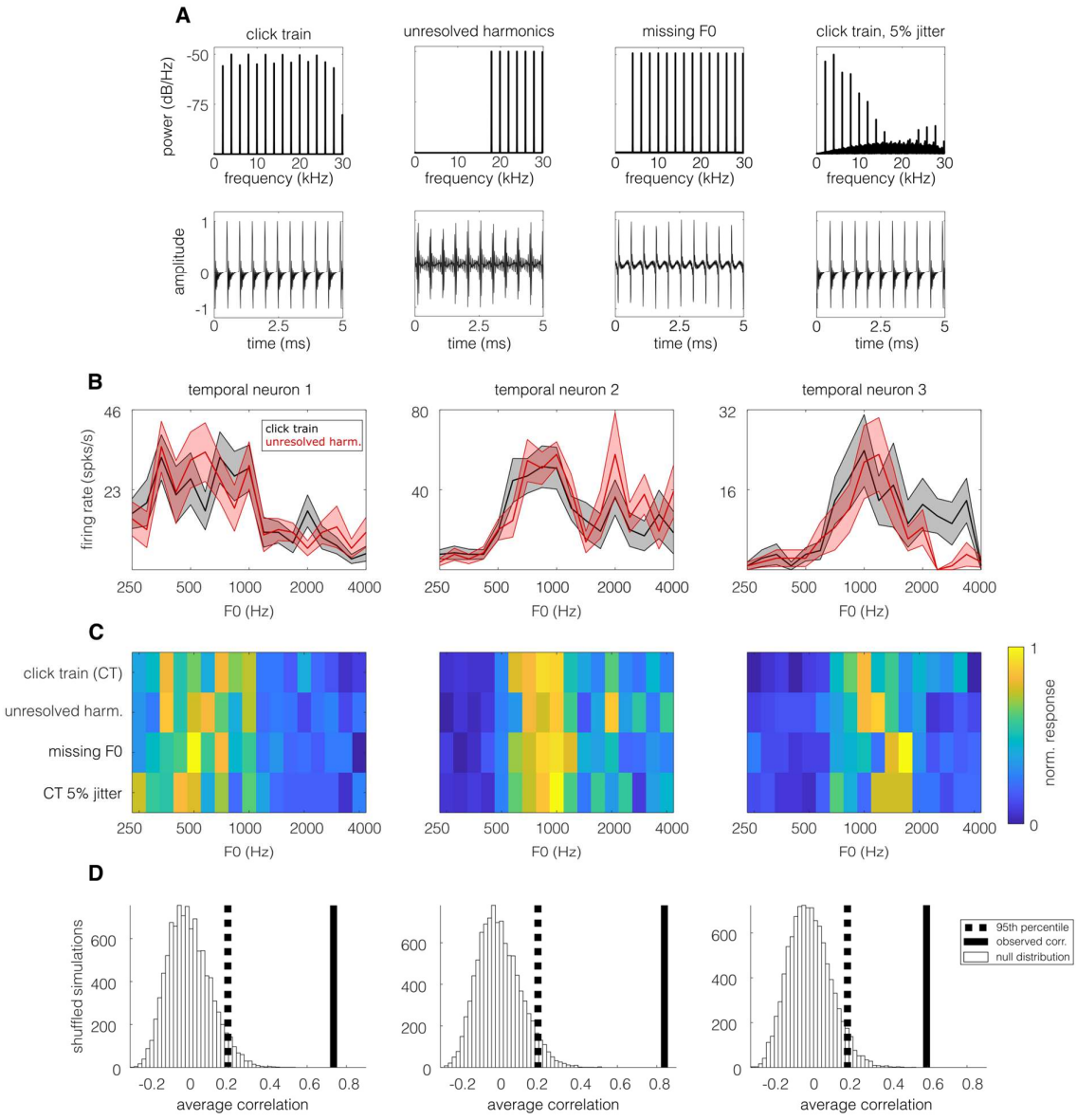

Figure 2.3. Temporal neurons derive F0 from unresolved harmonics. A) Spectra (top row) and waveforms (bottom row) of 4 different sound types used to classify temporal neurons (F0=2000 Hz). **B**) Each panel shows the trial-averaged spike rate response for an example temporal neuron as a function of F0. Shaded area shows the SEM. Temporal neurons show F0 sensitivity for click trains (black line) and unresolved harmonics (red line). **C**) The normalized trial-averaged responses to 4 types of stimuli that contain unresolved harmonics, plotted as in Figure 2.1C. **D**) Results of bootstrapped correlation test of F0-tuning similarity across stimuli, plotted as in Figure 2.1D. The same 3 example temporal neurons are shown in (**B**), (**C**), and (**D**).

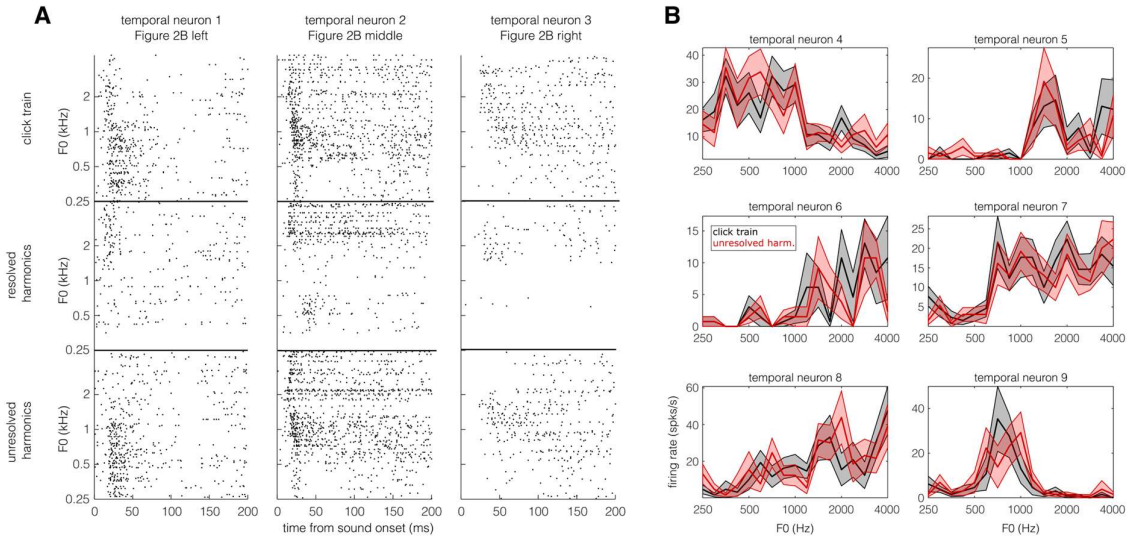

Figure 2.4. Raster plots of Figure 2.3 temporal neurons and additional examples. A) Raster plots for the 3 example temporal neurons depicted in Figure 2.3, shown for click trains (top), resolved harmonics (middle), and unresolved harmonics (bottom) in the first 200 ms after sound onset. Each black dot represents a spike. Responses are plotted according to the F0 of the acoustic stimulus. **B**) F0-tuning for 6 additional temporal neurons, shown as described in Figure 2.3B.

C. Temporal neurons are sensitive to phase manipulations

If temporal neurons primarily derive F0 from the sound's periodicity, they should be sensitive to manipulations of the sound's temporal fine structure. In the stimuli considered thus far, all harmonics were presented in phase. Here, we compare the responses of temporal neurons to 2 versions of the unresolved harmonic stimuli in which the temporal cues were disrupted through phase manipulations. In the alternating phase ("alt. phase") condition, the starting phase of consecutive harmonics alternated between sine and cosine phase. This creates a temporal periodicity at F0 and a second, weaker periodicity at 2F0 (Figure 2.5A, middle). We predicted that the tuning curves of temporal neurons might be shifted to a lower F0 in response to the alternating phase stimuli compared to the same sound presented in phase. In the randomized phase condition ("rand. phase"), the starting phase of each harmonic was individually randomized, in order to diminish the temporal periodicity at F0 and create a flatter temporal envelope (Figure 2.5A, right). We predicted that temporal neurons would show poor F0-tuning in response to the randomized phase stimulus compared to the same harmonics presented in phase. It is important to note that both these manipulations left the harmonic content of the

stimuli completely unaltered, as shown in the sound spectra (Figure 2.5A), so any changes in neuronal response would be due entirely to temporal sensitivity.

An example of a phase-sensitive temporal neuron is shown in Figure 2.5B. This neuron was tuned to a lower F0 for the alternating phase stimulus (orange line) compared to the in-phase version of this sound (black line), reflecting the increased periodicity of the alternating phase version. Furthermore, this neuron showed weaker responses to sounds with randomized phase, even when they were presented at the preferred F0 (green line). Therefore, this example neuron showed the phase sensitivity predicted for temporal neurons. Not all temporal neurons showed this level of phase sensitivity. A second example neuron shown in Figure 2.5C is one of the most phase-invariant temporal neurons in our dataset, with similar F0-tuning across all three unresolved harmonic stimuli.

Phase sensitivity was found to be a common, but not universal, feature of temporal neurons. Overall, they were consistently and significantly tuned to lower F0s in response to alternating phase stimuli compared to the same harmonics presented in phase (one-tailed paired t-test; $t(34)=5.9$, $p=6.2\times10^{-7}$; Figure 2.5D). Most temporal neurons were tuned to repetition rates between F0 and 2F0 (Figure 2.5D), likely because alternating phase stimuli provide conflicting temporal periodicity cues at these two rates. The phase manipulations also elicited weaker spiking responses compared to same harmonics presented in phase (1-way ANOVA, $F(2)=11.04$, $p=4.57\times10^{-5}$; Figure 2.5E). Therefore, the population of temporal neurons was sensitive to phase cues, even across sounds with identical spectral content.

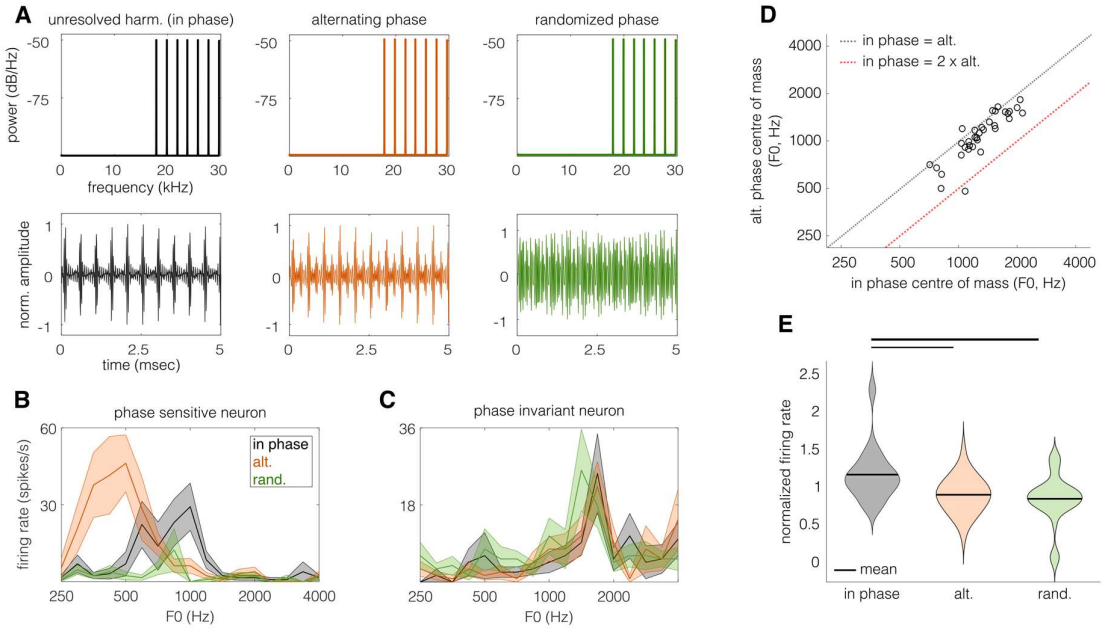

Figure 2.5. Temporal neurons are sensitive to phase manipulations. A) Spectra (top) and temporal waveforms (bottom) of stimuli in the 3 phase conditions are shown. The left panels show a tone complex containing high, unresolved harmonics, with each harmonic component presented in cosine phase (black). This stimulus has a strong temporal periodicity at F0. The middle panels show the same unresolved harmonic complex presented with "alternating phase" (orange). The right panels show the unresolved harmonic stimulus, with each harmonic component starting in randomized phase (green). **B**) F0-tuning of a phase-sensitive temporal neuron. The firing rate during each of the 3 unresolved harmonic stimuli is shown as a function of F0. A 2-way ANOVA showed main effects of phase manipulation or F0 x phase interactions in 23/35 (66%) of temporal neurons and 21/31 (68%) pitch neurons ($p<0.05$). **C**) F0-tuning of a phase-invariant neuron, plotted as in (**B**). **D**) The center of mass of a neuron's F0-tuning curve was defined as the weighted average of the trial-averaged responses elicited across all F0s presented. The scatterplot shows that most temporal neurons had a lower center of mass for alternating phase stimuli compared to the same stimuli presented in phase (paired t-test; $p<0.05$). Dotted lines show equal (black) or doubled (red) centers of mass for in phase versus alternating phase stimuli. **E**) Each temporal neuron's response to a given phase manipulation was taken as the average of its response at the best F0 (± 0.25 octaves) of the in phase F0-tuning curve. This response was then normalized across all phase manipulations. Neural response magnitude was smaller for the alternating and random phase stimuli than the in-phase stimuli (1-way ANOVA with Tukey-Kramer HSD test; thin line: $p<0.01$, thick line: $p<0.001$).

D. Pitch neurons encode F0 using both harmonic and temporal cues

Having identified harmonicity neurons that derive F0 from resolved harmonics and temporal neurons that derive F0 from the periodicity of unresolved harmonics, we next investigated whether there may exist more general "pitch neurons" that can derive F0 from both these acoustic cues. We defined putative pitch neurons as those

that were F0-sensitive for click trains, resolved harmonics, and unresolved harmonics (Figure 2.6A), combining the inclusion criteria of harmonicity and temporal neurons. We found 32 neurons that passed these criteria, and three examples are shown in Figure 2.6B (see also Figure 2.7). To determine if these putative pitch neurons could generalize their F0-tuning across a wide range of sounds, we used the bootstrapping approach described above to test if F0-tuning was similar across 5 stimulus timbres: click trains, resolved harmonics, unresolved harmonics, missing F0 sounds, and click trains with modest (5%) jitter (Figure 2.6A). While these 5 stimuli have widely different frequency spectra and temporal fine structure, they should all evoke a percept of pitch in both humans and ferrets (Walker et al., 2019).

Our bootstrapped correlation analysis showed that 31 of the 32 putative pitch neurons had statistically similar F0-tuning across all 5 stimuli ($p<0.05$; Figure 2.6D), even when we include sounds with non-overlapping frequency spectral (namely, resolved and unresolved harmonic stimuli). This evidence extends the existence of acoustically invariant “pitch neurons” beyond marmosets (Bendor and Wang, 2005), revealing a conserved cortical mechanism for pitch coding.

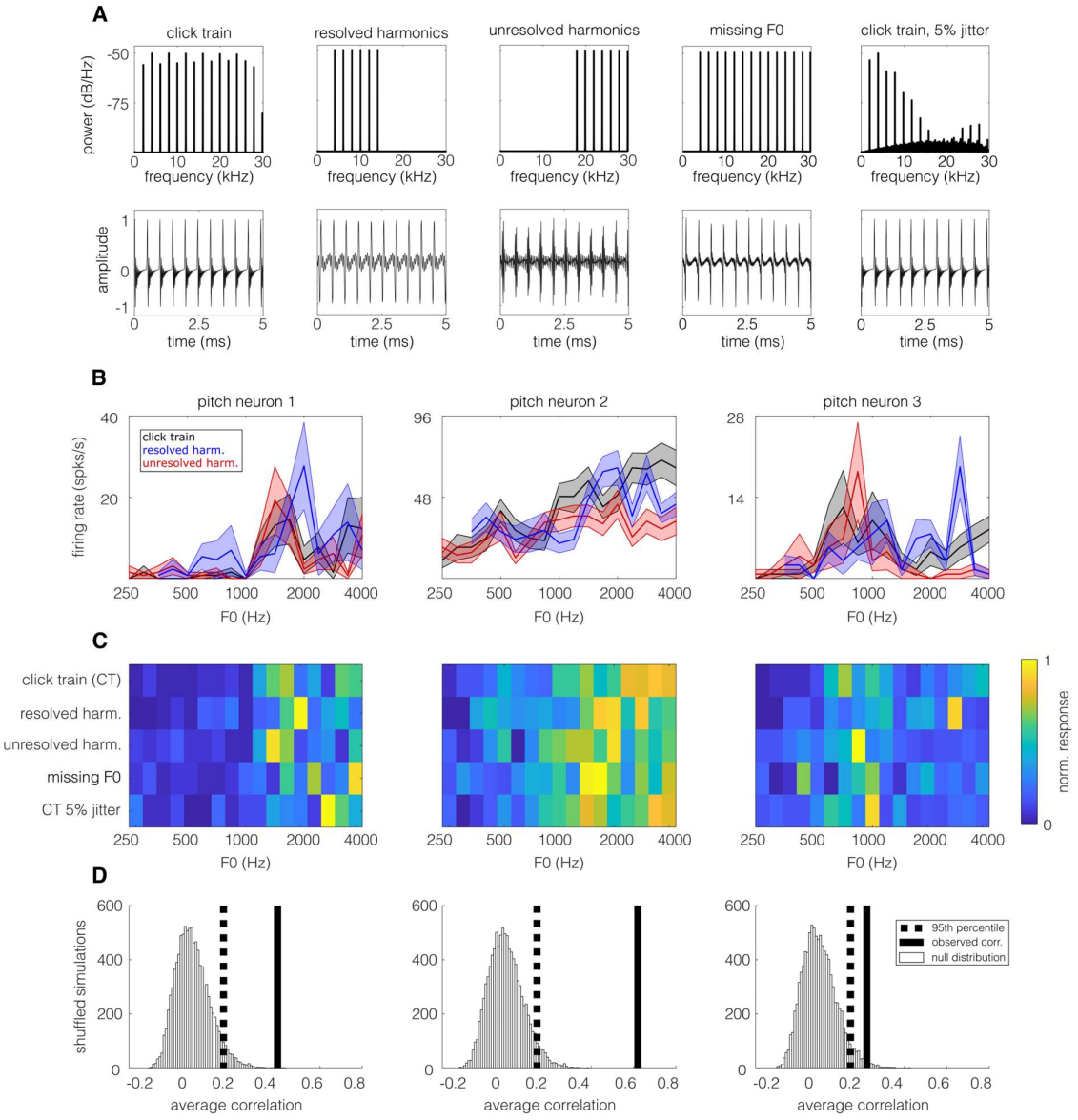

Figure 2.6. Pitch neurons can derive F0 from resolved and unresolved harmonics. A) Spectra (top row) and waveforms (bottom row) of 5 different sound types used to classify pitch neurons (F0=2000 Hz). **B**) Trial-averaged spike rate responses as a function of F0 for three example pitch neurons. Pitch neurons can derive F0 from resolved or unresolved harmonics and therefore have similar F0-tuning to click trains (black line), resolved harmonics (blue line), and unresolved harmonics (red line). **C**) Normalized trial-averaged responses to 5 types of stimuli that contain resolved or unresolved harmonics, or both, plotted as in Figure 2.1C. **D**) Results of bootstrapped correlation test of F0-tuning similarity across stimuli, plotted as in Figure 2.1D. The same 3 example temporal neurons are shown in (**B**), (**C**), and (**D**).

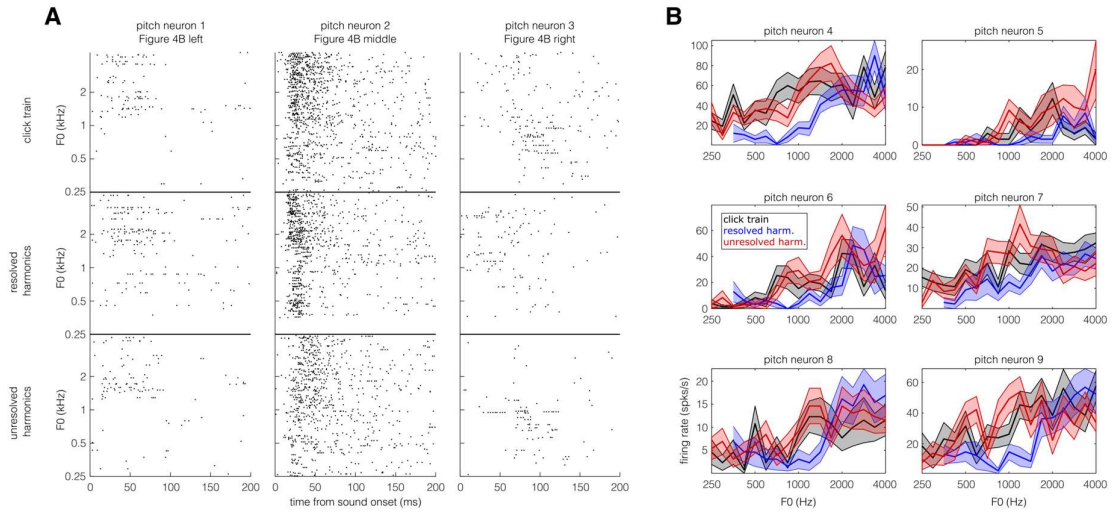

Figure 2.7. Raster plots of Figure 4 pitch neurons and additional examples. A) Raster plots for the 3 example pitch neurons depicted in Figure 2.6, shown for click trains (top), resolved harmonics (middle), and unresolved harmonics (bottom) in the first 200 ms after sound onset. Each black dot represents a spike. Responses are plotted according to the F0 of the acoustic stimulus. **B**) F0-tuning for 6 additional pitch neurons, shown as described in Figure 2.6B.

E. Sensitivity to pitch salience

If harmonicity and temporal neurons contribute to ferrets' perception of pitch, we might expect their response magnitude to decrease as pitch salience decreases. In human listeners (Pollack, 1968; Yost et al., 2005), pitch salience decreases as the temporal regularity and harmonic spectrum of a click train is degraded by randomly "jittering" the timing of individual clicks. Pitch neurons in the marmoset are sensitive to this jittering of click trains (Bendor and Wang, 2005). We investigated if pitch-selective neurons in ferrets are modulated by pitch salience, by presenting click trains with 5 levels of temporal jitter (0%, 5%, 10%, 20%, and 40% jitter; Figure 2.8A).

An example jitter-sensitive harmonicity neuron is shown in Figure 2.8B. Its spike rate at best F0 decreased as jitter increased. For comparison, Figure 2.8C shows a temporal neuron which was not sensitive to jitter; its tuning curve and response magnitude were consistent across jitter levels. Next, we examined if the populations of harmonicity and temporal neurons were jitter-sensitive overall. The spike rate response of each neuron was calculated for click trains presented within a half octave of best F0, and for each of the 5 levels of jitter. Responses of each neuron were normalized by the average response across all jitters presented to remove variance in

overall firing levels across neurons. We found that harmonicity (1-way ANOVA; $F(4)=3.77$, $p=0.0062$; Figure 2.8D), temporal ($F(4)=16.1$, $p=3.18\times10^{-11}$; Figure 2.8E), and pitch ($F(4)=13.6$, $p=1.8\times10^{-9}$; Figure 2.8F) neurons all showed weaker spiking responses as the level of jitter increased. This trend was evident in most individual pitch-selective neurons (Figure 2.9). Therefore, all three classes of pitch-selective neurons encoded temporal regularity in a manner consistent with a neural representation of pitch salience.

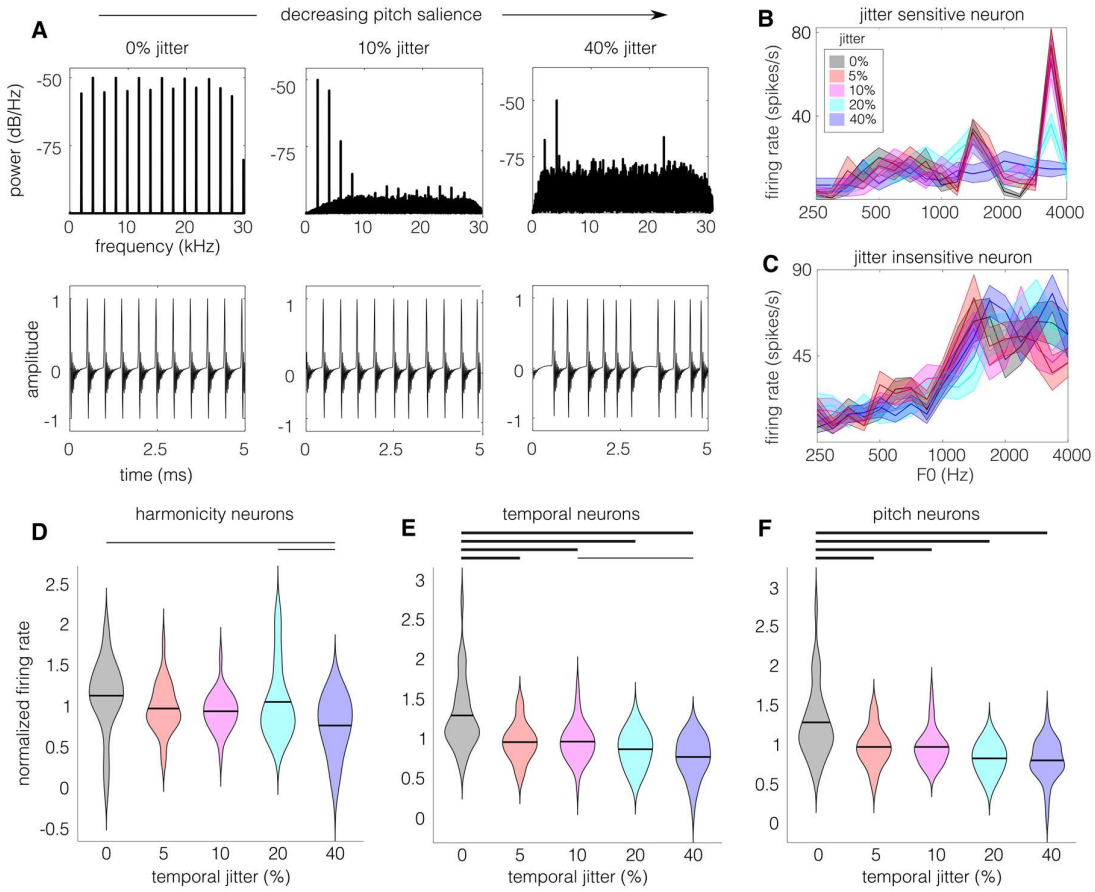

Figure 2.8. Temporal, harmonicity, and pitch neurons are sensitive to temporal regularity. A) Spectra (top row) and waveform (bottom row) of click trains with increasing temporal jitter between clicks (left to right). Jittering the temporal regularity of the clicks in the click train introduces noise in the harmonic spectrum (top row) and periodicity (bottom row), reducing the salience of the perceived pitch in human listeners (Pollack, 1968; Yost et al., 2005). **B**) F0-tuning of an example harmonicity neuron in response to click train stimuli with 0, 5, 10, 20, or 40% temporal jitter. The tuning curve is consistent for low percentages of jitter, but becomes flat for jitter above 20%. **C**) F0-tuning for an example temporal neuron, plotted as in (**B**). The F0-tuning of this neuron is not sensitive to the jitter manipulation. **D**) For each harmonicity neuron, the normalized response at best F0 (± 0.25 octaves) was calculated for each jitter level. The responses of neurons decreased as jitter level increased (1-way ANOVA with Tukey-Kramer HSD test; thin line: $p<0.05$, thick line: $p<0.001$). **E**) Jitter sensitivity at best F0 for temporal neurons, plotted as in (**D**). **F**) Jitter sensitivity at best F0 for pitch neurons, plotted as in (**D**).

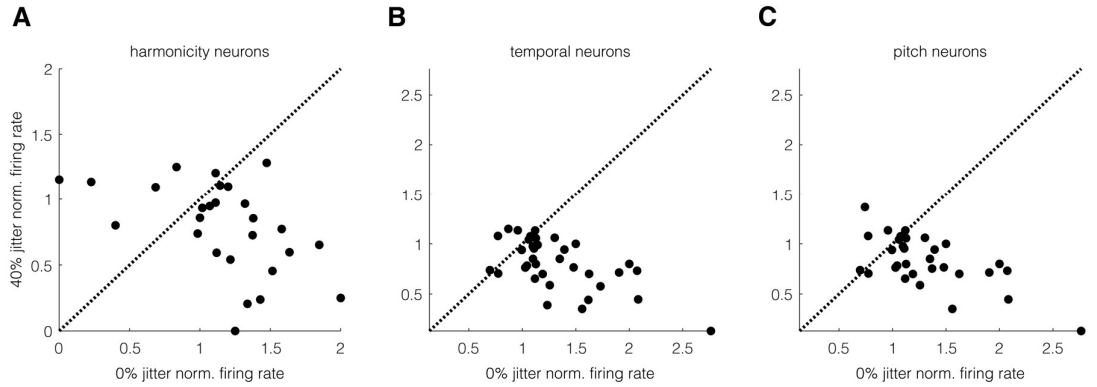

Figure 2.9. Firing rates are higher for regular than jittered click trains for most harmonicity, temporal, and pitch neurons. A-C) Normalized firing rates for all harmonicity neurons (**A**), temporal neurons (**B**), and pitch neurons (**C**) at 40% jitter compared to no jitter. Responses were normalized as described in Figure 2.8D. The jitter sensitivity of each neuron was assessed with a 2-way ANOVA on spike rate responses to click trains, with F0 and jitter condition (0, 5, 10, 20 and 40%) as independent variables. A neuron was considered jitter sensitive if it showed a main effect of jitter condition or an F0 x jitter interaction. We found that 15/27 (56%) harmonicity, 21/35 (60%) temporal, and 19/31 (61%) pitch neurons were significantly jitter sensitive ($p < 0.05$).

F. F0-tuning of pitch-selective neurons is not explained by pure tone frequency tuning or cochlear distortion artefacts

The above analyses have shown that a subpopulation of auditory cortical neurons can derive the fundamental frequency that we hear as pitch from resolved harmonics (harmonicity neurons), temporal periodicity (temporal neurons), or both these types of acoustical cues (pitch neurons). While pure tones elicit a clear percept of pitch in humans and ferrets (Walker et al., 2009), they do not contain harmonics from which to derive harmonic relations, and they only provide temporal periodicity cues in one frequency channel. Thus, they would not provide strong inputs for the type of harmonic and temporal neurons we have proposed here, which integrate F0 information across frequency channels. Previous studies have shown a poor correspondence between pure tone frequency tuning and preferences for the F0 of artificial vowel sounds in auditory cortex (Bizley et al., 2010), and pure tones are poor predictors of cortical responses to complex sounds more generally (Laudanski et al., 2012). Therefore, we hypothesize that pure tone responses may be a poor predictor of F0-tuning for complex sounds in harmonic, temporal, and pitch neurons.

We tested this hypothesis by comparing the F0-tuning of harmonicity, temporal, and pitch neurons measured in response to click trains (containing resolved harmonic and temporal periodicity cues), low harmonics (resolved harmonics only), high

harmonics (temporal periodicity only), and pure tones (energy at F0 only). Figure 2.10A shows the tuning curves of 3 example pitch neurons. The first example neuron shows considerable overlap in its pure tone frequency preference (double-peaked tuning for 1414 and 2828 Hz) and best F0 for complex sounds (F0 preference for one of these two frequencies for each complex sound). However, the other two example neurons display consistent F0 preferences across the 3 complex sounds, but non-overlapping pure tone frequency tuning. In fact, less than half of pitch-selective neurons were even responsive to pure tone stimuli (paired t-test, $p<0.05$; 70% harmonicity neurons, 40% temporal neurons, 39% pitch neurons).

Many pitch neurons showed multipeaked F0 tuning rather than a response to one single F0. Therefore, we computed correlations between F0 tuning curves to capture their shape across the full F0 range, rather than at only one best F0. Strong positive correlations indicate similar F0 preferences across stimuli, values near zero indicate unrelated tuning, and negative correlations indicate opposing preferences. The scatterplot in Figure 2.10B shows that correlations between complex stimuli cluster strongly in the positive range (top right quadrant), whereas correlations with pure tone tuning are centered near zero (Figure 2.10C). This finding is in line with our definition of pitch neurons, which required them to be tuned to both resolved and unresolved harmonics. These trends are also evident in the marginal distributions, which were positively skewed for click trains, but centered on zero correlation for pure tones.

When we examined only the peak of the F0 tuning curve, we found that the best F0 for click trains did not systematically relate to pure tone best frequency in the majority of harmonicity, temporal, or pitch neurons (Figure 2.11). Harmonicity neurons showed more similar frequency and F0 tuning than temporal and pitch neurons, but only 7 harmonicity neurons had the same best F0 and best frequency. Overall, these results suggest that the complex F0 tuning of our pitch-selective neurons differs from the mechanisms used to encode pure tone frequency.

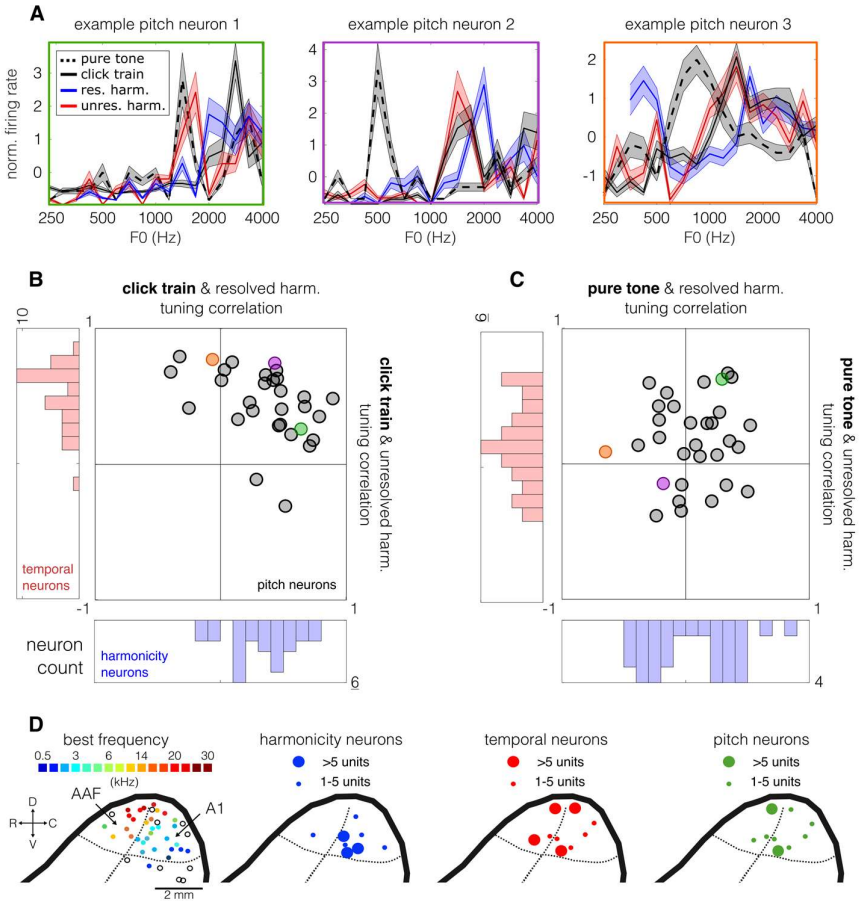

Figure 2.10. Pitch neurons′ F0-tuning for complex sounds does not correspond to pure tone frequency tuning. A) Pure tone frequency tuning and F0 tuning curves are shown for three 3 pitch neurons. Normalized trial-averaged spike rates (lines) are shown with SEM (shaded regions). The green, purple, and orange borders around these examples indicate their identity in (**B**) and (**C**). **B**) Axes indicate the correlation between F0-tuning measured with click trains and resolved harmonics (x-axis) or unresolved harmonics (y-axis). F0-tuning to click trains is positively correlated with F0-tuning to resolved harmonics in nearly all harmonicity neurons (blue histogram), and it is positively correlated with F0-tuning to unresolved harmonics in nearly all temporal neurons (red histogram). F0-tuning to click trains is positively correlated with F0-tuning for both low and unresolved harmonics in most pitch neurons (scatter plot; each dot is one pitch neuron). **C**) Results plotted as in (**B**), but examining correlations with pure tone frequency tuning, rather than click train F0-tuning. The distribution of correlations between frequency tuning and F0-tuning to resolved harmonics is centered around zero in harmonicity neurons (blue histogram). Similarly, temporal neurons′ frequency tuning is not correlated with F0-tuning to unresolved harmonics (red histogram). In pitch neurons, pure tone frequency tuning does not correlate systematically with F0-tuning to resolved or unresolved harmonics (scatter plot). **D**) Maps of ferret primary auditory cortex showing recording sites. Far left: each circle represents the site for each recording penetration in AAF (left) and A1 (right), combined across all 4 ferrets. The colour indicates the average best frequency of neurons recorded in the penetration, with white circles indicating regions with no significant frequency tuning (ANOVA, $p>0.05$). The 3 maps further right indicate the locations of harmonicity (blue), temporal (red) and pitch (green) neurons.

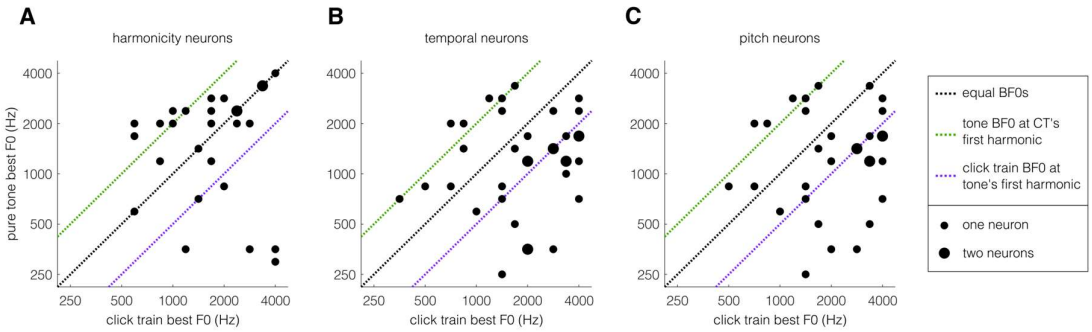

Figure 2.11. Click train F0-tuning does not correspond to pure tone frequency tuning. A-C) Pure tone best frequency versus click train best F0 for harmonicity neurons (**A**), temporal neurons (**B**), and pitch neurons (**C**). The dotted lines indicate: equal best F0s (black); pure tone best frequency equal to the first harmonic of the click train best F0 (green); and click train best F0 equal to the first harmonic of the pure tone best frequency (purple).

When a missing fundamental sound is presented to the ear, the active mechanism of outer hair cell amplification causes displacement of the basilar membrane at the position corresponding to F0 (Bowling et al., 2021; Verpy et al., 2008). Perceptually, psychophysical studies have shown that these cochlear distortion products can improve pitch detection in human subjects (Pressnitzer and Patterson, 2001; Smalt et al., 2012). Therefore, it is important to determine if cochlear distortion products contributed to the F0-tuning we observed in cortical neurons. Distortion products at F0 are 10-15 dB quieter than the harmonic components in a missing fundamental sound (Pressnitzer and Patterson, 2001), so we presented our resolved and unresolved harmonic stimuli in the presence of pink noise (5 dB below the harmonic components) to mask potential distortion products in the ferret cochlea. We found that harmonicity and temporal neurons retained their F0-tuning even in the presence of the pink noise masker (Figure 2.12). These results confirm that the pitch selectivity observed for missing F0 sounds was not due to cochlear distortion artefacts at F0.

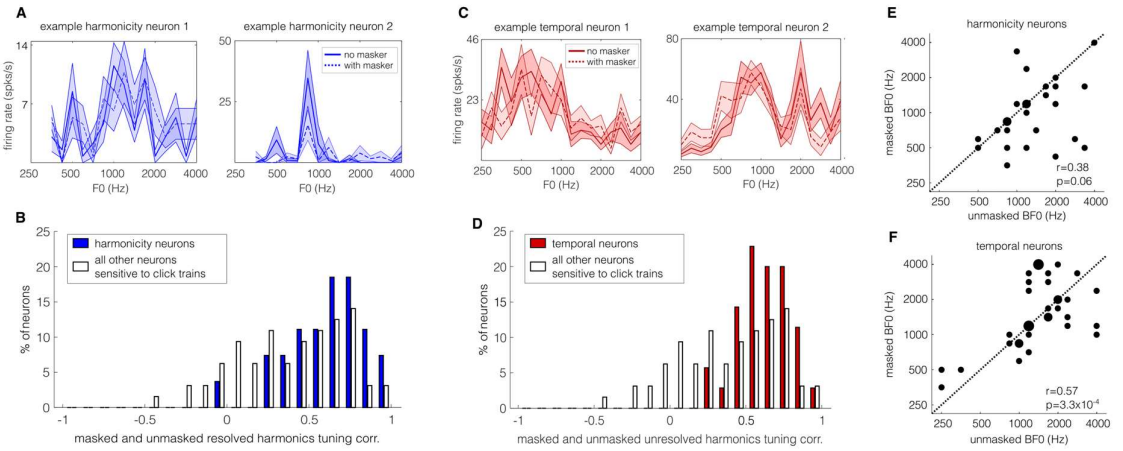

Figure 2.12. Distortion products do not explain observed F0-tuning. A) Each panel shows data from one example harmonicity neuron. Trial-averaged spike rates are plotted as a function of F0 for low harmonics presented with (dashed line) or without (solid line) a pink noise masker. **B**) Histograms show that the correlation between F0-tuning for masked and unmasked low harmonics within harmonicity neurons (blue bars) was higher than in other neurons that were F0-sensitive (but not pitch selective) (open bars; two sample t-test; t(89)=3.1, p=0.003). **C**) Each plot shows F0-tuning for masked (dashed line) and unmasked (solid line) high harmonics for two example temporal neurons, plotted as in (**A**). **D**) Plotted as in (**B**), histograms show the correlation between F0-tuning for masked and unmasked high harmonics within temporal neurons (red bars) was higher than in other F0-sensitive (but not pitch-selective) neurons (open bars; two sample t-test; t(97)=3.9, p=1.5x10^{-4}). **E**) For each harmonicity neuron, the scatter plot shows that the best F0 measured with unmasked and masked low harmonics were similar for most neurons (Pearson correlation values given in plot). Larger dots indicate two harmonicity neurons. **F**) Best F0 of temporal neurons were similar for unmasked and masked high harmonics (Pearson correlation values given in plot), plotted as in (**E**).

G. Anatomical locations of pitch-selective neurons

Given the evidence for an anatomically segregated region of pitch neurons in the marmoset auditory cortex (Bendor and Wang, 2005), we next explored where harmonicity, temporal, and pitch neurons were located across the surface of ferret auditory cortex. The auditory cortex of each animal was mapped based on anatomical landmarks (e.g. sulci), latency of neural responses to tones and noise bursts, and best frequencies (as described previously) (Bizley et al., 2009). Using this mapping, each neuron was localized to low- or high-frequency regions of A1 or AAF. The proportion of F0-sensitive neurons classified as harmonicity, temporal, or pitch neurons is mapped in each region is shown in Figure 2.10D and 2.13 (individual animal data) and summarized in Table 2.1. Only 1 F0-sensitive neuron was located in low-frequency AAF, so we omitted this region from further analysis.

To assess whether the three populations of neurons were homogeneously distributed across primary auditory cortex, we applied a chi-square test for homogeneity, with the anatomical regions as the categorical variable. We found that harmonicity neurons were distributed similarly across the three cortical regions, (low frequency A1, high frequency A1, and high frequency AAF), but were significantly more likely to be found in low frequency A1 ($\chi^2(2)=6.94$, $p=0.03$). Temporal and pitch neurons were significantly more prevalent in high-frequency A1 and AAF, particularly high-frequency AAF ($\chi^2(2)=32.6$, $p=8.1\text{x}10^{-8}$ for temporal neurons; $\chi^2(2)=16.6$, $p=2.5\text{x}10^{-4}$ for pitch neurons). In summary, we did not find strong evidence for a single pitch center in auditory cortex, although the proportion of harmonicity, temporal and pitch neurons varied between primary fields and across the tonotopic gradient.

cortical region	penetrations	sound-responsive neurons	F0-sensitive neurons (% of sound-responsive neurons)	harmonicity neurons (% of F0-sensitive neurons)	temporal neurons (% of F0-sensitive neurons)	pitch neurons (% of F0-sensitive neurons)
low-frequency A1	10	566	103 (18%)	23 (22%)	10 (10%)	11 (11%)
high-frequency A1	3	180	28 (16%)	2 (7%)	8 (29%)	6 (21%)
low-frequency AAF	2	33	1 (3%)	0	0	0
high-frequency AAF	3	93	33 (35%)	2 (6%)	17 (52%)	14 (42%)
TOTAL	**18**	**872**	**165**	**27**	**35**	**31**

Table 2.1. Distribution of neuron types across cortical penetrations in low and high regions of the two primary auditory cortical fields. We report the total number of sound-sensitive, F0-sensitive, harmonicity, temporal, and pitch neurons across all penetrations for each region.

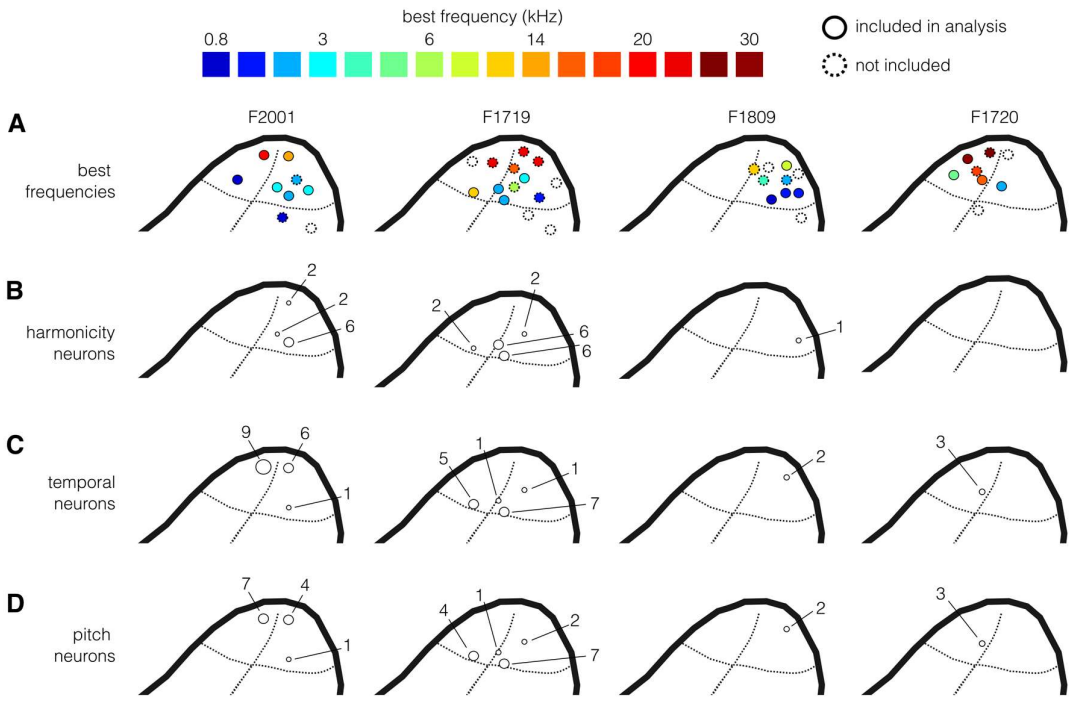

Figure 2.13. Distribution of harmonicity, temporal, and pitch neurons across primary auditory cortex. A) Best frequencies of each electrode penetration shown separately for each animal. Each circle shows the approximate anatomical location of the penetration, with the circle's color depicting the best frequency of that cortical region. White circles indicate a lack of clear pure tone frequency preference. Circles with a solid black border indicate penetrations which were included in the analyses in this paper. Circles with a dashed border indicate penetrations in which the stimuli described in this paper were not presented and were therefore not included in this paper. **B**) Absolute numbers of harmonicity neurons found in each penetration for each animal. Only penetrations in which harmonicity neurons were found are depicted. **C**) Absolute numbers of temporal neurons found in each penetration for each animal, as in (**B**). **D**) Absolute numbers of pitch neurons found in each penetration for each animal, as in (**B**).

2.6. Discussion

The fundamental frequency (F0) of harmonic sounds can be computed from either harmonic spacing or temporal periodicity (Oxenham, 2023, 2012). Computational models have demonstrated that a single spectrotemporal algorithm for F0 extraction is possible (Meddis and O'Mard, 1997; Plack et al., 2005), but human psychophysical studies instead suggest that pitch arises from the parallel processing of harmonic and temporal cues (Oxenham et al., 2011; Shackleton and Carlyon, 1994). Ferrets, like humans, can discriminate pitch shifts in complex sounds (Walker et al., 2009; Yin et al., 2010), although ferrets base their pitch judgments more heavily on temporal periodicity than resolved harmonic cues (Walker et al., 2019). Previous studies have shown that neurons throughout ferret auditory cortex can represent the F0 of artificial vowels, almost always in combination with other sound features (Bizley et al., 2010, 2009; Walker et al., 2011b), and their activity correlates with pitch judgments during behaviour (Bizley et al., 2013). However, it is not clear if neurons in ferret auditory cortex represent F0 invariantly across different stimuli, nor how they process harmonic and temporal cues to extract F0.

Here, we used high-density microelectrode recordings and a diverse stimulus set to identify three new neural subpopulations in ferret primary auditory cortex: harmonicity neurons that encode the F0 of resolved harmonics; temporal neurons tuned to F0 periodicity in unresolved harmonics; and pitch neurons that can extract F0 from both cue types. These populations showed properties previously only described in primate pitch neurons, including sensitivity to phase manipulations and temporal regularity, responses to missing fundamentals at F0, and invariance to cochlear distortion products (Bendor et al., 2012; Bendor and Wang, 2010, 2005). These F0 computations are well supported by findings of complex pitch discrimination in behaving ferrets (Walker et al., 2019, 2009), and could allow a listener to generalize pitch across a wide array of sounds, as humans do when recognizing a melody across different instruments.

A model of how harmonicity, temporal and pitch neurons may extract F0 from the representations of complex sounds in the cochlea is presented in Figure 2.14. Harmonicity neurons (H) could integrate inputs across place-coded representations of the resolved harmonic components of a sound. This mechanism has been previously described as a "harmonic sieve", with excitatory inputs from neurons

tuned to individual harmonics of F0 and inhibitory inputs from intermediate frequencies (Goldstein, 1973). Temporal neurons (T) may instead involve circuits that compute the periodicity of spike times in neurons tuned to unresolved harmonics. These inputs may also include neurons phase-locked to resolved harmonic components, as F0 can be calculated as the autocorrelation of spike times across neurons tuned to different harmonics (Licklider, 1951). Pitch neurons (P) can provide spectrally invariant representations of F0 by combining inputs from harmonic and temporal neurons tuned to a common F0. The auditory nerve representations of resolved harmonic and temporal periodicity described in Figure 2.14 have been previously observed by Delgutte and colleagues (Cariani and Delgutte, 1996). The evidence for the proposed dual processing of pitch cues at the level of auditory cortex is summarized below.

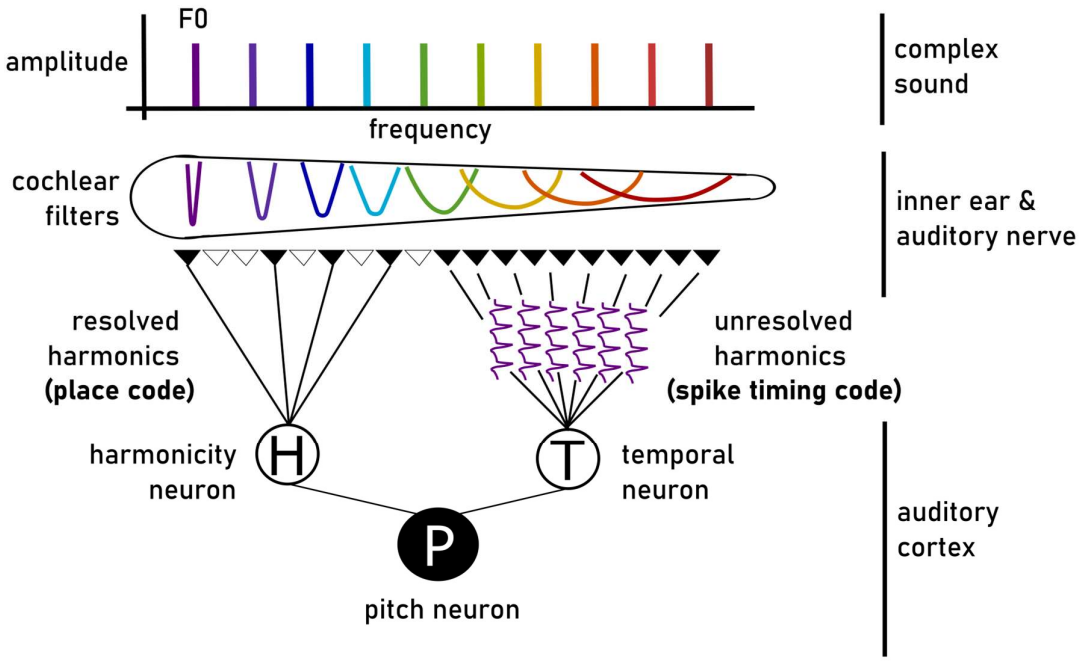

Figure 2.14. Schematic of proposed dual extraction of fundamental frequency (F0) from resolved and unresolved harmonics. Top: The spectrum of a harmonic tone complex is shown, with each frequency component in a different color. **Middle**: The frequency tuning curves of individual cochlear filters are shown along the unrolled basilar membrane, with color indicating the best frequency of each auditory nerve fiber. Because the tuning bandwidth of auditory nerve fibers is constant with respect to logarithmic frequency, tuning curves become exponentially broader with increasing linear frequency. Low numbered harmonics are termed "resolved" if only one harmonic falls within the tuning curve of a single auditory nerve fiber, and therefore F0 is represented with a place code as the pattern of harmonic activation across the population of resolved fibers. In contrast, auditory nerve fibers tuned to higher harmonics are likely to respond to multiple harmonics in the sound, so these are "unresolved" in the place code. Because multiple unresolved harmonics excite the same auditory nerve fiber, the spike times of auditory nerve fibers can phase lock to the sum of unresolved harmonics at F0 (purple traces show spike output over time). This produces an explicit temporal representation of F0 for resolved harmonics. **Bottom**: Our data show that there are distinct neural populations in auditory cortex to extract these F0 cues: harmonicity neurons (H) compute F0 from the place code of exclusively resolved harmonics; and temporal neurons (T) extract F0 from the temporal code of unresolved harmonics. These two neural populations may then converge on a third neural population, pitch neurons (P), which encode F0 irrespective of the types of acoustic cues present in the sound.

A. Dual pitch processing in auditory cortex

Our findings add to the evidence in non-human primates that cortical neurons can represent F0 using resolved harmonic or unresolved temporal envelope cues (Bendor et al., 2012; Song et al., 2016; Steinschneider et al., 1998). In macaques, populations of neurons in low-frequency A1 can represent resolved harmonics as a place code, while higher-frequency neurons time-lock to the temporal envelopes of unresolved harmonics (Fishman et al., 2013; Steinschneider et al., 1998). In marmosets, Wang and colleagues identified a "pitch area" on the low-frequency border of A1 and R, where individual neurons were tuned to the F0 of both pure tones and missing fundamentals (Bendor and Wang, 2010, 2005).

Within the marmoset pitch centre, a small subpopulation of pitch neurons were sensitive to alternating phase manipulations (Bendor et al., 2012). A separate subset of neurons with higher best F0s (>450 Hz) was insensitive to phase manipulations, but sensitive to resolved harmonic cues (Bendor et al., 2012). Ferret auditory cortex shows a similar division of labour in the present findings. Ferret temporal neurons showed sensitivity to phase manipulations, including a reduced response to both alternating phase and random phase harmonics. They showed similar tuning across complex periodic stimuli, even when only unresolved harmonics were presented. Harmonicity neurons generalized their F0-tuning across complex stimuli, but only for sounds containing resolved harmonics. As reported for marmoset pitch-selective neurons (Bendor and Wang, 2010, 2005), our temporal, harmonicity, and pitch neurons were sensitive to the temporal regularity of click trains, demonstrating a neural parallel to pitch salience (Pollack, 1968; Yost et al., 2005). Our results suggest that dual cortical processes for extracting F0 may be widespread across mammals. The present experiments were performed in anesthetized animals to control for the effects of attention and yield stable recordings throughout cortical layers during long stimulus pre sentations. The marmoset work was instead carried out in awake animals, and recordings were primarily made in layers 2/3. Pitch responses may differ across behavioral states (anesthetized, awake, and actively listening) and cortical layers, and these remain important questions for future study. It also remains possible that neurons tuned to a common F0 for simple tones and complex sounds may exist in non-primary regions of ferret auditory cortex.

B. Pitch-selective neurons in auditory cortex

Our results demonstrate that pitch-selective neurons exist in ferrets. As reported in marmosets, pitch-selective neurons were sensitive to the temporal regularity of jittered click trains, did not rely on cochlear distortion artefacts, and were sensitive to the phase of unresolved harmonic components (Bendor and Wang, 2010, 2005). Their overall prevalence was also comparable across these species: about 3.5% of sound-responsive neurons met our pitch neuron criteria, closely matching the ~3.6% reported in marmoset auditory cortex (Figure 2 of Bendor and Wang, 2005) (Bendor and Wang, 2005), although about one-third of neurons are pitch-selective in the marmoset pitch center (Bendor et al., 2012; Bendor and Wang, 2010, 2005). To our knowledge, this is the first demonstration of pitch-selective neurons outside primates.

There are two key differences between marmoset and ferret pitch neurons. First, marmoset pitch neurons were tuned to similar F0s for pure tones and missing fundamentals (Bendor et al., 2012; Bendor and Wang, 2005), while these preferences often differed in ferret pitch neurons. This may reflect our differences in classification criteria, as Bendor & Wang (2005) made this an inclusion criterion, while we did not. Our results also raise the possibility that pure-tone frequency and complex F0 extraction are mediated by distinct mechanisms. While a single tone does evoke a clear pitch percept, it has no harmonic structure, providing a weak input to our model of temporal and harmonicity neurons, which both pool information across harmonics (Figure 2.14).

A second difference is that pitch neurons were found distributed throughout primary auditory fields in ferrets. This contrasts with the local pitch area of marmosets (Bendor and Wang, 2005), but aligns with more distributed pitch representations reported in auditory cortex of macaques (Fishman et al., 2013; Norman-Haignere et al., 2019; Steinschneider et al., 1998) and humans (Allen et al., 2017; Berger et al., 2023; Gander et al., 2019; Griffiths et al., 2010). Even human fMRI studies showing evidence of a pitch center disagree on its location within auditory cortex (Norman-Haignere et al., 2019; Patterson et al., 2002; Plack et al., 2014). Overall, the bulk of evidence points toward a distributed network of pitch-selective cortical neurons. Anatomical clustering in marmosets may represent a species-specific specialization, or could result from the criterion for pitch neurons to generalize their F0 tuning to

pure tones. Our harmonicity neurons showed the most similar tuning for frequency and complex F0 (Figure 2.11), and were also more anatomically clustered in low frequency A1 (Figure 2.10D) than other pitch-selective neurons, so these may be the closest analog to marmoset pitch neurons.

The present experiments were performed in anaesthetized animals to control for the effects of attention and yield stable recordings throughout cortical layers during long stimulus presentations. The marmoset work was instead carried out in awake animals, and recordings were primarily made in layers 2/3. Pitch responses may differ across behavioural states (anaesthetized, awake, and actively listening) and cortical layers. Anaesthesia causes lower discharge rates of cortical neurons, and the responses of these neurons under anaesthesia are more likely to occur transiently at the sound's onset (Wang et al., 2005). Pure tone tuning has been reported to be more selective with smaller response bandwidths under anaesthesia (Gaese and Ostwald, 2001; Zurita et al., 1994). It is possible that cortical encoding of pitch may occur in a sustained encoding window and that pitch selectivity within single neurons may increase under awake conditions. These remain important questions for future study.

C. Conclusions and future directions

These findings establish that ferret auditory cortex implements dual pitch extraction strategies and contains neurons specialized for invariant pitch representation. This suggests that pitch perception relies on conserved cortical computations across mammals, including non-primates, and opens the door to study pitch in a wider range of species. For example, if mice were shown behaviourally to experience pitch perception, a toolbox of powerful genetic targeting techniques would be available to further dissect and manipulate pitch processing circuits, and *in vivo* intracellular recordings could be used to understand the synaptic mechanisms of pitch processing.

3. Population codes for pitch

3.1. Rationale

The findings reported in Chapter 2 provided the first evidence of pitch selective neurons outside of primates, however the prevalence of these neurons was low (~3%), consistent with the prevalence identified in Bendor and Wang, 2005 (Bendor and Wang, 2005). This low rate raises questions about the importance of these specific neurons to the perception of pitch: is it possible that the strongly hypothesis-driven approach of Chapter 2 has failed to capture the broader pitch encoding mechanism? Here, we employ a data-driven approach to the same dataset from Chapter 2 to identify the distribution of F0 information across the population using computational models trained to predict the F0 of each stimulus presentation. Using this approach, we can agnostically identify whether population-level F0 representations include the same small subset of pitch-selective neurons across a variety of sounds (as would be suggested by the results of Chapter 2), or whether distributed, non-overlapping neurons underpin these codes.

3.2. Contributions

Kerry Walker and Quentin Gaucher conceptualized the experiment, designed the stimuli, performed the craniotomy, recorded the electrophysiology, and pre-processed the data including spike sorting. Ben Willmore wrote software for linear decoding. Veronica Tarka wrote software for all analysis and produced all figures.

3.3. Introduction

Pitch is the tonal (‘low’ or ‘high’) quality of sound that allows us to experience melody in music and prosody in speech, and helps us to focus on a single voice in a crowded room (Assmann and Summerfield, 1987; Bird and Darwin, 1998; de Cheveigné, 1997; McPherson et al., 2022). Pitch perception is often impaired in those with ageing and hearing loss (Arehart, 1994; Hoekstra and Ritsma, 1977), even for people with healthy pure tone audiograms (Moore and Peters, 1992). Despite its importance to hearing, and an agreed role of auditory cortex in pitch perception, the cellular and circuit mechanisms that give rise to pitch perception remain poorly understood (Wang and Walker, 2012).

Most natural periodic sounds are composed of many frequencies, yet we experience their pitch as a single note at the fundamental frequency (F0). All frequency components of a pitch-evoking sound are integer multiples (*harmonics*) of F0. Lower harmonics are "resolved" as a population place code in the auditory nerve because the spacing between these harmonics is sufficiently large relative to the frequency tuning width of auditory nerve fibres (Moore and Gockel, 2011). The fundamental of higher "unresolved" harmonics is represented as a spike timing code, as auditory nerve fibre responses will phase-lock to periodicity at F0 (Cariani and Delgutte, 1996).

We perceive the pitch at F0 regardless of whether we hear resolved or unresolved harmonics (Carlyon, 1996a, 1996b; Shackleton and Carlyon, 1994), implying that the brain integrates these dual encoding strategies. The auditory cortex is a likely candidate for the site of this integration, as it is required for pitch judgements in humans (Stewart et al., 2006; Tramo et al., 2002; Zatorre, 1988) and animals (Whitfield, 1980). Studies in marmosets (Bendor et al., 2012; Bendor and Wang, 2010, 2005) and our recent work in ferrets (Tarka et al., 2025) have identified small subpopulations of *pitch-selective* neurons in auditory cortex that represent F0 derived from either resolved harmonics, periodicity in unresolved harmonics, or, in the case of specialized pitch neurons, invariantly from either cue type. However, a wealth of other studies in ferrets, macaques, and humans have further shown that the F0 of a given sound is well represented by neural populations distributed throughout auditory cortex (Allen et al., 2017; Berger et al., 2023; Fishman et al., 2013; Gander et al., 2019; Griffiths et al., 2010; Norman-Haignere et al., 2019; Steinschneider et al., 1998). These *pitch-sensitive* neurons are usually also sensitive to the timbre (i.e. the spectral identity) of sounds (Bizley et al., 2010, 2009; Walker et al., 2011b), so they may be unable to support, for example, recognizing the same musical note across different instruments. As many studies of F0 coding use only one type of sound, it is not clear to what extent cortical population codes are able to provide an invariant representation of pitch across resolved and unresolved pitch cues, or across sound timbres.

Population-level representations provide greater discriminability between different sensory features than is possible within single neurons due to the inherent noise in neuronal firing (Averbeck et al., 2006; Montijn et al., 2016; Paradiso, 1988). Because this greater resolution corresponds more closely with perceptual discriminability

(Jacobs et al., 2009), it is widely thought that population codes underly our perception of the external environment (Panzeri et al., 2015; Stanley, 2013). Though both resolved and unresolved harmonics can independently produce a percept of pitch, different species preferentially rely on one cue or the other when discriminating pitches (Klinge and Klump, 2010; Shofner and Chaney, 2013; Walker et al., 2019). It is unclear whether population representations of pitch in auditory cortex should represent the perceptual unity across both acoustic cues, or if population codes mirror the structural differences between these cues which is reflected in animals' behavioural preferences for one cue over the other.

To investigate the invariance of population F0 codes, we presented 10 stimulus timbres containing different combinations of F0 cues to anaesthetised ferrets while recording the spiking responses of large populations of individual neurons in the primary auditory cortex. We found that F0 can be decoded from population-level responses, and these codes are specific to the type of pitch cues available. Additionally, onset and offset responses both represent F0 in a cue-dependent manner, and in contrast to frequency tuning, F0-tuning does not simply linearly inverse across the two time windows. Finally, the activity of a small number (~20) of neurons is sufficient to accurately decode F0 from population responses, but these neurons are not essential to representing F0, suggesting a high degree of redundant F0 information across the population. These findings show that population-level representations of F0 are different depending on the available acoustic cues, and these representations differ mainly on the basis of the small subset of neurons that are most informative about F0.

3.4. Methods

A. Animals

Data were collected from 4 adult ferrets (*Mustela putorius furo*; 2 female; age: $\mu = 29.3$ weeks, $\sigma = 18.4$ weeks; Marshall BioResources, UK). All animal procedures in this study were approved by the local ethical review committee of the University of Oxford and performed under UK Home Office license.

B. Surgical procedure

The surgical procedure was performed as reported in Chapter 2.4B of this thesis.

C. Sound stimuli

We presented 10 sound types (*timbres*), each over a range of 15 F0s (354 - 4000Hz in ¼ octave increments). Each of the 10 timbres fell into one of 4 classes of pitch cue: spectral cues, temporal cues, combined spectral and temporal cues ("*combined cues*"), or pure tones (Figure 3.1).

The "*spectral cues*" stimulus contained only lower harmonics (Figure 3.1a) that are resolved in the ferret cochlea (Sumner and Palmer, 2012; Walker et al., 2019; "resolved harmonics"). We presented three types of "*temporal cues*" stimuli. All three contained the same set of unresolved harmonics up to 30 kHz, but these frequency components were either: all in cosine phase ("unresolved harmonics"; Figure 3.1b, left); in alternating sine and cosine phase, which doubles the periodicity of the sound ("alternating phase"; Figure 3.1b, middle); and in randomized phase, which diminishes the strong periodic peaks in the temporal waveform ("random phase"; Figure 3.1b, right). "*Combined cues*" stimuli were broadband sounds that contained both resolved and unresolved harmonics up to 30 kHz. The first timbre in this class, "missing fundamental" sounds, had energy at all harmonics, both resolved and unresolved, but no energy at F0 (Figure 3.1d, left). We also presented "click trains" which are generated from biphasic pulses at a rate of F0, producing all harmonics including F0 (Figure 3.1d, middle left). We varied the temporal regularity of these click trains by temporally jittering the timing of the individual clicks by either 0 (perfectly periodic), 5, 10, or 20% relative to the inter-click interval ("click train 5% jitter", etc; Figure 3.1d, middle through right). Stimuli with increased jitter are less temporally regular, and have a weaker pitch salience as perceived by humans (Pollack, 1968) and ferrets (unpublished data from our lab). Finally, we presented "pure tones", which contained energy at F0 only (Figure 3.1c).

All sounds were presented via closed field, in-ear earphones (Panasonic RP-HV094E-K with an attached otoscope tip) at 70 dB SPL_A for 300 ms followed by 400 ms silence (1 ferret), or for 200 ms followed by 600 ms silence (3 ferrets). Each timbre was presented 13 times at a given F0, and stimulus order was pseudo-randomized across timbre and F0. Linear onset and offset ramps of 25 ms were applied to every sound to prevent clicks.

At the start of recording from a new electrode penetration, 200 ms bursts of 30 dB broadband Gaussian noise (300 ms interstimulus interval) were used as search stimuli

to identify sound-responsive neurons. Once responsive units were detected, a pseudorandomized sequence of pure tones (100 ms duration, 500 ms interstimulus interval, 40-80 dB SPL_A in 10 dB steps, 0.5-40 kHz in 0.45 octave steps) were presented to evaluate frequency tuning, allowing localization of every microelectrode penetration within the tonotopic map. Finally, the complete set of pitch stimuli described above were presented.

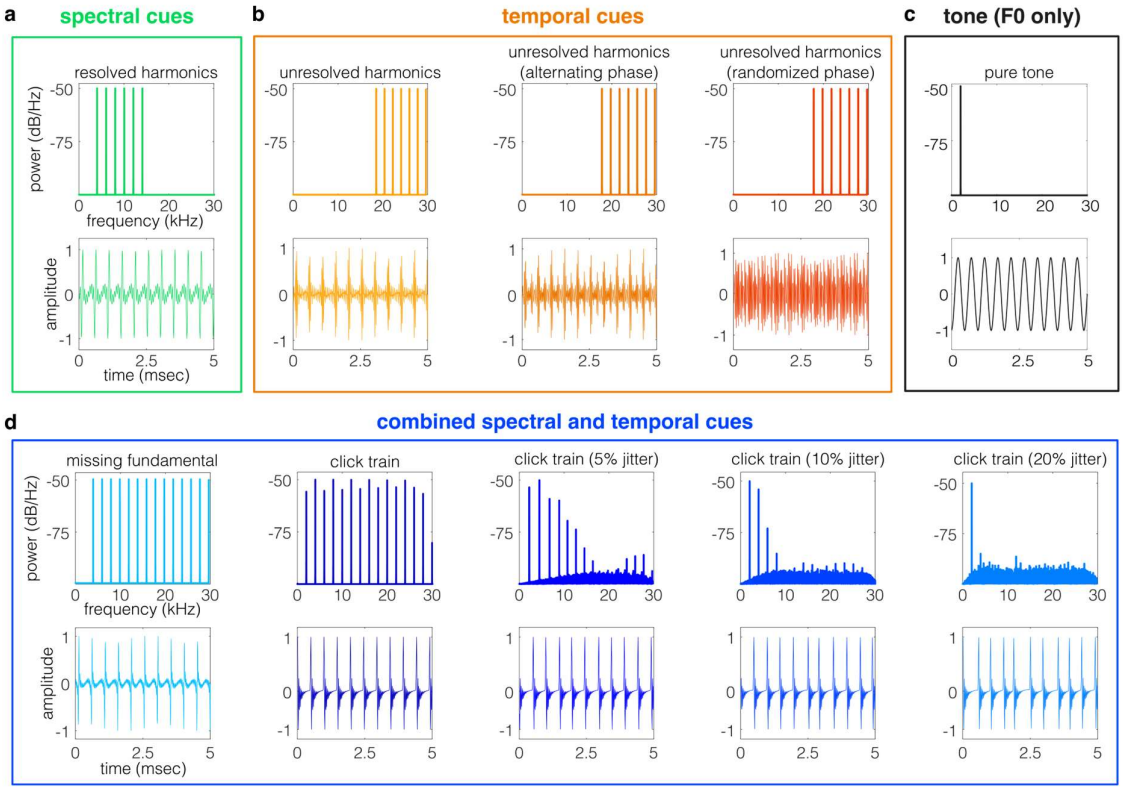

Figure 3.1. Stimuli. The spectrum (upper plot) and temporal waveform (lower plot) of all pitch stimuli are shown. Stimuli are coloured to indicate the 4 F0 cue classes. **a**) Spectral pitch cues are contained in resolved harmonics (green), which are place-coded in the auditory nerve. **b**) Unresolved harmonics (left) provide a temporal cue for pitch through spike timing in the auditory nerve, and these are doubled by alternating phases of neighbouring harmonics (middle) or weakened by randomizing phases across harmonics (left). **c**) The pure tone contains energy only at F0, with no harmonics over which spectral or temporal F0 cues can be integrated. **d**) Sounds with broad spectra contain both spectral and temporal pitch cues. The missing fundamental sound (left) contains energy at all harmonics except F0. Click trains contain energy at F0 and all harmonics. Click trains with 5, 10, and 20% temporal jitter have increasing spectral and temporal noise in their F0 cues with increasing jitter percentages.

D. Spike sorting

The neural recordings were processed offline, beginning with digital high-pass filtering at 150 Hz. To minimize shared electrical noise across channels, common average referencing was applied. Spiking events were subsequently identified and clustered using the Kilosort2 algorithm (https://github.com/MouseLand/Kilosort2) (Pachitariu et al., 2024, 2016). Manual curation of the resulting spike clusters was carried out in Phy (https://github.com/cortex-lab/phy) (Rossant et al., 2016). Units exhibiting consistent spike waveforms with low variability and a distinct refractory period in their autocorrelograms were classified as single units. Only these well-isolated single units were retained for further analysis.

E. Isolating sound-responsive neurons

We tested if each neuron was responsive to the pooled set of pitch stimuli using a paired-sample t-test ($\alpha = 0.05$). We compared the neuron's firing rate during the 150 ms of silence before the sound started with its firing rate during the 150 ms after sound onset, across all sound presentations. Response to sound offset was similarly tested in the 150 ms following sound offset. Only neurons that showed a significant response to sound onset or offset were included in further analysis.

F. Decoding F0

Neural responses were concatenated across the 18 recordings, and trial order was randomly shuffled for each neuron independently to minimize effects of simultaneous recording. A trial was taken as the 400 ms after sound onset, and spikes were binned at 10 ms, the smallest bin size before decoding performance significantly worsened (Figure 3.2). For one animal trials were 100 ms longer, so neural responses 170-270 ms after onset were removed in this animal in order to align sound offset responses across animals. Each neuron's firing rate was normalized by mean and variance across the whole data set. We used a linear model to decode F0 from a tensor of neuronal activity, X_{nbt}, where n is the neuron, b is the time bin in a trial, and t is the trial number (Figure 3.3a). The true F0 of each trial was defined as the target vector, y_t. The decoder produced a prediction of each trial's F0 from the sum of the element-wise product between the neuronal activity and a

matrix of weights, k, with one weight per neuron and time point. The decoder's predictions of each trial's F0 was therefore defined as $\hat{y}_t$ where

$$\hat{y}_t = a_0 + \sum_{b,n} X_{nbt} \cdot k_{nb}$$

[*eq. 3.1*]

Model parameters were optimized with L2 regularization using glmnet (https://glmnet.stanford.edu/articles/glmnet.html) to minimize mean square error between $\hat{y}_t$ and y_t. Trials were divided into 5 folds, each with 156 training trials and 39 non-overlapping test trials (with any leftover trials added to the test set of the fifth fold), and trials were randomly shuffled to promote an even sampling of F0s in each test set.

We trained a separate decoding model for each of the 10 timbres presented (Figure 3.3b). The fold with the lowest root mean square error (RMSE) between $\hat{y}_t$ and y_t was taken as the representative model for the given timbre. The decoding model for each timbre is therefore a 2-dimensional matrix, size $N \times B$, where N is the number of neurons and B is the number of time bins. Each value in this matrix is the coefficient for the activity of one neuron in one time bin, such that the sum of all values in the matrix multiplied by the neural activity on a single trial produces the estimation of that trial's F0. Model weights can therefore be considered a proxy for F0 information.

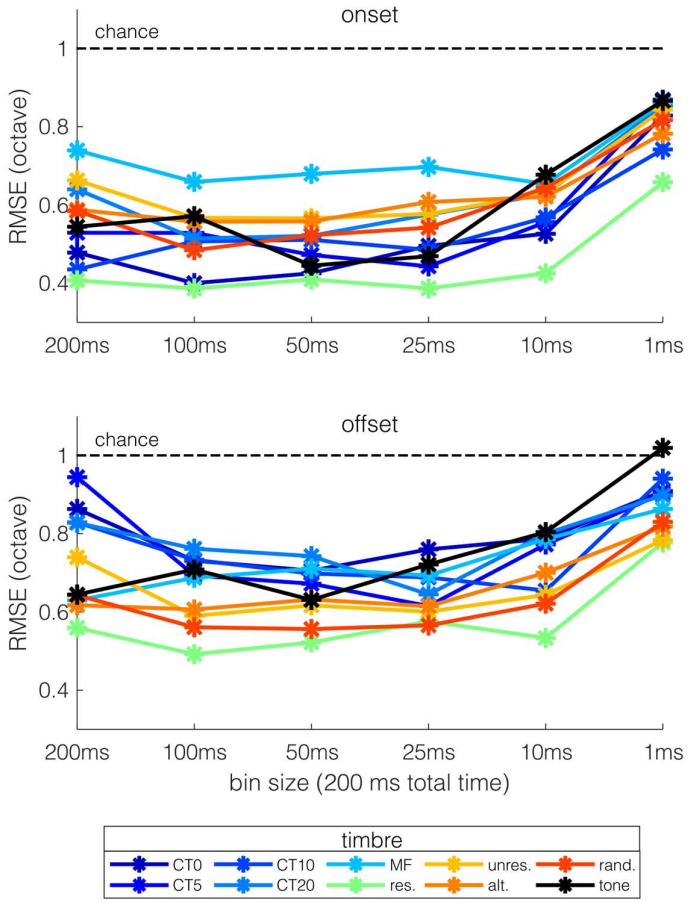

Figure 3.2. Decoding performance for different time bin sizes. Decoders trained on neural activity in the first 200 ms after sound onset (top) perform similarly when spikes are binned at 200, 100, 50, 25, and 10 ms across all timbres (indicated by different colours, see Figure 3.1). Decoding performance is worse at 1 ms bins across all timbres. Decoders trained on neural responses to sound offset (bottom) perform similarly for bin sizes of 100, 50, 25, and 10 ms across all timbres, and are in general worse for bin sizes of 200 ms and 1 ms. Reduced accuracy for 1 ms bins is likely to due to greater inter-trial response variability at this temporal resolution, as well as excessive model parameters causing noisier fits for decoding weights.

G. Principal component analysis

We performed principal component analysis on the kernels for each timbre's decoding kernel concatenated together. The kernels were concatenated across time, to produce one matrix of size $N \times (B * S)$, where N is the number of neurons, B is the number of time bins, and S is the number of stimuli (i.e. timbres).

H. Single neuron information

To assess how much a single neuron contributes to the overall F0 information in the population, we created an informativeness metric defined as the sum of the absolute value of its weights across all time bins in the decoding kernel, or for neuron n:

$$informativeness_n = \sum_b |K_{nb}|$$

[*eq. 3.2*]

3.5. Results

We examined how population codes in primary auditory cortex may represent the F0 of stimuli across a range of timbres. Each timbre was played at 15 different fundamental frequencies evenly sampled from 354–4000 Hz. We recorded responses of 1266 well-isolated single units ("neurons") from 4 ferrets (1077 in A1, 189 in AAF). This included 633 (50%) sensitive to both sound onset and offset, 234 (18%) sensitive to sound onset only, and 191 (15%) sensitive to sound offset only (paired-sample t-test, $p<0.05$). We then trained a linear model to decode the F0 of the sound presented on each trial from the neural responses of these 1085 neurons. The decoder was trained separately for each timbre (Figure 3.3a,b).

We found that F0 was linearly decodable from neural responses to all 10 timbres presented. F0 predictions of held-out test trials were on average 0.43 octaves from the true F0 across all timbres, and cross-correlations between target and predicted F0s were consistent on test trials across folds (Figure 3.4). The highest accuracy was achieved with resolved harmonics (Figure 3.3c, green) with test-set predictions on average 0.31 octaves from the true F0. The missing fundamental stimulus (Figure 3.3c, lightest blue), was the least predictable with decoder predictions on average 0.54 octaves from the true F0.

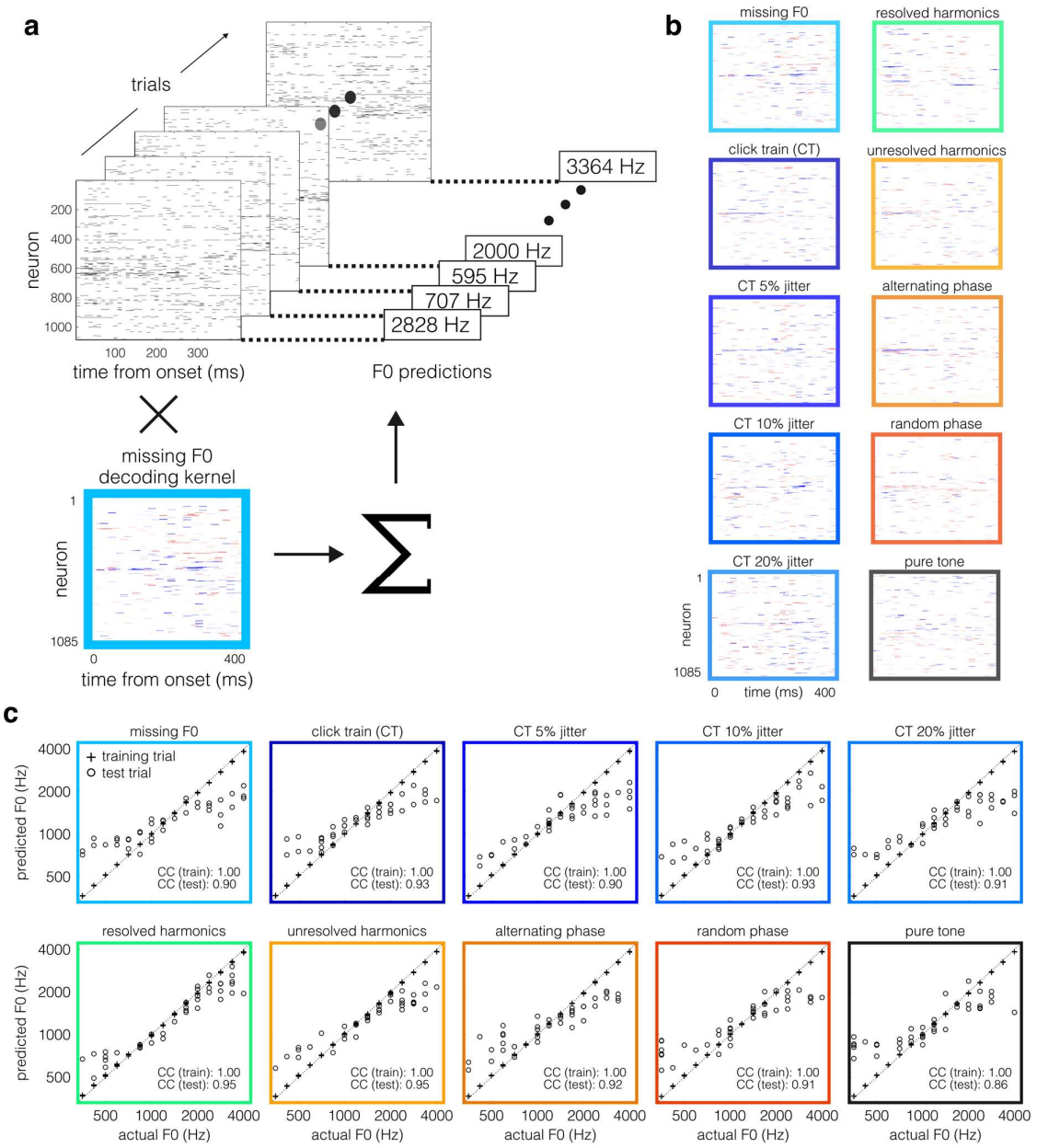

Figure 3.3. F0 is linearly decodable from the neural population. a) We decoded F0 on a trial-by-trial basis from the spiking responses of 1085 isolated single units. Spike counts on each trial were summed in 10 ms bins, taken from the onset of the sound until 200 ms following sound offset (400 ms total). A separate decoding model was trained for each of the 10 timbres presented. **b**) Decoding model weights ("kernels") trained for each of the 10 timbres. **c**) Each plot shows the decoding performance for the trials in the training set (crosses) and test set (open circles) for the training fold with the smallest RMSE for each timbre. Colours indicate pitch cue classes described in Figure 3.1.

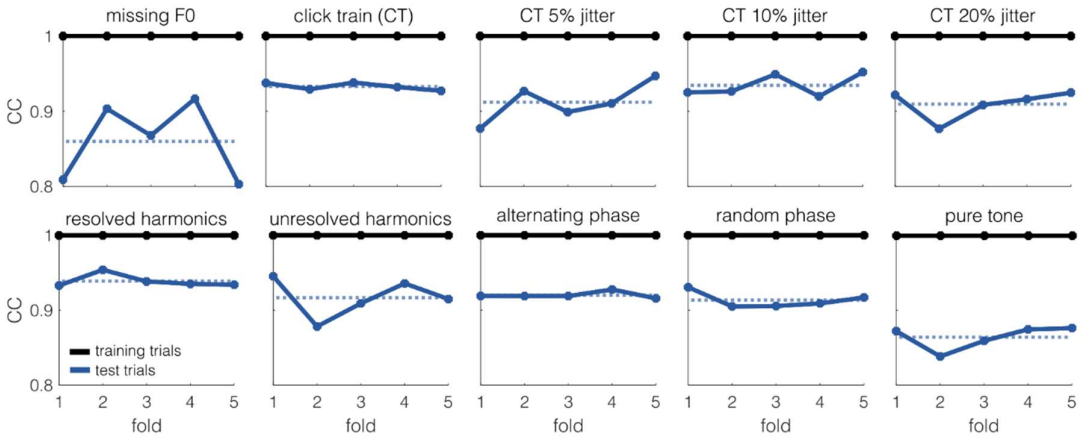

Figure 3.4. Decoding performance in all 5 training folds for each timbres. Each plot shows the cross-correlation (CC) between the actual F0 of the trial and the F0 predicted by the decoder. Trials used in the training set are shown in black, and the held-out test set is shown in blue. The average CC across all 5 folds is shown as a dashed line.

A. Population F0 codes generalize well within, but not across, classes of pitch cues

In the auditory nerve, the F0 of a sound's resolved harmonics are encoded via a tonotopic place code, while the F0 of unresolved harmonics are encoded via spike times that are phase-locked to the envelope periodicity. Sounds that evoke a common pitch but differ in timbre may contain different combinations of these cues, and here we investigate whether dual F0 encoding strategies are integrated by population codes at the level of primary auditory cortex. To test if the population F0 code could generalize across stimuli, we trained our decoder on one timbre and used the fitted weights to decode the F0 of another timbre. This approach was repeated for all possible pairings of training and testing timbres (Figure 3.5a). If the population pitch code had fully integrated spectral and temporal F0 cues, we would expect that any single timbre's decoding weights should decode the F0 of any other timbre in our set equally well.

Decoding accuracy was measured as the root mean square error (RMSE) across trials, where error is the distance in octaves between the predicted F0 and the actual F0 for a given trial. The decoding accuracy for each timbre pairing was normalized by dividing the decoding accuracy for the training timbre when tested on itself (i.e. normalized accuracies close to 1 indicate similar F0 representations for training and testing timbres). We found that for any given timbre, the trained decoding kernel most accurately predicts the F0 of only a subset of other timbres. There is a structure

to these regions of generalization, visible as 4 yellow squares in the cross-timbre decoding matrix in Figure 3.5b. Each region of yellow aligns with stimuli containing the same class of pitch cue, indicating that population F0 codes can generalize well within, but not across, these cue types. For example, decoders trained on a broad spectrum sound containing both resolved and unresolved harmonics can classify the F0 of other *combined cues* stimuli with high accuracy, but not those containing only one of these pitch cue types. Similarly, *unresolved harmonic* decoder weights can be used to classify the pitch of other sounds containing only unresolved harmonics, but performs poorly for resolved harmonic stimuli. Decoders trained on unresolved harmonics with randomized phase provide particularly poor F0 classification for other timbres (blue stripe in Figure 3.5b), which supports the important role of temporal periodicity in classifying the pitch of unresolved harmonics. We included only one stimulus type containing resolved harmonics, and its decoder does not generalize well to any other timbre. Interestingly, decoders trained to classify the frequency of pure tones, with energy only at F0, generalize even more poorly to other stimulus classes, suggesting that there are separate population codes for representing pure tone frequency and the pitch of complex sounds. Together, these findings suggest that F0 representations within populations of auditory cortical neurons are distinct for resolved and unresolved pitch cues, and the frequency of pure tones.

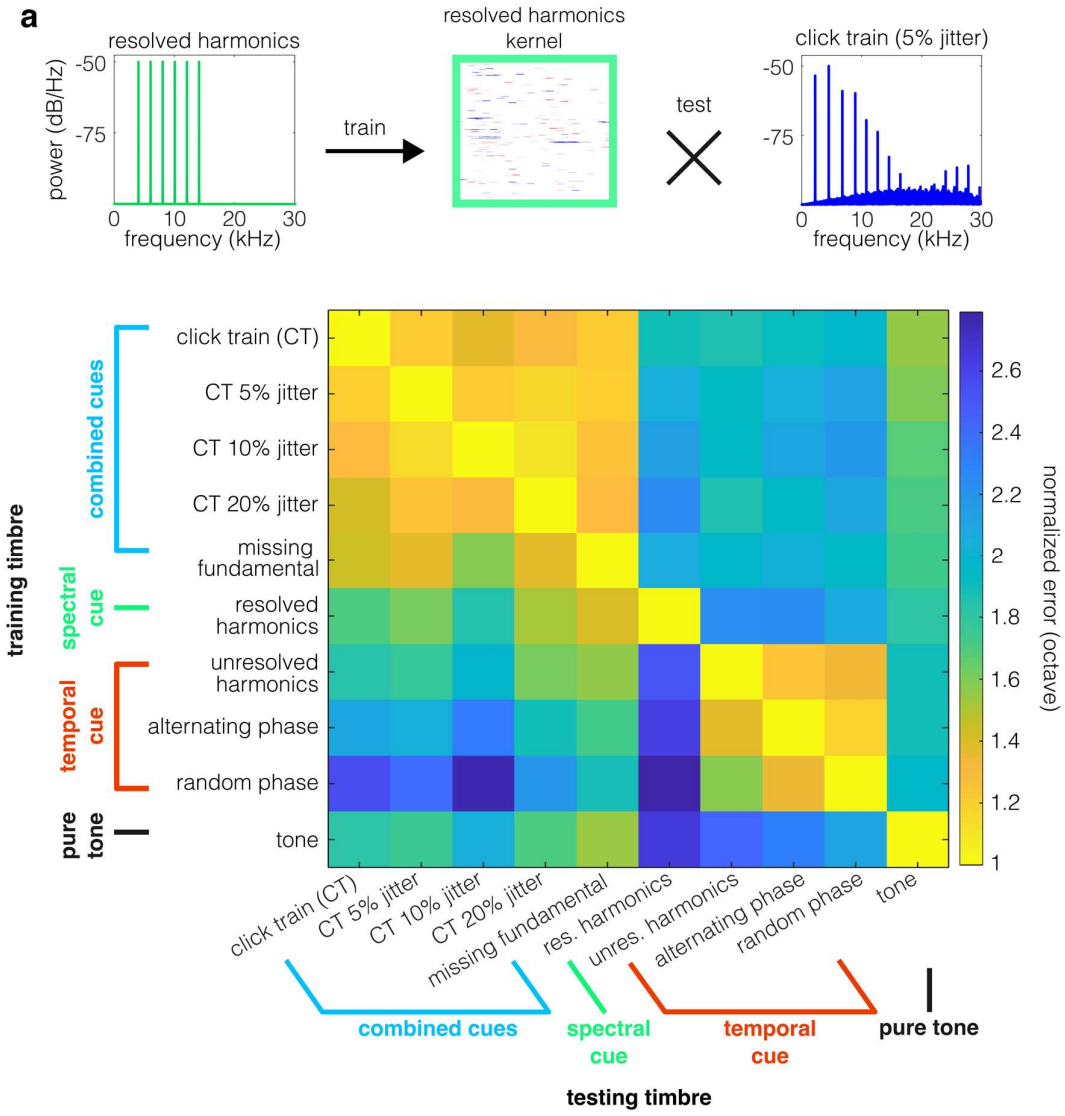

Figure 3.5. Different pitch cues are decoded by different population codes. a) Schematic of an example cross-timbre prediction. The F0-decoding model was trained separately on each timbre ("training stimulus", e.g. resolved harmonics) as described in Figure 2. The fitted model was then used to decode the F0 of a different timbre ("testing stimulus", e.g. click train with 5% jitter). Accurate cross-timbre predictions indicated that the neural populations represented F0 similarly in the training and testing stimuli. **b**) Cross-timbre F0 predictions for all possible pairings of training (y-axis) and testing (x-axis) stimuli. The colour scale shows the decoding prediction error for each test timbre relative to the training timbre (i.e. $E_{\text{cross timbre}} / E_{\text{same timbre}}$). Lower values indicate better relative decoding performance and therefore more similar F0 population codes between training and testing stimuli. Timbres that share F0 representations in the neural population clustered into four blocks of yellow values along the diagonal, corresponding to the four classes of pitch cues: resolved harmonics, unresolved harmonics, a combination of these two cues, and pure tones.

B. Population F0 codes are composed of non-redundant onset and offset responses

The decoder estimates weighting functions that vary across both neuron and time (Figure 3.3b), and the time-varying patterns of these weights indicate how neurons encode pitch in their responses. We used PCA to investigate whether large-scale trends in the temporal structure of these decoding weights diverge across different pitch cue classes. PCA was applied to a matrix comprised of every timbre's fitted decoding model concatenated across time. In this manner, the dimensionality reduction isolates informative temporal spiking patterns that can be compared across the 10 timbres.

The first three principal components explained 8.9% of the kernel variance (Figure 3.6), and these showed distinct temporal spiking patterns that varied across timbre. The first principal component primarily captured offset responses to broad spectrum timbres containing combined F0 cues (Figure 3.7a, left), while the second component largely captured onset responses to tones and offset responses to sounds containing only unresolved harmonics (Figure 3.7a, middle). The third principal component characterised a gradient in onset responses that differed between tones, stimuli containing only unresolved harmonics, and those containing resolved harmonics (Figure 3.7a, right). These results show that the temporal profile of population F0 codes differ across, but not within, the four pitch cue classes.

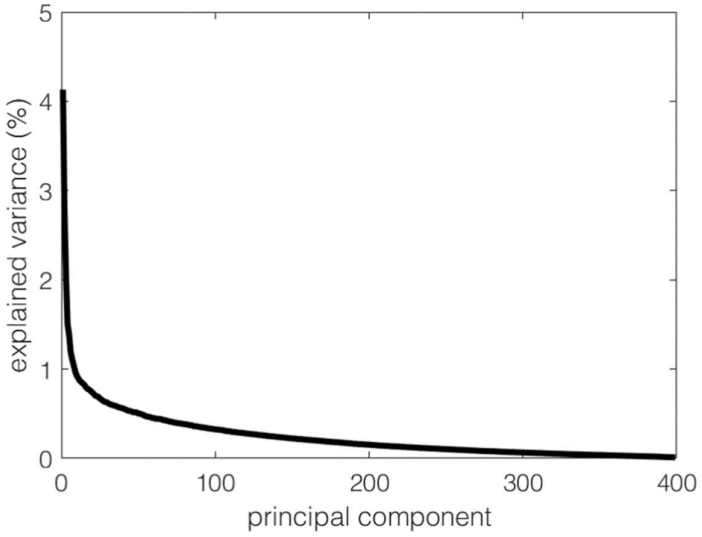

Figure 3.6. Explained variance from principal component analysis.

To investigate how the time course of the population code may be shared across stimulus timbres, we plotted the time trajectories of the fitted kernel for each stimulus type in three-dimensional principal component space (Figure 3.7b, left). These neural trajectories summarize the F0 information in the neural population at each time bin across the entire 400 ms trial. Trajectories that are close in PC space indicate a similar F0 representation in the neural responses at that time point, while divergent trajectories indicate different F0 representations. These principal component projections offer two interesting observations. First, there is a clustering of stimulus timbres into the four pitch cue classes, suggesting that the temporal evolution of these population codes is shared for stimuli that contain the same combinations of resolved and unresolved harmonic cues. Second, the trajectories of all stimuli have two deviations from the origin—one following sound onset, and a second following sound offset. The temporal spiking patterns identified by the PCA can therefore be broadly described as onset (0 – 200 ms) and offset (200 – 400 ms) responses, without much evidence for more complex patterns across fine time bins. In fact, the decoder performance remains relatively stable when spike rates are calculated in time bins of 10 – 200 ms, further demonstrating a lack of fine temporal spiking responses (Figure 3.2).

The unique trajectories of onset and offset codes observed in Figure 3.7b could in principle result from complementary F0 tuning during onset and offset responses to our complex sounds (i.e. offset responses could simply be rebound inhibition from onset responses, and therefore show inversed F0-tuning of onset responses). If this were the case, we would expect a high degree of redundancy between onset and offset responses, and the offset responses would contain little additional information for decoding pitch. To test this, we trained a F0-decoder separately for each timbre on the population spike count responses either during the onset window (total spike count from 0 – 200 ms after sound onset) or during the offset window (0 – 200 ms after sound offset). The fitted decoder was then tested on held-out responses either from the same response window (“within window”) or responses from the other window (“across window”). Decoder performance is shown in Figure 3.7c. Performance was generally poorer when kernels were used across windows, rather than within windows (Figure 3.7c; 2-way ANOVA, $F(1,1540) = 371.0$, $p = 3.08 \times 10^{-74}$). Within window performance was superior to across window performance for every complex timbre tested (Tukey-Kramer HSD test, $p<0.001$), but for pure tones decoder

performance was similar within and between response windows (Tukey-Kramer HSD test, p = 0.94). These results suggest that population codes for pure tones are highly redundant across onset and offset responses, consistent with offset responses resulting from inhibition rebound. In contrast, the onset and offset windows convey different, non-redundant population-coded information about the pitch of complex sounds.

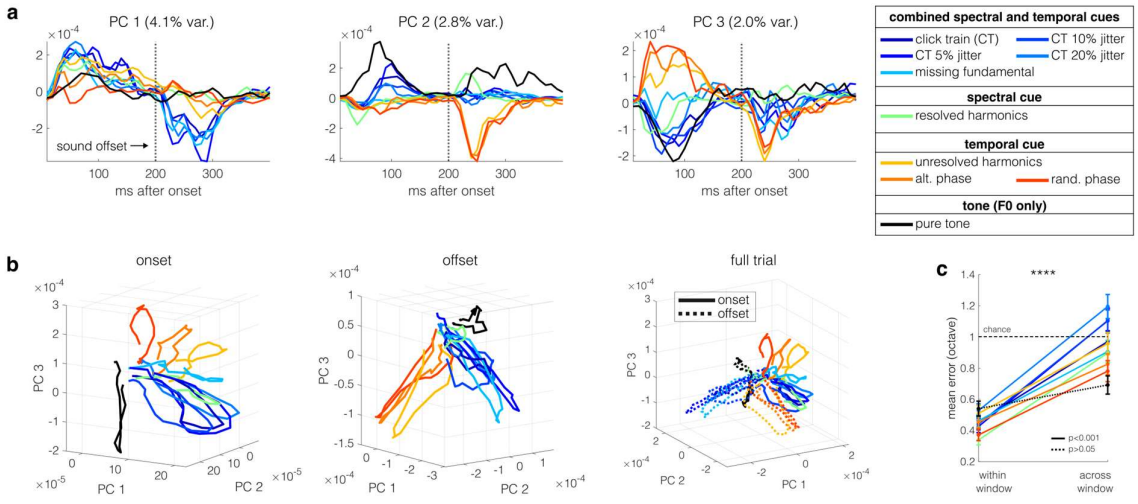

Figure 3.7. F0-decoding kernels separate by pitch-cue class in principal component space at sound onset and offset. a) First three principal components (PCs; left to right) of the PCA performed on all F0-decoding models together. Projections onto each PC summarize the activity of the entire neural population at a given time in the trial, isolated and coloured according to pitch-cue class (see Figure 3.1). **b**) The first three principal components as shown in **a**, plotted against each other and separated into trial onset (0-200 ms after sound onset, left panel), trial offset (0-200 ms after sound offset, middle panel), or the full trial (0-400 ms after sound onset, right panel) with onset and offset trajectories shown as solid and dotted lines respectively. The F0-decoding model trajectories occupy distinct areas of PC space depending on the pitch-cue category of the timbre, indicating that F0 information at the population level is encoded distinctly for each pitch cue class. Onset and offset trajectories also occupy distinct planes of PC space, suggesting that F0 may be represented differently at a sound's onset versus its offset. **c**) Relative performance (mean octave error) on held-out test trials of F0-decoding models trained and tested using either onset responses or offset responses ("within window"), versus models trained with onset responses and tested on offset responses or vice versa ("across window"). Errors were significantly lower when the models were tested across windows versus within windows (2-way ANOVA, ****p<0.0001), indicating that F0 representations are different at sound onset and offset. This is true for each timbre individually (Tukey-Kramer HSD test, p<0.001, solid lines) apart from pure tones (Tukey-Kramer HSD test, p>0.05, dashed line).

C. Different neural populations encode different pitch cue classes

The above analyses show that the F0 of different pitch cue classes are represented by different population responses. We next aimed to identify the source of this diversity. The F0 of different pitch cues may be carried by different subpopulations of neurons, or by the same group of neurons that are "read out" differently (i.e. have distinct F0 tuning curves for different timbres).

We first measured how much each neuron contributes to the population decoding performance by summing the absolute value of its fitted weights across all time bins in the kernel. Neurons were then ranked according to this metric of how "informative" they were for each timbre, where neurons ranked in the top 10% were considered highly informative. F0 information was concentrated in a small proportion of the overall population, with only about 20% of neurons contributing a comparable amount of information to the most informative neuron (Figure 3.8). We found that decoders trained on timbres within the same pitch cue class shared a higher proportion of informative neurons than across-class comparisons. For example, approximately 46% (50/109) of the most informative neurons were shared across click trains that were unjittered and 5% jittered, but only around 25% (27/109) of the neurons with largest weightings were shared between click trains and pure tones. This suggests that the key neuronal subpopulations that encode F0 for different pitch cue classes are more disjoint than those for similar pitch cues.

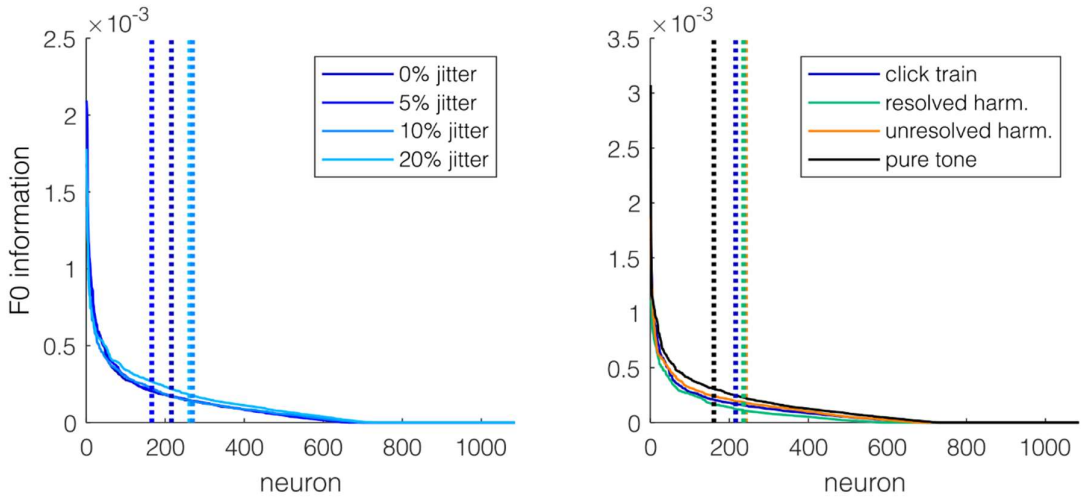

Figure 3.8. Distribution of the F0 information contributed by individual neurons. Each plot shows the amount of F0 information in each neuron in the population, where F0 information is defined as the sum of the neuron's unsigned weights in the decoding kernel. Colours indicate the timbres used in decoding (left: click trains, right: representative stimuli from each pitch cue class). Dashed lines show 90% of the most informative neuron's sum of weights.

The proportion of informative neurons shared between decoders for different timbre pairs is visualized as a matrix in Figure 3.9a, to facilitate comparison to the F0 generalization matrix in Figure 3.5b. Unexpectedly, the unresolved harmonic decoder shared many high-ranking neurons with most broad spectrum sounds (Figure 3.9b), although we previously found that these decoders do not generalize well between one another (Figure 3.5b). This may suggest that neurons can contribute to decoding the F0 of both sounds, but hold different F0-tuning curves between these pitch classes. We explored this possibility in Figure 3.9b, where the average correlation of F0-tuning is plotted for shared informative neurons, across all timbre pairings. As predicted, shared neurons showed more similar F0 tuning curves within a pitch cue class, but different F0 tuning curves across pitch classes. Interestingly, neurons that were informative for tones showed similar tuning for the F0 of complex sounds only if there was energy present at F0 (i.e. click trains). Together, these results show that even if neurons have different F0 tuning across pitch cue classes, they can contribute to population level codes for multiple classes.

As there were a substantial number of shared informative neurons across timbres, we asked how many of these top weighted neurons are needed to accurately decode F0. For each timbre, neurons were again ranked by informativeness (summed absolute decoder weights). Neurons were added one at a time to the decoding pool in

decreasing order of informativeness, and decoder performance was assessed. We found that approximately 20 of the most highly weighted neurons were sufficient to decode F0 at 90% of the accuracy of the entire population (Figure 3.9c, left).

Finally, we asked if F0 population decoding is critically dependent on the subset of neurons that are highly informative by performing a complementary analysis where we started with the whole population and removed neurons from the decoding pool one-by-one in decreasing order of informativeness. We found that decoding error increases nearly linearly with sequential removal of informative neurons, and the population retains enough information to decode pitch better than chance even after the 200 most informative neurons have been removed (Figure 3.9c, right). This is particularly surprising given that the vast majority of decoding weights are held by the ~200 most important neurons (Figure 3.8). Together, this suggests that small subsets of neurons capture the F0 information of the entire population, but there is a high degree of redundant F0 information in the population.

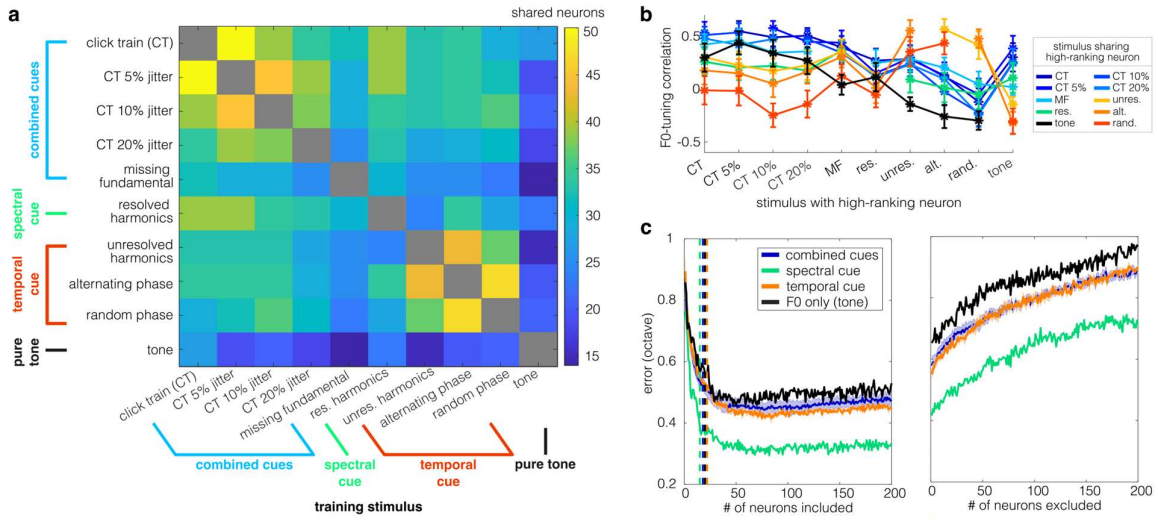

Figure 3.9. Different pitch cue classes share highly informative neurons. a) Matrix colours report the number of the top 10% (109) most informative neurons that are held in common between any two timbres. Two timbres share more neurons when they contain the same pitch cue (e.g. upper-lefthand yellow boxes corresponding to broad spectrum sounds). Sounds with temporal cues often share informative neurons with broad spectrum sounds as well. Spectral cues and pure tones do not share many informative neurons across pitch class. **b**) We compared the similarity of the F0-tuning (trial-averaged spike rates across each of 15 F0s presented) across different timbres in individual highly informative neurons by taking the correlation between the F0-tuning curves for the two timbres sharing a given high-ranking neuron. The average F0-tuning correlation amongst all shared neurons between any two timbres is shown as an asterisked dot along the coloured line corresponding to a single timbre. Neurons that rank highly for sounds with combined cues tend to have consistent F0-tuning (higher tuning correlation) for all sounds with combined cues, but have different F0-tuning (lower tuning correlation) for sounds with only temporal pitch cues. The inverse is true for neurons that rank highly for sounds with exclusively temporal cues. High-ranking neurons for sounds with exclusively spectral cues do not have common F0-tuning with any other timbre, while informative neurons for pure tones often have F0-tuning that correlates well with sounds with combined cues, but no other category. **c**) In the left plot, F0-decoding error is shown for decoding models trained on sequentially larger populations of neurons, starting from the most informative neuron and adding sequentially less informative neurons one by one. The right plot shows F0-decoding error for models trained on sequentially smaller populations of neurons, removing the most informative neurons first, followed by increasingly less informative neurons. Errors are shown for all timbres grouped by pitch cue class, with shaded areas indicating standard error of the mean across all timbres in one category (see Figure 3.1). Dotted vertical lines on left indicate the number of neurons needed to achieve 90% percent of the maximum accuracy for that pitch cue class.

3.6. Discussion

Harmonic spacing and temporal periodicity provide two acoustic cues to derive the F0 of a sound, and both humans and animals utilize both cues when making pitch judgements. Neurons in primary auditory cortex have been found to extract F0 from either spectral or temporal cues (Bendor et al., 2012; Fishman et al., 2013; Steinschneider et al., 1998), and a small subset of neurons encode F0 using either cue (Bendor and Wang, 2010, 2005; Tarka et al., 2025). Given the key role of primary auditory cortex in pitch perception (Stewart et al., 2006; Tramo et al., 2002; Warrier and Zatorre, 2004; Whitfield, 1980; Zatorre, 1988), it is not clear how these dual encoding mechanisms give rise to a unified perception of pitch across both spectral and temporal cues. Here, we investigated whether population-level representations of F0 show invariance across pitch cues which more closely align with the experience of pitch.

We used high-density microelectrode recordings in ferret primary auditory cortex to record 1085 isolated neurons while presenting timbres that contained either spectral or temporal cues for pitch, or both, and pure tones (Figure 3.1). Decoding models were trained for each timbre separately to predict the F0 of each sound using the weighted sums of the population activity (Figure 3.3). We found that F0 was linearly decodable for each timbre independently, but the decoding model for any given timbre only generalized to other timbres with the same pitch cues, or combination of pitch cues (Figure 3.5). Neural responses at both sound onset and offset contained information about the sound's F0, and population codes for F0 in these two windows were different from each other, particularly for complex sounds (Figure 3.7). Finally, small subsets of approximately 20 neurons were sufficient to capture the F0 information of the entire population, and these highly informative neurons were shared within but not across cue classes (Figure 3.9).

A. Dual pitch processing in auditory cortex

Our findings suggest that resolved harmonic and unresolved temporal cues are encoded differently across populations of neurons in primary auditory cortex, extending previous findings in ferrets and macaques that single neurons encode F0 selectively from either spectral or temporal cues (Bendor et al., 2012; Fishman et al., 2013; Steinschneider et al., 1998). This is in contrast to "pitch neurons" identified in

marmosets and ferrets which are tuned to the F0 of any periodic sound regardless of acoustic structure (Bendor and Wang, 2010, 2005; Tarka et al., 2025). Both resolved and unresolved harmonics contribute to the perception of pitch (Shackleton and Carlyon, 1994), but humans and marmosets preferentially employ resolved harmonics to make pitch judgements (Carlyon, 1996b, 1996a; Shackleton and Carlyon, 1994; Shofner and Campbell, 2012; Song et al., 2016), while other animals such as ferrets and gerbils rely more on unresolved harmonics (Klinge and Klump, 2010, 2009; Osmanski et al., 2013; Shofner and Chaney, 2013; Walker et al., 2019). We show that population representations of F0 reflect this perceptual asymmetry, leaving the neural underpinnings of unified pitch perception an open question.

We found that decoding models generalized particularly poorly to timbres with disrupted periodicity (alternating and random phase unresolved harmonics), underscoring the importance of temporal cues in ferrets' perception of pitch previously described (Walker et al., 2019). It is striking that population codes for pure tone frequency did not generalize to any other timbre (Figure 3.5), given that pure tones evoke a percept of pitch that is as salient as periodic complex sounds. This observation, however, aligns well with previous findings that the F0 of complex sounds and frequency are encoded differently in single neurons, even if these neurons respond in an F0-dependent manner to both complex sounds and tones (Tarka et al., 2025). Together, these findings suggest that primary auditory cortex integrates information across multiple frequency channels to represent F0, and this integration occurs separately for spectral (place coded) and temporal (rate coded) pitch cues.

B. F0-encoding in single neurons

While the dual-extraction model of pitch coding has been well-established in single neurons across several species (Bendor et al., 2012; Fishman et al., 2013; Steinschneider et al., 1998; Tarka et al., 2025), it is not clear how these single units combine to form population codes for F0. We found that subsets of approximately 20 neurons captured nearly all the pitch information in the overall population for a given timbre (Figure 3.9c), in line with previous findings that F0 can be decoded from small ensembles of 3-61 neurons at the same accuracy as ferrets' behavioural discrimination (Bizley et al., 2010). Furthermore, the removal of these highly informative neurons did not reduce decoding performance to chance, even after the

200 most informative neurons were removed (Figure 3.9c), suggesting a high degree of redundant F0 information across the cortical population. Neurons in these highly informative subsets were often shared between decoding models for timbres in the same pitch cue class, but were not shared between cue classes (Figure 3.9a). This suggests that neurons that encode F0 irrespective of timbre as reported in marmosets and ferrets (Bendor and Wang, 2010, 2005; Tarka et al., 2025) contribute less F0 information to the overall population than neurons specialized for a single pitch cue.

C. Sound offset responses

Our results showed that F0 was linearly decodable from neural responses to both sound onset and offset, and the population codes in these two windows were not linearly related and contained non-redundant information (Figure 3.7). As with onset responses, offset responses also encoded F0 in a cue-dependent manner. The role of offset responses in auditory perception and behaviour is not well-defined, but the auditory cortex has been identified as the first auditory station where offset responses are highly informative about stimulus identity (Lamothe et al., 2025). Though offset responses are often thought to simply reflect a release from inhibition following an excitatory onset response, studies have found a wide variety of relationships between onset and offset responses, ranging from highly similar (Fishman and Steinschneider, 2009; Ramamurthy and Recanzone, 2017) to largely independent (Scholl et al., 2010; Sollini et al., 2018). Our findings suggest that onset and offset responses are independently informative about F0, advancing theories that onset and offset responses arise through independent circuits in the auditory system, as in the visual system (Kopp-Scheinpflug et al., 2018).

Interestingly, onset and offset F0 codes were only distinct for complex sounds: pure tone decoding models generalized between onset and offset windows (Figure 3.7c), again suggesting that complex sounds evoke distinct circuits to pure tones in auditory cortex. These findings are the first evidence of distinct cortical offset response mechanisms for complex and pure tones, suggesting that future studies on offset responses may identify novel circuit dynamics underlying these responses by employing a battery of complex sounds to compare with existing findings from pure tones. Informative offset responses for complex sounds specifically may support the perceptual role sound terminations play in perceptual grouping (Bregman, 1990).

Future work could explore the role of offset responses by employing complex sounds with a range of amplitude modulation rates and depths as a parametrized proxy for speech terminations.

D. Summary

These findings establish that F0-encoding across populations of auditory cortical neurons is specific to spectral and temporal cues for pitch, and cortical representations of the F0 of complex sounds share few commonalities with pure tone frequency encoding. This supports the idea that pitch is encoded using two strategies at the level of auditory cortex, despite the perceptual unity across the two pitch cues. Furthermore, it highlights the necessity to prioritize the use of naturalistic complex sounds in studies on auditory cortex, as the circuits underlying stimulus encoding and behavioural correlates may be fundamentally distinct for pure tones and complex sounds.

4. Pitch as a selective listening cue

4.1. Rationale

In Chapters 2 and 3, I have reported single neurons and population codes for pitch. Given the critical role of pitch in selective listening (Ward et al., 2026), we next wished to ask how these cortical F0 codes may alter under active listening conditions. The lack of viable behavioural paradigms for studying selective listening in animal models prompted us to design a novel task design for ferrets that allows parametric control of pitch as the dominant acoustic cue available for auditory segregation. While neural data are not reported in this chapter, the behavioural design established here could be easily employed in conjunction with chronic neural recordings in future.

4.2. Contributions

Veronica Tarka conceptualized the experiment, wrote software for all data analysis, and produced all figures. Hardware and software for behavioural training were designed and implemented by Veronica Tarka, Artem Diuba, and Kerry Walker. Behavioural training and testing were performed by Veronica Tarka and Mya Hesketh-Bream. Rebecca Norris kindly supplied the illustrated graphic of a ferret.

4.3. Introduction

Natural environments are acoustically busy with many overlapping sounds from independent sources. Our ability to attend to one of these sound sources while ignoring others is described as *selective listening*. The brain is thought to achieve selective listening by grouping sound features from the same source into an auditory object, and enhancing or suppressing representations of this object in the auditory system depending on attentional demands. Pitch is among the strongest features used for selective listening in humans (Grossberg, 1996; Micheyl and Oxenham, 2010), but the role of pitch in auditory attention in non-human animals is largely unexplored.

Pitch is the tonal quality of sound on a low-to-high scale (American National Standards Institute, 1994), which we perceive at its fundamental frequency (F0). Sounds that evoke a pitch have waveforms with short, regularly repeating segments

(i.e. are *periodic*), and this repetition rate corresponds to F0. Periodic sounds are composed of frequency components (*harmonics*) that are integer multiples of F0. Many animals including macaques (Brosch et al., 2004), marmosets (Osmanski et al., 2013), gerbils (Klinge and Klump, 2009), and ferrets (Walker et al., 2019, 2009) have been shown to perceive pitch, and, like humans, are able to discriminate and order the F0 of sounds.

The perceptual segregation of overlapping sounds is strongly influenced by pitch in human listeners. Pitch separation makes it easier to identify simultaneously presented vowels (Assmann and Summerfield, 1994, 1990; Culling and Darwin, 1993; Zwicker, 1984) and follow a target sentence in a multi-talker environment (Brokx and Nooteboom, 1982; Brungart et al., 2001; Summers and Leek, 1998). Auditory streaming in humans is most often studied using ABA- streaming paradigms in which temporally interleaved low (A) and high (B) pitched tones are presented to form a repeating ABA- sequence. When A and B are close in frequency, listeners perceive a single integrated auditory stream with a "galloping" rhythm, while large frequency differences produce a percept of two segregated A and B streams (Bregman, 1990). This effect can be seen both for pure tone frequency (Bregman, 1990) and complex tone F0 (Bregman et al., 1990; Singh, 1987).

Few studies have explored selective listening in behaving animals due to the challenge of adapting the self-reporting paradigms used in human subjects into objective measurements of perception in animals. Frequency-based streaming behaviour has been demonstrated in ferrets (Ma et al., 2010) and monkeys (Izumi, 2002), but the exclusive use of pure tone stimuli in these studies limits insights into natural sounds such as speech. Here, we adapted the classical ABA- streaming paradigm to create a novel behavioural task that captures pitch-based auditory streaming in ferrets. This would allow us to determine if ferrets, like human listeners, segregate sound streams based on pitch cues. We found that the animals could use the pitch of both pure tones and complex sounds to segregate competing streams. Overall performance was worse when pitch information was degraded, and the animals were better able to segregate streams formed of complex tones than pure tones. This is the first evidence that non-human animals use the pitch of complex sounds for selective listening.

4.4. Methods

Three adult female ferrets (*Mustela putorius*: Marshall BioResources, UK) were trained in this study. Ferrets were trained in two sessions daily in blocks of 5 days, with a minimum of 24 hours of *ad libitum* water between training blocks. The animals were given water as a positive reinforcer during training. A minimum of 60 ml/kg of water was made available to each ferret daily, with any water not consumed during training supplied as wet food. All animal procedures were undertaken under UK Home Office license and approval from the local ethical review committee of the University of Oxford.

A. Training apparatus

Ferrets were trained in custom-built chambers with two wall-mounted spouts: a central trial-initiation spout and a reward spout to the right that delivered water. The animals' responses ("nose-pokes") at the central spout were detected via interruption of an infrared light-emitting diode and photodetector each positioned at either side of the spout's mouth. A loudspeaker (Visaton FRS 8) above the central spout delivered auditory stimuli, and all sounds were calibrated to produce a flat frequency response (± 2 dB) from 0.2 to 12 kHz through this speaker. Training chambers were fitted with acoustic foam on all flat surfaces apart from the floor to prevent reverberation. Custom MATLAB software (MathWorks, Inc.) with a TDT RP2 (chamber 1) or RM1 (chamber 2) signal processor and pre-amplifier (Tucker-Davis Technologies, Alachua, FL) to deliver auditory stimuli and monitor behaviour.

B. Stimuli

We presented a sequence of alternating A and B tones on each trial, with each tone 100 ms in duration and 50 ms of silence between tones (Figure 4.1). B tones were presented at a fixed F0 (600, 750, 900, or 1050 Hz; or only a subset of these for some task versions) for a single session, and A tones were presented at 3, 6, 11, or 14 semitones below the B tone (or a subset of these for some task versions). There are 12 semitones in one octave in the Western musical scale, and a difference of 1 semitone corresponds to the distance between two adjacent musical notes (e.g. between E and F). All tones were presented at 70 dB SPL_A, and had 5 ms onset and offset cosine ramps.

Oddball tones were sinusoidally amplitude modulated (SAM) at 40 Hz. SAM manipulations of the B tones were targets (B′) while SAM manipulations of A tones were distractors (A′). When the modulation rate is low (below ~60 Hz), amplitude modulation affects the "fluctuation" or "roughness" of a sound, but does not produce a secondary pitch in human listeners (Joris et al., 2004; Miller and Taylor, 1948). Given that our modulation rate was below this threshold and at the lowest end of the ferrets' hearing range (Kelly et al., 1986), the SAM modulation should produce a rhythmic "flutter" in the carrier tone for ferret listeners without introducing a secondary pitch percept at the modulating frequency.

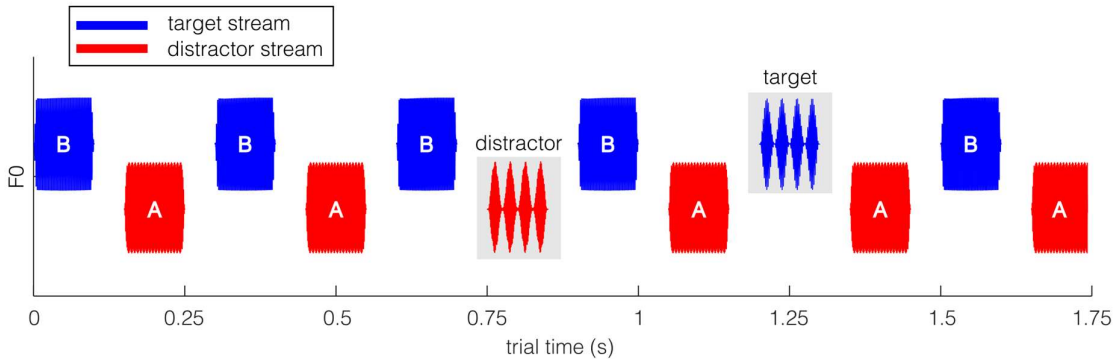

Figure 4.1. Example trial stimulus. On each trial, we presented two streams of 100 ms tones. Higher pitched B tones were presented at a constant F0 in a single session, and lower pitched A tones were presented at 3, 6, 11, or 14 semitones below the B tones. Ferrets were trained to respond within 500 ms of an oddball manipulation (sinusoid amplitude modulation) of a B tone to receive a water reward. Distractor oddballs were also presented in the A stream, which the animals were trained to ignore.

Three types of stimuli were used: 1) pure tones; 2) complex tones with fixed harmonic number; and 3) complex tones with fixed bandwidths (Figure 4.2). Any given session contained A and B tones drawn from only one of these categories. Pure tone stimuli were comprised of a single frequency component corresponding to the F0 of the tone (Figure 4.2a). Complex tones with fixed harmonic number always contained exactly 8 harmonics of F0 (Figure 4.2b). As a result, as the F0 difference between A and B tones increased, the spectral separation between the two sounds also increased. Complex tones with fixed bandwidth contained all harmonics of F0 within 0.2-12 kHz. A and B tones of this type therefore contained different numbers of harmonics, but covered the same spectral range (Figure 4.2c).

To selectively examine the role of pitch in the task, we created a control stimulus in which each harmonic frequency in the complex tones were independently jittered up or down in frequency by a random amount between 0-50% of F0 (Figure 4.9a). The F0

of the tone was also removed. This manipulation has been previously shown to impair pitch discrimination in humans (McPherson and McDermott, 2018). Each harmonic was adjusted independently from all other harmonics, but jitter values were constrained to produce a minimum of 30 Hz between neighbouring harmonics to prevent perceptible beating. A and B tones were jittered independently to each other, and a new, random jitter manipulation was applied on every trial.

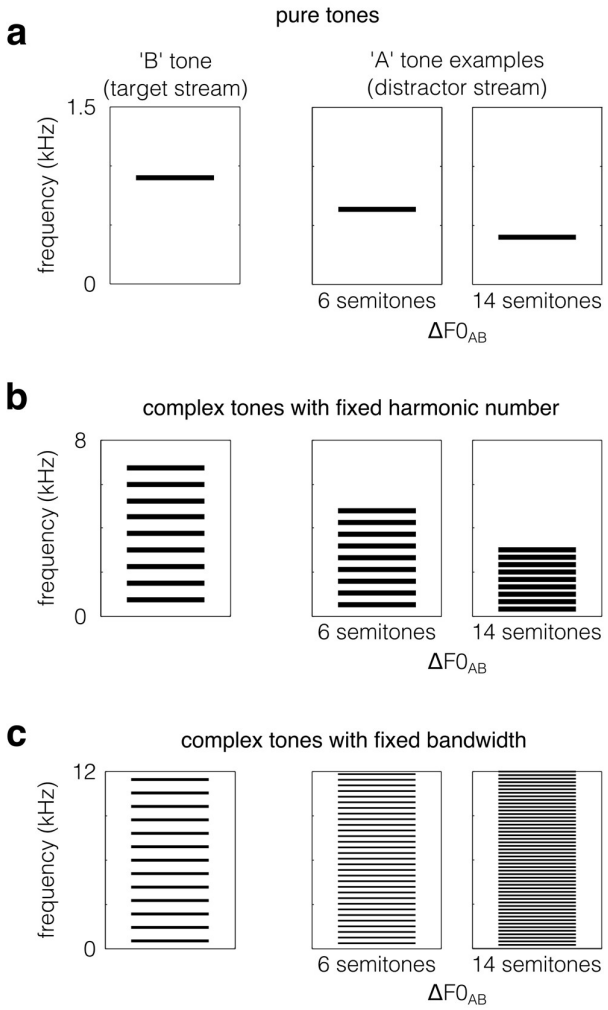

Figure 4.2. Stimulus examples for each task version. a) Pure tone stimuli contained energy at F0 only (horizontal bar). An example B tone is shown in the left panel, and two example A tones are shown in the middle and right panels. In addition to the 6 and 14 semitone pitch separations pictured here, we presented A tones at 3 and 9 semitones below $F0_B$. **b**) Example A and B tones as described in **a**, but for complex tones with fixed harmonic number which contained exactly 8 harmonics of F0. As the pitch separation between A and B tones increased, the difference in spectral range between the tones also increased. **c**) Example A and B tones as described in **a**, but for complex tones with fixed bandwidths. Tones in this task version contained all harmonics within 0.2-12 kHz, thereby covering the same spectral range regardless of the pitch difference between A and B tones.

C. Behavioural paradigm

Three naïve ferrets were trained to perform a go/no-go oddball detection task, with 7-15 months of training over 5 stages, and 3-9 months of behavioural testing. All training stages used complex tones with fixed harmonic number in 1 ferret (F2404), and fixed bandwidth in the other 2 ferrets.

Stage 1 (training): Animals were trained to nose-poke the central spout for a water reward delivered at the right spout. Initial sessions rewarded pokes of any duration, and the minimum poke duration was gradually increased over sessions until the animal was consistently maintaining their nose in the spout for at least 2 seconds. Minimum hold durations were randomized across a range of 500 ms to increase engagement and prepare the animals for variable reward timings in later training stages.

Stage 2 (training): We next introduced a SAM tone (70 dB, 100 ms duration) at one of three possible time points following trial initiation (600, 900, or 1200 ms). Ferrets initiated trials with a 200 ms nose-poke at the central spout, and were required to remove their nose from the spout ("release") within 500 ms of the SAM tone onset to receive a water reward at the right spout. A response within this window was termed a *hit*. A sequence of two or more hits in a row was rewarded with a doubled water reward size. Responses before sound onset were classified as *aborts*, while responses after the 500 ms window were classified as *misses*. An abort or miss was followed by a broadband noise burst (70 dB, 40 ms duration) to indicate an incorrect response, and no water was given. In 20% of trials, no tone was presented (*catch trials*), and animals were rewarded for maintaining their nose in the spout for 1700 ms (i.e. longer than the response window for the latest possible tone). Correct responses on catch trials were termed *correct rejections* and rewarded with water at the right spout. Following 3 consecutive sessions of performance >50%, the animal progressed to training stage 3.

Stage 3 (training): A sequence of repeating tones (100 ms, with 200 ms silence between tones) was presented on each trial, where one tone in this sequence was the target tone: a SAM tone at the same F0 as all other tones in the sequence. As in stage 2, the target SAM tone could occur at 600, 900, or 1200 ms after trial initiation, and a hit was achieved by releasing the spout within 500 ms from target tone onset (Figure 4.3a, left). False alarms, misses, and catch trials were classified as in stage 2. The

sequence of non-SAM tones was initially introduced at 35 dB with the target tone at 70 dB, in order to train the ferret to respond to the target. The level difference between target and non-target tones was gradually reduced over sessions by increasing the level of the non-targets to 70 dB, until the animal could discriminate the target using the SAM modulation without a level cue. Animals progressed to stage 4 following at least 3 consecutive sessions of performance over 50% correct responses without an intensity difference between target and non-target tones.

Stage 4 (training): Here, we added a second stream of tones ("A stream"), with an F0 19 semitones below the stream described in stage 3 ("B stream"). Higher pitched tones (B tones) and lower pitched tones (A tones) were interleaved, with each tone 100 ms in duration and 150 ms of silence between tones (Figure 4.1). As in stage 3, target tones were SAM B tones occurring at 600, 900, or 1200 ms, and responses were classified as in stages 2 and 3. However, to reduce overlap between response windows across the three target bins and discourage guessing, the response window was shortened to begin 150 ms after tone onset, reducing the response window duration to 350 ms. The A tones were initially introduced at 35 dB while target and non-target B tones remained at 70 dB, and the level difference between the two streams was reduced by increasing the level of A tones over sessions until the two streams were presented at 70 dB.

Stage 5 (testing): Testing began after the animal achieved >60% correct responses over 3 consecutive sessions in stage 4. On each trial, the animal was presented with a temporally interleaved sequence of A and B tones (both 70 dB), as in stage 4. B tones were presented at a single F0 for each session ($F0_B$), drawn at random from the set of 600, 750, 900, and 1050 Hz (or a subset of these for some task versions). The F0 of A tones ($F0_A$) varied randomly trial by trial, and were presented at either 3, 6, 11, or 14 semitones below $F0_B$. On each trial, one B tone was the target SAM tone, and this tone could occur at 600, 900, or 1200 ms, as in stages 2-4 (Figure 4.3b). Some trials also contained a SAM A tone (i.e. a distractor oddball), which could occur at any time before the target oddball that allowed non-overlapping response windows between target and distractor tones (Figure 4.3b). Withdrawal from the spout within the response window for the distractor oddball (150-500 ms after tone onset) was classified as a *false alarm (FA)* and was followed by a broadband noise burst (70 dB, 40 ms duration) without water reward (Figure 4.3a, middle left). Spout withdrawals outside the response windows for the target and distractor oddballs were termed

aborts (Figure 4.3a, right), while withdrawal after the target response window was termed a *miss* (Figure 4.3a, middle right). Aborts and misses did not result in a water reward and both were followed by broadband noise.

As in stages 2-4, 20% of trials contained no target oddball (catch trials; Figure 4.3b, right column). In these trials, ferrets were rewarded if they maintained their nose in the spout for at least 1700 ms after trial initiation, and a correct response on these trials was classified as a *correct rejection*. A subset (75%) of catch trials contained distractor oddballs.

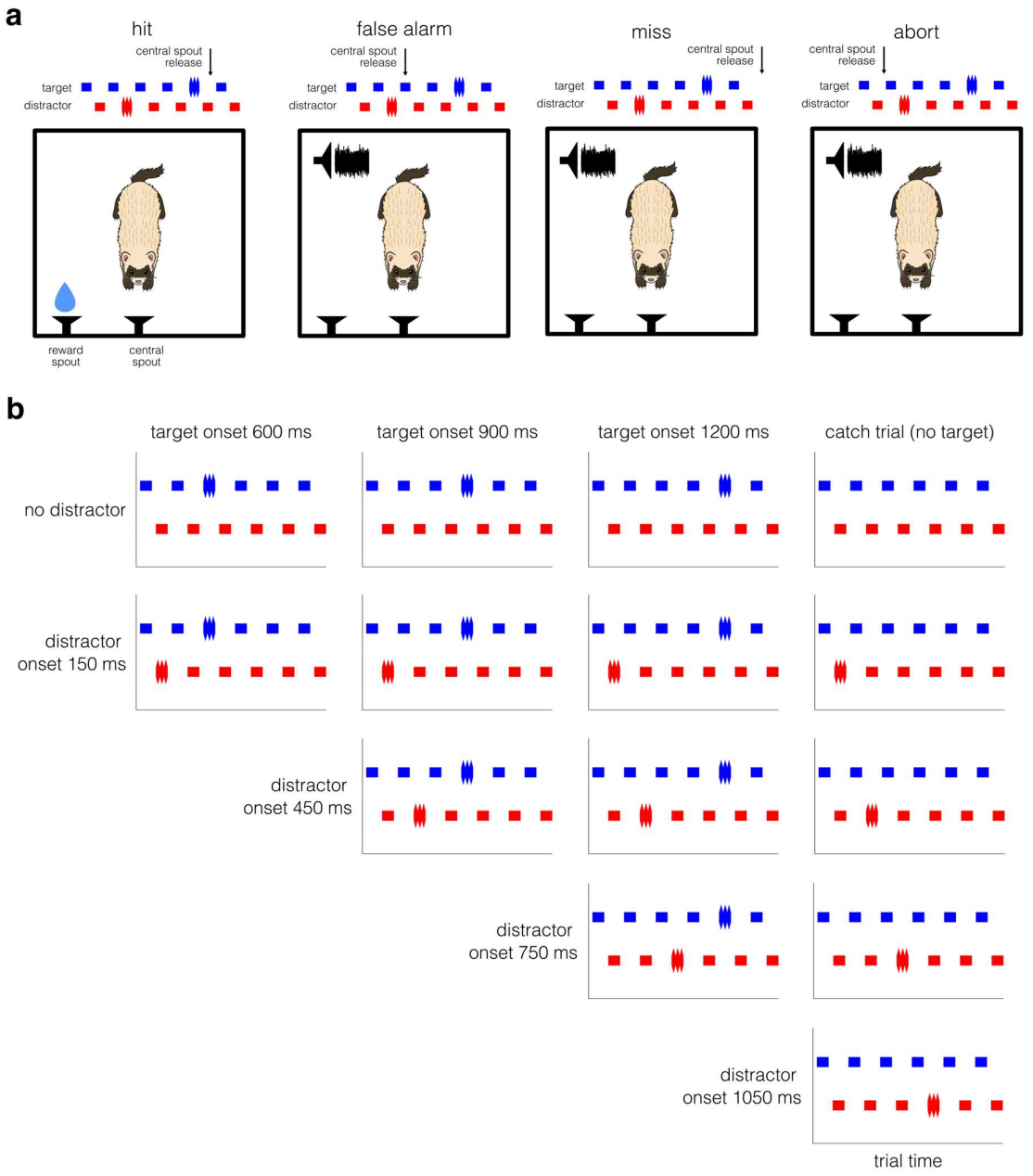

Figure 4.3. Example behavioural responses and trial types. **a**) Classifications of the animals' possible responses. Releasing the water spout within 500 ms of the target onset was classified a hit and was rewarded with water from the left spout (left panel). A response within 500 ms of the distractor onset was termed a false alarm (middle left panel). A response after the response window of the target was a miss (middle right panel), and a response before the onset of the first oddball was an abort (right panel). False alarms, misses, and aborts resulted in a short noise burst with no water reward. **b**) All possible combinations of oddball and distractor oddball onset times. Target oddballs could occur at one of three times (600, 900, or 1200 ms) or not at all (catch trials). Distractor oddballs could occur at any time before the target that allowed non-overlapping response windows for target and distractor.

D. Data analysis

Testing sessions were grouped into testing blocks of 2 consecutive days (4 sessions) of behavioural testing to analyse performance. This was done to pool enough trials to include all response types (hit, miss, false alarm, abort) across all trial types (4 target onset times x 3 semitone separations) to estimate d prime. D prime was defined as:

$$d' = z(H) - z(FA),$$

[*eq. 4.1*]

where z is the z-transform from the standard normal distribution, H is the hit rate, and FA is the false alarm rate.

4.5. Results

Three ferrets were trained in a go/no-go task to detect target oddballs and ignore distractor oddballs in a stream of alternating low (A) and high (B) pitched tones (Figure 4.1). The F0 of the B tones ($F0_B$) was held constant in a session, while the F0 of the A tones ($F0_A$) was either 3, 6, 11, or 14 semitones below $F0_B$ on any given trial. Target oddballs were sinusoidally amplitude modulated (SAM) B tones, while distractor oddballs were SAM A tones. The animals were trained to respond within 500 ms from the target onset and ignore distractor oddballs to receive a water reward.

In humans, the probability of segregating two auditory streams has been shown to increase with increasing pitch separation between the streams (Bregman, 1990; Bregman et al., 1990; Singh, 1987) and oddball detection within an ABA- stream is easier with increasing stream segregation (Micheyl et al., 2005a). We therefore anticipated that distractor oddballs would be easier to ignore as the perceptual distance between target and distractor streams increased, reflecting the ferrets' experience of the sequence of tones as two segregated A and B streams at greater F0 differences. This task therefore captures the animal's ability to segregate competing auditory streams, where higher task performance reflects better stream segregation.

We presented 3 versions of this task which varied in the acoustic structure of the tones: (1) pure tones, (2) complex tones where A and B tones each had 9 harmonic components ("*fixed harmonic number*"), and (3) complex tones where A and B tones contained all harmonics under 12 kHz ("*fixed bandwidth*") (Figure 4.2). All 3 animals

were able to perform the 3 versions of the task (Figure 4.4). Reaction times were clustered at the beginning of the target's response window for all task versions (Figure 4.4a), and d-prime was on average above 1 for all animals in all task versions (Figure 4.4b). Correct response rates were significantly different between the 3 versions (2-way ANOVA; $F(2,431) = 2.49$, $p = 3.65 \times 10^{-18}$), with significantly better performance in the 2 task versions with complex tones than the pure tone version (Tukey-Kramer HSD test, $p < 1 \times 10^{-8}$). Correct response rates were also higher for complex tones with a fixed number of harmonics than with fixed bandwidth (Tukey-Kramer HSD test, $p = 1.18 \times 10^{-6}$).

$F0_B$ varied across sessions to prevent overtraining the animals to a specific target F0. Overall rates of correct responses were not significantly different across $F0_B$ for pure tones (Figure 4.4b, left) or complex sounds with fixed harmonic number (Figure 4.4b, middle). However, correct response rates were significantly different across $F0_B$ for complex sounds with fixed bandwidth, though only one animal showed significant post-hoc differences in performance across $F0_B$s (Figure 4.4b, right).

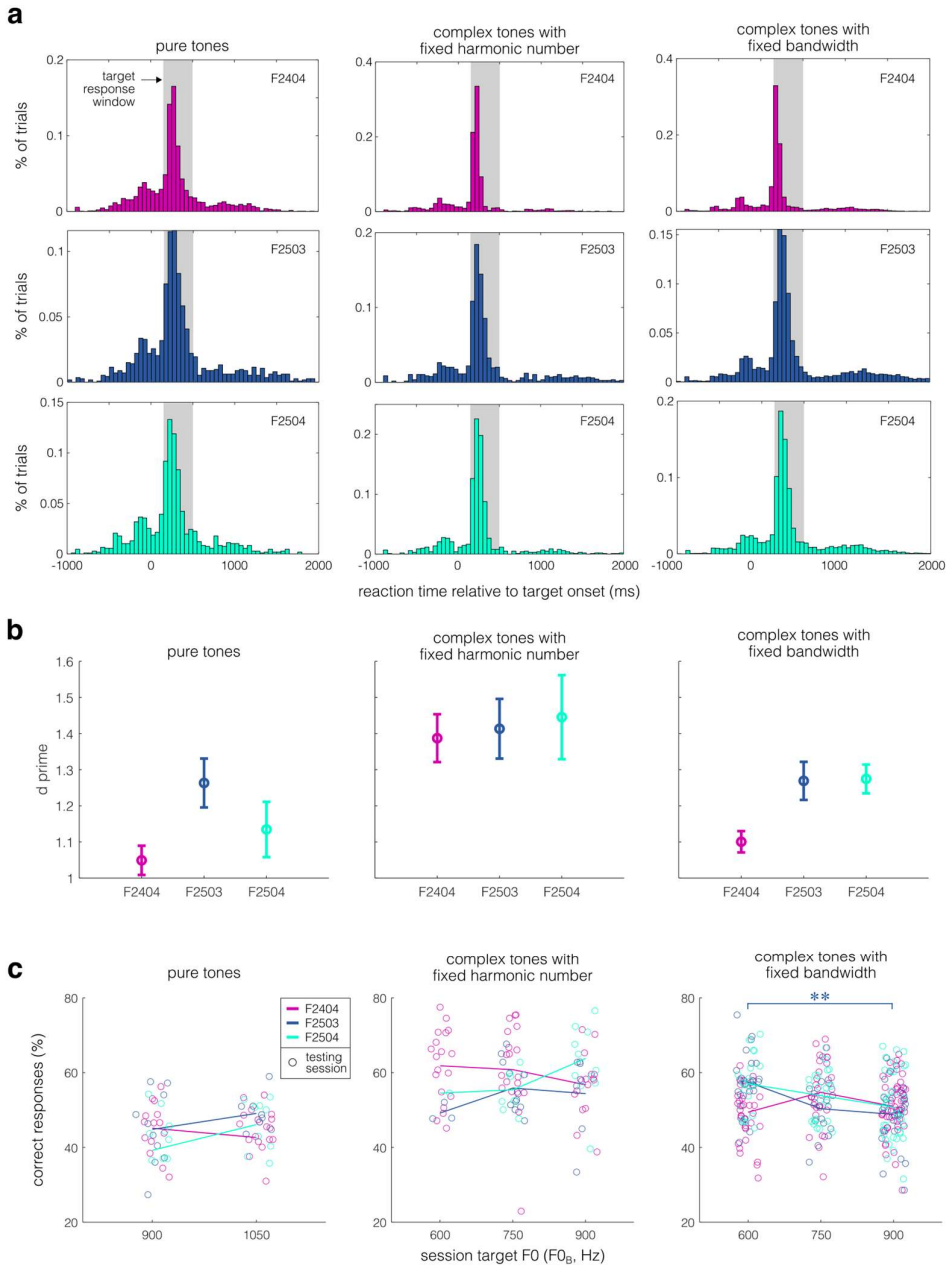

Figure 4.4. Reaction times and hit rates for the 3 task versions. **a**) Reaction times for the 3 animals (rows) for the 3 task versions (columns) are shown in each subpanel. Reaction times were aggregated across the 3 possible target onset times and are reported relative to the onset time of the target. **b**) Mean task performance (d prime) over all testing blocks is reported for each ferret in the three task versions (subplots left to right). **c**) Each subplot shows correct response rates for all testing sessions (open circles) across the different target F0s ($F0_B$) presented. Performance of the three animals is shown in different colours. Rates of correct responses were not significantly different across $F0_B$ for pure tones (left; 2-way ANOVA, $F(1,65) = 1.33$, $p = 0.25$) or complex tones with fixed harmonic number (middle; 2-way ANOVA, $F(2,88) = 0.3$, $p = 0.74$). Hit rates for complex tones with fixed bandwidth were significantly different (right; 2-way ANOVA, $F(2,263) = 7.4$, $p = 7.5 \times 10^{-4}$), but only one animal (F2503) had significantly different performance across $F0_B$ (Tukey-Kramer HSD test, 600 Hz versus 900 Hz, $^{**}p = 0.0042$).

A. Pure tones

In the pure tone version of the task, A and B tones were composed of a single frequency at F0 (Figure 4.5a). As described, $F0_B$ (target stream) was held constant in a session, while $F0_A$ (distractor stream) was one of 3, 6, 11, or 14 semitones below the target stream on any given trial. Performance across the 3 animals was significantly better at larger frequency separations (Figure 4.5b), indicating the animals found it easier to identify the target oddball and ignore the distractor oddballs when there was a larger frequency separation between them. This is consistent with the animals perceptually segregating the sounds with larger frequency differences. The performance advantage of larger frequency separations was most apparent when the target occurred in the latest time bin of the trial (Figure 4.5c). Trials with targets in the earliest time bin rarely resulted in false alarms (Figure 4.6). The stronger effect of semitone separation for targets in the latest time bin aligns with previous findings of a perceptual "build-up" of streaming in human listeners, wherein stream segregation becomes more likely over several seconds of stimulus presentation (Bregman, 1990). The effect of frequency separation on performance therefore became larger as perceptual segregation became more likely.

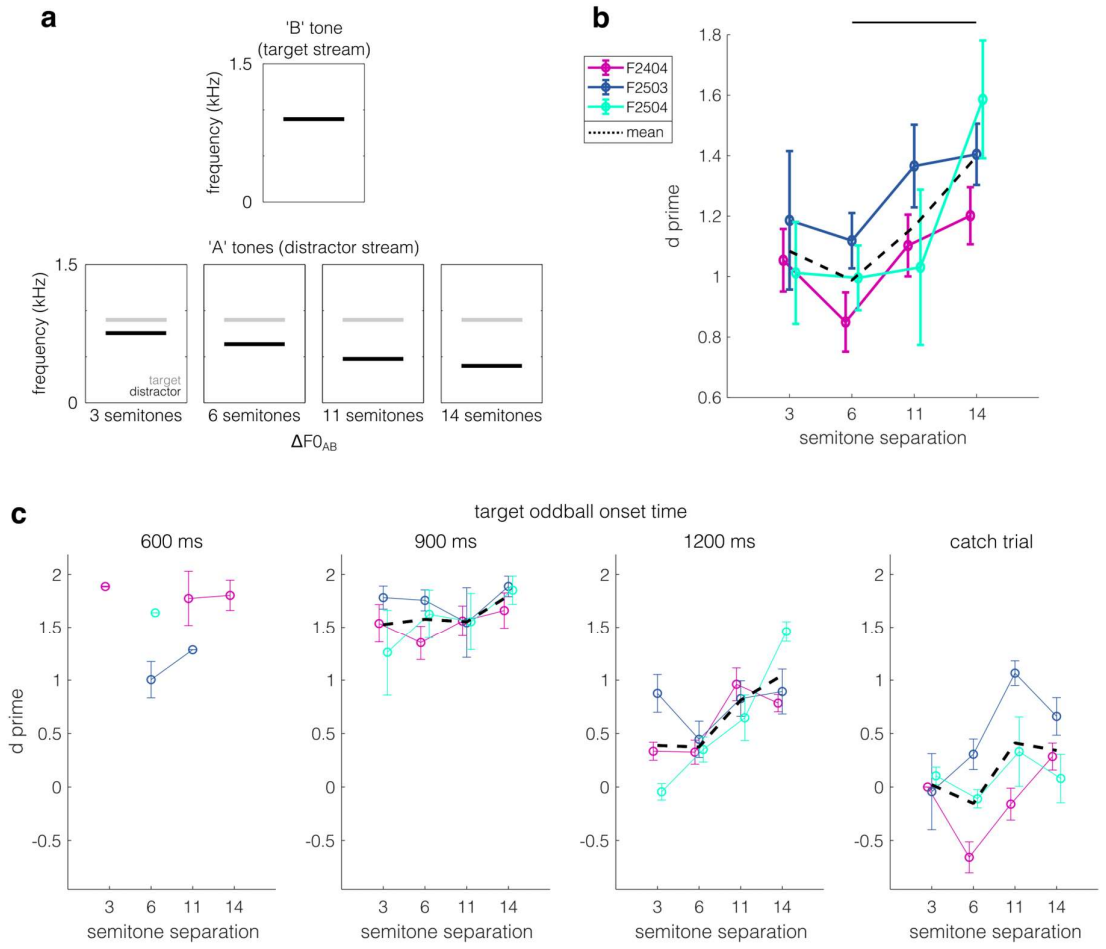

Figure 4.5. Stream segregation of pure tones. **a**) Schematic of stimulus tones. A (distractor) and B (target) streams were pure tones in this task version. $F0_B$ was held constant in a session, while $F0_A$ could be one of 3, 6, 11, or 14 semitones below $F0_B$ on a given trial. **b**) Performance (d prime) in the 3 animals (different colours) was significantly better with increasing pitch separation between streams (2-way ANOVA, $F(3,244) = 4.66$, $p = 0.0054$). Post-hoc comparison revealed a significant difference between performance at 6 and 14 semitones across animals (Tukey-Kramer HSD test, $p = 0.0034$; black horizontal line). Error bars show the standard error of the mean across testing blocks. **c**) Each subplot shows performance (d prime) across frequency separations for one of the 4 possible target onset times. Larger frequency differences between the competing streams provides the greatest advantage when the target is in the latest time bin (1200 ms, middle right plot). The sparsity of data in the leftmost plot reflects the absence of testing blocks in which d prime could be calculated due to a lack of false alarms in this early time window (Figure 4.6).

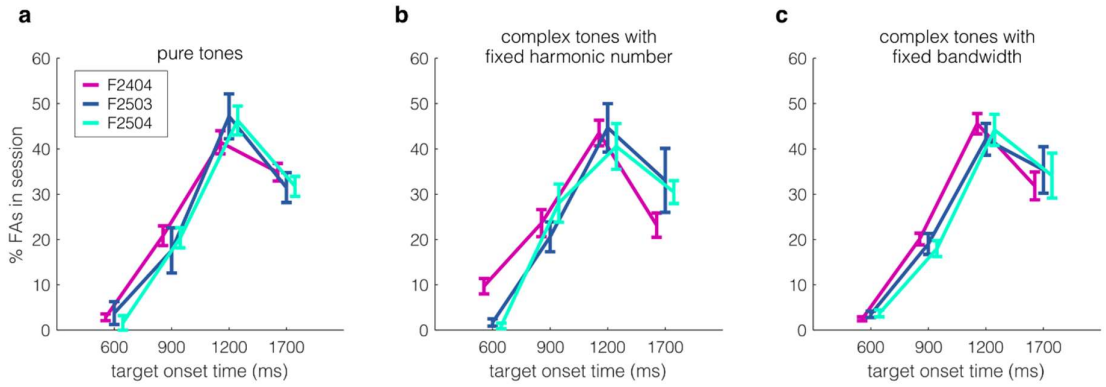

Figure 4.6. False alarm rates across target onset times. **a**) Each line reports the average percent of false alarm (FA) responses that occurred in a given target time bin across testing block for the pure tone task in one ferret. Percentages are separated by the target onset time. Error bars show the standard error of the mean. **b**) Probabilities of false alarm responses as reported in **a**, for complex tones with fixed harmonic number. **c**) False alarm probabilities as shown in **a**, for complex tones with fixed bandwidths.

B. Complex tones with fixed harmonic number

In this version of the task, A and B tones were complex sounds with 9 harmonics each (F0 and its first 8 harmonics) (Figure 4.7a). All other aspects of the task were as described previously. Because the absolute number of harmonics remained constant in all tones, the spectral separation of A and B tones also increased as the semitone separation increased (i.e. the frequency range covered by the tones became increasingly separated with greater pitch differences). This spectral bandwidth could provide an additional acoustic cue for sound segregation beyond pitch, as spectral envelope has been found to influence streaming in humans (Singh, 1987).

As with pure tones, performance in all 3 animals was significantly better with increasing semitone separation between the competing auditory streams (Figure 4.7b). Interestingly, performance was already significantly better at 6 semitones than 3 (Tukey-Kramer HSD test; $p = 6.41 \times 10^{-5}$) for these stimuli, whereas performance at these two pitch separations was statistically equivalent for pure tones ($p = 0.86$). This implies that stream segregation overall is easier for the animals in this task version than with pure tones. This cannot be explained trivially by the spectral separation cues available in this task version, as pure tones are even more spectrally separated. The stream segregation advantage of the "fixed harmonic number" tone complexes over pure tones must therefore be due at least in part to the combination of spectral

separation cues *and* complex frequency structure. The effect of pitch distance on task performance was again most obvious for the latest target time bin (1200 ms), though it was also present in earlier time bins (Figure 4.7c). The behavioural evidence for streaming build-up seen with pure tones was therefore not observed in this task version.

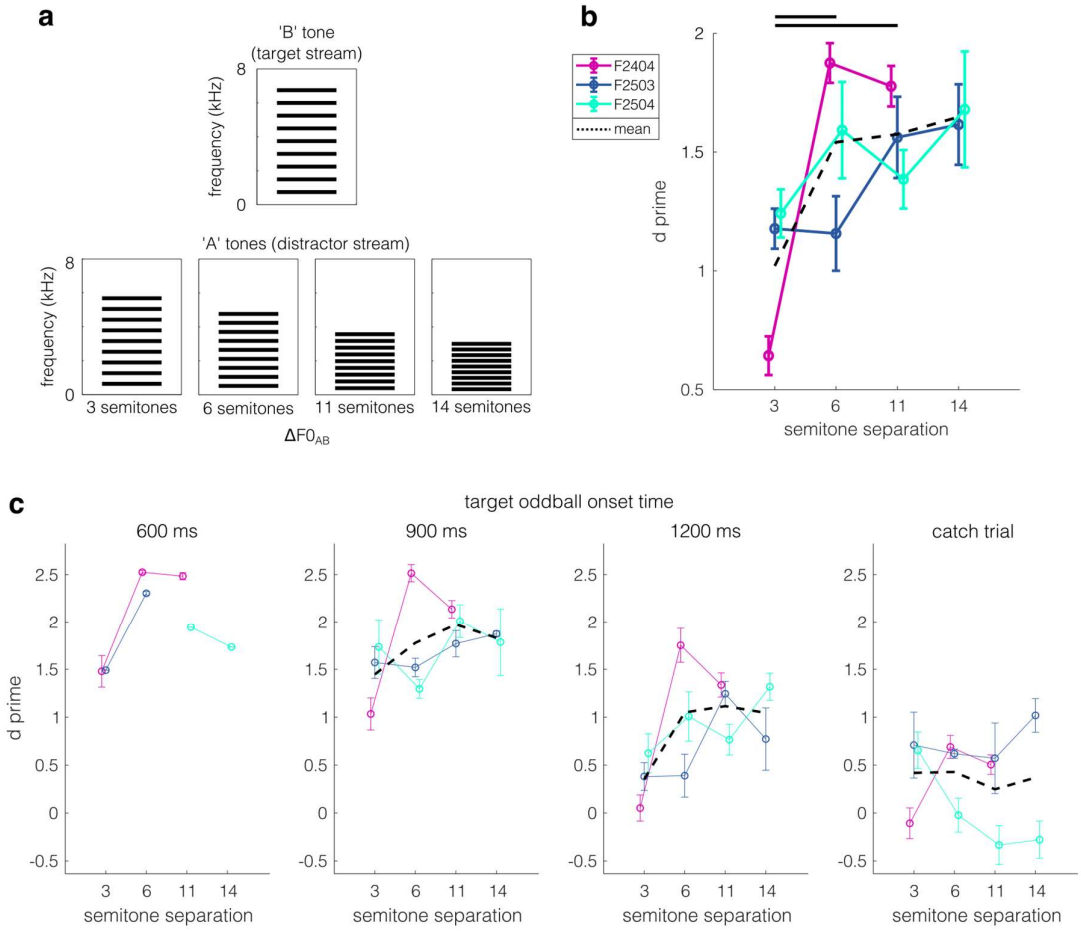

Figure 4.7. Stream segregation of complex tones with a fixed number of harmonics. a) Schematic of stimulus tones in this task version. A and B tones were complex sounds with 9 harmonic components each, and $F0_A$ was one of 3, 6, 11, or 14 semitones below $F0_B$. **b**) Overall performance across the 3 animals was significantly better with increasing pitch separation between the competing streams (2-way ANOVA, $F(2,76) = 16.34$, $p = 1.23 \times 10^{-6}$). Performance was worse at 3 semitones than 6 and 11 semitones separation (black horizontal bars; Tukey-Kramer HSD test; $p < 1 \times 10^{-4}$). **c**) Performance at each pitch separation for every possible target onset time, plotted as in Figure 4.5c.

C. Complex tones with fixed bandwidth

A and B tones in this task version were complex tones containing all harmonic components less than 12 kHz (Figure 4.8a). By maintaining a fixed bandwidth, the total number of harmonics in the tones changed depending on the F0, but the same spectral range (0.2-12 kHz) was always covered. This removed the spectral bandwidth cues present in the previous two versions of the task, meaning that F0 alone was the dominant acoustic cue available for streaming in this task version.

The pitch separation of the streams had a significant effect on the animals' performance (Figure 4.8b), but performance did not monotonically increase with larger separations as in the other task versions. Instead, performance was lower at 11 semitones than both 6 and 14 semitones. This suggests that stream segregation may have been influenced by an acoustic feature beyond pitch height alone, such as pitch chroma, as 11 semitones is near an octave, thereby making A and B similar in pitch chroma. Nevertheless, performance was best for 2 of the 3 ferrets at the largest pitch separation.

The effects of pitch separation on performance were clearest for targets that occurred in the latest time bin (Figure 4.8c), as was observed with pure tones. This evidence of the perceptual "build-up" of stream segregation was therefore observed in 2 of the 3 task versions.

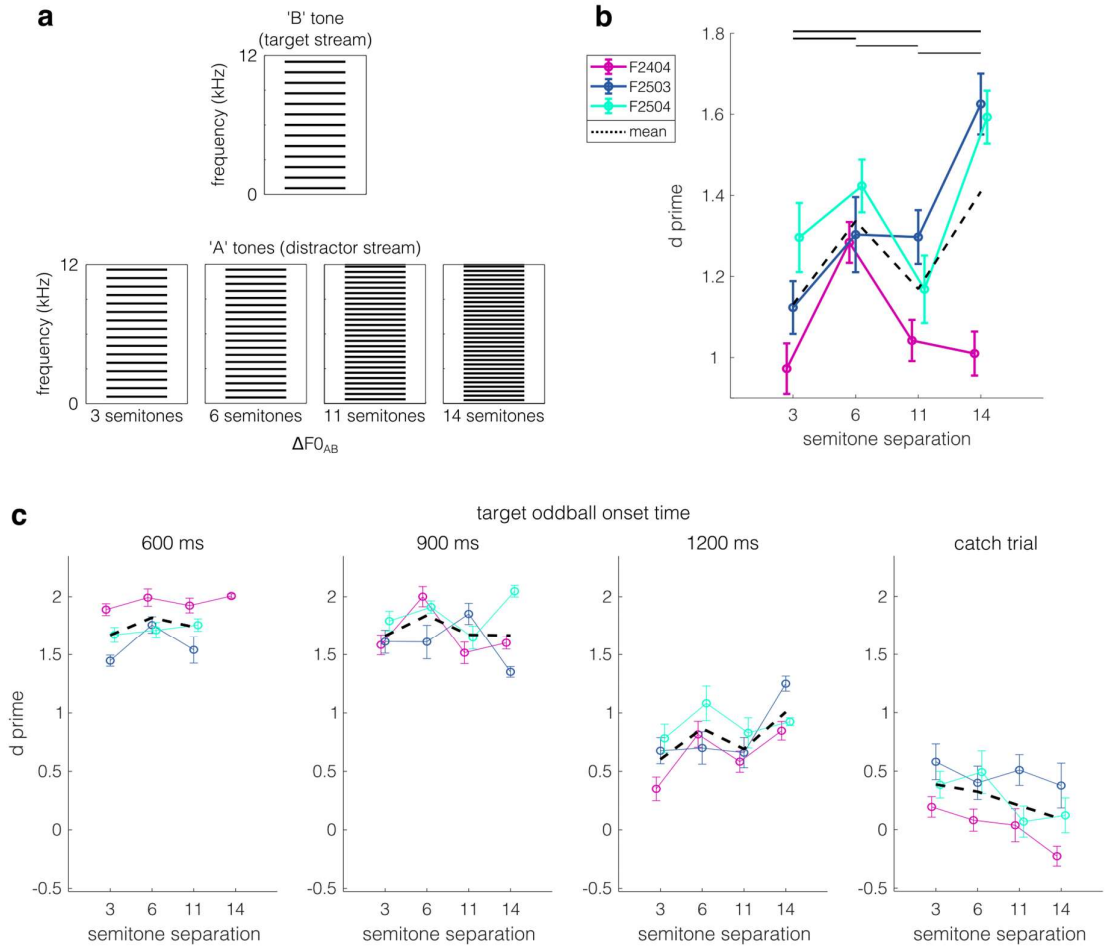

Figure 4.8. Stream segregation of complex tones with fixed bandwidths. **a**) Schematic of stimulus tones in this task version. A and B tones were complex sounds with all harmonic components under 12 kHz, and $F0_A$ was one of 3, 6, 11, or 14 semitones below $F0_B$. **b**) Overall performance across the 3 animals was significantly better with increasing pitch separation between the competing streams (2-way ANOVA, $F(3,244) = 7.35$, $p = 9.82 \times 10^{-5}$). Black bars indicate significant post-hoc comparisons (thin line $p < 0.05$, thick line $p < 0.01$; Tukey-Kramer HSD test). **c**) Performance at each pitch separation for every possible target onset time, plotted as in Figure 4.5c.

D. Complex tones with jittered harmonic frequencies

To confirm that the animals were using pitch to segregate the competing auditory streams, we introduced spectral jitter to the tone complexes in 50% of randomly interleaved trials. On these trials, the harmonic frequencies of complex tones with matched bandwidths were randomly and independently adjusted up or down by 0-50% in both the A and B tones, and F0 was removed entirely (Figure 4.9a). This manipulation disrupts both the spectral regularity and periodicity of the tone, and has been shown to strongly impair pitch discrimination in humans (McPherson and

McDermott, 2018), and should therefore impair the ferrets if they rely on pitch to perform this task. In this condition, A tones were only presented at 6, 11, or 14 semitones below $F0_B$, but all other aspects of the task were as described above in Chapter 4.5C.

We found that task performance was lower overall in jittered trials at all 3 semitone separations (Figure 4.9b), and larger semitone separations did not provide a performance advantage in jittered trials (Figure 4.9b; Tukey-Kramer HSD test, $p > 0.05$). Together, these findings suggest the animals found stream segregation significantly more difficult when pitch cues were removed.

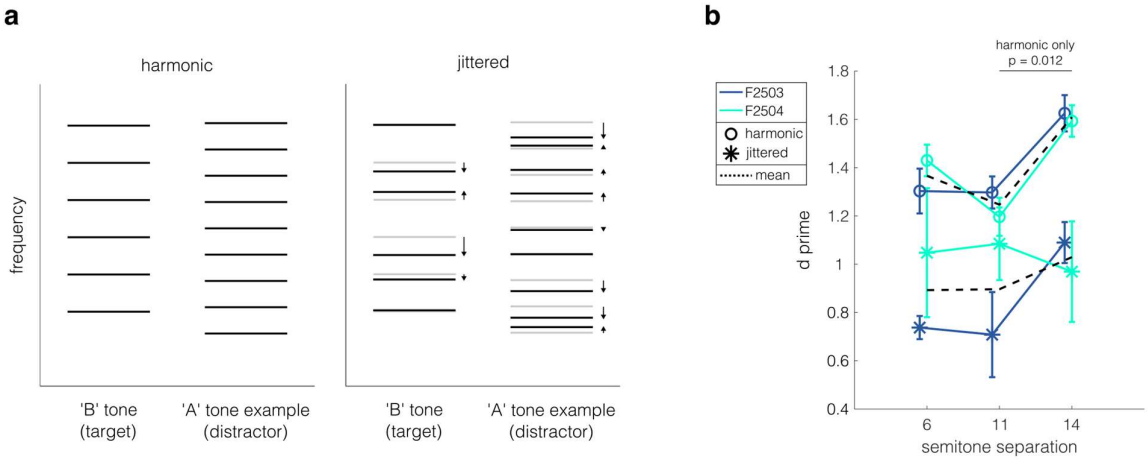

Figure 4.9. Stream segregation of jittered complex tones. **a**) Schematic of the stimuli used in this version of the task. 50% of trials presented streams formed of harmonic A and B tones (left), which were complex tones with fixed bandwidths as described in Figure 4.8a. The remaining trials presented A and B tones with spectrally jittered harmonics (right). On each trial, the harmonics of the A and B tones were independently and randomly adjusted higher or lower by 0-50% of F0 to disrupt spectral regularity and degrade the pitch information in the tones. **b**) Task performance (d prime) was lower overall in jittered trials for both ferrets (different colours) tested on the jittered stimuli (3-way ANOVA, $F(1,118) = 31.2$, $p = 1.26 \times 10^{-7}$). Semitone separation between competing streams was only advantageous to performance in harmonic trials (Tukey-Kramer HSD test; 11 vs 14 semitones, $p = 0.012$). Post-hoc comparisons revealed no significant difference between performance across semitone separations in jittered trials ($p > 0.05$).

4.6. Discussion

Pitch is a key attribute of hearing that has been shown to strongly influence auditory attention and stream segregation in humans (Grossberg, 1996; Micheyl and Oxenham, 2010), yet the role of pitch in selective listening in non-human animals has not been investigated. Here, we developed a novel two-tone streaming paradigm (Figure 4.1) to show that ferrets can use the pitch of both pure tones and complex sounds to segregate competing auditory streams in order to detect a SAM oddball in the target stream (Figures 4.5, 4.7, and 4.8). Degradation of pitch information through spectral jitter resulted in impaired overall performance on this streaming task and eliminated the performance advantage of a greater pitch separation between streams (Figure 4.9). We also report evidence of streaming "build-up" in these animals, as pitch separation effects were larger for targets presented later in trials (Figures 4.5c and 4.8c). These results are the first evidence of pitch-based selective listening in non-human animals.

Like humans, ferrets can perceive and discriminate the pitch of both pure tones and complex sounds (Walker et al., 2009), and our results show that they can use the pitch of both of these types of sounds to segregate auditory streams. This expands on previous demonstrations of frequency-based stream segregation using pure tones in ferrets (Ma et al., 2010) and monkeys (Izumi, 2002). The animals in our study were, however, better able to segregate streams of complex sounds than pure tones (Figure 4.3b), suggesting that the use of naturalistic sounds may be key for capturing auditory scene analysis. The performance disadvantage seen with pure tones could be due in part to a reduced salience of the oddball SAM manipulation when applied to pure tones, but such an effect should also improve the animals' ability to ignore distractor oddballs.

The animals could in theory achieve a correct response by guessing the timing of the target rather than listening for the oddball manipulation in the correction stream. While target timings may play a role in the animals' strategies for solving the task (for example, the probability of the target occurring increases as trial time increases), the overall performance of the animals cannot be explained exclusively by a timing strategy since this strategy does not vary in difficulty as a function of pitch separation between streams. In other words, a timing strategy would not reveal differences in performance dependant on the pitch separation between streams, as

the difficulty of reporting a fixed duration from trial onset does not require any active listening.

The animals were most able to segregate streams when they were composed of complex sounds that did not cover the same frequency range (Figure 4.7). This suggests that the spectral range of sounds may provide an additional cue for isolating independent sound sources. The use of spectral range as a segregation cue may be most evident for complex sounds, however, as the pure tone version of our task contained even greater spectral separation cues but, as previously discussed, produced the worst overall performance. When we presented streams of complex tones with equal spectral ranges, we observed a non-monotonic improvement of performance with increasing semitone separations between streams. One possibility is that pitch chroma influences segregation, at least when segregation is difficult. Pitch chroma describes the perceptual similarity of certain pitches along the Western musical scale (Walker et al., 2011a), and has been demonstrated in humans (Hoeschele et al., 2012), rhesus monkeys (Wright et al., 2000), dolphins (Ralston and Herman, 1995), and ferrets (Yin et al., 2010). Semitone separations of 3 and 11 are both chromatically closer to $F0_B$ than 6 semitones, which aligns with the performance advantage seen at 3 and 11 semitones over 6. Future work could isolate the role of pitch chroma in a selective listening paradigm using stimuli such as Shephard tones that specifically isolate pitch chroma and height cues.

Spectral jitter has been shown to impair pitch discrimination in human listeners (McPherson and McDermott, 2018; Popham et al., 2018), and we found that the introduction of spectral jitter to complex tones with fixed bandwidths (i.e. without spectral range cues) produced significantly worse performance overall and eliminated the performance advantage seen at larger semitone separations (Figure 4.9). This suggests that the harmonicity, and thus the pitch, of the competing streams was the dominant acoustic feature the animals used to segregate the auditory streams, rather than just the density or range of frequencies in the two sounds. The jitter manipulation we employed disrupted both spectral regularity and temporal periodicity, so it is not possible to draw conclusions about the relative importance of spectral versus temporal pitch cues in selective listening from these findings. However, studies in human listeners indicate that spectral pitch may play a stronger role in streaming (Grimault et al., 2000; Madsen et al., 2018), making this a logical next-step for future work.

These findings demonstrate that ferrets use pitch in selective listening, using a novel behavioural paradigm for studying auditory streaming in animals that can be compared directly to those previously used with human subjects. Future work can adapt this paradigm to explore a range of acoustic parameters that affect auditory streaming which have been widely studied in human listeners but not yet translated to animal models, such as timbre (Cusack and Roberts, 2000), amplitude modulation (Grimault et al., 2002), and tone presentation rate (Bregman, 1990). Developing these classic psychophysical tasks in animals allows us to explore the neuronal underpinnings of selective listening in individual neurons and neural circuits.

5. General discussion

5.1. Overview

Pitch is a salient characteristic of sound, and it plays a key role in selective listening in humans. Despite this central role, the neural mechanisms underlying pitch perception are poorly understood. In this thesis, I aimed to address the conflicting models for pitch encoding with *in vivo* electrophysiology in ferrets. I found evidence for pitch-selective neurons, a class of neurons which had not previously been found outside of marmosets, and I characterised a dual-encoding strategy for pitch in both single neurons and population codes. To explore the role of pitch in selective listening, I designed a novel paradigm to measure auditory streaming in ferrets. I found that ferrets can use pitch to segregate competing auditory streams, the first demonstration of pitch as an acoustic cue for selective listening outside of humans.

5.2. Dual-encoding mechanisms for pitch

The auditory system employs two strategies for encoding a sound's fundamental frequency (F0) from the earliest stages of the hierarchy. F0 can be represented through the pattern of responses to the individual harmonic components (rate-place code) or in the timing of responses phase-locked to the amplitude envelope (spike-timing code), and this dual-encoding strategy is evident at the level of the auditory nerve and cochlear nucleus (Chapter 1.3). Single neurons in primary auditory cortex have been identified which selectively derive F0 from either resolved or unresolved harmonics (Bendor et al., 2012; Fishman et al., 2013; Steinschneider et al., 1998), and in this thesis I identified this dual-encoding strategy in single neurons in ferret primary auditory cortex (Chapter 2.5A-B). I found that population codes for pitch were also specific to the acoustic cues available, as decoding models trained on responses to one acoustic cue did not predict the F0 of sounds with different cues (Chapter 3.5A). I found that a small number of neurons in ferret auditory cortex encoded F0 from either resolved or unresolved harmonics, or both (Chapter 2.5D), consistent with findings in the marmoset (Bendor and Wang, 2010, 2005). In both the ferret and marmoset, these pitch-selective neurons were rare (~3% of all sound-responsive neurons recorded).

The pitch of resolved and unresolved harmonics is perceptually the same, which implies that the bifurcated representations of these two acoustic cues for pitch

should unite at some level of the auditory hierarchy to underpin this perception. While the pitch-selective neurons identified in marmosets (Bendor and Wang, 2010, 2005) and ferrets (Chapter 2.5D) mirror this perceptual unity, their relative scarcity make it challenging to model how they may drive invariant pitch perception. This is especially true given that this invariant encoding is not reflected at the level of population codes, which are regarded as the basis of perception (Panzeri et al., 2015; Stanley, 2013). Future work could elucidate the neural mechanisms of invariant pitch perception through lesion studies that selectively knock-out or inhibit cortical regions in order to motivate neural recordings that better target regions most likely to contain an invariant code for pitch.

5.3. Anatomical distribution of pitch-selective neurons

The pitch-selective neurons identified in marmosets in Bendor and Wang, 2005 were enriched in the low frequency region of primary auditory cortex that borders secondary auditory areas. They posited that this anatomical region may be the "pitch area" where pitch is encoded in auditory cortex (Bendor and Wang, 2005). In contrast, several studies conducted in macaques and humans have instead identified anatomically distributed pitch representations (Allen et al., 2017; Berger et al., 2023; Fishman et al., 2013; Gander et al., 2019; Griffiths et al., 2010; Kikuchi et al., 2019; Norman-Haignere et al., 2019; Steinschneider et al., 1998). The pitch-selective neurons we identified in ferret primary auditory cortex (Chapter 2.5) were anatomically distributed, consistent with the findings in macaques and humans. The harmonicity neurons we identified in Chapter 2.5A were more anatomically clustered in low frequency A1 than the other pitch-selective neurons, and this region may be analogous to the pitch area described by Bendor and Wang (Bendor and Wang, 2005). Harmonicity neurons were also more likely to show similar tuning between pure tone tuning and complex F0-tuning (Chapter 2.5, Figure 2.11), suggesting that the criterion employed by Bendor and Wang for pitch selective neurons to share tuning between pure tones and complex sounds may bias the location of pitch-selective neurons towards low frequency regions of primary auditory cortex. The cortical penetrations from which we recorded were biased towards low frequency A1, and the yield of active single neurons from each penetration varied greatly. It is therefore possible that a replication of this experiment wherein single neurons are recorded

from an even sampling of locations across primary auditory cortex may reveal a bias for pitch-selective neurons to be located in an isolated region.

5.4. Pure tone pitch

Pure tones evoke a salient pitch percept, meaning that any neural code for F0 should in theory represent the pitch of pure tones as well as complex tones. Indeed, pure tone F0-tuning (i.e. frequency tuning) was part of the inclusion criteria for the seminal finding of pitch-selective neurons in marmosets (Bendor and Wang, 2005). In contrast with the pitch-selective neurons defined by Bendor and Wang, the frequency tuning of the pitch-selective neurons I identified in ferret auditory cortex did not relate to their F0-tuning to complex sounds (Chapter 2.5F). Similarly, a decoding model trained to predict the pitch of stimuli from neural responses to pure tones did not accurately predict the pitch of complex sounds (Chapter 3.5A). Together this suggests that the pitch of pure tones is not encoded via the same mechanisms as complex sounds at the level of primary auditory cortex. This is consistent with our motivating hypothesis for Chapter 2: neurons must integrate either the place code of resolved harmonics or the temporal code of unresolved harmonics, or both, to encode F0 (Chapter 2.6). Pure tones are trivially encoded in the tonotopic map of primary auditory cortex and would therefore not evoke sufficient excitatory drive for F0-encoding neurons that specifically integrate responses across multiple neurons to derive F0.

I also found that pure tones evoked population codes for pitch at sound offset which were redundant to those at sound onset, in contrast to the non-redundant onset/offset codes evoked by all complex sounds presented (Chapter 3.5B). This suggests that population codes for frequency are different in nature to population codes for F0 in primary auditory cortex, consistent with our hypothesis that F0-encoding is mechanistically different for pure tones and complex sounds in single neurons. Future work could expand on these findings by measuring neural responses to amplitude modulated sounds which parameterize the rate and depth of sound onset and offset, and comparing responses to pure tones and complex sounds under these conditions.

5.5. Pitch and selective listening

Pitch is a key acoustic cue for auditory segregation in humans (Grossberg, 1996; Micheyl and Oxenham, 2010), but this has not been demonstrated in any other animal. Many animals perceive pitch (Brosch et al., 2004; Klinge and Klump, 2009; Osmanski et al., 2013; Walker et al., 2019, 2009), and streaming behaviour has been demonstrated in birds (Dent et al., 2016; Hulse et al., 1997; MacDougall-Shackleton et al., 1998), goldfish (Fay, 1998), ferrets (Ma et al., 2010), and monkeys (Izumi, 2002). Frequency-based auditory streaming has been demonstrated in ferrets (Ma et al., 2010) and monkeys (Izumi, 2002), and birds have been shown to use spectral envelope cues to integrate birdsong syllables into a song (Dent et al., 2016). I demonstrated that ferrets can use the pitch of both pure tones and complex sounds to segregate competing auditory streams using a novel paradigm that adapts the classical ABA- streaming paradigm often employed with human subjects (Chapter 4.5A-C). The ferrets performed better when the competing streams were composed of complex sounds rather than pure tones, suggesting that auditory scene analysis may be best captured using naturalistic complex sounds. It should be noted that this task, and all other paradigms used in animal models to date, capture streaming under the influence of attention because the tasks demand the animals listen for a specific acoustic cue to report. This could in theory limit direct comparisons between findings in behaving animals and those in humans who report perception without instructions or task demands.

The neural correlates of ABA- streaming have been explored in passively listening and anaesthetised animals (Bee and Klump, 2004; Elhilali et al., 2009; Fishman et al., 2001; Itatani and Klump, 2009; Micheyl et al., 2005b; Pressnitzer et al., 2008), but they have never been measured in behaving animals. The behavioural paradigm I have reported opens the possibility for future work to characterise the neural mechanisms supporting auditory attention under conditions that can be directly compared with human psychophysical findings. Additionally, this paradigm can be readily adapted to measure a variety of acoustic cues for streaming that have already been demonstrated in humans, such as stimulus presentation rate (Bregman, 1990), timbre (Cusack and Roberts, 2000), and amplitude modulation (Grimault et al., 2002). Though the role of pitch in auditory attention in ferrets has been established in a general sense in this thesis, studies in humans have characterized this role in greater detail using a variety of acoustic manipulations (Singh, 1987; Vliegen et al., 1999;

Vliegen and Oxenham, 1999). Future studies can utilize the behavioural paradigm described in Chapter 4 in conjunction with stimuli from the human psychophysical corpus to more fully map the correspondence between selective listening in humans and non-human animals, a critical step for relating the invasive neural measurements that are only technically feasible in animal models back to human listeners.

Appendix

Spike sorting quality metrics

This appendix includes average waveforms, inter-spike intervals (ISIs), and average firing rates over the recording for each harmonicity neuron and temporal neuron identified in any penetration recorded from animal F2001 (6 penetrations total). If no neurons are shown for a given penetration, no harmonicity or temporal neurons were identified in this penetration.

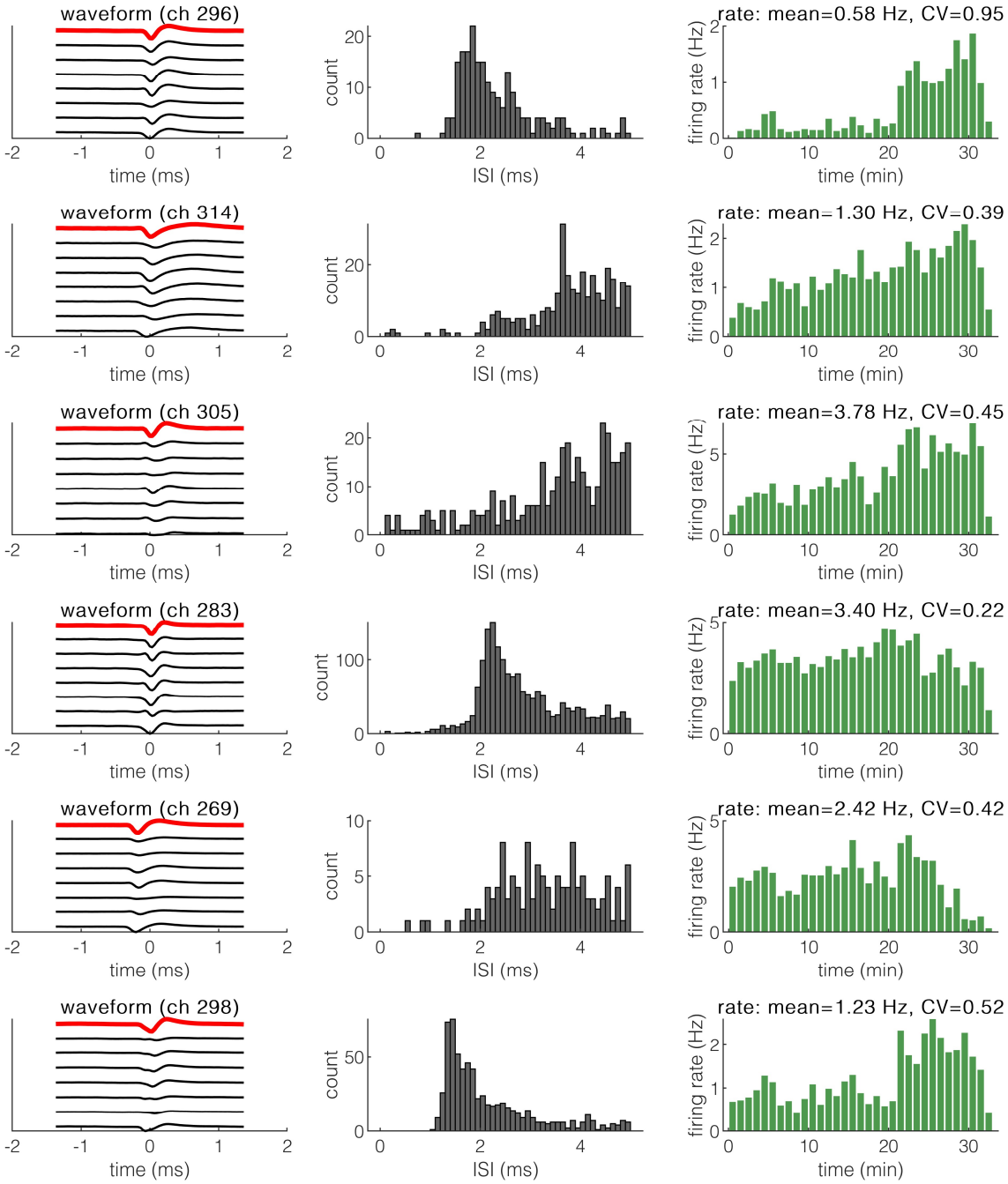

Figure A1. Spike sorting quality metrics for harmonicity neurons identified in penetration 2 for animal F2001. Each row displays metrics for a single harmonicity neuron. The first column displays the average waveforms for the channel with the highest amplitude in the cluster (red) and the 7 nearest channels (black). The second column displays inter-spike intervals (ISI) between 0 and 5 ms. The third column shows the average firing rate across the duration of the recording. The mean firing rate over the recording and the coefficient of variation (CV) for the firing rate are reported in the title of the subplot.

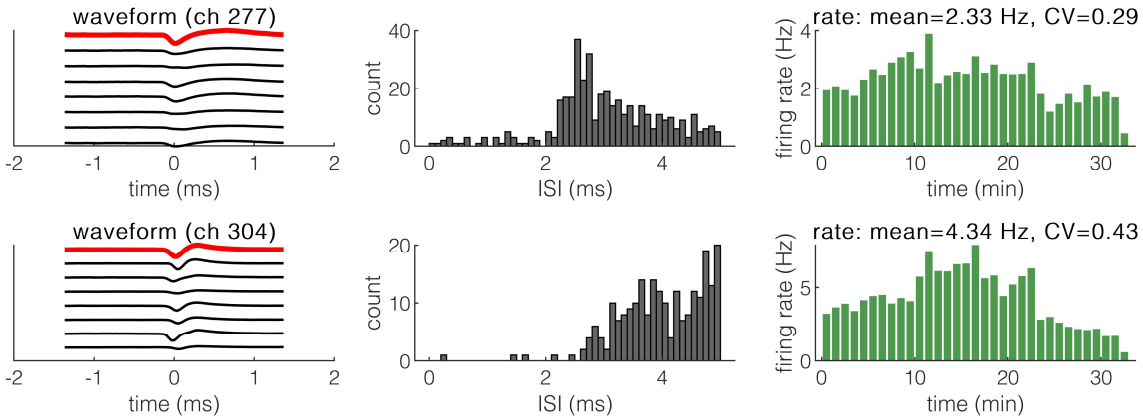

Figure A2. Spike sorting quality metrics for harmonicity neurons identified in penetration 3 for animal F2001. All plots are as described in Figure A1.

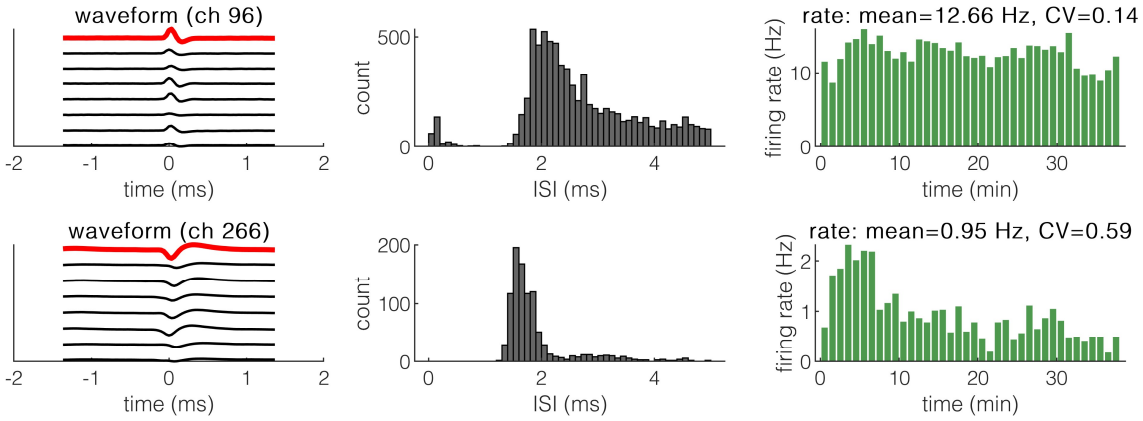

Figure A3. Spike sorting quality metrics for harmonicity neurons identified in penetration 4 for animal F2001. All plots are as described in Figure A1.

Temporal neurons

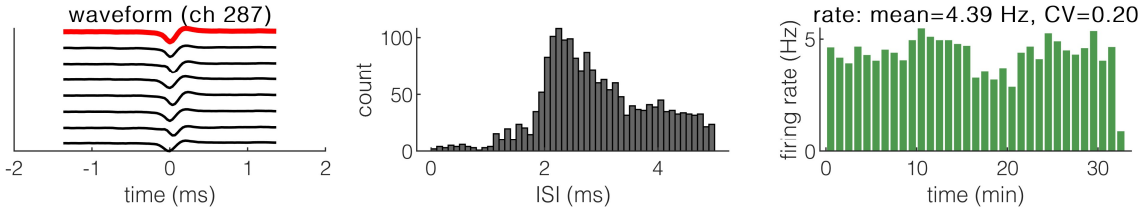

Figure A4. Spike sorting quality metrics for temporal neurons identified in penetration 4 for animal F2001. All plots are as described in Figure A1.

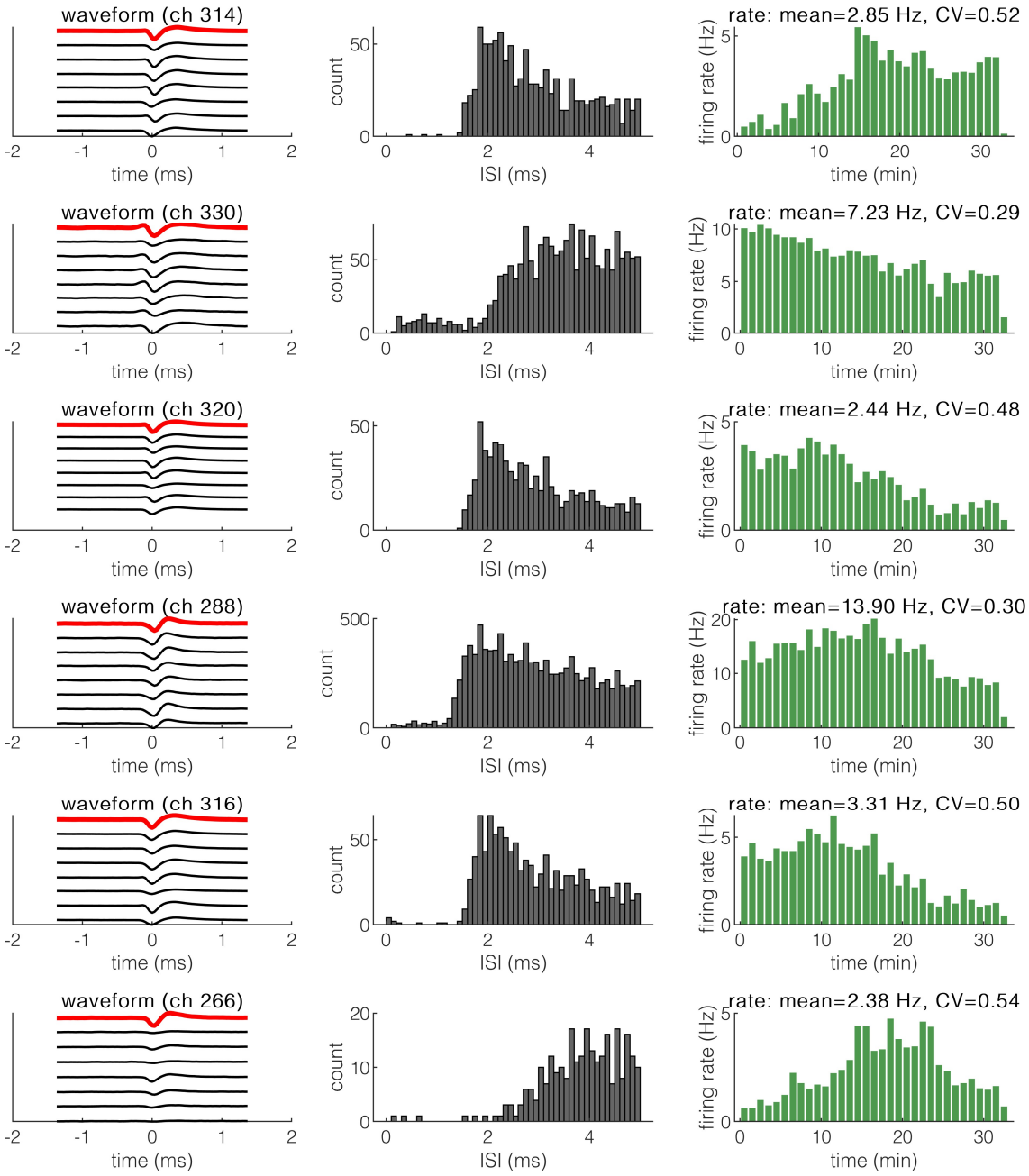

Figure A5. Spike sorting quality metrics for temporal neurons identified in penetration 3 for animal F2001. All plots are as described in Figure A1.

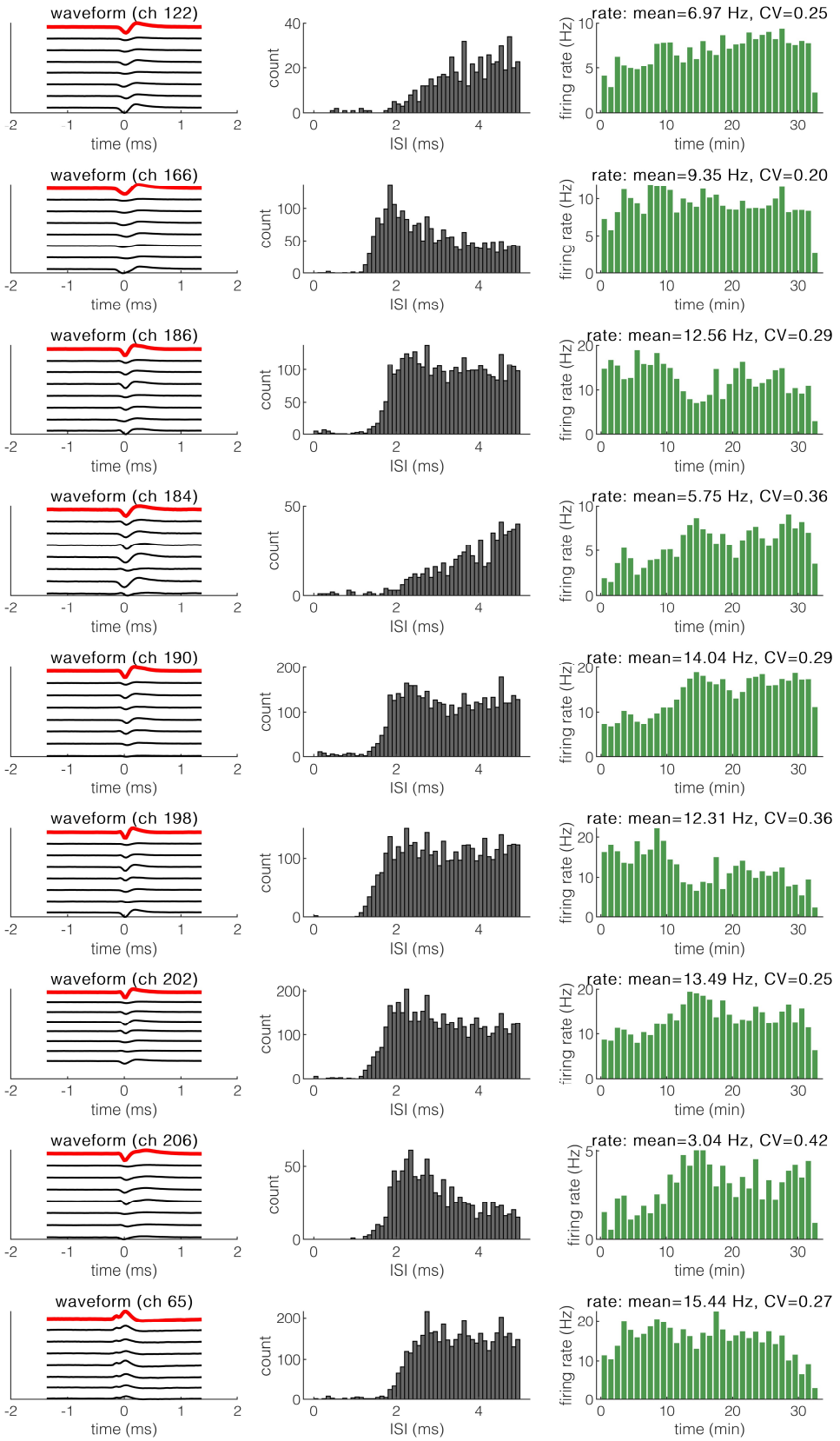

Figure A6. Spike sorting quality metrics for temporal neurons identified in penetration 8 for animal F2001. All plots are as described in Figure A1.

References

Allen, E.J., Burton, P.C., Olman, C.A., Oxenham, A.J., 2017. Representations of Pitch and Timbre Variation in Human Auditory Cortex. Journal of Neuroscience 37, 1284–1293. https://doi.org/10.1523/JNEUROSCI.2336-16.2016

ANSI-S1.1, 1994. Acoustical Terminology. Acoustical Society of America.

Arehart, K.H., 1994. Effects of harmonic content on complex-tone fundamental-frequency discrimination in hearing-impaired listeners. The Journal of the Acoustical Society of America 95, 3574–3585. https://doi.org/10.1121/1.409975

Assmann, P.F., Summerfield, Q., 1994. The contribution of waveform interactions to the perception of concurrent vowels. The Journal of the Acoustical Society of America 95, 471–484. https://doi.org/10.1121/1.408342

Assmann, P.F., Summerfield, Q., 1990. Modeling the perception of concurrent vowels: Vowels with different fundamental frequencies. The Journal of the Acoustical Society of America 88, 680–697. https://doi.org/10.1121/1.399772

Assmann, P.F., Summerfield, Q., 1987. Perceptual segregation of concurrent vowels. The Journal of the Acoustical Society of America 82, S120. https://doi.org/10.1121/1.2024632

Averbeck, B.B., Latham, P.E., Pouget, A., 2006. Neural correlations, population coding and computation. Nature Reviews Neuroscience 7, 358–366. https://doi.org/10.1038/nrn1888

Bee, M.A., Klump, G.M., 2004. Primitive Auditory Stream Segregation: A Neurophysiological Study in the Songbird Forebrain. Journal of Neurophysiology 92, 1088–1104. https://doi.org/10.1152/jn.00884.2003

Bendor, D., Osmanski, M.S., Wang, X., 2012. Dual-Pitch Processing Mechanisms in Primate Auditory Cortex. Journal of Neuroscience 32, 16149–16161. https://doi.org/10.1523/JNEUROSCI.2563-12.2012

Bendor, D., Wang, X., 2010. Neural Coding of Periodicity in Marmoset Auditory Cortex. Journal of Neurophysiology 103, 1809–1822. https://doi.org/10.1152/jn.00281.2009

Bendor, D., Wang, X., 2005. The neuronal representation of pitch in primate auditory cortex. Nature 436, 1161–1165. https://doi.org/10.1038/nature03867

Berger, J.I., Gander, P.E., Kikuchi, Y., Petkov, C.I., Kumar, S., Kovach, C., Oya, H., Kawasaki, H., Howard, M.A., Griffiths, T.D., 2023. Distribution of multiunit pitch responses recorded intracranially from human auditory cortex. Cerebral Cortex 33, 9105–9116. https://doi.org/10.1093/cercor/bhad186

Bird, J., Darwin, C.J., 1998. Effects of a difference in fundamental frequency in separating two sentences, in: Palmer, A., Rees, A., Summerfield, A.Q., Meddis, R. (Eds.), Psychophysical and Physiological Advances in Hearing. Whurr, London, pp. 263–269.

Bizley, J.K., Walker, K.M.M., 2010. Sensitivity and selectivity of neurons in auditory cortex to the pitch, timbre, and location of sounds. The Neuroscientist 16, 453–469. https://doi.org/10.1177/1073858410371009

Bizley, J.K., Walker, K.M.M., King, A.J., Schnupp, J.W.H., 2010. Neural Ensemble Codes for Stimulus Periodicity in Auditory Cortex. Journal of Neuroscience 30, 5078–5091. https://doi.org/10.1523/JNEUROSCI.5475-09.2010

Bizley, J.K., Walker, K.M.M., Nodal, F.R., King, A.J., Schnupp, J.W.H., 2013. Auditory cortex represents both pitch judgments and the corresponding acoustic cues. Current Biology 23, 620–625. https://doi.org/10.1016/j.cub.2013.03.003

Bizley, J.K., Walker, K.M.M., Silverman, B.W., King, A.J., Schnupp, J.W.H., 2009. Interdependent Encoding of Pitch, Timbre, and Spatial Location in Auditory Cortex. Journal of Neuroscience 29, 2064–2075. https://doi.org/10.1523/JNEUROSCI.4755-08.2009

Bowling, T., Wen, H., Meenderink, S.W.F., Dong, W., Meaud, J., 2021. Intracochlear distortion products are broadly generated by outer hair cells but their contributions to otoacoustic emissions are spatially restricted. Scientific Reports 11, 13651. https://doi.org/10.1038/s41598-021-93099-7

Bregman, A.S., 1990. Auditory Scene Analysis: The Perceptual Organization of Sound. The MIT Press. https://doi.org/10.7551/mitpress/1486.001.0001

Bregman, A.S., Liao, C., Levitan, R., 1990. Auditory grouping based on fundamental frequency and formant peak frequency. Canadian Journal of Psychology 44, 400–413. https://doi.org/10.1037/h0084255

Brokx, J.P.L., Nooteboom, S.G., 1982. Intonation and the perceptual separation of simultaneous voices. Journal of Phonetics 10, 23–36. https://doi.org/10.1016/S0095-4470(19)30909-X

Brosch, M., Selezneva, E., Bucks, C., Scheich, H., 2004. Macaque monkeys discriminate pitch relationships. Cognition 91, 259–272. https://doi.org/10.1016/j.cognition.2003.09.005

Brungart, D.S., Simpson, B.D., Ericson, M.A., Scott, K.R., 2001. Informational and energetic masking effects in the perception of multiple simultaneous talkers. The Journal of the Acoustical Society of America 110, 2527–2538. https://doi.org/10.1121/1.1408946

Cariani, P.A., Delgutte, B., 1996. Neural correlates of the pitch of complex tones. I. Pitch and pitch salience. Journal of Neurophysiology 76, 1698–1716. https://doi.org/10.1152/jn.1996.76.3.1698

Carlyon, R.P., 1996a. Encoding the fundamental frequency of a complex tone in the presence of a spectrally overlapping masker. The Journal of the Acoustical Society of America 99, 517–524. https://doi.org/10.1121/1.414510

Carlyon, R.P., 1996b. Masker asynchrony impairs the fundamental-frequency discrimination of unresolved harmonics. The Journal of the Acoustical Society of America 99, 525–533. https://doi.org/10.1121/1.414511

Culling, J.F., Darwin, C.J., 1993. Perceptual separation of simultaneous vowels: Within and across-formant grouping by F0. The Journal of the Acoustical Society of America 93, 3454–3467. https://doi.org/10.1121/1.405675

Cusack, R., Roberts, B., 2000. Effects of differences in timbre on sequential grouping. Perception & Psychophysics 62, 1112–1120. https://doi.org/10.3758/BF03212092

de Cheveigné, A., 1997. Concurrent vowel identification. III. A neural model of harmonic interference cancellation. The Journal of the Acoustical Society of America 101, 2857–2865. https://doi.org/10.1121/1.419480

Dent, M.L., Martin, A.K., Flaherty, M.M., Neilans, E.G., 2016. Cues for auditory stream segregation of birdsong in budgerigars and zebra finches: Effects of location, timing, amplitude, and frequency. The Journal of the Acoustical Society of America 139, 674–683. https://doi.org/10.1121/1.4941322

Elhilali, M., Ma, L., Micheyl, C., Oxenham, A.J., Shamma, S.A., 2009. Temporal Coherence in the Perceptual Organization and Cortical Representation of Auditory Scenes. Neuron 61, 317–329. https://doi.org/10.1016/j.neuron.2008.12.005

Fay, R.R., 1998. Auditory stream segregation in goldfish (*Carassius auratus*). Hearing Research 120, 69–76. https://doi.org/10.1016/S0378-5955(98)00058-6

Fishman, Y.I., Micheyl, C., Steinschneider, M., 2013. Neural Representation of Harmonic Complex Tones in Primary Auditory Cortex of the Awake Monkey. Journal of Neuroscience 33, 10312–10323. https://doi.org/10.1523/JNEUROSCI.0020-13.2013

Fishman, Y.I., Reser, D.H., Arezzo, J.C., Steinschneider, M., 2001. Neural correlates of auditory stream segregation in primary auditory cortex of the awake monkey. Hearing Research 151, 167–187. https://doi.org/10.1016/S0378-5955(00)00224-0

Fishman, Y.I., Steinschneider, M., 2009. Temporally dynamic frequency tuning of population responses in monkey primary auditory cortex. Hearing Research 254, 64–76. https://doi.org/10.1016/j.heares.2009.04.010

Fishman, Y.I., Steinschneider, M., Micheyl, C., 2014. Neural Representation of Concurrent Harmonic Sounds in Monkey Primary Auditory Cortex: Implications for Models of Auditory Scene Analysis. Journal of Neuroscience 34, 12425–12443. https://doi.org/10.1523/JNEUROSCI.0025-14.2014

Gaese, B.H., Ostwald, J., 2001. Anesthesia Changes Frequency Tuning of Neurons in the Rat Primary Auditory Cortex. Journal of Neurophysiology 86, 1062–1066. https://doi.org/10.1152/jn.2001.86.2.1062

Gander, P.E., Kumar, S., Sedley, W., Nourski, K.V., Oya, H., Kovach, C.K., Kawasaki, H., Kikuchi, Y., Patterson, R.D., Howard, M.A., Griffiths, T.D., 2019. Direct electrophysiological mapping of human pitch-related processing in auditory cortex. Neuroimage 202, 116076. https://doi.org/10.1016/j.neuroimage.2019.116076

Goldstein, J.L., 1973. An optimum processor theory for the central formation of the pitch of complex tones. The Journal of the Acoustical Society of America 54, 1496–1516. https://doi.org/10.1121/1.1914448

Griffiths, T.D., Hall, D.A., 2012. Mapping Pitch Representation in Neural Ensembles with fMRI. Journal of Neuroscience 32, 13343–13347. https://doi.org/10.1523/JNEUROSCI.3813-12.2012

Griffiths, T.D., Kumar, S., Sedley, W., Nourski, K.V., Kawasaki, H., Oya, H., Patterson, R.D., Brugge, J.F., Howard, M.A., 2010. Direct recordings of pitch responses from human auditory cortex. Current Biology 20, 1128–1132. https://doi.org/10.1016/j.cub.2010.04.044

Grimault, N., Bacon, S.P., Micheyl, C., 2002. Auditory stream segregation on the basis of amplitude-modulation rate. The Journal of the Acoustical Society of America 111, 1340–1348. https://doi.org/10.1121/1.1452740

Grimault, N., Micheyl, C., Carlyon, R.P., Arthaud, P., Collet, L., 2000. Influence of peripheral resolvability on the perceptual segregation of harmonic complex tones differing in fundamental frequency. The Journal of the Acoustical Society of America 108, 263–271. https://doi.org/10.1121/1.429462

Grossberg, S., 1996. Pitch-Based Streaming in Auditory Perception, in: Musical Networks: Parallel Distributed Perception and Performance. MIT Press, Cambridge, MA, pp. 117–140.

Hartmann, W.M., McAdams, S., Smith, B.K., 1990. Hearing a mistuned harmonic in an otherwise periodic complex tone. The Journal of the Acoustical Society of America 88, 1712–1724. https://doi.org/10.1121/1.400246

Hoekstra, A., Ritsma, R.J., 1977. Perceptive hearing loss and frequency selectivity, in: Evans, E.F., Wilson, J.P. (Eds.), Psychophysics and Physiology of Hearing. Academic Press, London, pp. 263–271.

Hoeschele, M., Weisman, R.G., Sturdy, C.B., 2012. Pitch chroma discrimination, generalization, and transfer tests of octave equivalence in humans. Attention, Perception, & Psychophysics 74, 1742–1760. https://doi.org/10.3758/s13414-012-0364-2

Houtsma, A.J.M., Goldstein, J.L., 1972. The Central Origin of the Pitch of Complex Tones: Evidence from Musical Interval Recognition. The Journal of the Acoustical Society of America 51, 520–529. https://doi.org/10.1121/1.1912873

Hulse, S.H., MacDougall-Shackleton, S.A., Wisniewski, A.B., 1997. Auditory scene analysis by songbirds: Stream segregation of birdsong by European starlings (Sturnus vulgaris). Journal of Comparative Psychology 111, 3–13. https://doi.org/10.1037/0735-7036.111.1.3

Itatani, N., Klump, G.M., 2009. Auditory Streaming of Amplitude-Modulated Sounds in the Songbird Forebrain. Journal of Neurophysiology 101, 3212–3225. https://doi.org/10.1152/jn.91333.2008

Izumi, A., 2002. Auditory stream segregation in Japanese monkeys. Cognition 82, B113–B122. https://doi.org/10.1016/S0010-0277(01)00161-5

Jacobs, A.L., Fridman, G., Douglas, R.M., Alam, N.M., Latham, Peter.E., Prusky, G.T., Nirenberg, S., 2009. Ruling out and ruling in neural codes. Proceedings of the National Academy of Sciences 106, 5936–5941. https://doi.org/10.1073/pnas.0900573106

Joris, P.X., Carney, L.H., Smith, P.H., Yin, T.C., 1994. Enhancement of neural synchronization in the anteroventral cochlear nucleus. I. Responses to tones at the characteristic frequency. Journal of Neurophysiology 71, 1022–1036. https://doi.org/10.1152/jn.1994.71.3.1022

Joris, P.X., Schreiner, C.E., Rees, A., 2004. Neural Processing of Amplitude-Modulated Sounds. Physiological Reviews 84, 541–577. https://doi.org/10.1152/physrev.00029.2003

Jun, J.J., Steinmetz, N.A., Siegle, J.H., Denman, D.J., Bauza, M., Barbarits, B., Lee, A.K., Anastassiou, C.A., Andrei, A., Aydın, Ç., Barbic, M., Blanche, T.J., Bonin, V., Couto, J., Dutta, B., Gratiy, S.L., Gutnisky, D.A., Häusser, M., Karsh, B., Ledochowitsch, P., Lopez, C.M., Mitelut, C., Musa, S., Okun, M., Pachitariu, M., Putzeys, J., Rich, P.D., Rossant, C., Sun, W.-L., Svoboda, K., Carandini, M., Harris, K.D., Koch, C., O'Keefe, J., Harris, T.D., 2017. Fully integrated silicon probes for high-density recording of neural activity. Nature 551, 232–236. https://doi.org/10.1038/nature24636

Karsh, B., 2017. SpikeGLX.

Kelly, J.B., Kavanagh, G.L., Dalton, J.C.H., 1986. Hearing in the ferret (*Mustela putorius*): Thresholds for pure tone detection. Hearing Research 24, 269–275. https://doi.org/10.1016/0378-5955(86)90025-0

Kikuchi, Y., Kumar, S., Baumann, S., Overath, T., Gander, P.E., Sedley, W., Patterson, R.D., Petkov, C.I., Griffiths, T.D., 2019. The distribution and nature of responses to broadband sounds associated with pitch in the macaque auditory cortex. Cortex 120, 340–352. https://doi.org/10.1016/j.cortex.2019.07.005

Kim, D., Leonard, G., 1988. Pitch-period following response of cat cochlear nucleus neurons to speech sounds, in: Duifhuis, H., Horst, J., Wit, H. (Eds.), Basic Issues in Hearing. Academic Press, London, pp. 252–260.

Klinge, A., Klump, G., 2010. Mistuning detection and onset asynchrony in harmonic complexes in Mongolian gerbils. The Journal of the Acoustical Society of America 128, 280–290. https://doi.org/10.1121/1.3436552

Klinge, A., Klump, G.M., 2009. Frequency difference limens of pure tones and harmonics within complex stimuli in Mongolian gerbils and humans. The Journal of the Acoustical Society of America 125, 304–314. https://doi.org/10.1121/1.3021315

Kopp-Scheinpflug, C., Sinclair, J.L., Linden, J.F., 2018. When Sound Stops: Offset Responses in the Auditory System. Trends in Neurosciences 41, 712–728. https://doi.org/10.1016/j.tins.2018.08.009

Krishna, B.S., Semple, M.N., 2000. Auditory temporal processing: responses to sinusoidally amplitude-modulated tones in the inferior colliculus. Journal of Neurophysiology 84, 255–273. https://doi.org/10.1152/jn.2000.84.1.255

Lamothe, C., Bagur, S., Gosselin, E., Bathellier, B., 2025. Sound offset responses become highly informative in the auditory cortex. The Journal of Physiology n/a. https://doi.org/10.1113/JP288416

Langner, G., Albert, M., Briede, T., 2002. Temporal and spatial coding of periodicity information in the inferior colliculus of awake chinchilla (Chinchilla laniger). Hearing Research 168, 110–130. https://doi.org/10.1016/s0378-5955(02)00367-2

Langner, G., Schreiner, C.E., 1988. Periodicity coding in the inferior colliculus of the cat. I. Neuronal mechanisms. Journal of Neurophysiology 60, 1799–1822. https://doi.org/10.1152/jn.1988.60.6.1799

Laudanski, J., Edeline, J.-M., Huetz, C., 2012. Differences between spectro-temporal receptive fields derived from artificial and natural stimuli in the auditory cortex. PLoS One 7, e50539. https://doi.org/10.1371/journal.pone.0050539

Licklider, J.C.R., 1951. A duplex thory of pitch perception. Experientia 7, 128–134. https://doi.org/10.1121/1.1917296

Ludwig, K.A., Miriani, R.M., Langhals, N.B., Joseph, M.D., Anderson, D.J., Kipke, D.R., 2009. Using a common average reference to improve cortical neuron recordings from microelectrode arrays. Journal of Neurophysiology 101, 1679–1689. https://doi.org/10.1152/jn.90989.2008

Ma, L., Micheyl, C., Yin, P., Oxenham, A.J., Shamma, S.A., 2010. Behavioral measures of auditory streaming in ferrets (Mustela putorius). Journal of Comparative Psychology 124, 317–330. https://doi.org/10.1037/a0018273

MacDougall-Shackleton, S.A., Hulse, S.H., Gentner, T.Q., White, W., 1998. Auditory scene analysis by European starlings (Sturnus vulgaris): Perceptual segregation of tone sequences. The Journal of the Acoustical Society of America 103, 3581–3587. https://doi.org/10.1121/1.423063

Madsen, S.M.K., Dau, T., Moore, B.C.J., 2018. Effect of harmonic rank on sequential sound segregation. Hearing Research 367, 161–168. https://doi.org/10.1016/j.heares.2018.06.002

McPherson, M.J., Grace, R.C., McDermott, J.H., 2022. Harmonicity aids hearing in noise. Attention, Perception, & Psychophysics 84, 1016–1042. https://doi.org/10.3758/s13414-021-02376-0

McPherson, M.J., McDermott, J.H., 2018. Diversity in pitch perception revealed by task dependence. Nature Human Behaviour 2, 52–66. https://doi.org/10.1038/s41562-017-0261-8

Meddis, R., O'Mard, L., 1997. A unitary model of pitch perception. The Journal of the Acoustical Society of America 102, 1811–1820. https://doi.org/10.1121/1.420088

Merzenich, M.M., Knight, P.L., Roth, G.L., 1975. Representation of cochlea within primary auditory cortex in the cat. Journal of Neurophysiology 38, 231–249. https://doi.org/10.1152/jn.1975.38.2.231

Micheyl, C., Carlyon, R.P., Cusack, R., Moore, B.C.J., 2005a. Performance measures of auditory organization, in: Pressnitzer, D., De Cheveigné, A., McAdams, S., Collet, L. (Eds.), Auditory Signal Processing. Springer, New York, NY, pp. 202–210. https://doi.org/10.1007/0-387-27045-0_25

Micheyl, C., Oxenham, A.J., 2010. Pitch, harmonicity and concurrent sound segregation: Psychoacoustical and neurophysiological findings. Hearing Research 266, 36–51. https://doi.org/10.1016/j.heares.2009.09.012

Micheyl, C., Tian, B., Carlyon, R.P., Rauschecker, J.P., 2005b. Perceptual Organization of Tone Sequences in the Auditory Cortex of Awake Macaques. Neuron 48, 139–148. https://doi.org/10.1016/j.neuron.2005.08.039

Miller, G.A., Taylor, W.G., 1948. The Perception of Repeated Bursts of Noise. The Journal of the Acoustical Society of America 20, 171–182. https://doi.org/10.1121/1.1906360

Montijn, J.S., Meijer, G.T., Lansink, C.S., Pennartz, C.M.A., 2016. Population-Level Neural Codes Are Robust to Single-Neuron Variability from a Multidimensional Coding Perspective. Cell Reports 16, 2486–2498. https://doi.org/10.1016/j.celrep.2016.07.065

Moore, B.C.J., Glasberg, B.R., Peters, R.W., 1986. Thresholds for hearing mistuned partials as separate tones in harmonic complexes. The Journal of the Acoustical Society of America 80, 479–483. https://doi.org/10.1121/1.394043

Moore, B.C.J., Glasberg, B.R., Peters, R.W., 1985. Relative dominance of individual partials in determining the pitch of complex tones. The Journal of the Acoustical Society of America 77, 1853–1860. https://doi.org/10.1121/1.391936

Moore, B.C.J., Gockel, H.E., 2011. Resolvability of components in complex tones and implications for theories of pitch perception. Hearing Research, Annual Reviews 2011 276, 88–97. https://doi.org/10.1016/j.heares.2011.01.003

Moore, B.C.J., Peters, R.W., 1992. Pitch discrimination and phase sensitivity in young and elderly subjects and its relationship to frequency selectivity. The Journal of the Acoustical Society of America 91, 2881–2893. https://doi.org/10.1121/1.402925

Norman-Haignere, S.V., Kanwisher, N., McDermott, J.H., Conway, B.R., 2019. Divergence in the functional organization of human and macaque auditory cortex revealed by fMRI responses to harmonic tones. Nature Neuroscience 22, 1057–1060. https://doi.org/10.1038/s41593-019-0410-7

Osmanski, M.S., Song, X., Wang, X., 2013. The Role of Harmonic Resolvability in Pitch Perception in a Vocal Nonhuman Primate, the Common Marmoset (*Callithrix jacchus*). The Journal of Neuroscience 33, 9161–9168. https://doi.org/10.1523/JNEUROSCI.0066-13.2013

Oxenham, A.J., 2023. Questions and controversies surrounding the perception and neural coding of pitch. Frontiers in Neuroscience 16, 1074752. https://doi.org/10.3389/fnins.2022.1074752

Oxenham, A.J., 2012. Pitch Perception. Journal of Neuroscience 32, 13335–13338. https://doi.org/10.1523/JNEUROSCI.3815-12.2012

Oxenham, A.J., Micheyl, C., Keebler, M.V., Loper, A., Santurette, S., 2011. Pitch perception beyond the traditional existence region of pitch. Proceedings of the National Academy of Sciences 108, 7629–7634. https://doi.org/10.1073/pnas.1015291108

Pachitariu, M., Sridhar, S., Pennington, J., Stringer, C., 2024. Spike sorting with Kilosort4. Nature Methods 21, 914–921. https://doi.org/10.1038/s41592-024-02232-7

Pachitariu, M., Steinmetz, N., Kadir, S., Carandini, M., Kenneth D., H., 2016. Kilosort: realtime spike-sorting for extracellular electrophysiology with hundreds of channels. bioRxiv. https://doi.org/10.1101/061481

Panzeri, S., Macke, J.H., Gross, J., Kayser, C., 2015. Neural population coding: combining insights from microscopic and mass signals. Trends in Cognitive Sciences 19, 162–172. https://doi.org/10.1016/j.tics.2015.01.002

Paradiso, M.A., 1988. A theory for the use of visual orientation information which exploits the columnar structure of striate cortex. Biological Cybernetics 58, 35–49. https://doi.org/10.1007/BF00363954

Patterson, R.D., Uppenkamp, S., Johnsrude, I.S., Griffiths, T.D., 2002. The Processing of Temporal Pitch and Melody Information in Auditory Cortex. Neuron 36, 767–776. https://doi.org/10.1016/S0896-6273(02)01060-7

Plack, C.J., Barker, D., Hall, D.A., 2014. Pitch coding and pitch processing in the human brain. Hearing Research 307, 53–64. https://doi.org/10.1016/j.heares.2013.07.020

Plack, C.J., Fay, R.R., Oxenham, A.J., Popper, A.N. (Eds.), 2005. Pitch: Neural Coding and Perception, Springer Handbook of Auditory Research. Springer New York, New York, NY. https://doi.org/10.1007/0-387-28958-5

Pollack, I., 1968. Detection and Relative Discrimination of Auditory "Jitter." The Journal of the Acoustical Society of America 43, 308–315. https://doi.org/10.1121/1.1910780

Popham, S., Boebinger, D., Ellis, D.P.W., Kawahara, H., McDermott, J.H., 2018. Inharmonic speech reveals the role of harmonicity in the cocktail party problem. Nature Communications 9, 2122. https://doi.org/10.1038/s41467-018-04551-8

Pressnitzer, D., Patterson, R.D., 2001. Distortion products and the perceived pitch of harmonic complex tones, in: Breebart, D.J., Houtsma, A.J.M., Kohlrausch, A., Prijs, V.F., Schoonoven, R. (Eds.), Physiological and Psychophysical Bases of Auditory Function. Shaker, Maastricht, pp. 97–104.

Pressnitzer, D., Sayles, M., Micheyl, C., Winter, I.M., 2008. Perceptual Organization of Sound Begins in the Auditory Periphery. Current Biology 18, 1124–1128. https://doi.org/10.1016/j.cub.2008.06.053

Ralston, J.V., Herman, L.M., 1995. Perception and generalization of frequency contours by a bottlenose dolphin (Tursiops truncatus). Journal of Comparative Psychology 109, 268–277. https://doi.org/10.1037/0735-7036.109.3.268

Ramamurthy, D.L., Recanzone, G.H., 2017. Spectral and spatial tuning of onset and offset response functions in auditory cortical fields A1 and CL of rhesus

macaques. Journal of Neurophysiology 117, 966–986. https://doi.org/10.1152/jn.00534.2016

Rees, A., Palmer, A.R., 1988. Rate-intensity functions and their modification by broadband noise for neurons in the guinea pig inferior colliculus. The Journal of the Acoustical Society of America 83, 1488–1498. https://doi.org/10.1121/1.395904

Roberts, B., Brunstrom, J.M., 1998. Perceptual segregation and pitch shifts of mistuned components in harmonic complexes and in regular inharmonic complexes. The Journal of the Acoustical Society of America 104, 2326–2338. https://doi.org/10.1121/1.423771

Rossant, C., Kadir, S.N., Goodman, D.F.M., Schulman, J., Hunter, M.L.D., Saleem, A.B., Grosmark, A., Belluscio, M., Denfield, G.H., Ecker, A.S., Tolias, A.S., Solomon, S., Buzsaki, G., Carandini, M., Harris, K.D., 2016. Spike sorting for large, dense electrode arrays. Nature Neuroscience 19, 634–641. https://doi.org/10.1038/nn.4268

Scholl, B., Gao, X., Wehr, M., 2010. Nonoverlapping Sets of Synapses Drive On Responses and Off Responses in Auditory Cortex. Neuron 65, 412–421. https://doi.org/10.1016/j.neuron.2010.01.020

Schreiner, C., Langner, G., 1988. Periodicity coding in the inferior colliculus of the cat II. Topographical organisation. Journal of Neurophysiology 60, 1823–1840.

Shackleton, T.M., Carlyon, R.P., 1994. The role of resolved and unresolved harmonics in pitch perception and frequency modulation discrimination. The Journal of the Acoustical Society of America 95, 3529–3540. https://doi.org/10.1121/1.409970

Shofner, W.P., Campbell, J., 2012. Pitch strength of noise-vocoded harmonic tone complexes in normal-hearing listeners. The Journal of the Acoustical Society of America 132, EL398–EL404. https://doi.org/10.1121/1.4757697

Shofner, W.P., Chaney, M., 2013. Processing pitch in a non-human mammal (Chinchilla laniger). Journal of comparative psychology (Washington, D.C. : 1983) 127, 142–153. https://doi.org/10.1037/a0029734

Singh, P.G., 1987. Perceptual organization of complex-tone sequences: A tradeoff between pitch and timbre? The Journal of the Acoustical Society of America 82, 886–899. https://doi.org/10.1121/1.395287

Smalt, C.J., Krishnan, A., Bidelman, G.M., Ananthakrishnan, S., Gandour, J.T., 2012. Distortion products and their influence on representation of pitch-relevant information in the human brainstem for unresolved harmonic complex tones. Hearing Research 292, 26–34. https://doi.org/10.1016/j.heares.2012.08.001

Sollini, J., Chapuis, G.A., Clopath, C., Chadderton, P., 2018. ON-OFF receptive fields in auditory cortex diverge during development and contribute to directional sweep selectivity. Nature Communications 9, 2084. https://doi.org/10.1038/s41467-018-04548-3

Song, X., Osmanski, M.S., Guo, Y., Wang, X., 2016. Complex pitch perception mechanisms are shared by humans and a New World monkey. Proceedings of the National Academy of Sciences 113, 781–786. https://doi.org/10.1073/pnas.1516120113

Stanley, G.B., 2013. Reading and writing the neural code. Nature Neuroscience 16, 259–263. https://doi.org/10.1038/nn.3330

Steinschneider, M., Reser, D.H., Fishman, Y.I., Schroeder, C.E., Arezzo, J.C., 1998. Click train encoding in primary auditory cortex of the awake monkey: Evidence for two mechanisms subserving pitch perception. The Journal of the Acoustical Society of America 104, 2935–2955. https://doi.org/10.1121/1.423877

Stewart, L., Von Kriegstein, K., Warren, J.D., Griffiths, T.D., 2006. Music and the brain: disorders of musical listening. Brain 129, 2533–2553. https://doi.org/10.1093/brain/awl171

Summers, V., Leek, M.R., 1998. F0 Processing and the Separation of Competing Speech Signals by Listeners With Normal Hearing and With Hearing Loss. Journal of Speech, Language, and Hearing Research 41, 1294–1306. https://doi.org/10.1044/jslhr.4106.1294

Sumner, C.J., Palmer, A.R., 2012. Auditory nerve fibre responses in the ferret. European Journal of Neuroscience 36, 2428–2439. https://doi.org/10.1111/j.1460-9568.2012.08151.x

Sumner, C.J., Wells, T.T., Bergevin, C., Sollini, J., Kreft, H.A., Palmer, A.R., Oxenham, A.J., Shera, C.A., 2018. Mammalian behavior and physiology converge to confirm sharper cochlear tuning in humans. Proceedings of the National Academy of Sciences 115, 11322–11326. https://doi.org/10.1073/pnas.1810766115

Tarka, V.M., Gaucher, Q., Walker, K.M.M., 2025. Pitch selectivity in ferret auditory cortex. https://doi.org/10.1101/2025.11.03.686220

Tramo, M.J., Shah, G.D., Braida, L.D., 2002. Functional Role of Auditory Cortex in Frequency Processing and Pitch Perception. Journal of Neurophysiology 87, 122–139. https://doi.org/10.1152/jn.00104.1999

Verpy, E., Weil, D., Leibovici, M., Goodyear, R.J., Hamard, G., Houdon, C., Lefèvre, G.M., Hardelin, J.-P., Richardson, G.P., Avan, P., Petit, C., 2008. Stereocilin-deficient mice reveal the origin of cochlear waveform distortions. Nature 456, 255–258. https://doi.org/10.1038/nature07380

Vliegen, J., Moore, B.C.J., Oxenham, A.J., 1999. The role of spectral and periodicity cues in auditory stream segregation, measured using a temporal discrimination task. The Journal of the Acoustical Society of America 106, 938–945. https://doi.org/10.1121/1.427140

Vliegen, J., Oxenham, A.J., 1999. Sequential stream segregation in the absence of spectral cues. The Journal of the Acoustical Society of America 105, 339–346. https://doi.org/10.1121/1.424503

Walker, K.M.M., Bizley, J.K., King, A.J., Schnupp, J.W.H., 2011a. Cortical encoding of pitch: Recent results and open questions. Hearing Research, Auditory Cortex: Current Concepts in Human and Animal Research 271, 74–87. https://doi.org/10.1016/j.heares.2010.04.015

Walker, K.M.M., Bizley, J.K., King, A.J., Schnupp, J.W.H., 2011b. Multiplexed and robust representations of sound features in auditory cortex. The Journal of Neuroscience 31, 14565–14576. https://doi.org/10.1523/JNEUROSCI.2074-11.2011

Walker, K.M.M., Gonzalez, R., Kang, J.Z., McDermott, J.H., King, A.J., 2019. Across-species differences in pitch perception are consistent with differences in cochlear filtering. eLife 8, e41626. https://doi.org/10.7554/eLife.41626

Walker, K.M.M., Schnupp, J.W.H., Hart-Schnupp, S.M.B., King, A.J., Bizley, J.K., 2009. Pitch discrimination by ferrets for simple and complex sounds. The Journal of the Acoustical Society of America 126, 1321–1335. https://doi.org/10.1121/1.3179676

Wang, X., Lu, T., Snider, R.K., Liang, L., 2005. Sustained firing in auditory cortex evoked by preferred stimuli. Nature 435, 341–346. https://doi.org/10.1038/nature03565

Wang, X., Walker, K.M.M., 2012. Neural Mechanisms for the Abstraction and Use of Pitch Information in Auditory Cortex. Journal of Neuroscience 32, 13339–13342. https://doi.org/10.1523/JNEUROSCI.3814-12.2012

Ward, J., Tarka, V.M., Diuba, A., Walker, K.M.M., 2026. Pitch at the Cocktail Party: A Comparative Approach to Studying Selective Attention. Biology 15, 618. https://doi.org/10.3390/biology15080618

Warrier, C.M., Zatorre, R.J., 2004. Right temporal cortex is critical for utilization of melodic contextual cues in a pitch constancy task. Brain 127, 1616–1625. https://doi.org/10.1093/brain/awh183

Whitfield, I.C., 1980. Auditory cortex and the pitch of complex tones. The Journal of the Acoustical Society of America 67, 644–647. https://doi.org/10.1121/1.383889

Winter, I.M., 2005. The Neurophysiology of Pitch, in: Plack, C.J., Fay, R.R., Oxenham, A.J., Popper, A.N. (Eds.), Pitch: Neural Coding and Perception. Springer, New York, NY, pp. 99–146. https://doi.org/10.1007/0-387-28958-5_4

Wright, A.A., Rivera, J.J., Shyan, M., Hulse, S.H., Neiworth, J.J., 2000. Music perception and octave generalization in rhesus monkeys. Journal of Experimental Psychology 129, 291–307. https://doi.org/10.1037/0096-3445.129.3.291

Yin, P., Fritz, J.B., Shamma, S.A., 2010. Do ferrets perceive relative pitch? The Journal of the Acoustical Society of America 127, 1673–1680. https://doi.org/10.1121/1.3290988

Yost, W.A., Mapes-Riordan, D., Shofner, W., Dye, R., Sheft, S., 2005. Pitch strength of regular-interval click trains with different length "runs" of regular intervals. The Journal of the Acoustical Society of America 117, 3054–3068. https://doi.org/10.1121/1.1863712

Zatorre, R.J., 1988. Pitch perception of complex tones and human temporal-lobe function. The Journal of the Acoustical Society of America 84, 566–572. https://doi.org/10.1121/1.396834

Zurita, P., Villa, A.E.P., de Ribaupierre, Y., de Ribaupierre, F., Rouiller, E.M., 1994. Changes of single unit activity in the cat's auditory thalamus and cortex associated to different anesthetic conditions. Neuroscience Research 19, 303–316. https://doi.org/10.1016/0168-0102(94)90043-4

Zwicker, U.T., 1984. Auditory recognition of diotic and dichotic vowel pairs. Speech Communication 3, 265–277. https://doi.org/10.1016/0167-6393(84)90023-2