Journal of the British Academy
Volume 13, Issue 3
Article a31

RESEARCH ARTICLE

Voice conversion and cloning: psychological and ethical implications of intentionally synthesising familiar voice identities

Carolyn McGettigan*, Steven Bloch, Cennydd Bowles, Tanvi Dinkar, Nadine Lavan, Jonathan Chaim Reus and Victor Rosi

ABSTRACT

Voice identity conversion and cloning technologies use artificial intelligence to generate the auditory likeness of a specific human talker's vocal identity. Given the deeply personal nature of voices, the widening availability of these technologies brings both opportunities and risks for human society. This article outlines key concepts and findings from psychological research on self-voice and other-voice perception that have a bearing on the potential impacts of synthetic voice likenesses on human listeners. Additional insights from speech and language therapy, human–computer interaction, ethics, and the law are incorporated to examine the broader implications of emergent and future voice cloning technologies.

Keywords voice identity, voice synthesis, artificial intelligence, voice perception, psychology, ethics

Published: 11 September 2025

* Corresponding author.
E-mail: c.mcgettigan@ucl.ac.uk

Citation
McGettigan, C., Bloch, S., Bowles, C., Dinkar, T., Lavan, N., Reus, J.C. & Rosi, V. (2025), 'Voice conversion and cloning: psychological and ethical implications of intentionally synthesising familiar voice identities', *Journal of the British Academy*, 13(3): a31
https://doi.org/10.5871/jba/013.a31

Published by The British Academy.

Introduction

Human voices are dynamic expressions of identity. The mere presence of a voice can lend a sense of personhood to machines (Abercrombie *et al*. 2023), while hearing the familiar voices of other humans can soothe and reassure (Seltzer *et al*. 2010, 2012). Voice conversion and cloning, which use artificial intelligence to synthesise a human talker's vocal identity, can now be done at no or low cost, and on the basis of a small amount of audio input data. Thus, where self-voice synthesis was once mainly used in quite narrow contexts, a person's partially or fully synthesised voice identity can today be readily applied in a wide range of settings, from virtual assistants to audiobook readers and chatbots. Given the deeply personal nature of voices (Sidtis & Kreiman 2012), we need a better understanding of the implications of personalised voice synthesis technologies for human listeners and their relationships with voices in everyday life.

Voice conversion and cloning

Voice conversion and cloning are terms used to describe the application of AI to create novel audio that is recognisable as the speech of a specific person—in other words, an 'audio deepfake' of a person's vocal identity. In some respects,

voice identity replicas have been an everyday reality since the advent of voice recording and editing technologies first allowed individual voices to be reproduced by machines. However, to flexibly generate novel utterances in a speaker's voice requires much greater sophistication than, for example, concatenating individual recordings of words and phrases. Primarily with the aim of providing individualised synthetic voices to patients whose physical voice was compromised or lost, developments in voice technology during the early 21st century found ways to computationally learn aspects of a speaker's voice quality (the acoustic properties that made them uniquely recognisable) and use these to create a text-to-speech (TTS) synthesiser bearing a likeness to that speaker. Such approaches were typically implemented in augmentative and alternative communication (AAC) devices for use in everyday communication by patients who have lost the use of their physical voice due to illness (for example, motor neurone disease (MND); also known as ALS) or injury (for retrospective accounts, see Mills *et al.* 2014, Veaux *et al.* 2013, Yamagishi *et al.* 2012). These early models, while potentially transformative for users, often required hours of donated audio recordings from the original speaker (a process known as 'voice banking') for the computer model to adequately learn and reproduce the identity-specific characteristics of the voice. The resulting synthetic voices typically also bore some limitations, both in their likeness to the original talker, and in their perceived naturalness (Mckelvey *et al.* 2012).

With advances in deep learning and generative artificial intelligence (AI), the 2010s and 2020s have seen dramatic and accelerating changes in voice synthesis capabilities, with implications for AAC (Judge & Hayton 2022) and other use cases. In one approach, known as voice conversion, models can be employed to learn the acoustics of one speaker and transfer these to recordings from another speaker, thus 'grafting' a new identity onto pre-existing speech or vocal audio. This form of 'speech-to-speech' synthesis can be applied to human speech in (almost) real time (for example, via online conferencing or streaming platforms). Other methods, more commonly known as voice cloning, instead typically use 'text-to-speech' methods to generate audio bearing the learned acoustic features of a target voice identity. Here, the output speech does not need to be present in existing recordings, or be produced by a donor speaker, but rather is determined by the user's text input (with output latencies dependent on the input length; for a general introduction to these methods, see Hutiri *et al.* 2024).

There are some immediate implications of the broad differences between conversion and cloning. On the one hand, the fact that conversion is applied to pre-existing human speech audio allows for aspects of naturalness and context-appropriateness (for example, emotion or speech rate) to be retained in the synthesised speech output, while these may not be guaranteed in fully generative cloning approaches. On the other hand, the outputs of conversion are limited to, and constrained by, the existing audio to which conversion is applied, and therefore on the labour of the human speaker(s) producing the donor speech materials—in contrast, voice cloning offers theoretically limitless amounts of possible speech output generated from text prompts. In this paper, we will focus

on both methods' common approach of replicating voice identity characteristics to consider the psychological impacts of this approach within different use cases.

What is very powerful about some of the latest voice identity synthesis approaches is that they no longer require bespoke model training or fine-tuning to generate a specific voice identity—voice conversion or cloning can be achieved by mapping a small amount of voice input audio (sometimes as little as 3 seconds; Microsoft Research (n.d.)) to an existing learned speaker space (Arik *et al.* 2018; Jia *et al.* 2018). Utterances generated from such a 'zero-shot learning' approach can bear startling perceptual similarity to the original talker, with even fully generative models producing fluent and naturalistic speech intonation in the talker's own and other languages. In the case of purely generative speech synthesis, synthesised pauses and breaths further aid in creating the illusion of hearing an authentic human speech recording (Abercrombie *et al.* 2023). Some of the most cutting-edge voice conversion and cloning models are commercially available via user-friendly web-based interfaces and application programming interfaces (APIs) (for example, from ElevenLabs, Microsoft (VALL-E), Speechify, Meta (Audiobox)), while other open-source voice cloning methods are accessible to users with appropriate expertise and computing hardware (for example, YourTTS, Tortoise TTS, OpenVoice).

Whether using voice conversion or cloning techniques, the possibility of synthesising a person's voice identity has thus rapidly opened to a global user base of professionals and general publics (Federal Trade Commission 2019; Hutiri *et al.* 2024). For example:

- A social media influencer might synthesise their own voice for ease of generating new online content (Zhang *et al.* 2021).
- A university lecturer might decide to generate personalised voiceovers for their teaching materials in multiple languages for a diverse student audience (Dao *et al.* 2021; Pérez *et al.* 2021).
- A parent may wish to apply their own voice identity to audiobooks to be played as bedtime stories for their children (Epp *et al.* 2017).
- A voiceover artist may be interested in using a synthesised version of their voice to maximise their commercial reach (given appropriate legal protections (Cieslak 2024)).

When thinking about these intentional uses of voice conversion or cloning, some key questions begin to arise for the voice owner.

- First, *who* gets to hear and use my synthesised voice? At a voice owner's discretion, their voice identity could be made available to audiences ranging from intimate (for example, used only by the self, and/or selected close relatives) to more general but context-defined (for example, a lecturer's voice used only by their university), to completely widespread (for example, a voice donated to an open-access database for use by anyone).
- Second, *how* will my voice be used? A person may be happy for their voice to be used as an audiobook reader of children's books, but not to read books describing graphic violence; they may be happy to be the voice of Amazon

Echo's weather reports, but not to voice a personal chatbot within the same device.
- Third, *for how long* will my voice be used? A university lecturer may prefer that their employer ceases to create new content in their voice once they have resigned or retired; a person making their will might prefer to place limits on how their inheritors use their voice data after they are deceased.

In purely practical terms, the state-of-the-art is still not without its limitations. Most widely available voice cloning models generate speech in the style of reading text aloud and, despite good overall naturalness, there are still shortcomings in the appropriate synthesis of context-relevant emotional prosody and speaking styles (Kolekar *et al*. 2024). Similarly, the success of cloning and conversion in terms of perceived accuracy is dependent on the composition of the training dataset underpinning the model's functionality, which may lack sufficient representation of minority languages, non-standard accents, identities, and speaking styles (Barnett 2023)—larger, speaker-specific input data is required to fine-tune models to obtain more personalised results (thus placing an added burden on minoritised talkers to provide suitable data). Finally, as mentioned above, the relative ease of implementing and using voice conversion or cloning technology is currently correlated with financial cost—the more technically accessible web-based models are typically made available on a subscription basis, with limits on the amount of material that can be generated and downloaded per unit time. When the subscription is paused, so is access to the clone, and thus it is not always possible to 'own' one's cloned voice for use in a truly permanent and flexible way (for example, integration into word processing software, social media apps, or AAC hardware). However, there have been very recent initiatives to widen accessibility even to commercial tools—for example, ElevenLabs' Impact Program which offers free licences to individuals with MND/ALS and 'social good' partners working in sectors such as education (ElevenLabs 2024a).

Given the rapid changes in the sophistication of voice identity synthesis outlined above, we conclude that it is reasonable to expect that the technological capacity for high-quality, low-latency, and inclusive voice identity conversion and cloning will continue to advance rapidly, with accessibility following suit. Thus, it is already time to consider a world in which the synthesised vocal identities of ourselves and other people are available for use in our everyday lives. Here, therefore, we will use findings from psychology, neuroscience, and allied literatures as a scaffold to address questions (such as those around *who*, *how*, and *for how long*, as outlined above) and form predictions about the possible impacts of voice conversion and cloning on human perception and experience. Where available, we will integrate evidence from existing empirical research on personalised voice technologies, including our own preliminary findings, although we note that this literature is still in its infancy and thus our overall perspective must necessarily be more speculative than evaluative or conclusive.

Here, we focus on the speaking voice as the predominate modality of human vocal expression and the main mechanism for human sociality and social

organisation. While acknowledging that synthesis of the sung voice as it relates to personal identity is also of great importance, particularly for artists (Josan 2024), there may be important distinctions in the treatment of our questions in the context of the singing voice that go beyond the scope of the current paper—for example, the personal singing voice is strongly linked to creative expression, and singers may have higher-stakes involvement in economies that commodify voice and vocal identity.

Later parts of the discussion will broaden the focus to consider multidisciplinary insights on the moral, ethical, and legal issues associated with voice conversion and cloning, culminating in speculative consideration of the cloned voice as a part of digital afterlives. Throughout, our focus will be on intentional and legal uses of these technologies, rather than on issues around deepfaking and identity misrepresentation, which are actively discussed elsewhere (for example, Barnett 2023, Hutiri *et al.* 2024). Our discussion is nonetheless relevant to existing and emerging legislation, in terms of how this might be shaped to minimise certain ethical risks to both the human owners/donors of voice data, and to the other stakeholders implicated in applications of voice identity synthesis technologies.

The human voice

The human voice is a dynamic audio signal that can be used flexibly to express thoughts, emotions, and mental states via both verbal (that is, speech) and non-verbal (for example, laughter or sighs) vocal behaviours (Belin *et al.* 2004; Lavan *et al.* 2019; Scott & McGettigan 2016). The physical sound of the voice arises from the vibration of air molecules passing through the human vocal tract. The vibration of the vocal folds within the larynx (the 'source') generates a largely periodic signal (that is, one with pitch) that is modulated by both the morphology of the static structures of the vocal tract (for example, hard palate) as well as the dynamic positioning of the articulators including the lips, jaw, and soft palate (the 'filter' (Fant 1971)). As a vocal behaviour, human speech requires exquisite precision and coordination of the source and filter to execute a rapid stream of vowels and consonants that are recognisable and comprehensible to human listeners; within this, changes to the velocity of airflow, the rate and quality of vocal fold vibrations, and the rate and extent of articulations, can add variation in the perceived energy, linguistic focus, and emotional content of the spoken signal (broadly termed 'prosody'). However, while acoustic correlates of voice quality—or timbre—within speech and other vocal behaviours can be measured and quantified, it has remained challenging to adequately relate these physical properties of vocal sounds to our complex perceptual experiences of voices [that is, to cross the 'timbral abyss' (Kreiman 2024)].

The overall shape of the vocal apparatus, as well as the ways in which its moveable parts are engaged, varies widely across and within speakers. Between-speaker variations are strongly influenced by sexual maturity: children have shorter vocal tracts and vocal folds than adults, yielding voices that sound smaller and higher-pitched than adult voices. In adults, the effects of testosterone during male puberty mean that post-pubertal males have on average

longer vocal tracts as well as longer and thicker vocal folds than adult females, with the perceptual consequence that males tend to sound larger and lower-pitched than females (Cartei & Reby 2013; Fitch & Giedd 1999). The language(s) and accent(s) spoken by talkers will add a host of additional differences in the physical dynamics of speech behaviours and their acoustic and perceptual correlates across groups of people. At the level of the individual, we can also see variation in the shape and the dynamics of vocal behaviours that underpin each speaker's unique vocal character and repertoire (Lim *et al*. 2021). Moreover, the idiosyncrasies of a person's speech are not fixed—speakers can volitionally modulate their speech in a variety of ways, including learning to speak additional languages, disguising their vocal identity (including, but not limited to, expert voice artistry), and dynamically adjusting speech depending on the acoustic, communicative, and social contexts (for example, Aziz-Zadeh *et al*. 2010, Cartei *et al*. 2012, 2019, Guldner *et al*. 2020, 2024, Hazan & Baker 2011, Hughes *et al*. 2014, Pisanski & Reby 2021, Sorokowski *et al*. 2019).

What are the implications of between-speaker and within-speaker variations for personalised voice synthesis? In terms of replicating voice qualities that can map onto a given speaker, the difficulty of the task increases with the degree of individuation required—thus, while it is has for some time been very straightforward to convert or synthesise a voice to sound broadly like an adult male human, it becomes more complex to make that voice sound specifically like our first author's oldest male friend from Derry in Northern Ireland when he's in a bad mood. In the context of the state of the art, this has a lot to do with how models are trained and fine-tuned: in order to be able to replicate the diversity of speakers in the human population, the datasets upon which models are trained will need to include a fair amount of that diversity, and a biased model will produced biased outcomes (for example, the friend from Northern Ireland becomes cloned as a male-sounding speaker with a General American accent). The related issue, noted above, is that a perfect reproduction of a given speaker under all possible situations requires additional data about how that person might sound under different acoustic/communicative/social pressures, including both their linguistic habits (vocabulary, word choice, and syntactic and pragmatic preferences) and how these become manifest in their speech acoustics. For that, personalised voice synthesis models need more person-specific data, which requires substantially greater input from both speakers and the model developers.

Perceiving voices, perceiving people: implications for hearing voice clones

Within psychology and neuroscience, as well as allied literatures including phonetics, there is a wealth of existing research findings on voice identity processing, from how we discriminate, recognise, and identify individuals and how this is affected by familiarity, to how we evaluate a person's broader physical, psychological, and social characteristics from the sound of their voice.

More applied research, from these and other fields—most notably speech and language therapy and human–computer interaction—considers more directly how synthesised voices might affect the day-to-day experience and decision-making of the technology's human end users.

While the most sophisticated state-of-the-art voice synthesis models can now generate impressive likenesses of specific speakers, the outputs of some models (for example, zero-shot cloning) can still produce poorer likenesses of some speakers than of others (see ElevenLabs 2024b). Thus, we can consider here the extent to which the relative accuracy of personalised voice identity synthesis might meaningfully interact with other aspects of human voice processing, in positive and negative ways.

Perceiving people from recordings of human voices

The most basic question to ask when considering the utility and impact of personalised voice identity synthesis is: Can human listeners perceive voice identities accurately, sufficient to potentially benefit from personalised synthesis? Broadly speaking, yes: there is good evidence that adults generally have the ability to detect unique voice identities from recordings of speech and other vocalisations (Belin *et al.* 2011; Kreiman & Sidtis 2011; Lavan *et al.* 2019; Mathias & von Kriegstein 2014; Scott & McGettigan 2016). These abilities have been tested via a variety of experimental paradigms, including voice identity discrimination (for example, making same/different identity judgements on pairs of vocal stimuli), voice recognition (for example, making familiarity judgements), and voice identification (for example, naming or classifying voices).

Given that many intentional users of voice banking and cloning technology are doing so in order to hear—and be heard by—familiar people, later sections of this paper will be dedicated to considering the familiar voices of the self and others. However, work on the perception of *unfamiliar* voices is also informative about how voices are perceived and evaluated by human listeners. This is because, although recognition and identification are not possible for completely unfamiliar voices, listeners can still perceive a substantial amount of information from the sound of an unknown human voice. It has been shown that listeners can, from even a very brief (<1 s) sample, form impressions of speaker sex, height, weight, personality traits (for example, trustworthiness or dominance), and social characteristics (for example, professionalism or educatedness; for a review, see Lavan & McGettigan 2023)). These percepts often show high agreement across listeners (Lavan 2023; Lavan & Sutherland 2024; McAleer & Belin 2018; McAleer *et al.* 2014; Mileva & Lavan 2023), and can influence decision-making in applied contexts such as elections (Klofstad *et al.* 2012; Tigue *et al.* 2012). Indeed, the mere presence of a voice can increase the perceived humanity (and conversely, decrease the *de*humanisation) of other people, from job candidates to political opponents (Schroeder & Epley 2015, 2016; Schroeder *et al.* 2017). These studies are instructive for our current purpose because they illustrate that voices can be powerful triggers of quite rich mental representations of other people. Therefore, we continue our discussion by

considering how personalised synthetic voices might affect human listeners, and how this might be modulated by listeners' familiarity with the original speakers.

Perceiving people from synthetic voices and artificial agents

The evidence outlined above shows that, when a human voice is heard, a *person* is heard, and the nature of that percept has the potential to shape our inter-personal interactions. Extending this idea, then, we can propose that, when a synthetic voice sounds sufficiently human, it has the potential to engage the same perceptual and evaluative processes. Previous findings already report that human listeners' evaluations of (non-personalised) synthetic voices vary along similar social trait dimensions to those reported for human voice impressions (Shiramizu *et al*. 2022). However, two recent developments in the literature on personalised voice synthesis—specifically, voice cloning—are highly relevant to our current discussion. First, it has been shown by ourselves and other authors that human listeners struggle to accurately categorise, or discriminate between, authentic speech recordings and voice clones when the (cloned) talkers are not familiar to them (Barrington *et al.* 2025; Lavan *et al*. 2025; Rosi *et al*. 2025), also rating human recordings and cloned audio as similarly 'real' (Lavan *et al*. 2025). Second, two of our own studies have reported that voice clones of unfamiliar human identities convey personality traits differently compared with authentic recordings of those human talkers—for example, by sounding more dominant, competent, or trustworthy (Lavan *et al.* 2025; Rosi *et al.* 2025). That is, voice clones can not only sound like humans, but potentially like *superior* humans, compared with the sound of genuine human voices.

Researchers from psychology and human–computer interaction have explored the broader idea of person perception from non-human agents by considering the impacts of adding human-like voices to machines (including robots) and artificial agents. Seaborn *et al.* (2021) identify what they term a 'vocaloid shift' in research on human–agent interaction—a recent increased interest in voice, influenced by the proliferation of voiced agents such as Amazon's Alexa and Apple's Siri. Their systematic review of voice in human–agent interaction identifies support for the Computer Are Social Actors theory (Nass *et al.* 1994), stating that 'as a general rule, people seem to unthinkingly treat voice-based agents as people' (p. 29). Abercrombie, Cercas-Curry, Dinkar, *et al.* (2023) further examine the effects of anthropomorphism in dialogue systems, which are pertinent to the current discussion as one obvious way in which a cloned voice might be integrated into a machine: overall, the authors' findings suggest that the addition of humanising characteristics to non-human entities—including names, lists of traits and hobbies, and voices—affects the way in which these entities are evaluated and understood. Specifically, the agent becomes anthropomorphised, such that the human interacting with it begins to afford it the personhood and cognitive capacity of another human. It is argued that the humanising power of voices and speech creates competing pressures for the use of voices in machines—on the one hand, human users may have more naturalistic and rewarding interactions with a robot or agent that sounds more humanlike, while on the other hand there is increased risk of negative human

experience if the performance of the agent undershoots the assumed capability of its anthropomorphised persona (Abercrombie *et al.* 2023).

The personification of machines not only influences perception but also has consequences for behaviour—the recent trend for obviously gendered and often subservient presentations of dialogue systems (Faber 2020) has evoked verbal abuse from human users, including gendered slurs and sexual harassment (see Cercas Curry & Rieser 2018, Cercas Curry *et al.* 2021), raising the possibility that the anthropomorphising of distinctly non-human entities is creating a dangerous practice ground for abusive human–human interactions. In other situations, human-like artificial agents may engender levels of trust and self-disclosure that are inappropriate to a human–computer interaction and could lead to emotional dependencies on entities that have no capacity for human emotion or empathy (Andries & Robertson 2023; Ki *et al.* 2020; Pitardi & Marriott 2021). With these considerations in mind, Abercrombie *et al.* make several recommendations for the design and research of dialogue systems incorporating speech that will prove useful to our current discussion. For system design, they encourage designers to recognise humans' tendency to personify machines and in turn consider the appropriateness of adding anthropomorphising voices that may make the boundary between human and machine unclear for human users. For researchers of AI and human–computer interaction, the authors recommend actively investigating both the impacts of design features that enhance anthropomorphism and those that seek to limit it. In relation to the latter, they cite Wilson & Moore (2017), who found that the creation of voices for cartoon characters and robots can target congruence between an entity's voice and role (for example, within a movie) by using acoustic profiles that are distinctly *not* humanlike—that is, with careful voice design, some socially appropriate or desirable qualities of a spoken interaction with an artificial agent can be preserved while simultaneously reinforcing that the agent is definitely not human.

Perceiving the synthesised voices of familiar people

When considering the use of personalised voice synthesis, which seeks to maximise not only the humanlike qualities of a voice but its likeness to a specific known human identity, the considerations and recommendations raised by Abercrombie *et al.* (2023) become yet more relevant. Familiar voices, and in particular those of personally-familiar people like friends and relatives, are recognised with high accuracy compared with lab-learned identities (Kanber *et al.* 2022). This has implications for the acceptability of synthesised voices by the people who hear them, because it is likely that a highly personally-familiar listener would have a well-defined appreciation of whether a clone offered a good and recognisable likeness to the sound of the original speaker's voice. Indeed, our recent study investigating the perception of cloned voices showed that—at least for the zero-shot model we tested—listeners were highly accurate at detecting clones (versus human recordings) of voices they knew well (that is, their own voice and the voice of a friend) and showed the opposite preferences to unfamiliar listeners when rating the personalities of cloned versus recorded

voice samples (Rosi *et al.* 2025). At the same time, personality ratings of cloned voices became arguably more favourable (that is, higher in attractiveness, competence, and trustworthiness) with increasing perceived similarity between the clone and the original voice, suggesting that increased accuracy of cloning may indeed be linked to positive perceptual evaluations.

There are other quantifiable benefits to hearing a familiar voice: In challenging listening situations with overlapping talkers, studies have repeatedly shown that listeners are better able to comprehend speech from familiar-voice targets (Bradlow & Pisoni 1999; Domingo *et al.* 2020; Holmes & Johnsrude 2020; Johnsrude *et al.* 2013; Nygaard & Pisoni 1998), even if that voice has been rendered less recognisable via acoustic manipulations (Holmes *et al.* 2018). There are potential opportunities in this area for personalised voice synthesis. For example, if speech-to-speech conversion or text-to-speech cloning of a familiar-voice identity can be achieved to create personally-familiar audiobook narration, or health-related reminders from a vocal assistant, this could provide a benefit for users' perceptual encoding of the voice-mediated information. In such applications, zero-shot TTS functionality in particular would additionally reduce the burden on the original speaker to create extensive recordings. However, such contextualised benefits need to be demonstrated empirically.

Taking a different angle, other researchers have started to examine whether there is any neurobiological basis to the colloquial expression 'it's so good to hear your voice' (McGettigan 2015). Sidtis & Kreiman (2012) argue for the central significance of the familiar voice recognition for human experience, citing its evolutionary importance via the evidence for familiar voice patterns in other species, the primacy of the mother's voice in human infancy, as well as the 'whole-brain' nature of voice recognition given that a familiar voice pattern unlocks not only identification but a host of associated biographical knowledge, memories, and emotional associations with the voice owner. Empirical work has supported the notion that familiar voices can have beneficial impacts on the listener's physiological state—in studies with mother–daughter dyads, the sound of the mother's voice increased release of oxytocin (which is associated with interpersonal bonding) in daughters (Seltzer *et al.* 2010) and reduced daughters' stress-related cortisol levels (Seltzer *et al.* 2012). Similarly, neurobiological investigations of children's responses to hearing their mother's voice have shown evidence for greater engagement of brain regions associated with rewarding experience when hearing the mother's voice compared with other, unrelated, female voices (Abrams *et al.* 2016, 2022). Most recently, in a study focusing on celebrity voices, we have shown that listeners will work harder to hear a personally-valued voice than other voices and sounds, and that this is reflected in the elevated engagement of the brain's reward and motivation systems during the anticipation and experience of voices (Kanber *et al.* 2025). It may well be that there is some functional interplay between the rewarding qualities of personally-familiar (and valued) voices and the perceptual benefits reported above, which merits investigation. In reading research, there is already evidence that the level of enjoyment of a book synopsis, as an index of reward, is predictive of the accuracy of comprehension of the text (Bains *et al.* 2023)—the

use of a rewarding familiar voice identity in audiobook listening may similarly be able to enhance the linguistic encoding of speech.

Existing observations from clinical settings speak to the broader issue of the desirability and acceptability of synthesised voices to personally-familiar listeners. Linse *et al.* (2018) note that patients with MND and caregivers alike find it problematic when the quality of a synthetic voice is poor, but point to a shortage of research about whether more personalised approaches are indeed valuable. Nonetheless, anecdotal reports endorse the potential importance of both the perception and retention of personally familiar voices by loved ones: Benson (2021) describes a case study of a patient with MND whose wife was fearful of forgetting what her husband sounded like, and who valued his banked voice because it gave her and their children the opportunity to 'mitigate against some of the losses' created by the disease (Benson 2021, p. 114). In an analysis of structured interviews about voice banking, Cave & Bloch (2021) found that significant communication partners of people living with MND felt that voice banking would not only preserve the voice owner's sense of identity, but might contribute positively to their personal and working relationships as well as maintaining social networks. Given the current state of the art and its variable outcomes, the extent to which the similarity of a cloned voice to its original speaker might matter for 'unlocking' the perceptual and rewarding benefits of hearing that identity is currently unknown, and will need to be established (though our findings with non-clinical experimental participants suggest a positive relationship (Rosi *et al.* 2025)). It should also be noted that the sound of a voice is only one element of the communicative context for a potential user living with an illness or disability: Authors writing about disability urge caution when hailing new technologies as assistive solutions without including disabled people themselves in the conversation (for example, Alper (2017) on voice AAC and Jackson *et al.* (2022) on the 'Disability Dongle').

Evidence outlined earlier shows that the mere presence of voice may lead to humans engaging with non-human agents as if they were real people. Other work considers the integration of real human voices into technologies such as virtual assistants. In one study, listeners exposed to personally-familiar (recorded) voices within an Alexa-type application experienced stronger senses of co-presence and evaluated these voices more favourably compared with a synthesised (unfamiliar) voice, while at the same time finding the familiar voices more eerie than the synthetic voice (Chan *et al.* 2021). There is to date very little evidence of how these effects might manifest in the context of personalised voice synthesis: However, in a very small follow-up study embedding *cloned* familiar voices within a virtual assistant, Chan and colleagues reported overall positive responses, but also that users were very sensitive to mismatches in accent and speaking style between the original voice and its synthetic version. We speculate that familiar voices may attenuate or accentuate the perceived personhood of a machine, depending on the sophistication of the application: while personalised calendar reminders or familiar voice assistance may generate positive feeling and enhance productivity, contexts such as conversational AI chatbots may be rejected for their obvious lack of authenticity.

The (synthesised) self-voice: a special case of familiar voice perception

The self-voice, being the product of one's own actions, is typically the voice with which an individual has the most auditory experience. It is highly familiar, highly embodied, and is ultimately the auditory signature of a person's expressed identity. Participants are less susceptible to perceptual illusions in their own voice (Aruffo & Shore 2011), show stronger motor-induced suppression in the auditory cortex when they expect to hear their own voice (Johnson *et al.* 2021; Knolle *et al.* 2019) and exhibit self-voice sensitivity in the brain's evoked responses as measured with EEG (Conde *et al.* 2015; Graux *et al.* 2013, 2015). Within any discussion of personalised voice synthesis, it is therefore crucial to consider the impacts on the owner of the voice that is synthesised.

Self-voice perception is complicated by the fact that there is an important distinction between the self-voice we hear while speaking and any externally reproduced version of it (whether an echo, a digital recording heard through a loudspeaker, or a clone): namely, the experience we have of our voice as we speak is multi-modal because we not only receive auditory input from the sound of the voice in the air but also via the conduction of the sound through the bones and tissues of the body (see Orepic *et al.* 2023). Nonetheless, high levels of perceptual sensitivity to the self-voice have been shown both when it is presented via air-conduction (Candini *et al.* 2014) or via bone-conduction (Orepic *et al.* 2023). This has implications for the experience of the *synthesised* self-voice. For example, listeners might experience lower perceptual acceptability for clones of the self-voice than for clones of other familiar voices, because we usually experience our own voice identity via an additional modality that is not typically available when hearing the voices of others. Initial empirical findings contest this proposal: we recently found no evidence of differences in the detectability of voice cloning, or in ratings of similarity between a clone and the 'real' voice, when comparing listeners' perception of self-voice clones with clones of a friend's voice (Rosi *et al.* 2025).

Liking of the self-voice has also received special attention in the literature. While some accounts suggest that the recorded speaking voice is relatively disliked by listeners because of the sensory mismatch between embodied and audio-only perception (for example, Shuster & Durrant 2003), other work claims a self-voice enhancement bias (Hughes & Harrison 2013; Peng *et al.* 2019, 2020, 2021). Thus, for synthesised voices, what sounds 'like me' may or may not be preferred, relative to the sounds of synthesised others' voices. In our recent work, we found some suggestion that self-voice similarity is indeed preferred, where personality ratings of attractiveness, competence, trustworthiness (and to a weaker extent, dominance) were higher for cloned voice samples that sounded more similar to the self (Rosi *et al.* 2025).

Given the human significance of the voice (Nathanson 2017) and the expression of identity as a basic human right for users of AAC (Wofford *et al.* 2022), access to personalised synthetic voices in clinical contexts could elevate patient autonomy, have clear positive benefit to patient well-being, and afford greater equality of opportunity (Nathanson 2017). These conditions of respect

for autonomy, beneficence, and justice are three of what are often considered the foundational principles of bioethics (Beauchamp & Childress 2001), meaning their successful attainment carries significant moral value that we should not discount. However, in both clinical and non-clinical use cases for synthetic voices, there are possibilities for personalisation without fully replicating the original voice identity—for example, users may be satisfied with a synthesised voice that has the same gender and regional accent as their original voice (Sutton *et al.* 2019). In other cases, individuals using a synthetic voice for various aspects of communication or pastimes [for example, creating social media content or online gaming (Zhang *et al.* 2021)] might prefer to adopt a different vocal persona to their original voice. Recent work in experimental psychology has demonstrated the possibility for multiple recorded and embodied self-voices—including self-voice clones—to co-exist in the self-concept as multiple versions of 'me' and receive prioritisation in perception (Payne *et al.* 2021, 2024; Rosi *et al.* 2024). These studies have to date been conducted with healthy participants, for whom the original self-voice is still available and there is no immediate prospect of it being lost. It will be beneficial to investigate these prioritisation effects with people for whom using a synthetic voice is personally more significant, and not only in clinical contexts—for example, teachers, actors, and media professionals may face choices about the use of personalised voice synthesis that have direct implications for their professional practice and their self-identity within that.

'Giving' voices, taking data: inclusivity, ethics, and legality of voice conversion and cloning

The previous sections outlined key findings from voice research and their potential implications for the use of personalised voice synthesis. Here, we consider some broader issues concerning the nature of personalised AI that are of relevance to voice conversion and cloning.

Data ownership: who owns my voice?

When it comes to voices, and voice data, the question of ownership can be considered from a range of perspectives. Here we touch upon some of the philosophical, ethical, and legal issues, with a specific focus on implications for voice cloning and personalised voice synthesis. Please see also Watt, Harrison, and Cabot-King (2020) for a more detailed examination of the legal considerations around voice ownership in general.

Assuming adequate informed consent for personalised voice synthesis can be provided (where ethicists may be called upon to define and articulate the conditions for this), in an intentional voice banking or cloning context the speaker will have donated recordings volitionally, and thus it might be argued that they have 'given' their voice for synthesis. However, despite the strong attachment of self-identity to voice, there is a philosophical consideration about whether a person's synthesised voice really is 'their' voice in the same way as

their embodied, physical voice. Even audio recordings of voices are, strictly speaking, replicas of a physical sound, and a resynthesised clone is even further removed than that (regardless of its likeness to the original voice).

On this issue, Napolitano (2022) raises an important point of conflict between what personalised voice synthesis technologies actually do and how they are sold to clinical and non-clinical populations of end users: While philosophers might argue that a disembodied recording/clone of a voice coming out of a loudspeaker or headphones is not the speaker's actual voice, voice tech companies simultaneously use language that invokes concepts of ownership (for example, 'giving back' a voice that has been 'lost'). If a voice can be 'given back' by a company to a consumer who has 'lost' it, this could imply that the company—at least temporarily—owns something of the consumer's to which the consumer themself has no access. This potentially raises concerns about personal data ownership in the context of data protection law. Napolitano moreover suggests that a framing of cloned voices as 'owned' voices runs the risk of promoting a culture of ableism based on a medical model of disability, where voices 'corrupted by disability' (Napolitano 2022: 202) must be restored to their pristine original versions by corrective voice technologies. It will be useful to monitor and evaluate the language used around voices and ownership in order to preserve a distinction between the voice as an embodied behaviour versus audio recordings/syntheses that can exist independent of bodily agency (for example, within a voice assistant).

Nonetheless, when it comes to considerations of legal protections, voice recordings can be considered personal data under the 2018 General Data Protection Regulation. Thus, when thinking about the voice recordings that would be used in voice cloning or conversion, the agent of the embodied voice is entitled to protections under such frameworks. Globally, emergent and proposed legislations about AI use and specifically regarding deepfakes (see Rouse Editor 2024) also attempt to prevent the unlawful use of cloning or deepfaking to cause harm (for example, by defrauding, misleading, or defaming through the unauthorised impersonation of individuals). We note here that, given the likelihood of rapid change in this area, we can only offer a snapshot from the time of writing. However, for example, the EU's AI Act will require disclosure when AI is being used to create a resemblance of another person. Other proposed regulation of deepfakes has faced resistance—for example, from those who wish to enshrine the rights to free speech [for example, in the form of political parody (Ray 2024)]. Otherwise, when it comes to the voices of living individuals, commercial service providers' Terms of Service may already regulate against the use of cloning and other personalised voice synthesis technologies without the express permission of the user.

We can therefore speculate that living individuals may eventually have robust protection against the unauthorised use of their voice recordings to produce personlised synthetic voices. However, it remains to be seen how nuanced these protections may be. Norms and expectations around privacy are highly influenced by context: Understanding privacy not as a singular, immutable right but as a process by which individuals and groups negotiate their boundaries

(Altman 1975), we may see that an individual might consent to use of their synthetic voice identity in an educational or healthcare context, they may not wish their synthesised voice to be used in other settings.

The question of voice data ownership becomes somewhat more complex in death. Historically, laws have been enacted in some US states that allow for the prevention of celebrity likenesses being used posthumously for commercial purposes, but in ways that have not captured the rights of deceased private individuals (Roberts 2023; see also Bassett 2019). In 2023, Roberts explained that simple text messages and emails were potentially the intellectual property of the author, but could be subject to Fair Use, for example to create a chatbot whose messages are educational in nature. Therefore, without suitable modifications to the legislation, relying on copyright, intellectual property, or trademark law may not offer robust protection for deceased private individuals. To protect against unauthorised 'resurrection', Roberts instead suggested an expansion of existing legislation allowing pre-deceased individuals to grant nominated 'digital fiduciaries' access to their digital assets after their death [see recent implementations of this idea by major social media sites like Facebook (Birnhack & Morse 2022)]. For example, a person making their will could stipulate how their digital assets should be used after their death—to potentially include the person's wishes as to whether they allow their assets to be used in voice synthesis.

Data availability and inclusivity: AI bias and (under)representation

The current reality of many published AI-assisted voice synthesis models is that they are likely trained on a preponderance of English (Barnett 2023), despite some concerted efforts to address this [for example, Mozilla's Common Voice (Mozilla n.d.)]. Thus, where a user who speaks (a standard variant of) UK or US English might obtain a voice clone that closely matches both their voice quality and speaking style, others may experience a salient accent mismatch. Such bias in AI models has been documented elsewhere, in terms of the outputs of both analytical and generative models with relevance to gender, race, and disability (for example, AI Now Institute 2019, Angwin *et al.* 2016, Buolamwini & Raji 2019, Nicoletti & Bass 2024, Zhou *et al.* 2024). For a listener determining the acceptability of a personalised synthetic voice, a mismatch in perceived accent may be at least as influential as an inaccurate reproduction of the individual's voice characteristics. Indeed, patients with foreign accent syndrome—a dysarthria that affects the perceived nativeness of their accent—describe acute loss or change in their sense of self and identity due to the othering nature of their speech disorder [Miller *et al.* (2011); see also Shadden (2005) for a broader discussion of aphasia as 'identity theft']. Sociolinguists Sutton *et al.* (2019) highlight the importance of personal histories and socio-cultural contexts in shaping the way a person speaks and sounds, calling for both individuation and diversification in the types of voices designed for use in voice–user interfaces. They argue that having a limited number of accents available where synthetic voices are used not only negatively impacts equity and accessibility of these technologies for users from under-represented accent groups, but also

contributes to the homogenisation and standardisation of voices in our everyday lives.

However, there may be a trade-off to consider: given prejudices against minoritised accents (and their users) are still at large in our society (Levon *et al.* 2021), the wide availability of synthetic voices also increases the risk of harm—for example, through appropriation and negative stereotyping (Zhang *et al.* 2021). Future developments in synthetic voice design will need to consider how to balance the desirability of increased access to, and personalisation of, synthetic voices against the vulnerability of natural and naturalistic voices to social biases and stereotypes. This may indeed require some form of educational intervention—just as Amazon Kids' polite mode and Google's Pretty Please functions have encouraged children to say 'please' and 'thank you' to voice assistants (Ribino 2023), adults could be trained to address their own unconscious biases in the design, creation, and use of synthetic voices to represent themselves, other people, and machines.

It seems plausible that harms around minoritised voice identities could be mitigated through proper regulation and careful design. If this were achieved, the autonomy, beneficence, and justice benefits mentioned above in relation to clinical cases may plausibly outweigh the moral concerns around voice conversion and cloning.

Cloning voices, (re)creating 'people': the prospect of digital avatars and afterlives

The voice is only one modality through which an individual can be represented: when combined with other data—for example, the appearance, language, knowledge, and personality of an existing individual—there exists the possibility to create a multimodal, generative representation of a whole persona. Forms of digital avatars have been used to great effect in entertainment, educational, and awareness-raising contexts, from an avatar of the deceased UK celebrity Bob Monkhouse in a cancer charity ad campaign, to a holographic representation of Tupac Shakur in 2012 (Bassett 2019), and the recent Abba Voyage experience in which 'Abbatars' replicate younger versions of the band members (Matthews & Nairn 2023). Widening access to technologies that can both replicate and generate novel output will open the possibility for both public figures and private individuals to be digitally replicated. Thus, an educator might agree to teach into their retirement via a digital avatar, while an in-demand celebrity might provide simultaneous 'virtual' appearances at multiple events. These sorts of applications will inevitably call for further legal consideration beyond the mere disclosure of AI use—for example, around personal liability and rights to intellectual property.

Perhaps the most explored application of a more complete digital personhood is the idea of achieving 'digital immortality' by creating a digital self that could exist after death. As well as being the subject of academic research (for example, Savin-Baden & Burden 2019), an emergent Digital Afterlife Industry (Öhman &

Floridi 2018) subjects this idea to commercial exploitation (for example, Project December, HereAfter). Commercial enterprises such as Project December indeed already offer the possibility to 'resurrect' a deceased person as a posthumous textbot or 'thanabot' (from 'thanatology', the study of death) by training a generative language model with text originally produced by the deceased (Henrickson 2023). The experience of using—or even hearing about—interactions with such thanabots can be emotionally very powerful (Henrickson 2023). Thanabots also raise implications for how future humans might engage with grief: Morse (2023) notes how posthumous digital communication technologies present the opportunity to emphasise a 'continuing bond' with deceased loved one, rather than a need to move on or let go.

What would happen if a thanabot exhibited not only the linguistic habits of a known human, but also spoke with their synthesised voice? Bassett (2021) interviewed a man called Noah, who created what he called 'Dadbot'—a text-based thanabot built from an oral history collected before his father died. Noah explicitly mentions his reticence about making the bot too realistic for fear that it might become 'uncanny' or 'creepy', using the specific example of the voice:

> *what if you could have a synthesis of his … real voice … and … I knew … I know that's a line out there and when I think about an animated version of him for instance that just gives me the creeps … personally* (Bassett 2021: 819).

Indeed, Bassett's (2019) work with 'digital inheritors'—people with access to digital remains of a loved one such as social media accounts, images, and recordings—brings up some powerful considerations about Noah's 'line' between pleasant and unsettling perceptions of digital afterlives. Several of Bassett's interviewees reported positive feelings of comfort and happiness from listening to digital audio of their deceased loved ones—for example, voicemails, birthday messages, voice recordings—in line with psychological arguments about the richness and personal value of familiar voices (McGettigan 2015; Sidtis & Kreiman 2012). However, in contrast, others reported hearing the deceased's voice as too painful and difficult, with one interviewee reporting that she still enjoyed watching old videos of her loved one but strictly with the audio muted. Quotes from these more reluctant listeners identify something about the dynamic nature of voices that is at odds with the knowledge that their original owners have died: the person 'seems more alive', and the sound of the voice is 'too direct' and 'unadulterated' (Bassett 2019: 162). In the cases described by Bassett, the interviewees had control over whether, and how often, to listen to the audio of their loved ones. However, it is not always possible to have this control—for example, if a person chooses to record and/or synthesise audio messages that are timed for delivery to friends and family after their death.

It is clear that any application of intentional voice cloning or conversion that would make it possible to create an infinite supply of lifelike posthumous communications, whether one-way (that is, pre-designed messages sent from the

deceased) or two-way [that is, thanabots and avatars (Bassett 2019)], bears serious moral and ethical consideration. Among the issues highlighted by Hollanek & Nowaczyk-Basiñska (2024) are impacts on human dignity of both data donors and service interactants, potential for psychological harm through the impeding of proper grieving, and the prospect that a convincing thanabot may so influence an end user that it effectively subverts autonomic choice. As noted earlier in this paper, the inclusion of voice has the potential to accentuate the realism of human–computer interactions, and the impact of this may be yet more strongly felt when the only available version of a loved one is their posthumous digital clone. And yet, as Morse (2023) notes about posthumous digital communication technologies: 'Often, the media coverage … highlights the innovation involved, regardless of its impact on grief processes' (p. 2). An important source of instruction in this space will be to consider the experiences of people living with MND and their engagement with the process of voice banking: already at the point of banking, these individuals know that they may have only a few years to live, and thus the whole process is contextualised within that framework (Benson 2021). There may be opportunities to take learnings and strategies from these more acute clinical contexts and translate them into wider discussions about generative personalised audio as a form of digital legacy.

Conclusions and recommendations

In this paper, we have examined the implications of personalised voice synthesis from the perspective of primarily psychological research conducted across numerous disciplines. We have synthesised prior work on human and non-human voice perception with the latest findings on perception of AI-assisted voice synthesis, to offer predictions and speculations about how the nature of highly humanlike and personalised synthetic voice identities might impact on human experience, and we have linked this to broader considerations around the legality and morality of attempts to replicate, 'return', or resurrect human vocal identities via generative AI.

We report evidence that the naturalistic features of voices can now be convincingly recreated through voice synthesis technologies, to the point where synthetic voice audio cannot be discriminated from human speech recordings (Barrington *et al.* 2025; Lavan et al. 2025; Rosi *et al.* 2025), yet can also sound more appealing and competent than genuine human voices (Rosi *et al.* 2025). When considering familiar listeners and the prospect of intentional self-voice synthesis, there are still limitations on success—for example, poor replication of the talker's accent and other idiosyncrasies can lead to easy detection of synthetic content and associated costs in its social evaluation (Rosi *et al.* 2025). We recommend that, to improve inclusivity and representation, existing models must pay closer attention to representation of different languages and accent identities through obtaining larger and more diverse training data. This is necessary not only to mitigate against inequities of access to personalised voice synthesis, but also to counteract the over-representation of norms (for example, prestige accents) in our auditory world.

When it comes to applications of personalised voice synthesis, we are clear that experiments with healthy, typically younger adults, do not capture the experiences of technology users in applied settings, clinical or otherwise. Therefore, despite our findings indicating the importance of self-similarity in how listeners evaluate a voice clone's social qualities (Rosi *et al.* 2025), we suggest collecting data on how much the perceived accuracy of replicating a specific identity actually matters for the acceptability and utility of a clone in different applied contexts: Where 'perfection' is costly or unattainable, we propose that synthetic voices bearing a resemblance to the most important general identity characteristics (for example, gender, accent) of a speaker may be sufficient for a speaker's purposes.

At the same time, limits on the success of current voice cloning tools should not form the ceiling of our expectations. Thus, applications of personalised voice synthesis that currently seem infeasible may quickly become a reality. We have outlined the potential for both rewards and risks to speakers and listeners—harnessing familiar-voice advantages for spoken language comprehension will have different implications to using a personalised chatbot to simulate social contact with a loved one. Recognising the particularly humanising and rewarding properties of personally significant voices (Seltzer *et al.* 2010, 2012; Kanber *et al.* 2025), we recommend that society must apply particular care around how digital voice identities are 'given' and inherited, and the terms of use surrounding them: While legislation and regulation may move quickly to prevent the unauthorised use of voice cloning and protect the speaker's personal data, guidance on legal and intentional uses and their impacts on both speakers and listeners will likely lag behind. Findings from clinical contexts (for example, voice banking) are particularly pertinent here, as they are uniquely relevant for the synthesis of audio selves: these make clear that supporting human users to intentionally adopt voice technology must go beyond merely providing (or selling) it, to consider the complex psychosocial processes informing both speakers and listeners (Benson 2021; Cave & Bloch 2021).

In sum, we argue that responsible design and implementation of personalised voice synthesis requires an urgent expansion of research on the perceptual properties and psychological impacts of synthetic voices across diverse use cases, alongside a recognition of the learnings already made from clinical contexts, and, more broadly, a societal commitment to incorporating these findings in how we discuss, adopt (or not), and regulate the use of artificial voices in human settings.

Acknowledgements

This work was funded by a British Academy Mid-Career Fellowship (MCFSS23\230112) awarded to CM. CM also acknowledges funding from a Leverhulme Trust Research Leadership Award (RL-2016-013).

References

Abercrombie, G., Cercas Curry, A., Dinkar, T., Rieser, V. & Talat, Z. (2023), 'Mirages: on anthropomorphism in dialogue systems', in *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing* (Singapore, Association for Computational Linguistics), 4776–90. https://doi.org/10.18653/v1/2023.emnlp-main.290

Abrams, D.A., Chen, T., Odriozola, P., Cheng, K.M., Baker, A.E., Padmanabhan, A., Ryali, S., Kochalka, J., Feinstein, C. & Menon, V. (2016), 'Neural circuits underlying mother's voice perception predict social communication abilities in children', *Proceedings of the National Academy of Sciences*, 113(22): 6295–300. https://doi.org/10.1073/pnas.1602948113

Abrams, D.A., Mistry, P.K., Baker, A.E., Padmanabhan, A. & Menon, V. (2022), 'A neurodevelopmental shift in reward circuitry from mother's to nonfamilial voices in adolescence', *Journal of Neuroscience*, 42(20): 4164–73. https://doi.org/10.1523/JNEUROSCI.2018-21.2022

AI Now Institute (2019, November 29). *Disability, Bias, and AI - Report*. AI Now Institute. https://ainowinstitute.org/publication/disabilitybiasai-2019

Alper, M. (2017), *Giving Voice: Mobile Communication, Disability, and Inequality* (Cambridge, MA, The MIT Press). https://doi.org/10.7551/mitpress/10771.001.0001

Altman, I. (with Internet Archive) (1975), *The Environment and Social Behavior: Privacy, Personal Space, Territory, Crowding* (Monterey, CA, Brooks/Cole Pub. Co.). http://archive.org/details/environmentsocia0000altm

Andries, V. & Robertson, J. (2023), 'Alexa doesn't have that many feelings: children's understanding of AI through interactions with smart speakers in their homes', *Computers and Education: Artificial Intelligence*, 5: 100176. https://doi.org/10.1016/j.caeai.2023.100176

Angwin, J., Larson, J., Mattu, S. & Kirchner, L. (2016, May 23). *Machine Bias*. ProPublica. https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

Arik, S., Chen, J., Peng, K., Ping, W. & Zhou, Y. (2018), 'Neural voice cloning with a few samples', in *Advances in Neural Information Processing Systems, Vol. 31* (Red Hook, NY, Curran Associates, Inc.) https://papers.nips.cc/paper_files/paper/2018/hash/4559912e7a94a9c32b09d894f2bc3c82-Abstract.html

Aruffo, C. & Shore, D.I. (2011), 'Self-voice, but not self-face, reduces the McGurk effect', *I-Perception*, 2(8): 772. https://doi.org/10.1068/ic772

Aziz-Zadeh, L., Sheng, T. & Gheytanchi, A. (2010), 'Common premotor regions for the perception and production of prosody and correlations with empathy and prosodic ability', *PLoS One*, 5(1): e8759. https://doi.org/10.1371/journal.pone.0008759

Bains, A., Spaulding, C., Ricketts, J. & Krishnan, S. (2023), 'Using a willingness to wait design to assess how readers value text', *NPJ Science of Learning*, 8(1): 17.

Barnett, J. (2023), 'The ethical implications of generative audio models: a systematic literature review', in *Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society* (New York, Association for Computing Machinery), 146–61. https://doi.org/10.1145/3600211.3604686

Barrington, S., Cooper, E.A. & Farid, H. (2025). People are poorly equipped to detect AI-powered voice clones. arXiv. https://doi.org/10.48550/arXiv.2410.03791

Bassett, D.J. (2019). *You only live twice: a constructivist grounded theory study of the creation and inheritance of digital afterlives*, Ph.D. Thesis. University of Warwick. Available from: http://webcat.warwick.ac.uk/record=b3488708~S15

Bassett, D.J. (2021), 'Ctrl+Alt+Delete: the changing landscape of the uncanny valley and the fear of second loss', *Current Psychology*, 40(2): 813–21. https://doi.org/10.1007/s12144-018-0006-5

Beauchamp, T.L. & Childress, J.F. (2001), *Principles of Biomedical Ethics* (Oxford, Oxford University Press).

Belin, P., Fecteau, S. & Bédard, C. (2004), 'Thinking the voice: neural correlates of voice perception', *Trends in Cognitive Sciences*, 8(3): 129–35. https://doi.org/10.1016/j.tics.2004.01.008

Belin, P., Bestelmeyer, P.E.G., Latinus, M. & Watson, R. (2011), 'Understanding voice perception', *British Journal of Psychology*, 102(4): 711–25. https://doi.org/10.1111/j.2044-8295.2011.02041.x

Benson, J. (2021), 'Saving lost voices: a toolkit for preserving communicative identity', in *Clinical Cases in Dysarthria* (London, Routledge).

Birnhack, M. & Morse, T. (2022), 'Digital remains: property or privacy? *International Journal of Law and Information Technology*, 30(3): 280–301. https://doi.org/10.1093/ijlit/eaac019

Bradlow, A.R. & Pisoni, D.B. (1999), 'Recognition of spoken words by native and non-native listeners: talker-, listener-, and item-related factors', *The Journal of the Acoustical Society of America*, 106(4): 2074–85. https://doi.org/10.1121/1.427952

Buolamwini, J. & Raji, I.D. (2019). Actionable auditing: investigating the impact of publicly naming biased performance results of commercial AI products. https://dspace.mit.edu/handle/1721.1/123456

Candini, M., Zamagni, E., Nuzzo, A., Ruotolo, F., Iachini, T. & Frassinetti, F. (2014), 'Who is speaking? Implicit and explicit self and other voice recognition', *Brain and Cognition*, 92: 112–7. https://doi.org/10.1016/j.bandc.2014.10.001

Cartei, V. & Reby, D. (2013), 'Effect of formant frequency spacing on perceived gender in pre-pubertal children's voices', *PLoS One*, 8(12): e81022. https://doi.org/10.1371/journal.pone.0081022

Cartei, V., Cowles, H.W. & Reby, D. (2012), 'Spontaneous voice gender imitation abilities in adult speakers', *PLoS One*, 7(2): e31353. https://doi.org/10.1371/journal.pone.0031353

Cartei, V., Garnham, A., Oakhill, J., Banerjee, R., Roberts, L. & Reby, D. (2019), 'Children can control the expression of masculinity and femininity through the voice', *Royal Society Open Science*, 6(7): 190656. https://doi.org/10.1098/rsos.190656

Cave, R. & Bloch, S. (2021), 'Voice banking for people living with motor neurone disease: views and expectations', *International Journal of Language & Communication Disorders*, 56(1): 116–29. https://doi.org/10.1111/1460-6984.12588

Cercas Curry, A. & Rieser, V. (2018), '#MeToo Alexa: how conversational systems respond to sexual harassment', in Alfano, M., Hovy, D., Mitchell, M. & Strube, M. (eds), *Proceedings of the Second ACL Workshop on Ethics in Natural Language Processing* (New Orleans, LA, Association for Computational Linguistics), 7–14. https://doi.org/10.18653/v1/W18-0802

Cercas Curry, A., Abercrombie, G. & Rieser, V. (2021), 'ConvAbuse: data, analysis, and benchmarks for nuanced abuse detection in conversational AI', in Moens, M.-F., Huang, X., Specia, L. & Yih, S.W. (eds), *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing* (Punta Cana, Association for Computational Linguistics), 7388–403. https://doi.org/10.18653/v1/2021.emnlp-main.587

Chan, S.W.T., Gunasekaran, T.S., Pai, Y.S., Zhang, H. & Nanayakkara, S. (2021), 'KinVoices: using voices of friends and family in voice interfaces', *Proceedings of the ACM on Human–Computer Interaction*, 5(CSCW2): 1–25. 446. https://doi.org/10.1145/3479590

Cieslak, M. (2024, July 31). Video games strike rumbles on in row over AI. BBC News. https://www.bbc.com/news/articles/czvxz11gl57o

Conde, T., Gonçalves, Ó.F. & Pinheiro, A.P. (2015), 'Paying attention to my voice or yours: an ERP study with words', *Biological Psychology*, 111: 40–52. https://doi.org/10.1016/j.biopsycho.2015.07.014

Dao, X.-Q., Le, N.-B. & Nguyen, T.-M.-T. (2021), 'AI-powered MOOCs: video lecture generation', in *Proceedings of the 2021 3rd International Conference on Image, Video and Signal Processing* (New York, Association for Computing Machinery), 95–102. https://doi.org/10.1145/3459212.3459227

Domingo, Y., Holmes, E. & Johnsrude, I.S. (2020), 'The benefit to speech intelligibility of hearing a familiar voice', *Journal of Experimental Psychology: Applied*, 26(2): 236–47. https://doi.org/10.1037/xap0000247

ElevenLabs (2024a, March 1). On a mission to help 1 million people reclaim their voice. ElevenLabs. https://elevenlabs.io/impact-program

ElevenLabs (2024b, April 19). Why does my voice or accent not sound correct after cloning? ElevenLabs. https://help.elevenlabs.io/hc/en-us/articles/13434263477137-Why-does-my-voice-or-accent-not-sound-correct-after-cloning

Epp, C.D., Munteanu, C., Axtell, B., Ravinthiran, K., Aly, Y. & Mansimov, E. (2017), 'Finger tracking: facilitating non-commercial content production for mobile e-reading applications', in *Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services* (New York, Association for Computing Machinery), 1–15. https://doi.org/10.1145/3098279.3098556

Faber, L.W. (2020), *The Computer's Voice: From Star Trek to Siri* (Minneapolis, MN, University of Minnesota Press). https://doi.org/10.5749/j.ctv1bzfnsv

Fant, G. (1971), *Acoustic Theory of Speech Production: With Calculations Based on X-Ray Studies of Russian Articulations* (Berlin, Boston, Walter de Gruyter).

Federal Trade Commission (2019, October 22). You don't say: an FTC workshop on voice cloning technologies. Federal Trade Commission. https://www.ftc.gov/news-events/events/2020/01/you-dont-say-ftc-workshop-voice-cloning-technologies

Fitch, W.T. & Giedd, J. (1999), 'Morphology and development of the human vocal tract: a study using magnetic resonance imaging', *The Journal of the Acoustical Society of America*, 106(3 Pt 1): 1511–22. https://doi.org/10.1121/1.427148

Graux, J., Gomot, M., Roux, S., Bonnet-Brilhault, F., Camus, V. & Bruneau, N. (2013), 'My voice or yours? An electrophysiological study', *Brain Topography*, 26(1): 72–82. https://doi.org/10.1007/s10548-012-0233-2

Graux, J., Gomot, M., Roux, S., Bonnet-Brilhault, F. & Bruneau, N. (2015), 'Is my voice just a familiar voice? An electrophysiological study', *Social Cognitive and Affective Neuroscience*, 10(1): 101–5. https://doi.org/10.1093/scan/nsu031

Guldner, S., Nees, F. & McGettigan, C. (2020), 'Vocomotor and social brain networks work together to express social traits in voices', *Cerebral Cortex*, 30: 6004–20. https://doi.org/10.1093/cercor/bhaa175

Guldner, S., Lavan, N., Lally, C., Wittmann, L., Nees, F., Flor, H. & McGettigan, C. (2024), 'Human talkers change their voices to elicit specific trait percepts', *Psychonomic Bulletin & Review*, 31(1): 209–22. https://doi.org/10.3758/s13423-023-02333-y

Hazan, V. & Baker, R. (2011), 'Acoustic-phonetic characteristics of speech produced with communicative intent to counter adverse listening conditions', *The Journal of the Acoustical Society of America*, 130(4): 2139–52. https://doi.org/10.1121/1.3623753

Henrickson, L. (2023), 'Chatting with the dead: the hermeneutics of thanabots', *Media, Culture & Society*, 45(5): 949–66. https://doi.org/10.1177/01634437221147626

Hollanek, T. & Nowaczyk-Basiñska, K. (2024), 'Griefbots, deadbots, postmortem avatars: on responsible applications of generative AI in the digital afterlife industry', *Philosophy & Technology*, 37(2): 63. https://doi.org/10.1007/s13347-024-00744-w

Holmes, E. & Johnsrude, I.S. (2020), 'Speech spoken by familiar people is more resistant to interference by linguistically similar speech', *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 46(8): 1465–76. https://doi.org/10.1037/xlm0000823

Holmes, E., Domingo, Y. & Johnsrude, I.S. (2018), 'Familiar voices are more intelligible, even if they are not recognized as familiar', *Psychological Science*, 29(10): 1575–83. https://doi.org/10.1177/0956797618779083

Hughes, S.M. & Harrison, M.A. (2013), 'I like my voice better: self-enhancement bias in perceptions of voice attractiveness', *Perception*, 42(9): 941–9. https://doi.org/10.1068/p7526

Hughes, S.M., Mogilski, J.K. & Harrison, M.A. (2014), 'The perception and parameters of intentional voice manipulation', *Journal of Nonverbal Behavior*, 38(1): 107–27. https://doi.org/10.1007/s10919-013-0163-z

Hutiri, W., Papakyriakopoulos, O. & Xiang, A. (2024), 'Not my voice! A taxonomy of ethical and safety harms of speech generators', in *Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency* (New York, Association for Computing Machinery), 359–76. https://doi.org/10.1145/3630106.3658911

Jackson, L., Haagaard, A. & William, R. (2022, April 19). Disability dongle. Platypus. https://blog.castac.org/2022/04/disability-dongle/

Jia, Y., Zhang, Y., Weiss, R.J., Wang, Q., Shen, J., Ren, F., Chen, Z., Nguyen, P., Pang, R., Moreno, I.L. & Wu, Y. (2018), 'Transfer learning from speaker verification to multispeaker text-to-speech synthesis', in *Proceedings of the 32nd International Conference on Neural Information Processing Systems* (New York, Curran Associates Inc.), 4485–95.

Johnson, J.F., Belyk, M., Schwartze, M., Pinheiro, A.P. & Kotz, S.A. (2021), 'Expectancy changes the self-monitoring of voice identity', *European Journal of Neuroscience*, 53(8): 2681–95. https://doi.org/10.1111/ejn.15162

Johnsrude, I.S., Mackey, A., Hakyemez, H., Alexander, E., Trang, H.P. & Carlyon, R.P. (2013), 'Swinging at a cocktail party: voice familiarity aids speech perception in the presence of a competing voice', *Psychological Science*, 24(10): 1995–2004. https://doi.org/10.1177/0956797613482467

Josan, H.H.S. (2024), *AI Voice Cloning: Copyright, Innovation, and Artists' Rights* (Centre for International Governance Innovation). https://canadacommons.ca/artifacts/11341542/ai-and-deepfake-voice-cloning/12230558/

Judge, S. & Hayton, N. (2022), 'Voice banking for individuals living with MND: a service review', *Technology and Disability*, 34(2): 113–22. https://doi.org/10.3233/TAD-210366

Kanber, E., Lavan, N. & McGettigan, C. (2022), 'Highly accurate and robust identity perception from personally familiar voices', *Journal of Experimental Psychology. General*, 151(4): 897–911. https://doi.org/10.1037/xge0001112

Kanber, E., Roiser, J.P. & McGettigan, C. (2025), 'Personally-valued voices engage reward-motivated behaviour and brain responses', *Social Cognitive and Affective Neuroscience*, 20(1): nsaf056. https://doi.org/10.1093/scan/nsaf056

Ki, C.-W. (Chloe), Cho, E. & Lee, J.-E. (2020), 'Can an intelligent personal assistant (IPA) be your friend? Para-friendship development mechanism between IPAs and their users', *Computers in Human Behavior*, 111: 106412. https://doi.org/10.1016/j.chb.2020.106412

Klofstad, C.A., Anderson, R.C. & Peters, S. (2012), 'Sounds like a winner: voice pitch influences perception of leadership capacity in both men and women', *Proceedings of the Royal Society B: Biological Sciences*, 279(1738): 2698–704. https://doi.org/10.1098/rspb.2012.0311

Knolle, F., Schwartze, M., Schröger, E. & Kotz, S.A. (2019), 'Auditory predictions and prediction errors in response to self-initiated vowels', *Frontiers in Neuroscience*, 13: 1146. https://doi.org/10.3389/fnins.2019.01146

Kolekar, S.S., Richter, D.J., Bappi, M.I. & Kim, K. (2024), 'Advancing AI voice synthesis: integrating emotional expression in multi-speaker voice generation', in *2024 International Conference on Artificial Intelligence in Information and Communication (ICAIIC)* (Piscataway, NJ, IEEE), 193–8. https://doi.org/10.1109/ICAIIC60209.2024.10463204

Kreiman, J. (2024), 'Information conveyed by voice quality', *The Journal of the Acoustical Society of America*, 155(2): 1264–71. https://doi.org/10.1121/10.0024609

Kreiman, J. & Sidtis, D. (2011), *Foundations of Voice Studies*, 1st ed. (Chichester, John Wiley & Sons, Ltd.). https://doi.org/10.1002/9781444395068

Lavan, N. (2023), 'The time course of person perception from voices: a behavioral study', *Psychological Science*, 34(7): 771–83. https://doi.org/10.1177/09567976231161565

Lavan, N. & McGettigan, C. (2023), 'A model for person perception from familiar and unfamiliar voices', *Communications Psychology*, 1(1): 1–11. https://doi.org/10.1038/s44271-023-00001-4

Lavan, N. & Sutherland, C.A.M. (2024), 'Idiosyncratic and shared contributions shape impressions from voices and faces', *Cognition*, 251: 105881. https://doi.org/10.1016/j.cognition.2024.105881

Lavan, N., Burton, A.M., Scott, S.K. & McGettigan, C. (2019), 'Flexible voices: identity perception from variable vocal signals', *Psychonomic Bulletin & Review*, 26(1): 90–102. https://doi.org/10.3758/s13423-018-1497-7

Lavan, N., Irvine, M., Rosi, V. & McGettigan, C. (2025), *Voice deep fakes sound realistic but not (yet) hyperrealistic*, OSF. https://doi.org/10.31234/osf.io/jqg6e_v2

Levon, E., Sharma, D., Watt, D.J.L., Cardoso, A. & Ye, Y. (2021), 'Accent bias and perceptions of professional competence in England', *Journal of English Linguistics*, 49(4): 355–88. https://doi.org/10.1177/00754242211046316

Lim, Y., Toutios, A., Bliesener, Y., Tian, Y., Lingala, S.G., Vaz, C., Sorensen, T., Oh, M., Harper, S., Chen, W., Lee, Y., Töger, J., Monteserin, M.L., Smith, C., Godinez, B., Goldstein, L., Byrd, D., Nayak, K.S. & Narayanan, S.S. (2021), 'A multispeaker dataset of raw and reconstructed speech production real-time MRI video and 3D volumetric images', *Scientific Data*, 8(1): Article 1. https://doi.org/10.1038/s41597-021-00976-x

Linse, K., Aust, E., Joos, M. & Hermann, A. (2018), 'Communication matters—pitfalls and promise of hightech communication devices in palliative care of severely physically disabled patients with amyotrophic lateral sclerosis', *Frontiers in Neurology*, 9: 603. https://doi.org/10.3389/fneur.2018.00603

Mathias, S.R. & von Kriegstein, K. (2014), 'How do we recognise who is speaking? *Frontiers in Bioscience (Scholar Edition)*, 6(1): 92–109. https://doi.org/10.2741/s417

Matthews, J. & Nairn, A. (2023), 'Holographic ABBA: examining fan responses to ABBA's virtual "live" concert', *Popular Music and Society*, 46(3): 282–303. https://doi.org/10.1080/03007766.2023.2208048

McAleer, P. & Belin, P. (2018), in Fruholz, S. & Belin, P. (eds), *The Perception of Personality Traits From Voices* (Oxford University Press), 585–606. https://global.oup.com/academic/product/the-oxford-handbook-of-voice-perception-9780198743187?cc=gb&lang=en&#

McAleer, P., Todorov, A. & Belin, P. (2014), 'How do you say 'hello'? Personality impressions from brief novel voices', *PLoS One*, 9(3): e90779. https://doi.org/10.1371/journal.pone.0090779

McGettigan, C. (2015), 'The social life of voices: studying the neural bases for the expression and perception of the self and others during spoken communication', *Frontiers in Human Neuroscience*, 9: 129. https://www.frontiersin.org/articles/10.3389/fnhum.2015.00129

Mckelvey, M., Evans, D.L., Kawai, N. & Beukelman, D. (2012), 'Communication styles of persons with ALS as recounted by surviving partners', *Augmentative and Alternative Communication*, 28(4): 232–42. https://doi.org/10.3109/07434618.2012.737023

Microsoft Research (n.d.). VALL-E. Retrieved 7 October 2024. Available from: https://www.microsoft.com/en-us/research/project/vall-e-x/

Mileva, M. & Lavan, N. (2023), 'How quickly can we form a trait impression from voices? *Journal of Experimental Psychology. General*.

Miller, N., Taylor, J., Howe, C. & Read, J. (2011), 'Living with foreign accent syndrome: insider perspectives', *Aphasiology*, 25(9): 1053–68. https://doi.org/10.1080/02687038.2011.573857

Mills, T., Bunnell, H.T. & Patel, R. (2014), 'Towards personalized speech synthesis for augmentative and alternative communication', *Augmentative and Alternative Communication*, 30(3): 226–36. https://doi.org/10.3109/07434618.2014.924026

Morse, T. (2023), 'Digital necromancy: Users' perceptions of digital afterlife and posthumous communication technologies', *Information, Communication & Society*, 27(2): 240–56. https://doi.org/10.1080/1369118X.2023.2205467

Mozilla (n.d.). Common Voice. Retrieved 1 November 2024. Available from: https://commonvoice.mozilla.org/en/about

Napolitano, D. (2022), 'Voice cloning and the socio-cultural challanges of assistive technology', in *ICCHP-AAATE 2022 Open Access Compendium 'Assistive Technology, Accessibility and (e)Inclusion' Part II* (Linz, Association ICCHP). https://doi.org/10.35011/icchp-aaate22-p2-25

Nass, C., Steuer, J. & Tauber, E.R. (1994), 'Computers are social actors', in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (New York, Association for Computing Machinery), 72–8. https://doi.org/10.1145/191666.191703

Nathanson, E. (2017), 'Native voice, self-concept and the moral case for personalized voice technology', *Disability and Rehabilitation*, 39(1): 73–81. https://doi.org/10.3109/09638288.2016.1139193

Nicoletti, L. & Bass, D. (2024, August 22). Humans Are Biased. Generative AI Is Even Worse. Bloomberg.Com. https://www.bloomberg.com/graphics/2023-generative-ai-bias/

Nygaard, L.C. & Pisoni, D.B. (1998), 'Talker-specific learning in speech perception', *Perception & Psychophysics*, 60(3): 355–76. https://doi.org/10.3758/BF03206860

Öhman, C. & Floridi, L. (2018), 'An ethical framework for the digital afterlife industry', *Nature Human Behaviour*, 2(5): 318–20. https://doi.org/10.1038/s41562-018-0335-2

Orepic, P., Kannape, O.A., Faivre, N. & Blanke, O. (2023), 'Bone conduction facilitates self-other voice discrimination', *Royal Society Open Science*, 10(2): 221561. https://doi.org/10.1098/rsos.221561

Payne, B., Lavan, N., Knight, S. & McGettigan, C. (2021), 'Perceptual prioritization of self-associated voices', *British Journal of Psychology*, 112(3): 585–610. https://doi.org/10.1111/bjop.12479

Payne, B., Addlesee, A., Rieser, V. & McGettigan, C. (2024), 'Self-ownership, not self-production, modulates bias and agency over a synthesised voice', *Cognition*, 248: 105804. https://doi.org/10.1016/j.cognition.2024.105804

Peng, Z., Wang, Y., Meng, L., Liu, H. & Hu, Z. (2019), 'One's own and similar voices are more attractive than other voices', *Australian Journal of Psychology*, 71(3): 212–22. https://doi.org/10.1111/ajpy.12235

Peng, Z., Hu, Z., Wang, X. & Liu, H. (2020), 'Mechanism underlying the self-enhancement effect of voice attractiveness evaluation: Self-positivity bias and familiarity effect', *Scandinavian Journal of Psychology*, 61(5): 690–7. https://doi.org/10.1111/sjop.12643

Peng, Z., Hu, Z., Wang, X., Jiao, T., Li, H. & Liu, H. (2021), 'Gender context modulation on the self-enhancement effect of vocal attractiveness evaluation', *PsyCh Journal*, 10(6): 858–67. https://doi.org/10.1002/pchj.472

Pérez, A., Díaz-Munío, G.G., Giménez, A., Silvestre-Cerdà, J.A., Sanchis, A., Civera, J., Jiménez, M., Turró, C. & Juan, A. (2021), 'Towards cross-lingual voice cloning in higher education', *Engineering Applications of Artificial Intelligence*, 105: 104413. https://doi.org/10.1016/j.engappai.2021.104413

Pisanski, K. & Reby, D. (2021), 'Efficacy in deceptive vocal exaggeration of human body size', *Nature Communications*, 12(1): 968. https://doi.org/10.1038/s41467-021-21008-7

Pitardi, V. & Marriott, H.R. (2021), 'Alexa, she's not human but … Unveiling the drivers of consumers' trust in voice-based artificial intelligence', *Psychology & Marketing*, 38(4): 626–42. https://doi.org/10.1002/mar.21457

Ray, S. (2024, October 3). Federal Judge Halts California's new anti-deepfakes law—Musk says its 'score one' for free speech. Forbes. https://www.forbes.com/sites/siladityaray/2024/10/03/federal-judge-halts-californias-new-anti-deepfakes-law-musk-says-its-score-one-for-free-speech/

Ribino, P. (2023), 'The role of politeness in human–machine interactions: a systematic literature review and future perspectives', *Artificial Intelligence Review*, 56(1): 445–82. https://doi.org/10.1007/s10462-023-10540-1

Roberts, R.J. (2023), 'You're only mostly dead: protecting your digital ghost from unauthorized resurrection', *Federal Communications Law Journal*, 75(2): 273–96.

Rosi, V., Payne, B. & McGettigan, C. (2024), 'Effects of self-similarity and self-generation on the perceptual prioritisation of voices', *OSF*, https://doi.org/10.31234/osf.io/aqt4n

Rosi, V., Soopramanien, E. & McGettigan, C. (2025), 'Perception and social evaluations of cloned and recorded voices: effects of familiarity and self-relevance', *Computers in Human Behavior: Artificial Humans*, 4: 100143. https://doi.org/10.1016/j.chbah.2025.100143

Rouse (ed.) (2024, September 4). *AI-generated deepfakes: What does the law say*? Rouse. https://rouse.com/insights/news/2024/ai-generated-deepfakes-what-does-the-law-say/

Savin-Baden, M. & Burden, D. (2019), 'Digital immortality and virtual humans', *Postdigital Science and Education*, 1(1): 87–103. https://doi.org/10.1007/s42438-018-0007-6

Schroeder, J. & Epley, N. (2015), 'The sound of intellect: speech reveals a thoughtful mind, increasing a job candidate's appeal', *Psychological Science*, 26(6): 877–91. https://doi.org/10.1177/0956797615572906

Schroeder, J. & Epley, N. (2016), 'Mistaking minds and machines: How speech affects dehumanization and anthropomorphism', *Journal of Experimental Psychology: General*, 145(11): 1427–37. https://doi.org/10.1037/xge0000214

Schroeder, J., Kardas, M. & Epley, N. (2017), 'The humanizing voice: speech reveals, and text conceals, a more thoughtful mind in the midst of disagreement', *Psychological Science*, 28(12): 1745–62. https://doi.org/10.1177/0956797617713798

Scott, S. & McGettigan, C. (2016), 'The voice: from identity to interactions', in *APA Handbook of Nonverbal Communication* (Washington, DC, American Psychological Association), 289–305. https://doi.org/10.1037/14669-011

Seaborn, K., Miyake, N.P., Pennefather, P. & Otake-Matsuura, M. (2021), 'Voice in human–agent interaction: a survey', *ACM Computing Surveys*, 54(4): 81, 1–43. https://doi.org/10.1145/3386867

Seltzer, L.J., Ziegler, T.E. & Pollak, S.D. (2010), 'Social vocalizations can release oxytocin in humans', *Proceedings of the Royal Society B: Biological Sciences*, 277(1694): 2661–6. https://doi.org/10.1098/rspb.2010.0567

Seltzer, L.J., Prososki, A.R., Ziegler, T.E. & Pollak, S.D. (2012), 'Instant messages vs. speech: hormones and why we still need to hear each other', *Evolution and Human Behavior*, 33(1): 42–5. https://doi.org/10.1016/j.evolhumbehav.2011.05.004

Shadden, B. (2005), 'Aphasia as identity theft: theory and practice', *Aphasiology*, 19(3–5): 211–23. https://doi.org/10.1080/02687930444000697

Shiramizu, V.K.M., Lee, A.J., Altenburg, D., Feinberg, D.R. & Jones, B.C. (2022), 'The role of valence, dominance, and pitch in perceptions of artificial intelligence (AI) conversational agents' voices', *Scientific Reports*, 12(1): Article 1. https://doi.org/10.1038/s41598-022-27124-8

Shuster, L.I. & Durrant, J.D. (2003), 'Toward a better understanding of the perception of self-produced speech', *Journal of Communication Disorders*, 36(1): 1–11. https://doi.org/10.1016/S0021-9924(02)00132-6

Sidtis, D. & Kreiman, J. (2012), 'In the beginning was the familiar voice: personally familiar voices in the evolutionary and contemporary biology of communication', *Integrative Psychological & Behavioral Science*, 46: 146–59. https://doi.org/10.1007/s12124-011-9177-4

Sorokowski, P., Puts, D., Johnson, J., Żółkiewicz, O., Oleszkiewicz, A., Sorokowska, A., Kowal, M., Borkowska, B. & Pisanski, K. (2019), 'Voice of authority: professionals lower their vocal frequencies when giving expert advice', *Journal of Nonverbal Behavior*, 43(2): 257–69. https://doi.org/10.1007/s10919-019-00307-0

Sutton, S.J., Foulkes, P., Kirk, D. & Lawson, S. (2019), 'Voice as a design material: sociophonetic inspired design strategies in human–computer interaction', in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (New York, Association for Computing Machinery), 1–14. https://doi.org/10.1145/3290605.3300833

Tigue, C.C., Borak, D.J., O'Connor, J.J.M., Schandl, C. & Feinberg, D.R. (2012), 'Voice pitch influences voting behavior', *Evolution and Human Behavior*, 33(3): 210–6. https://doi.org/10.1016/j.evolhumbehav.2011.09.004

Veaux, C., Yamagishi, J. & King, S. (2013), 'Towards personalised synthesised voices for individuals with vocal disabilities: voice banking and reconstruction', in *Proceedings of the Fourth Workshop on Speech and Language Processing for Assistive Technologies* (Association for Computational Linguistics), 107–111. https://aclanthology.org/W13-3917

Watt, D., Harrison, P.S. & Cabot-King, L. (2020), 'Who owns your voice? Linguistic and legal perspectives on the relationship between vocal distinctiveness and the rights of the individual speaker', *International Journal of Speech, Language and the Law*, 26(2): 137–80. https://doi.org/10.1558/ijsll.40571

Wilson, S. & Moore, R. (2017), 'Robot, alien and cartoon voices: implications for speech-enabled systems', in *Proceedings of the 1st International Workshop on Vocal Interactivity In-and-between Humans, Animals and Robots (VIHAR-2017)*. 40–4. https://vihar-2017.vihar.org/

Wofford, M.C., Ogletree, B.T. & De, N.T. (2022), 'Identity-focused practice in augmentative and alternative communication services: a framework to support the intersecting identities of individuals with severe disabilities', *American Journal of Speech-Language Pathology*, 31(5): 1933–48. https://doi.org/10.1044/2022_AJSLP-21-00397

Yamagishi, J., Veaux, C., King, S. & Renals, S. (2012), 'Speech synthesis technologies for individuals with vocal disabilities: voice banking and reconstruction', *Acoustical Science and Technology*, 33(1): 1–5. https://doi.org/10.1250/ast.33.1

Zhang, L., Jiang, L., Washington, N., Liu, A.A., Shao, J., Fourney, A., Morris, M.R. & Findlater, L. (2021), 'Social media through voice: synthesized voice qualities and self-presentation', *Proceedings of the ACM on Human–Computer Interaction*, 5(CSCW1): 1–21. 161. https://doi.org/10.1145/3449235

Zhou, M., Abhishek, V., Derdenger, T., Kim, J. & Srinivasan, K. (2024). Bias in generative AI. arXiv. https://arxiv.org/abs/2403.02726v1

About the authors

Carolyn McGettigan is a psychologist and neuroscientist researching the perception and expression of voices, including aspects of speech, emotions, and identity. They are Chair in Speech and Hearing Sciences at UCL, and a principal investigator in the Vocal Communication Laboratory (https://vocolab.net). In 2023–24, Professor McGettigan held a Mid-Career Fellowship from The British Academy, investigating the psychology and ethics of voice clones.

Steven Bloch is a speech and language therapist and conversation analyst researching interactions between people with progressive speech loss, arising from conditions like motor neurone disease, and their communication partners. He is Chair in Communication and Social Interaction at UCL, Head of the Department of Language and Cognition, and co-lead of the UCL Better Conversations lab. E-mail: s.bloch@ucl.ac.uk

Cennydd Bowles is a technology ethicist, author of Future Ethics, and recently a Fulbright Scholar at Elon University, North Carolina. His research interests include manipulation and deception within AI systems, the use of futures methods to provoke moral imagination, and the ethics of product experimentation. E-mail: cennydd@cennydd.com

Tanvi Dinkar is a Research Associate at Heriot-Watt University, researching safety and ethics in generative AI. She is particularly interested in how online spaces affect offline lives, especially for women and girls, including work on online gender-based violence, deceptive human-likeness of chatbots, and AI-driven beauty standards. She completed her PhD in Computer Science at Institut Polytechnic with a Marie-Curie Scholarship, and holds MSc degrees in Linguistics and Informatics, both from the University of Edinburgh. E-mail: t.dinkar@hw.ac.uk

Nadine Lavan is a Senior Lecturer in the Department of Psychology at Queen Mary University of London. Nadine's work examines voice perception, with a focus on exploring how listeners makes sense of who they are talking to. As part of this work, Nadine has therefore asked questions such as: how do listeners recognise a familiar person from their voice, how do we become familiar with a voice, and how do we form first impressions?

Jonathan Chaim Reus (he/they) is a multidisciplinary artist, musician, and researcher exploring how sonic technologies (particularly AI and data-driven voice technologies) transform bodies, communities, and identities. He is an affiliate researcher at the University of Sussex's Experimental Music Technologies Lab, the Sussex Digital Humanities Lab, and the University of Iceland's Intelligent Instruments Lab. Jonathan was recently the creative lead on the S+T+ARTS EU-funded DADAsets project, and received the KONTINUUM commission for generative radio art. E-mail: j.reus@sussex.ac.uk

Victor Rosi is a researcher in cognitive science and audio signal processing, exploring how identity—particularly gender—is expressed and perceived through the voice. He currently holds a British Academy Postdoctoral Fellowship, studying the acoustic and perceptual cues of gender-diverse voices. E-mail: v.rosi@ucl.ac.uk