Original Research Article

BIG DATA & SOCIETY

Big Data & Society
October–December: 1–15

DOI: 10.1177/20539517241297889
journals.sagepub.com/home/bds
Sage

Sounding out voice biometrics: Comparing and contrasting how the state and the private sector determine identity through voice

Daniel Leix Palumbo[1] and **Robert Prey**[1,2]

Abstract

The voice biometrics industry is promised today as a new center of digital innovation. Tech companies and state agencies are massively investing in speech recognition and analysis systems, pushed by the belief that the acoustics of voice contain unique individual characteristics to convert into information and value through artificial intelligence. This article responds to this current development by exploring the under-researched datafication of the auditory realm to reveal how the sound of voice is emerging as a site for identity construction by both states and corporations. To do so, we look at two different case studies. First, we examine a patent granted to the streaming service Spotify, which aims to improve the platform's music recommendation system by analyzing users' speech. Second, we discuss the use of voice biometrics in German asylum procedures, where the country of origin of undocumented asylum seekers is determined through accent analysis. Through these seemingly distinct case studies, we identify not only the common assumptions behind the rationale for adopting voice biometrics, but also important differences in the way the private sector and the State determine identity through the analysis of the sounding voice. These two entities are rarely examined together and are often conflated when addressing practices of auditory surveillance. Thus, our comparative and contrastive approach contributes to existing scholarship that questions the claimed efficiency and ethics of voice biometrics' extractive practices, further defining the operations and assumptions of the private sector and the State.

Keywords

Voice biometrics, voice, sound, datafication, identity, comparative case study

Introduction

Voice has regained central importance as a medium of interaction through networked technologies and as a site of value creation (Gallego, 2021). We are in the era of the voice technology revolution (Miller, 2022) and witnessing the emergence of the voice biometrics industry. Voice biometrics converts non-verbal elements of speech into digital data, combining vocal tract physiology and speaking behavior to identify individuals or map unique characteristics. Benefitting from advancements in machine learning, tech corporations and state agencies are heavily investing in this field, with the global market projected to reach about USD 10 billion by 2028 (Wadhwani and Ganka, 2022).

Current and potential uses of this technology vary widely. In the private sector, banks, insurance companies, and call centers increasingly use voice authentication to verify speakers and connect them to accounts, as this is considered a more secure gateway than pins and passwords for transactions or other operations (Turow, 2022). Call centers also use voice biometrics to monitor mood, detecting distress or interest in both agents and customers, and suggesting when to be more empathetic or when to slow down in order to improve customer relations (Simonite, 2018). Companies like Amazon and Google release patents aiming to detect demographic information, health, and emotions from voice (Turow, 2022), marketing these solutions at international AI voice summits (Burgess, 2022).

[1]Centre for Media and Journalism Studies, University of Groningen, Groningen, The Netherlands
[2]Oxford Internet Institute, University of Oxford, Oxford, GB

Corresponding author:
Daniel Leix Palumbo, Centre for Media and Journalism Studies, University of Groningen, Broerstraat 5, 9712 CP Groningen, The Netherlands.
Email: d.f.g.leix.palumbo@rug.nl

The State also increasingly relies on voice biometrics, after having fostered the development of forensic practices for speaker identification in the last century (Li and Mills, 2019). Edward Snowden's revelations disclosed how the US National Security Agency (NSA) used voice biometrics to identify speakers during the Global War on Terrorism. The system's algorithms were fed data captured by millions of phone conversations, internet calls, and video conferences both in the USA and overseas (Kofman, 2018). Interpol uses BATVOX (Kang, 2022), a voice biometrics technology developed by Nuance Communication, and, in 2014, launched the Speaker Identification Integrated Project (SiiP), a European initiative that created the first international voice biometrics database for law enforcement (Jansen et al., 2021).

The common denominators in these different uses are: (1) the belief that the acoustic aspects of voice contain unique characteristics about an individual, and (2) that AI can extract and convert these into information and value. At its core is the notion that voice reveals one's essential and authentic identity (Kang, 2022), and that digital data can objectively quantify voice (van Dijk, 2014). This reframed socio-technical imaginary drives the global multi-billion-dollar voice biometrics industry (Kang, 2022), treating voice as a gold mine for hyper-personalized targeting and state control. While we are still at an early stage, the technologies have been developed and the patents granted (Turow, 2022). The prevalence of virtual assistants has already normalized to give away voice according to Joseph Turow (2022), who warns of this subtle, pervasive surveillance, yet largely ignored by the press and policymakers.

Behind the seductive convenience offered by voice biometrics lies the amplification of long-standing forms of discrimination. The normative beliefs and assumptions of the voice biometrics industry treat voice, body, and identity as fixed, correlatable objects. This leads to reductionist categories of identity based on the detection of so-called "innate" physiological characteristics (Kang, 2022). For example, law enforcement's current use of voice biometrics for criminal profiling poses particular risks to marginalized and over-scrutinized social groups (Jansen et al., 2021).

In this article, we contribute to critical investigations of the voice biometrics industry by re-examining the sounding voice. By "sounding voice," we mean the auditory elements that complement language, such as pitch, intonation, pace, accent, amplitude, and frequency, through which various studies have approached voice as an embodied and sociocultural phenomenon (Cavarero, 2005; Eidsheim, 2019; LaBelle, 2014; Neumark, 2010). The qualities of voice go beyond what is said: extralinguistic elements of voice express affective and political dimensions (Kanngieser, 2012). These dimensions gain increasing importance for top-down processes of datafication, thereby extending the "crisis of voice" discussed in recent scholarly work (Chouliaraki and Zaborowski, 2017; Couldry, 2010; Georgiou, 2018). Voice is considered in this literature as a process for giving an account of oneself (Butler, 2005). Voice allows individuals to narrate their world from their embodied position (Couldry, 2010). We refer to this understanding of voice as "the narrative voice." While sound is not the only way to "voice" oneself (non-verbal people can also narrate their being), we understand the sounding voice as a site of the narratable self. A sounding utterance is the creation of a world—it is a performative, constituted, and relational act (Butler, 2005; Kanngieser, 2012) that reductionist assumptions informing the use of voice biometrics do not recognize.

While building on previous work questioning the claimed efficiency of voice biometrics' extractive practices, we believe there is a need to define commonalities and distinctions between how the State and the private sector employ voice biometrics. We thus build our argument from two different case studies. In the first case, we look at a Spotify patent for the detection of information from voice. This technology would allow the Swedish company to gather demographic information and even emotional states from its users' voices in order to further personalize the listening experience. The global leader in audio streaming, Spotify provides us with a representative case study of the broader "voice intelligence industry" (Couldry and Turow, 2022) that is expanding the techniques and assumptions behind voice surveillance deep into our homes and our everyday lives. We will analyze both the patent and Spotify's evolving privacy policies. Given the opacity that surrounds technology companies, the analysis of patents and privacy policies offers researchers a means through which to make sense of technological developments (Daim, et al. 2006; Delfanti and Frey, 2021; Viera Magalhães and Couldry, 2021). Tracing updates to privacy policies helps us track corporate strategies over time. Patents, on the other hand, can be considered containers of a company's future aspirations.

In the second case, we look at how the German Bundesamt für Migration und Flüchtlinge (BAMF, or federal office for migration and refugees) uses voice biometrics to analyze accents in order to determine the country of origin of undocumented asylum seekers and their eligibility for asylum. Analyzing language to assess the legitimacy of asylum applicants' claims is not new. Since the 1990s, border agencies of many countries have hired language experts to assess individual cases under forensic linguistics programs, such as the Language Analysis for the Determination of Origin (LADO) (Pfeifer, 2023a). To speed up processing, Germany has been the first and only country in Europe to shift to an automated solution by adopting voice biometrics (Ozkul, 2023). To analyze this case, we examine official brochures and research publications by the BAMF, slides from an official presentation at the Council of Europe's 2020 conference, information released on the agency's website and internal

documents released following parliamentary inquiries. We also use documents released following freedom of information requests made through the online platform FragdenStaat[1]. Finally, we rely on information gathered by the journalist Anna Biselli who has been conducting important investigative work on this case.

Explaining and justifying the case studies

Before we review discussions about the politics of what we call the "narrative voice" and the "sounding voice," we first need to briefly explain and justify our choice of two seemingly divergent case studies. Following more recent comparative case study methods to research new media (Bartlett and Vavrus, 2017; Hallin, 2020), we choose these two case studies to uncover shared assumptions and logics inherent in the development of voice biometrics, and used to promote these technologies to various actors. At the same time, this research design demonstrates how different actors in the private sector and the State adopt the same technology but develop distinct practices based on respective contexts of use, allowing conceptual distinctions to be drawn. Specifically, we show how Spotify and BAMF have two very different ways of understanding and deriving identity through voice biometrics.

Thus, while we build on existing research questioning the efficiency and ethics of voice biometrics' extractive practices, our contribution lies in our comparative approach. By identifying commonalities and distinctions between the private sector and the State's use of the same technology, we disentangle distinct uses, which are often conflated.

Naturally, the limits of our approach are inherent to the comparative case study approach. The very choice of which case studies to examine and compare influences the findings and resulting arguments. We are also aware that Spotify cannot be representative of the whole private sector, just as the BAMF cannot stand for the State. Nevertheless, Spotify is not only the leading global audio streaming company, but is also a driving force in the development of voice-enabled human–computer interaction. Spotify thus offers insights into potential future developments in the practices and assumptions around voice biometrics in the private sector. Similarly, BAMF, as the first to adopt voice biometrics for determining the country of origin in asylum procedures and a promoter of tech solutions with partner countries, highlights how state agencies may incorporate voice biometrics in the future.

What is at stake in both cases is the value of voice in both its narrative and sounding dimensions. In the following review, we highlight the importance of the politics of these dimensions before analyzing our case studies.

Voice as a process and as a value

In the context of politics, the narrative voice is traditionally regarded as the representational channel of self-expression and determination (Arendt, 1958). Following Nick Couldry (2010), we define voice as the process of giving an account of oneself, of narrating one's sense of the self, one's experience of the world and the conditions under which one lives. A democratic system that values voice provides the conditions for such an account not only to be expressed, but to be also taken into consideration (Couldry, 2010). Couldry argues that contemporary neoliberal democracies are oxymoronic forms of democracy because they do not set the conditions for particular voices to be taken into consideration.

And yet, in this framework which Couldry defines as the contemporary crisis of voice (2010), there is a continuous call for more voices in the public sphere. That alone, however, does not equate with recognition—with having one's voice valued. An example is that of subaltern voices like those of migrants in European media. Migrants are rarely narrators of their own stories. They are often stripped of any form of personhood, as they are portrayed as either the threatening others or the vulnerable victim (Chouliaraki and Zaborowski, 2017). There is no value given to these voices as those of unique political subjects. In this article we add to this debate by analyzing how the datafication of the sounding voice as an extractive practice contributes to undermining the political value of voice.

The sound of voice: Acoustics matter

Voice, however, is much more than narrative, discourse, or theme. When we listen to voices, we focus on their acoustics in an act of capture (Capelletti, 2020). Even if we do not see the source of a voice, we assume we can know about its origin through sound: informed by our sociocultural expectations, we distinguish the voice of an elderly man from a teenage boy, recognize whether someone is happy or not, infer someone's ethnicity or geographic provenience, and so on (Eidsheim, 2019; Weidman, 2015). Such assumptions rely on enduring beliefs of a resilient "folk theory" (Kang, 2023) that sees the sounding voice as a prior innate, stable, unmediated object, that, in opposition to semantics, is culturally independent and can reveal one's inner essential and natural identity (Eidsheim, 2019).

However, voice is an inherently fluid phenomenon, and different variables affect its sonic expression. On a physiological level, voice is produced by a collection of bodily processes and organs that are themselves not static: the whole vocal apparatus undergoes various physical changes caused by factors, such as aging, nutrition, health issues, and inhalation that affect its sounding utterances (2019). Most importantly, these bodily components self-adjust in accordance with cultural performance: the sounding voice reiterates and reflects the sociocultural environment in which it participates. Speakers adopt, often beyond consciousness, techniques to train and arrange their vocal apparatus to produce and stylize the sounds

they emit according to the various communities they belong to (Boland, 2010; Eidsheim, 2019; Kanngieser, 2012). In particular, they adapt their speech patterns according to whom they interact with, and the surrounding situational, circumstantial and emotional variables may impact the acoustics of the vocalizations (Kanngieser, 2012). It is thus productive to think about a sounding utterance as an intersubjective experience, where one's vocalizations are most often unpredictable and connected to the context and those to whom we relate. Milla Tiainen (2013) introduces "constitutive relationality" as a concept to frame this ontological fluidity of voice and to understand it as a composition that comes into being through different social, cultural, and physiological relations as well as an event that repeatedly emerges and occurs in different forms.

This fluid understanding of voice refuses essentialist categorization based on ideas of authenticity and nature in connection to identity, but it does not deny uniqueness in the acoustics of voice. As understood in the framework above, voice is a socially grounded process (Couldry, 2010), and the sounding voice carries traces of our lived experience. The uniqueness of one's voice is then in the distinct trajectory of each voice because the way we each encounter the world is unique (Couldry, 2010). To value voice is then to also recognize the inherent plurality present within the acoustics of each individual voice. In the following section, we analyze how the rising voice biometrics industry values the sounding voice.

Sounding out the voice biometrics industry

The acousmatic question—"Who's this? Who's speaking?" (Eidsheim, 2019)—drives today's voice biometrics industry. The bloom of tech startups and companies of this nascent business see intonations, accents, and frequencies as the gateway to greater insights into identities and desires. In particular, it is believed that by capturing the sounding voice, it is possible to spot something otherwise unrevealed by words—to detect "truths" that one would not want to share or even be conscious of (Turow, 2021).

Besides selling authentication tools, the industry is preparing to commercialize a broader set of extractive processes engaged in the analysis and processing of the sounding voice (Gallego, 2021; Kang, 2023; Turow, 2021). There is academic research that looks at how to detect from voice biomarkers early symptoms of diseases, such as Parkinson's and Alzheimer's (Cummins et al., 2017; Fagherazzi et al., 2021; Latif et al., 2021), emotional states (Akçay and Oğuz, 2020), and demographics (Reubold et al., 2010; Simpson, 2009). The goal of the voice biometrics industry is to integrate and advance this knowledge to develop products for the marketplace. An example is Amazon's Halo wristband launched in 2020, which includes among its features the monitoring of voice tone to detect health and assess emotional states (Hurel and Couldry, 2022).

In his critical account of the voice biometrics industry, Edward Kang (2023) nevertheless questions the effectiveness of such extractive practices. By using the analogy of race as a social technology (Benjamin, 2019), he analyzes how the industry advances a sociotechnical imaginary (Kang, 2022), which renders voice an analytical object that can be linked to body and identity, further reinforcing socially constructed classifications of difference. As such, voice—an extremely fluid phenomenon as we described in the previous section—needs to be made "learnable," which means that it must be translated from a qualitative problem into a quantitative machine learning framework (Kang, 2023). This ontological reduction simplifies voice as a fixed mathematical object from which to identify patterns to correlate to other highly complex qualitative variables, such as emotion, gender, accent, intent, etc., which need to be made learnable in turn. These negotiations cause a series of ontological dissonances within the chain of decisions made in developing voice profiling machine learning systems, highlighting the weak foundations and epistemological limits of the voice biometrics industry's intended extractive practices (2023).

Similar criticisms have been leveled at forensic linguist investigations in courtroom cases or in the domain of migration, as in the already mentioned case of LADO. Scholars in sociolinguistics indicated how monoglossic and homogenetic language ideologies affect these practices and may have detrimental consequences on the cases by oversimplifying the nature of language as being static and context- and exchange-free (Campbell, 2013; Eades et al., 2023). In addition, Michelle Weitzel (2018) emphasizes how behind the allure of science and human expertise, these seemingly objective linguistic analyses present the semantic and acoustic features of voice as evidence to make different forms of assessment while fundamentally disregarding individual subjectivity and undermining one's right to voice oneself.

The turn to voice biometrics is symptomatic of what Slavoj Zizek identifies as "the demise of symbolic efficiency" (1999) in post-modern culture. In an era of information surfeit, one is increasingly aware and reflexive of the instability and constructed character of representation, as well as the absence of absolute knowable truths (Andrejevic, 2013). Consequently, the awareness of the multitude of fleeting and perspectival narratives leads to a culture of suspicion toward discourse and strives to chase "truths" that have not yet been mediated by representation (2013). This suspicion results in outsourcing processes of knowledge and sense-making to automated systems, such as voice biometrics, which bypass the untrustworthiness of language to detect in the sounding voice "the essential and unmediated truth" of who one is and what one

desires. Thus, at the core of the project of legitimating voice biometrics is a "post-comprehension strategy" (Andrejevic, 2013). The content of one's account is considered unreliable and not important, what matters is instead that which can be seen, observed and documented from the body or one's behaviors.

In this regard, the underlying assumptions that drive voice biometrics share a loose affinity with behaviorism —the scientific movement that peaked in the mid-20th century. Behaviorism rejected introspection in favor of "objective" explanations of how and why people do what they do–derived solely from observable and measurable behavior (Moore, 1999). Like early behaviorists, proponents of voice biometrics mistrust self-reports. By prioritizing the "sounding voice" over the "narrative voice," voice biometrics meet the behaviorist "rule" of shifting attention from inner states to overt behavior (de Jong and Prey, 2022). As we will see in the sections that follow, this has potential implications for how we "voice" our identities, and even our everyday fleeting desires.

In what follows, we contribute to the critique of voice biometrics and reductionist attempts to determine identity. We however emphasize the distinct ways that power informs how identity is understood and how identity is derived from voice. The operations of the State and the private sector are often conflated when addressing practices of auditory surveillance. In the following two case studies, we not only explicate commonalities in the rationale for adopting voice biometrics, but also indicate important differences in the way the State and the private sector determine identity through the analysis of the sounding voice.

The value of the voice in everyday life

In 2021, the music streaming powerhouse Spotify was granted a patent for a method of processing voice and background audio signals. The patent—"Identification of taste attributes from an audio signal" (Hulaud, 2021)—would allow the company to recommend music and other contents based on the analysis of features it identifies in a Spotify user's voice.

To better grasp the rationale behind the development of this patent, Spotify's turn to voice needs to first be situated within a longer history of attempts at music taste prediction and personalization. In the early days of recommendation system development, users were coaxed into providing explicit feedback about their music preferences through ratings, "thumbs," or by asking the user for demographic information, such as gender or age (de Jong and Prey, 2022). Over time, a different approach gained prominence —one that focused on tracking implicit behaviors, such as user interactions with streams or other traceable activities (Jannach et al., 2018; Seaver, 2019: 430). This shift occurred due to the realization that explicit user data like ratings are unreliable. Explicit ratings exhibit significant variability based on time and context: a user might rate a song three stars one day and five stars the next. Moreover, explicit data are relatively scarce as they demand considerable effort and time from users. In contrast, implicit behavioral data prove adept at predicting future user actions and are more readily available, making them easier to gather (Ekstrand and Willemsen, 2016). The challenge, thus, is to glean enough information from the user without them needing to actively tell the service what they want to listen to. It is within this context that we can better understand the role of voice biometrics within the music recommendation sector. Voice is seen to provide the ideal vehicle through which to access one's desires or "inner state."

Spotify is not alone in turning to voice biometrics in the highly competitive music and audio streaming market. However, voice was not always prioritized as a source of data at Spotify. An analysis of the company's evolving privacy policies over the past decade reveals that the first mention of "voice data" does not appear in Spotify's privacy policy until 2018 (coincidentally or not, the same year the patent is first applied for). In 2020, a separate "voice control policy" is introduced, wherein Spotify describes in detail what voice data it collects and what it does with these data. In the following year, the term "voice assistants" appeared in the privacy policy for the first time as potential third-party sources of personal data. Finally, a June 2022 update of the policy expands the definition of "voice data" from audio clips to also include transcripts of recordings.

The pattern is thus clear: voice has assumed an increasingly central role within the data accumulation and analysis strategy at Spotify. The specific claims made about voice biometrics can be further assessed through an analysis of Spotify's patent application. The patent describes how taste attributes can be derived from an audio signal as follows: a microphone first picks up audio signals of voice and background noises; an audio signal processor then converts the voice signals to digital data for storage or further processing; and a speech recognition process[2] is performed on the voice signals, while background noises are further processed to retrieve environmental metadata. It is claimed that this process can identify emotions, gender, age, accent, and "numerous other characterizations and classifications" through the voice signals (Hulaud, 2021: 7). The voice content and environmental metadata are fed through a computer processor to generate taste attributes. These taste attributes are then used to determine preferences for media content, such as a particular song or artist (see Figure 1).

The patent offers a number of options for classifying a user's emotional state through their voice. These include a model, called Parrott's emotions by groups, which uses a tree-structured list of emotions with levels (e.g. primary, secondary, tertiary) and their corresponding confidence

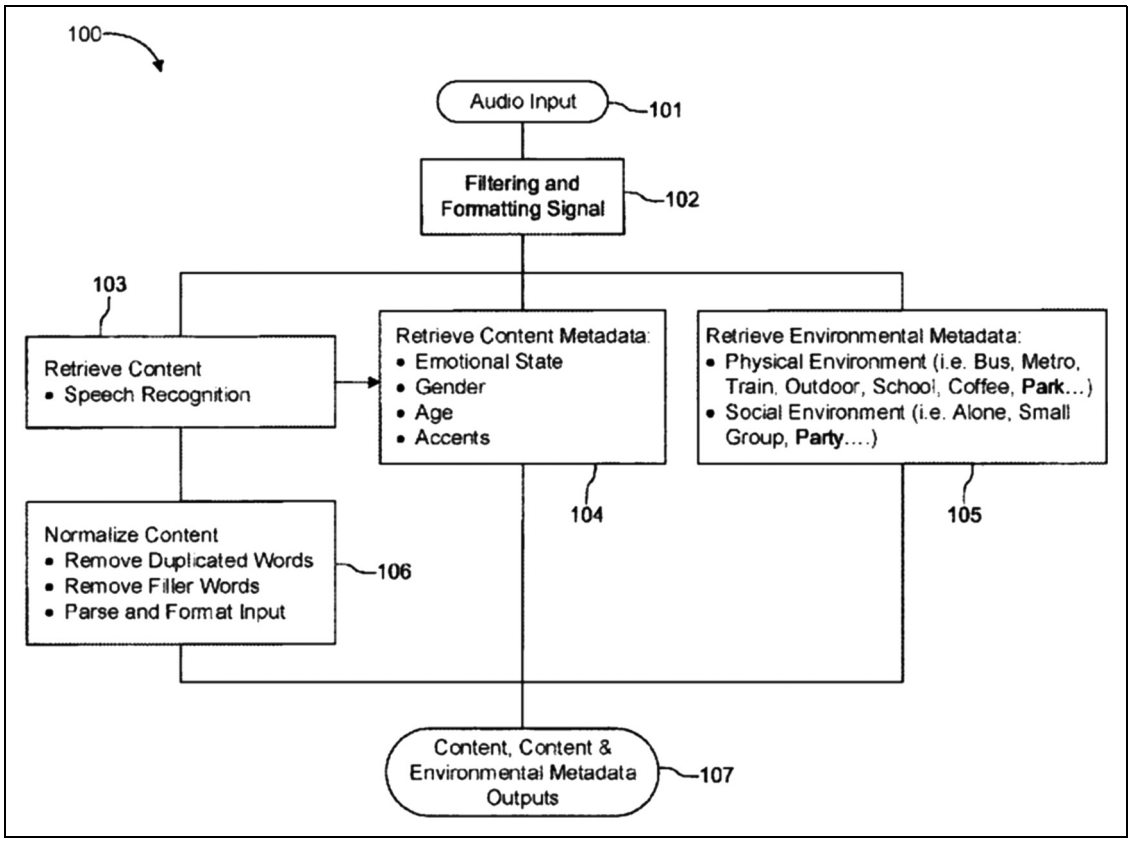


Figure 1. Flow diagram illustrating the process of determining attributes from voice and background audio (Hulaud, 2021).

measures. The example provided in the patent is: ″emotions″: {″primary″: ″joy″, "confidence″:0.876, ″secondary″: ″cheerfulness″, ″confidence″:0.876, ″tertiary″: ″delight″, ″confidence″:0.876}. As an alternative, a simpler approach could group emotions into happy, angry, afraid, sad or neutral categories (for ex: ″emotions″: {″type″: ″happy″, ″confidence″:0.876}). Finally, it is suggested that prosodic aspects of speech (e.g. intonation, stress, rhythm, etc.) also facilitate the detection and categorization of the emotional state of a speaker.

The patent also explains how "gender" can be classified: "gender-related information" can be extracted from a speech signal and represented by a set of vectors, called a feature (Hulaud, 2021: 5). For example, frequency differences in the audio signal can be classified by gender. According to the patent, age can be "roughly determined" from voice according to a combination of vocal tract length and pitch. Various approaches can also be used to determine the accent of a voice. For example, acoustic, phonotactic, prosodic, or other types of properties can be extracted with a language/accent verification system. As the patent states, emotions, gender, age, and accent "are merely examples," and the method could be expanded to detect for more attributes of a voice (Hulaud, 2021: 7).

A few months after Spotify was granted this patent, a coalition of over 180 musicians and global human rights organizations signed a letter titled "Dear Spotify: don't manipulate our emotions for profit." The letter called on the company to abandon the technology, charging it with "covert manipulation and monitoring" of users (Access Now, 2021). In its reply, Spotify stated that it had not yet employed the patent nor did it have plans to do so (Spotify, 2021). Regardless, patents, such as this one, provide "a site of inquiry into value" (Aradau and Blanke, 2022: 116). They reveal the broader interests and imaginaries that drive a field. As Viera Magalhães and Avella (2023: 2) put it:

> Patent applications might or might not offer insight into which technologies companies actually use. Yet, as both a corporate performance and strategy to secure monetizable intellectual rights, they illuminate the horizons and directions of technology development within these organizations.

The patenting of this particular technology demonstrates how voice is considered to be a potentially useful terrain upon which to extract valuable data now or in the future. Whether or not this particular patent is employed, Spotify

is only one of many tech companies turning to voice biometrics and, in the process, expanding the techniques of voice surveillance deep into our homes and our everyday lives (Couldry and Turow, 2022).

As companies like Spotify jockey for position among their competitors while at the same time attempting to attract greater ad spend from marketers, voice becomes merely the latest in a long line of claims to provide the most detailed and complete customer profile. The scientific consensus is by no means settled when it comes to determining demographic or emotional information from voice. Moreover, there appear to be critical blind spots that may serve to exacerbate already-existing disparities—speech recognition does not work as well for women's voices, for example (Bajorek, 2019). Whether or not claims about scientific validity are plausible is somewhat beside the point though. If companies and marketers act according to the belief that they are valid, potentially harmful and discriminatory assumptions may be embedded over time. For example, Couldry and Turow (2022: 16) remind us that "[t]here is a longer history of how judgements based on sound have been tied to racial discrimination, but voice intelligence could embed such ties more insidiously and pervasively than before."

A misguided song recommendation that is made on the basis of an assumed gender or emotional state may be a seemingly trivial case of misrecognition. However, the assumption it is based upon is certainly significant: that one's identity, as derived from biometric characteristics of voice, may be in contradiction with what a speaker actually says. As Couldry and Turow (2022: 10) put it, "voice starts to "speak double"— that is, in two potentially conflicting registers, only one of which (expressive voice) can be under the speaker's control."

Of course, there is a critical distinction between voice biometrics as employed by Spotify and our next case, the German border agency, BAMF. Most critically, refugee claimants have no agency in submitting to these methods.

The value of the voice at the border

While the deployment of Spotify's patent seems to remain only an aspiration of a broader imaginary, voice biometrics are currently in actual use in European asylum procedures, but largely outside of scrutiny. Borders and refugee camps have long functioned as testing grounds to experiment with new technologies before opening their use on the general population (Leurs, 2023). As migration became a highly politicized topic during the period of the so-called "European refugee crisis," the "Fortress Europe" approach has materialized not only in the enforcement of deterrence policies, or the outsourcing of borders control to external non-EU third countries, but also in the incremental adoption of data-driven surveillance technologies (Dijstelbloem, 2021) that compose the "digital border" (Chouliaraki and Georgiou, 2022) of Europe. It is within this broader context that we need to situate the processes of digitalization initiated by the BAMF and its adoption of IT tools of identity-making (Beckmann and Biselli, 2020; Witteborn, 2022), including voice biometrics.

Germany faced a significant influx of asylum seekers, exceeding 1.2 million by 2015/2016, revealing flaws in its reception system (Tangermann, 2017). According to BAMF (2018), despite efforts to bolster staffing and tighten regulations, processing delays persisted due to asylum applicants' lack of identity documentation. Consequently, BAMF initiated a digitalization strategy centered on employing IT tools for identity verification (Kreienbrink, 2018).

The flagship of these tools is the voice biometric system that the BAMF decided to fully introduce in 2017 after a period of trial in the city of Bamberg (Leix Palumbo, 2024; Ozkul, 2023; Pfeifer, 2023b; Vieira de Oliveira, 2021). Named "dialect identification assistance system" (DIAS), the tool is used to analyze the accent and dialect of asylum seekers without documents or whose documentation is considered not valid or counterfeit so as to establish their country of origin (FragDenStaat, 2017). The rationale for introducing this tool was to assist the BAMF decision-makers in establishing the identity of undocumented and/or untrusted asylum seekers while speeding up the processing of the applications by filtering out those coming from so-called "safe countries of origin." In addition, the tool serves to provide evidence for justifying the return of the refused asylum applicants to their countries of origin; countries which require proof of identity before they will offer to help with the deportations (Tangermann, 2017).

BAMF's DIAS is based on the third-party software Nuance Speech Suite licensed by Microsoft's Nuance Communication Inc. (Deutscher Bundestag, 2022).[3] The majority of the training data are purchased from the University of Pennsylvania's Linguistic Data Consortium and a small part from the German crowdsourcing company Clickworker GmbH, while another part of the training data consists of speech samples self-developed by the BAMF (Deutscher Bundestag, 2022). At the system's core are the so-called language models, namely, statistical models of different languages and dialects. Machine learning techniques classify the languages and dialects of the training data by identifying statistics and patterns, such as similarities in the frequencies of certain phonemes and their combination (Germany's Presidency of the Council of the European Union, 2020). The result of such classification is the language model, which is a collection of statistics for phonemes, sounds, and acoustic characteristics for each language. As we describe more precisely in the following paragraphs, this works as a reference model against which the asylum seeker's speech sample is matched to identify the country of origin. These technical details of the system introduced by the BAMF reveal a process of

identification based on an analysis, which largely focuses on the acoustic characteristics of language, in opposition to that of a human linguist who would also include in the analysis the syntactical, lexical, morphological, and grammatical aspects (Germany's Presidency of the Council of the European Union, 2020). The establishment of identity is thus based on the capture of the sounding voice.

The practice of identification through the DIAS works as follows (see Figure 2): during the application procedure, the BAMF personnel enters the administrative data connected to the asylum applicant and calls an in-house number to start the identification process; the asylum applicant is then invited to verbally describe in 2 minutes a picture over the phone in the fullest detail possible and with no breaks; the description is recorded and saved in a central file repository (FragDenStaat, 2017); after the speech sample is matched against the software's language models, a report assesses the language/dialect of the asylum applicant in probability percentages (see Figure 3); and, finally, the report is included in an electronic case file as one of the sources of information used to assist the case worker in the overall assessment of eligibility for asylum.

At the moment, the DIAS is solely geared toward assessing Arab-speaking and Persian-speaking asylum applicants (see Figure 4). There is a plan to create a Kurdish language model in the future, as well as models for further languages and dialects of other regions (Deutscher Bundestag, 2022; Germany's Presidency of the Council of the European Union, 2020). From the numbers released during two parliamentary inquiries of 2018 and 2022, the importance that the DIAS has gained as a tool used to assist decisions on asylum is clearly visible (Deutscher Bundestag, 2018, 2022). In the first 3 months after its introduction in 2017, the DIAS was used in 3681 cases. In 2020, the DIAS was used 9923 times, and in 2021, this usage increased to 15,052 times (Deutscher Bundestag, 2022). In total, the DIAS has confirmed the identity of asylum applicants in around 76% of the cases, while it has been used in 24% of the cases to disprove claims (Deutscher Bundestag, 2022).

The DIAS is increasingly becoming a standardized step to assist decision-makers in assessing the eligibility for asylum despite the system's limitations, as acknowledged by the BAMF. The detection rate for Arabic dialects was around 80% between 2017 and 2020 before being improved to 85% in 2021 (Deutscher Bundestag, 2022). For the newly introduced Persian dialects, BAMF is even less reliable: the detection rate reported for Dari and Farsi is 73%, and 77% for Pashto (Deutscher Bundestag, 2022).[4]

The decision-makers received training on how to use the DIAS in a few internal courses and manuals organized by the BAMF, but they have no expertise in linguistics or speech technology. This raises the question of whether they are aware of the limitations of the tool or have the ability to make sense of contradicting or uncertain results (Ozkul, 2023). The BAMF states in this respect that the DIAS is not a replacement for the analysis of language by a trained linguist, but rather a complementary assistant tool that does not determine the final decision but is at disposal of the decision-makers (Germany's Presidency of the Council of the European Union, 2020). In case there are still doubts about the country of origin of the asylum applicant, a longer recording of 30 minutes can be arranged and sent to one language expert for a more in-depth assessment (Ozkul, 2023). It is, however, unclear whether all the uncertain and contradicting results go through this second evaluation.

According to BAMF, the DIAS is not meant to replace the role of the asylum seekers' narratives. If an asylum applicant's identity is contradicted by the DIAS, they have the opportunity to challenge the decision in the following interviews. However, journalist Anna Biselli has reported stories of different asylum applicants who were unjustly rejected due to decisions made by BAMF's DIAS (2018). In particular, the concern is that decision-makers will just blindly follow the results produced by the DIAS when the workload is high, and there is pressure for the BAMF personnel to make decisions quickly. Despite these concerns, BAMF's aim at the moment is to make the automated analysis of language through voice biometrics the new standard; outsourcing to language experts and linguists only when necessary (Germany's Presidency of the Council of the European Union, 2020).[5]

However, the use of the DIAS in asylum procedures relies on an erroneous assumption, which sees language as a marker of geographic origin (Leix Palumbo, 2024; Pfeifer, 2023b; Vieira de Oliveira, 2021). The sounding voice of a speaker changes dramatically throughout life. In particular, the way we speak is very likely to change if we have lived in many different places (Rosenhouse, 2013). Furthermore, the same region is usually inhabited by different communities of speakers and one's language, dialect, or accent can vary much depending on one's social environment (Campbell, 2013). The sounding voice is also not simply accumulative. As previously described, the way we speak is not simply derived from the imitation of parents, relatives, friends, and other people we engage with: various unpredictable contextual, physical, and emotional factors can also influence the acoustics of the sounding voice (Eidsheim, 2019). This series of factors makes establishing a country of origin through voice analysis a problematic and error-prone process of identity determination, whether it is done by a voice biometric system or a commissioned linguist (Abu Hamdan, 2016; Pfeifer, 2023a). We could argue instead that the sounding voice, rather than an identifier of one's country of origin, is instead an outcome of one's life.

Thus, underpinning the use of the DIAS in asylum procedures is the assumption that migrants have only one unique dimension of linguistic socialization (Bellanova

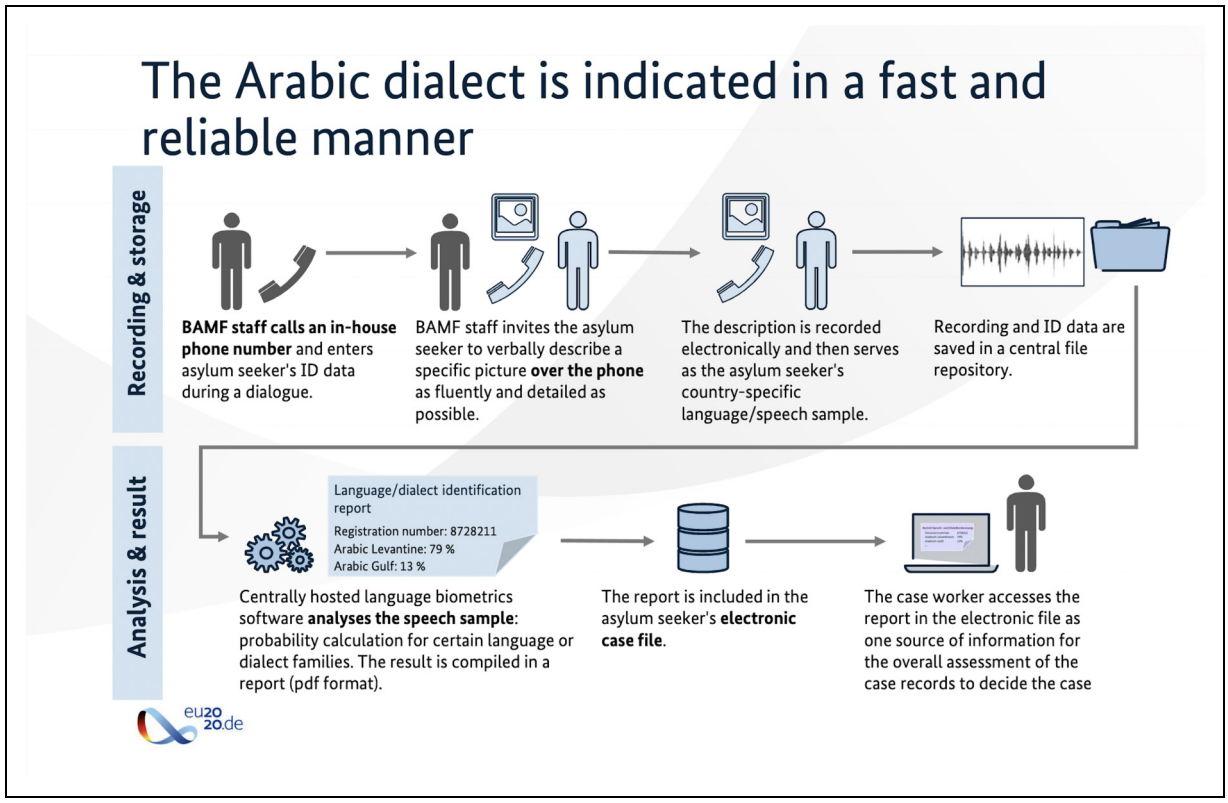


Figure 2. Illustration of voice biometrics' use in asylum procedures (Germany's Presidency of the Council of the European Union, 2020).

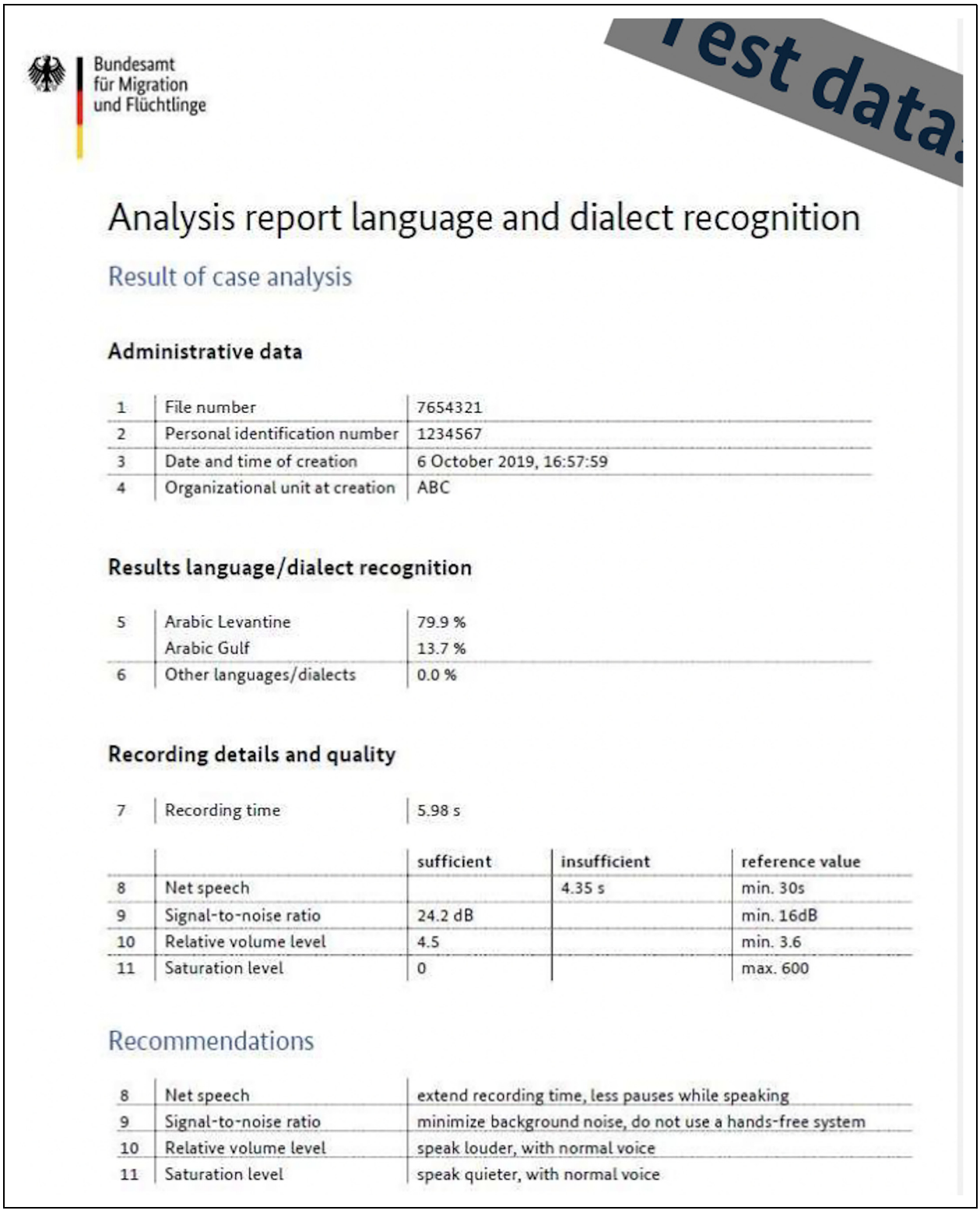
and Fuster, 2019). By operating with this reductionist view, voice biometrics do not only impede attempts to claim asylum, they also reflect and mutually reinforce the objectification of migrants that has for long operated at a symbolic and discursive level (Ahmed, 2000; Chouliaraki and Georgiou, 2022; De Genova, 2013). Voice biometrics erase the series of relations that compose a sounding voice and advance a fetishized notion of the migrant as someone who arrives through a direct route from their country of origin. This denies the erratic and lengthy reality of journeys through third countries and refugee camps. It furthermore reproduces racialized otherness by putting forward the idea that asylum seekers escape areas with a stable and monolingual community.

The dominant rationality informing the use of voice biometrics in asylum procedures does not then recognize the plurality inherent in the sounding voice. In turn, this reductionist view does not value voice as the process of giving an account of oneself. Voice biometrics replace the role of narrative biographical history in the assessment of asylum applications with the extraction of biometric information (Ajana, 2010). In enforcing this shift, the core belief underpinning the voice biometrics industry is promoted—the idea that the sounding voice can allow the detection of unfiltered truths, which would otherwise be obscured by words.

Discussion

As detailed above, both private companies and the State are turning to the sounding voice as a resource for algorithmic decision-making. There are also other motives behind the use of voice biometrics by both of the parties we analyzed. Spotify develops a patent, regardless of its actual use, to demonstrate to investors that it is at the forefront of technological advancements to better understand users and their desires. The BAMF purchases voice biometrics to affirm itself as a leader in innovation and the European digital society (BAMF, 2018). Moreover, the BAMF promotes its use of voice biometrics to other parties in the EU as a techno-solution to fix the flaws of its reception system in the context of the so-called "European refugee crisis" (Germany's Presidency of the Council of the European Union, 2020).

Spotify and the BAMF represent two very different case studies, which reflect distinct modes of power exercised through the datafication of the sounding voice. Nonetheless, these two cases are not only united by the use of the same technology, voice biometrics, but by a logic of mistrust. This logic claims that you are not who you say you are, and that you do not want what you say you want. Instead, your voice betrays what you really



Bundesamt
für Migration
und Flüchtlinge

Test data

Analysis report language and dialect recognition

Result of case analysis

Administrative data

1	File number	7654321
2	Personal identification number	1234567
3	Date and time of creation	6 October 2019, 16:57:59
4	Organizational unit at creation	ABC

Results language/dialect recognition

5	Arabic Levantine	79.9 %
	Arabic Gulf	13.7 %
6	Other languages/dialects	0.0 %

Recording details and quality

7	Recording time	5.98 s		
		sufficient	**insufficient**	**reference value**
8	Net speech		4.35 s	min. 30s
9	Signal-to-noise ratio	24.2 dB		min. 16dB
10	Relative volume level	4.5		min. 3.6
11	Saturation level	0		max. 600

Recommendations

8	Net speech	extend recording time, less pauses while speaking
9	Signal-to-noise ratio	minimize background noise, do not use a hands-free system
10	Relative volume level	speak louder, with normal voice
11	Saturation level	speak quieter, with normal voice

Figure 3. Voice biometrics' result sample (Germany's Presidency of the Council of the European Union, 2020).

want, and who you really are. Processes of algorithmic profiling through voice biometrics are used to gain access to a pre-discursive truth supposedly present in the sounding voice, bypassing one's "unreliable" narrative accounts.

Decision-making in both our cases needs numbers and percentages derived from a valid source to act. In line with the episteme of behaviorism, such sources of objective evidence cannot exist in self-reports, but need to be

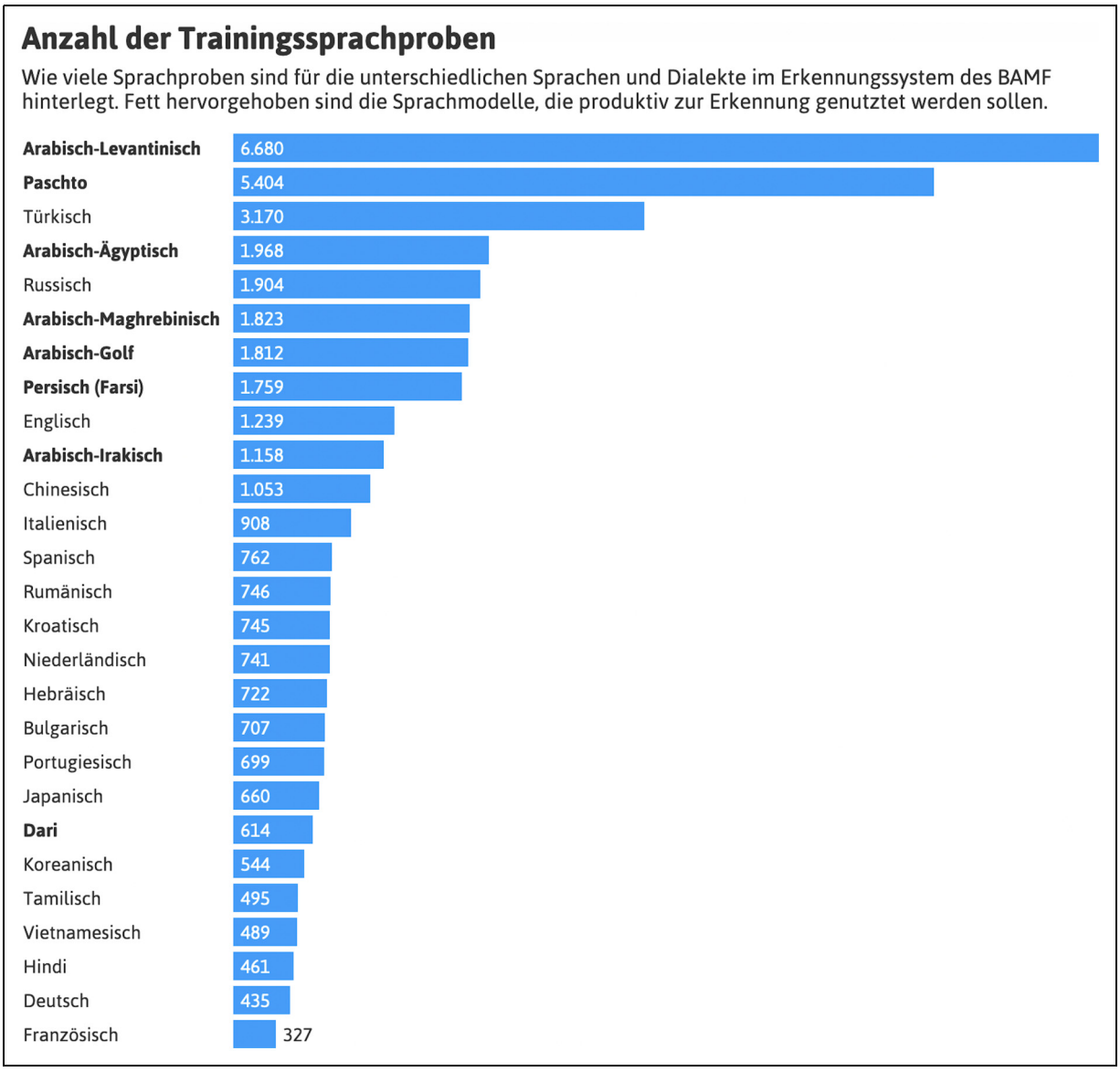


Figure 4. Voice biometrics' training data (Germany's Presidency of the Council of the European Union, 2020).

observable and measurable in overt behavior. Both the BAMF and the Spotify patent seek to make decisions by undermining the narrative of their subjects. The knowledge generated by voice biometrics is considered superior to the self-conscious truth spoken by the subject. As such, instead of an account of oneself, voice morphs into a resource to extract for capitalistic production and State control.

Although united by this logic of mistrust, the cases of Spotify and the BAMF present on the other hand important differences in how the State and the private sector use voice biometrics. They both adopt voice biometrics to derive identity from the sounding voice, the former to better recommend music and the latter to determine country of origin in order to assess asylum claims. However, precisely how identity is understood and derived in the two cases is crucially different. Spotify's patent—like other consumer-facing technologies that employ voice biometrics—fully embraces the fluidity of identity. For Spotify, any listener's musical identity is continuously changing and dependent on context. They therefore promise through voice biometrics to *enact* identity (Law, 2008; Ruppert, 2011) through a continuous process of identification. BAMF's DIAS, on the other hand, understands identity as more fixed and permanent. Rather than enacting identity, it *constructs* identity (Law, 2008): identity for BAMF is seen as a building that needs to be constructed from objective and reliable (voice data) materials.

Users of Spotify or other consumer-facing services often eagerly accept voice biometrics into their everyday lives. If the patent we analyzed is one day employed, Spotify will

promote the idea that it will permit listeners to gain a much more intimate understanding of their musical identity. As Couldry and Turow (2022: 13) argue, for users to willingly submit voice data for collection and analysis, a "deep seduction" is required; one that derives from "the idea that this is done to know human beings better." In submitting to this seduction and inviting voice biometrics into our everyday lives, we run the risk of depriving ourselves from developing our own self-knowledge and ability to "voice" our music (and other) taste preferences. It could be said that while BAMF employs the DIAS to identify "strangers," the use of voice biometrics to recommend music makes us strangers to ourselves.

However, while we have already made a clear distinction between the two cases in terms of agency, it is also worth stressing the divergent stakes of decisions made in these two cases. A Spotify user may disagree with the recommendations or relinquish listening decisions to the system. Voice biometrics as employed by BAMF, however, provides border control with a final fixed identification which may impact one's right to asylum.

Conclusion

In our critical account of voice biometrics, we have detected an extension of the contemporary crisis of voice. New forms of datafied knowledge production that target the sounding voice paradoxically undermine the narrative accounts offered through self-reports of identity. Voice biometrics' extractive practices play a role in the process of commodification of voice, reducing it to pure economic form to mine the sounding voice in the chase of "truths" that undermine any spoken meaning. As such, these practices deny the very political value of voice as the process of giving an account of oneself. They give no meaning to the lived experience of the subject and their sense of the self, and they prevent one's account from having any impact on the hidden decisions that are taken on their behalf.

We do not intend to indicate causality between the use of voice biometrics and this undermining of the narrative voice, as the extractive practices we have described in this article are but the result of current economic and political formations. We are also not arguing that the sounding voice does not contain resourceful information, which should never be analyzed. Current research on the potential uses of voice biometrics in healthcare to detect earlier symptoms of diseases is only one example of the beneficial uses of this technology, further highlighting the necessity of a critical debate that has recently arisen (Couldry and Turow, 2022; Jansen et al., 2021; Kang, 2023; Sterne, 2022; Turow, 2021).

Such debate, however, will have to be founded on an understanding of the sounding voice as a carrier of lived experience. Our analysis of the two cases has not only highlighted a reductionist perspective on the sounding voice and the dismissal of individuals' self-representations, but also emphasized how this erosion of the narratable self is central to the underlying logic of voice biometrics. Our analysis has also drawn attention to the two distinct understandings of identity through which the State and the private sector employ voice biometrics. We believe that this distinction offers the potential for future research in the field of auditory surveillance, which to-date has tended to conflate the operations of the State and private companies. In addition, the comparative approach adopted in this study can be applied to draw commonalities and distinctions also in the study of technologies other than voice biometrics, offering a deeper understanding of how different institutions employ similar technologies.

Declaration of conflicting interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the Nederlandse Organisatie voor Wetenschappelijk Onderzoek (grant number PGW.22.027).

ORCID iDs

Daniel Leix Palumbo https://orcid.org/0009-0005-1620-0212
Robert Prey https://orcid.org/0000-0002-2339-5872

Notes

1. These documents disclose the training manuals provided to the BAMF personnel, numbers on the use of the technology in asylum procedures and figures regarding its error rate, and information regarding the technical details of the technology and its functioning.
2. According to the patent, speech recognition can be performed using existing methodologies, such as hidden Markov models (HMM), dynamic time warping (DTW)-based speech recognition, neural networks, Viterbi decoding, and deep feedforward and recurrent neural networks.
3. Nuance Communication Inc. is a major service provider in the voice biometrics industry, from the development of voice tools for healthcare to fraud detection to audio forensic technologies in the field of security.
4. It is unclear, however, what that percentage of unsuccessful recognition of language stands for. Does it refer to cases where the report results were recognizable by the decision-makers as being obviously wrong (an Arab speaker was assessed as German or English), or to cases where further scrutiny by the decision-makers in the hearings reconsidered the result produced by the DIAS? It remains ambiguous also what happens in those cases in which the result produced is less obvious, for instance, when the probability percentages of two variations of Arabic are quite similar. How do the decision-makers deal with these uncertainties?

5. The main efforts by the BAMF are directed at establishing collaboration with other border agencies in the EU in order to define a common process built around voice biometrics. The aim of the BAMF is to create through such collaboration an interoperable database similar to that of EURODAC for fingerprints, which would allow the border agencies to exchange speech samples, combining the efforts for training and improving the voice biometric software. At the moment, the BAMF has exchanged its anonymized training data with the Dutch border agency IND as part of a pilot project, and it has also exchanged information with various interested partners, including Austria, Finland, Greece, Lithuania, Norway, Sweden, and Switzerland (Biselli, 2022). Clearly, the current response is to improve the system's efficiency by trying to incorporate as much quality training data as possible, so as to improve the present detection rates and reduce error percentages.

References

Abu Hamdan L (2016) *[inaudible] A Politics of Listening in 4 Acts*. London: Sternberg Press.

Access Now (2021) Dear spotify: Don't manipulate our emotions for profit'. Available at: https://www.accessnow.org/spotify-tech-emotion-manipulation/ (accessed 17 May 2023).

Ahmed S (2000) *Strange Encounters: Embodied Others in Post-Coloniality*. London: Routledge.

Ajana B (2010) Recombinant identities: Biometrics and narrative bioethics. *Journal of Bioethical Inquiry* 7(2): 237–258.

Akçay MB and Oğuz K (2020) Speech emotion recognition: Emotional models, databases, features, preprocessing methods, supporting modalities, and classifiers. *Speech Communication* 116: 56–76.

Andrejevic M (2013) *Infoglut: How Too Much Information Is Changing the Way We Think and Know*. New York: Routledge.

Aradau C and Blanke T (2022) *Algorithmic Reason: The New Government of Self and Other*. Oxford: OUP.

Arendt H (1958) *The Human Condition*. Chicago: University of Chicago press.

Bajorek JP (2019) Voice recognition still has significant race and gender biases. Harvard Business Review, 10 May. Available at: https://hbr.org/2019/05/voice-recognition-still-has-significant-race-and-gender-biases (accessed 23 May 2023).

BAMF (2018) *Digitisation Agenda 2020. Success Stories and Future Digital Projects at the Federal Office for Migration and Refugees (BAMF)*. Nuremberg: Federal Office for Migration and Refugees.

Bartlett L and Vavrus F (2017) Comparative case studies: An innovative approach. *Nordic Journal of Comparative and International Education (NJCIE)* 1(1): 5–17.

Beckmann L and Biselli A (2020) *Invading Refugees' Phones: Digital Forms of Migration Control in Germany and Europe*. Berlin: Gesellschaft für Freiheitsrechte e.V.

Bellanova R and González Fuster G (2019) Composting and computing: On digital security compositions. *European Journal of International Security* 4(3): 345–365.

Benjamin R (2019) *Race After Technology: Abolitionist Tools for the New Jim Code*. Cambridge and Medford, MA: Polity.

Biselli A (2018) Eine software des BAMF bringt Menschen in Gefahr. Vice, 20 August. Available at: https://www.vice.com/de/article/a3q8wj/fluechtlinge-bamf-sprachanalyse-software-entscheidet-asyl (accessed 24 March 2023).

Biselli A (2022) BAMF weitet automatische Sprachanalyse aus. Netzpolitik, 5 September. Available at: https://netzpolitik.org/2022/asylverfahren-bamf-weitet-automatische-sprachanalyse-aus/ (accessed 24 March 2023).

Boland P (2010) Sonic geography, place and race in the formation of local identity: Liverpool and Scousers. *Geografiska Annaler: Series B, Human Geography* 92(1): 1–22.

Burgess M (2022) The race to hide your voice. Wired, 1 June. Available at: https://www.wired.com/story/voice-recognition-privacy-speech-changer/ (accessed 17 March 2023).

Butler J (2005) *Giving an Account of Oneself*. New York: Fordham University Press.

Campbell J (2013) Language analysis in the United Kingdom's refugee status determination system: Seeing through policy claims about 'expert knowledge'. *Ethnic and Racial Studies* 36(4): 670–690.

Capelletti M (2020) How the sirens lost their wings: On vocality and silencing. EX NUNC Journal - On Mediterranean. Available at: https://www.ex-nunc.org/exnjournal-mattia-capelletti (accessed 6 June 2023).

Cavarero A (2005) *For More Than One Voice: Toward a Philosophy of Vocal Expression*. Stanford, CA: Stanford University Press.

Chouliaraki L and Georgiou M (2022) *The Digital Border: Migration, Technology, Power*. New York: NYU Press.

Chouliaraki L and Zaborowski R (2017) Voice and community in the 2015 refugee crisis: A content analysis of news coverage in eight European countries. *International Communication Gazette* 79(6–7): 613–635.

Couldry N (2010) *Why Voice Matters: Culture and Politics After Neoliberalism*. London: Sage.

Couldry N and Turow J (2022) Market-driven voice profiling: A framework for understanding. *Advertising & Society Quarterly* 23(3): 1–15.

Cummins N, Schmitt M, Amiriparian S, et al. (2017) "You sound ill, take the day off": Automatic recognition of speech affected by upper respiratory tract infection. In: Annual International Conference of the IEEE Engineering in Medicine and Biology Society, July 2017, pp. 3806–3809: IEEE Engineering in Medicine and Biology Society.

Daim TU, Rueda G, Martin H, et al. (2006) Forecasting emerging technologies: Use of bibliometrics and patent analysis. *Technological Forecasting and Social Change* 73(8): 981–1012.

De Genova N (2013) Spectacles of migrant 'illegality': The scene of exclusion, the obscene of inclusion. *Ethnic and Racial Studies* 36(7): 1180–1198.

de Jong M and Prey R (2022) The behavioral code: Recommender systems and the technical code of behaviorism. In: Cressman D (eds) *The Necessity of Critique. Andrew Feenberg and the Philosophy of Technology*. Cham: Springer International Publishing, 143–159.

Delfanti A and Frey B (2021) Humanly extended automation or the future of work seen through Amazon patents. *Science, Technology, & Human Values* 46(3): 655–682.

Deutscher Bundestag (2018) Drucksache 19/6647: Einsatz von IT-Assistenzsystemen im Bundesamt für Migration und Flüchtlinge. Available at: https://dserver.bundestag.de/btd/19/066/1906647.pdf (accessed 23 March 2023).

Deutscher Bundestag (2022) Drucksache 20/3238: Einsatz von Dialekterkennungssoftware im Bundesamt für Migration und Flüchtlinge. Available at: https://dserver.bundestag.de/btd/20/032/2003238.pdf (accessed 23 March 2023).

Dijstelbloem H (2021) *Borders as Infrastructure: The Technopolitics of Border Control*. Cambridge, MA and London: MIT Press.

Eades D, Fraser H and Heydon G (2023) *Forensic Linguistics in Australia: Origins, Progress and Prospects*. Cambridge, UK: Cambridge University Press.

Eidsheim NS (2019) *The Race of Sound: Listening, Timbre, and Vocality in African American Music*. Durham and London: Duke University Press.

Ekstrand MD and Willemsen MC (2016) Behaviorism is not enough: Better recommendations through listening to users. In: Proceedings of the 10th ACM Conference on Recommender Systems, New York, USA, 7 September 2016, pp. 221–224. New York: Association for Computing Machinery.

Fagherazzi G, Fischer A, Ismael M, et al. (2021) Voice for health: The use of vocal biomarkers from research to clinical practice. *Digital Biomarkers* 5(1): 78–88.

FragDenStaat (2017) Integriertes Identitätsmanagement-Plausibilisierung, Datenqualität und Sicherheitsaspekte. Einführung in die neuen IT-Tools. Available at: https://fragdenstaat.de/dokumente/9653-schulung_idms_bamf/ (accessed 22 March 2023).

Gallego JI (2021) The value of sound: Datafication of the sound industries in the age of surveillance and platform capitalism. *First Monday* 26(7): 1–21.

Georgiou M (2018) Does the subaltern speak? Migrant voices in digital Europe. *Popular Communication* 16(1): 45–57.

Germany's Presidency of the Council of the European Union 2020 (2020) Session I: Language recognition and name transcription. Available at: https://migrationnetwork.un.org/sites/g/files/tmzbdl416/files/docs/cdr_slides_tks_dias_common_language_analysis.pdf (accessed 21 March 2023).

Hallin D (2020) Comparative research, system change, and the complexity of media systems. *International Journal of Communication* 14(12): 5775–5786.

Hulaud S (2021) U.S. Patent No. 10,891,948. Washington, DC: U.S. Patent and Trademark Office.

Hurel LM and Couldry N (2022) Colonizing the home as data-source: Investigating the language of Amazon skills and Google actions. *International Journal of Communication* 16(20): 5184–5204.

Jannach D, Lerche L and Zanker M (2018) Recommending based on implicit feedback. In: Brusilovsky P and He D (eds) *Social Information Access: Systems and Technologies*. Cham: Springer International Publishing, 510–569.

Jansen F, Sánchez-Monedero J and Dencik L (2021) Biometric identity systems in law enforcement and the politics of (voice) recognition: The case of SiiP. *Big Data & Society* 8(2): 1–13.

Kang EB (2022) Biometric imaginaries: Formatting voice, body, identity to data. *Social Studies of Science* 52(4): 581–602.

Kang EB (2023) Ground truth tracings (GTT): On the epistemic limits of machine learning. *Big Data & Society* 10(1): 1–12.

Kanngieser A (2012) A sonic geography of voice: Towards an affective politics. *Progress in Human Geography* 36(3): 336–353.

Kofman A (2018) Forget about Siri and Alexa—When it comes to voice identification, the "NSA reigns supreme". The Intercept, 19 January. Available at: https://theintercept.com/2018/01/19/voice-recognition-technology-nsa/ (accessed 10 March 2023).

Kreienbrink A (2018) Restriction, pragmatic liberalisation, modernisation: Germany's multifaceted response to the "refugee crisis". In: Sirkeci I, Lana de Freitas Castro E and Sezgi Sözen Ü (eds) *Migration Policy in Crisis*. London: Transnational Press London, 31–51.

LaBelle B (2014) *Lexicon of the Mouth: Poetics and Politics of Voice and the Oral Imaginary*. London and New York: Bloomsbury.

Latif S, Qadir J, Qayyum A, et al. (2021) Speech technology for healthcare: Opportunities, challenges, and state of the art. *IEEE Reviews in Biomedical Engineering* 14: 342–356.

Law J (2008) On sociology and STS. *The Sociological Review* 56(4): 623–649.

Leix Palumbo D (2024) The weaponization of datafied sound: The case of voice biometrics in German asylum procedures. In: Leurs K and Ponzanesi S (eds) *Doing Digital Migration Studies: Theories and Practices of the Everyday*. Amsterdam: Amsterdam University Press, 303–322.

Leurs K (2023) *Digital Migration*. London: Sage.

Li X and Mills M (2019) Vocal features: From voice identification to speech recognition by machine. *Technology and Culture* 60(2): 129–160.

Miller E (2022) How Voice Technology Will Drive Business Communications In 2022. Forbes, 4 February. Available at: https://www.forbes.com/sites/forbestechcouncil/2022/02/04/how-voice-technology-will-drive-business-communications-in-2022/ (accessed 20 June 2023).

Moore J (1999) The basic principles of behaviorism. In: Thyer B (eds) *The Philosophical Legacy of Behaviorism*. Dordrecht: Springer, 41–68.

Neumark N, Gibson R and Van Leeuwen T2010) *Voice: Vocal aesthetics in digital arts and media*. Cambridge, MA and London: MIT Press.

Ozkul D (2023) Automating Immigration and Asylum: The Uses of New Technologies in Migration and Asylum Governance in Europe. Report, Refugee Studies Centre, University of Oxford, UK, January.

Pfeifer M (2023a) The native ear: Accented testimonial desire and asylum. In: Rangan P, Saxena A, Tharoor Srinivasan R and Sundar P (eds) *Thinking with an Accent. Toward a New Object, Method, and Practice*. California: University of California Press, 192–207.

Pfeifer M (2023b) Your Voice is (Not) Your Passport. In: Sounding Out!. Available at: https://soundstudiesblog.com/2023/06/12/your-voice-is-not-your-passport/ (accessed 13 June 2023).

Reubold U, Harrington J and Kleber F (2010) Vocal aging effects on F0 and the first formant: A longitudinal analysis in adult speakers. *Speech Communication* 52(7–8): 638–651.

Rosenhouse J (2013) Assessing acoustic features in the speech of asylum seekers. *Acoustical Society of America* 19(1): 1–9.

Ruppert E (2011) Population objects: Interpassive subjects. *Sociology* 45(2): 218–233.

Seaver N (2019) Captivating algorithms: Recommender systems as traps. *Journal of Material Culture* 24(4): 421–436.

Simonite T (2018) This Call May Be Monitored for Tone and Emotion. Wired, 19 March. Available at: https://www.wired.com/story/this-call-may-be-monitored-for-tone-and-emotion/ (accessed 10 March 2023).

Simpson AP (2009) Phonetic differences between male and female speech. *Language and Linguistics Compass* 3(2): 621–640.

Spotify (2021) 'Letter to Access Now'. Available at https://www.accessnow.org/wp-content/uploads/2021/04/Spotify-Letter-to-Access-Now-04–15-2021-.pdf (accessed 17 May 2023).

Sterne J (2022) Is machine listening listening? *Communication +1* 9(1): 1–4.

Tangermann J (2017) *Documenting and Establishing Identity in the Migration Process: Challenges and Practices in the German Context; Focused Study by the German National Contact Point for the European Migration Network (EMN)—Working Paper 76*. Nuremberg: Federal Office for Migration and Refugees.

Tiainen M (2013) Revisiting the voice in media and as medium — New materialist propositions. *NECSUS. European Journal of Media Studies* 2(2): 383–406.

Turow J (2021) *The Voice Catchers: How Marketers Listen in to Exploit Your Feelings, Your Privacy, and Your Wallet*. New Haven, CT: Yale University Press.

Van Dijck J (2014) Datafication, dataism and dataveillance: Big data between scientific paradigm and ideology. *Surveillance & Society* 12(2): 197–208.

Vieira de Oliveira PJS (2021) "…the table was set, and we were never dead": On the persistence of colonial listening in Germany. *MAST* 2(2): 89–101.

Viera Magalhães JC and Avella H (2023) Moderating through emotions: Technologies of content mood-eration and the shifting foundations of speech governance. In: AoIR2023: The 24th Annual Conference of the Association of Internet Researchers, USA, Philadelphia, PA, 18–21 October. AoIR Selected Papers of Internet Research.

Viera Magalhães JC and Couldry N (2021) Giving by taking away: Big tech, data colonialism and the reconfiguration of social good. *International Journal of Communication* 15: 343–362.

Wadhwani P and Ganka S (2022) Voice Recognition Market Size | 2022–2028 Share Report. Report, Global Market Insights Inc., USA, March.

Weidman A (2015) Voice. In: Novak D and Sakakeeny M (eds) *Keywords in Sound*. Durham: Duke University Press, 232–246.

Weitzel MD (2018) Audializing migrant bodies: Sound and security at the border. *Security Dialogue* 49(6): 421–437.

Witteborn S (2022) Digitalization, digitization and datafication: The" three D". *Transformation of Forced Migration Management. Communication, Culture and Critique* 15(2): 157–175.

Žižek S (1999) *The Ticklish Subject: The Absent Centre of Political Ontology*. New York: Verso.