From Voice to Self: An Integrative Framework on Self-Voice Processing

Pavo Orepic [1] & Ana P. Pinheiro [2]

[1] Linguistic Research Infrastructure, University of Zurich, Switzerland

[2] Faculdade de Psicologia, CICPSI, Universidade de Lisboa, Lisboa, Portugal

Contact: pavo.orepic@uzh.ch, appinheiro@psicologia.ulisboa.pt

Abstract

The self-voice plays a fundamental role in communication and identity, yet remains a relatively neglected topic in psychological science. As AI-generated and digitally manipulated voices become more common, understanding how individuals perceive and process their own voice is increasingly important. Disruptions in self-voice processing are implicated in several clinical conditions, including psychosis, autism, and personality disorders, highlighting the need for integrative models to explain self-voice across contexts. However, research faces two major challenges: a methodological one – replicating the bone-conducted acoustics that shape natural self-voice perception, and a conceptual one – a persistent bias toward treating the self-voice as purely auditory. To address these gaps, we propose a framework decomposing the self-voice into five interacting components: auditory, motor, memory, multisensory integration, and self-concept. We review the functional and neural basis of each component and suggest how they converge within distributed brain networks to support coherent self-voice processing. This integrative framework aims to advance theoretical and translational work by bridging psychology, neuroscience, clinical research, and voice technology in the context of emerging digital voice environments.

Keywords

Introduction

Our voice is a uniquely personal signal that manifests our identity in both overt and subtle ways. It reflects a distinctive configuration of auditory-acoustic features that convey physical attributes such as age, sex, and body size, alongside psychological and social traits, including aspects of one's personality (Sidtis & Zäske, 2021). Like a vocal fingerprint, the self-voice thus serves as a powerful marker of individuality. It is the voice we hear most frequently throughout our lives since we are exposed to it every time we speak. It also functions as a primary vehicle for communication and self-expression, influencing how others perceive us and how we perceive ourselves. Notably, disruptions in self-voice processing have been implicated in clinical conditions such as psychosis (Pinheiro et al., 2017), autism (Chakraborty & Chakrabarti, 2015), and personality disorders (Orepic, Iannotti, et al., 2023), suggesting that self-voice perception plays a broader role in psychological functioning. Despite its relevance, the self-voice has received surprisingly little systematic attention in the cognitive sciences and in clinical research, especially compared to other self-related stimuli like the self-face or body.

We argue that progress in self-voice research has been limited by two major challenges. The first is methodological: natural self-voice perception involves bone conduction, wherein sound vibrations travel through cranial structures, modifying how one's voice is perceived internally (Reinfeldt et al., 2010). This differs substantially from the air-conducted recordings typically used in experiments, which sound unfamiliar or even aversive to participants (Gur & Sackeim, 1979). Consequently, studies using digital self-voice stimuli may evoke unnatural perceptual and emotional responses, introducing confounds in both behavioral and neuroimaging paradigms. The second challenge is conceptual: people typically perceive their own voice as a purely auditory phenomenon, often unaware of the broader sensory and motor processes that support this perception. This auditory-centric perspective is mirrored in much of the self-voice literature, which has historically emphasized acoustic features over other

modalities (see Box 2). However, natural self-voice processing is inherently multimodal (Orepic, Kannape, et al., 2023), integrating auditory, motor, proprioceptive, and tactile feedback during speech production. These co-occurring signals likely shape how we identify our voice as our own, distinguishing it from other voices in the environment.

The study of self-voice is particularly timely, as advances in artificial intelligence (AI) increasingly enable seamless and realistic voice modifications. Altering how we hear our own voice can lead to short-term changes in cognition, emotion, and behavior. Similar to the *Proteus effect* – where individuals adapt their behavior to align with their digital avatars (Yee & Bailenson, 2007) – emerging evidence [1] shows that even brief exposure to a modified self-voice can influence personality traits (Fang et al., 2024), emotional states (Aucouturier et al., 2016; Costa et al., 2018), and social attitudes (Arakawa et al., 2021). For example, making one's voice sound calmer can reduce anxiety during interpersonal conflict (Costa et al., 2018), while aging one's voice with AI can reduce biases towards the elderly (Arakawa et al., 2021). Similar, but longer-lasting effects are observed in clinical contexts such as gender transition or laryngectomy, where changes to the self-voice reshape self-perception over time (Bickford et al., 2019; Bultynck et al., 2017). These findings highlight a powerful insight: modifying our voice changes not only how we sound but also how we perceive ourselves. Understanding how the brain integrates these altered signals opens new directions for neuroscience research, clinical intervention, and the design of voice-based technologies.

To address the gaps in self-voice research and guide future studies, we propose a new framework that decomposes self-voice processing into five interacting 'building blocks': auditory, motor, memory, multisensory integration, and self-concept (see Figure 1). We first

---

[1] This evidence should, nevertheless, be considered preliminary, as most findings derive from preprints (Fang et al., 2024) or small-sample exploratory designs without peer review or replication (e.g., Arakawa et al., 2021; Costa et al., 2018).

describe the historical and disciplinary context of self-voice research and discuss the special neurocognitive status of the self-voice. We then outline the functional and neural basis of each building block and examine how they may converge to support a unified sense of vocal selfhood. This integrative framework emphasizes translational relevance, particularly in light of emerging technologies – such as AI-generated voice cloning and digital communication platforms – that are reshaping how individuals perceive their own voice.

By reframing the self-voice as a multidimensional construct, this review aims to advance a new line of inquiry that bridges theoretical, methodological, and applied domains in an increasingly digital world.
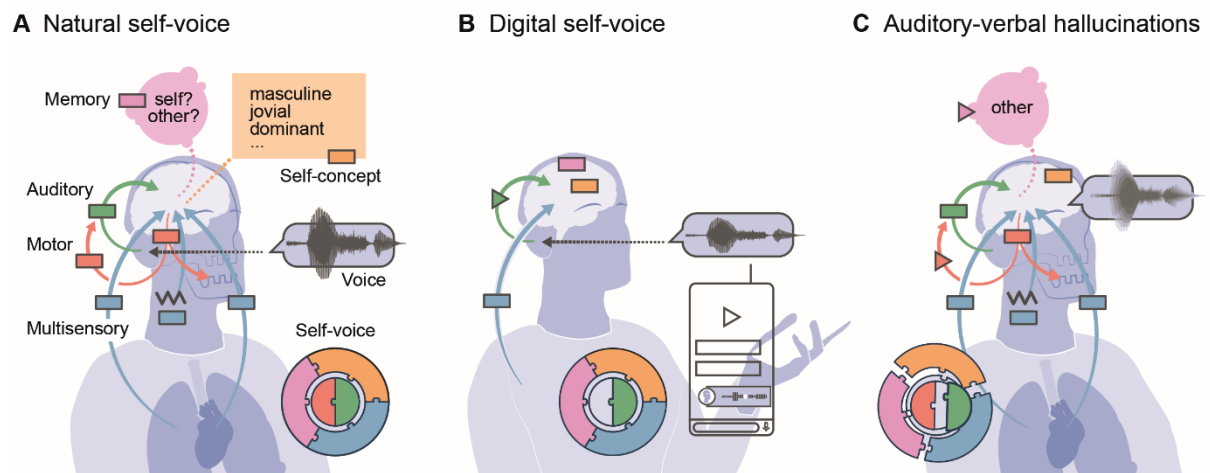

Figure 1. Self-voice building blocks. **A)** In natural self-voice perception (during speaking), all five blocks combine to form a complete self-voice representation (puzzle inset). The Auditory block (green) processes the acoustic signal (voice waveform). The Motor block (red) generates the voice (downward arrow) and sends sensorimotor predictions to the Auditory block (upward arrow) to stabilize perception. The Memory block (pink) provides an internal self-voice template for recognizing whether a voice is self- or externally generated. The Multisensory Integration block (blue) incorporates exteroceptive (e.g., bone-conducted vibrations) and interoceptive cues (e.g., respiration). The Self-concept block (orange) aligns the perceived voice with personality and social identity. **B)** Digital self-voices lack Motor involvement (incomplete puzzle) and bone-conducted cues, altering perceived voice quality (changed waveform; triangle in the Auditory block). **C)** In auditory-verbal hallucinations, the blocks fail to integrate properly (distorted puzzle), producing a hallucinated voice (blurry waveform). Disruptions in sensorimotor prediction (triangle in red upward arrow) and in the Memory block (pink triangle) contribute to self-other voice confusion.

Is the Self-Voice Special?

As the sound we hear most often, a central marker of identity, and a vital tool for communication, the self-voice has a unique cognitive and neural status. This distinctiveness is not only evident when compared to other voices (Graux et al., 2015; Kaplan et al., 2008) but also when contrasted with other self-related stimuli, such as the self-face (Aruffo & Shore, 2012; Hughes & Nicholson, 2010). Crucially, the special processing of the self-voice cannot be fully explained by familiarity alone (Graux et al., 2015; Kaplan et al., 2008; Nakamura et al., 2001), suggesting it engages additional mechanisms beyond those typically involved in processing familiar stimuli.

The self-voice compared to other voices

Behavioral and neuroimaging studies consistently support the idea that the self-voice is functionally distinct from other voices (see Tables 1 and 2 and Figure 2). The few existing functional neuroimaging studies have shown that listening to recordings of one's own voice activates specific brain regions more strongly than listening to others. These include the right anterior cingulate cortex and left inferior frontal cortex (self *vs.* unfamiliar voice; Allen et al., 2005), the right inferior frontal sulcus and parainsular cortex (self *vs.* familiar voice; Nakamura et al., 2001), and the right inferior frontal gyrus (self *vs.* familiar voice; Kaplan et al., 2008). These findings suggest a degree of anatomical specialization for self-voice perception, particularly in right-lateralized frontoinsular networks.

Further evidence for right-hemisphere involvement comes from lesion and lateralization studies. Rosa et al. (Rosa et al., 2008) found a selective advantage for left-hand responses (right-hemisphere processing) in a self-other voice discrimination task using voice morphs. Similarly, right-hemisphere lesions were associated with greater deficits in explicit (*vs.*

implicit) self-voice recognition (Candini et al., 2018), underscoring the importance of this hemisphere in consciously accessing vocal self-identity.

Electrophysiological studies also show early and sustained differentiation of the self-voice. Event-related potential (ERP) studies have revealed enhanced amplitudes for the self-voice compared to familiar and unfamiliar voices, beginning within 100 ms of voice onset and extending into later processing stages (Graux et al., 2013; Pinheiro, Rezaii, Nestor, et al., 2016; Pinheiro et al., 2023). These effects are seen in early (e.g., N1) and later components (e.g., P3, Late Positive Potential; see Table 2), supporting a self-prioritization effect that is modulated by task demands, such as explicit attention to speaker identity (Pinheiro et al., 2023). Differences in task design can yield seemingly contradictory findings across studies (e.g., opposite effects on the same ERP component). Such discrepancies are expected, as different tasks engage distinct cognitive processes that modulate auditory regions in different ways (e.g., excitatory *vs*. inhibitory influences). Importantly, however, a consistent pattern remains: within each study, ERPs reliably differentiate self-voice from other-voice stimuli (see Supplementary Material for additional discussion). These ERP findings suggest that the self-voice captures attention automatically and is subject to sustained, elaborative processing, paralleling the prioritization observed for emotionally salient stimuli (Hajcak & Foti, 2020). Psychophysiological data reinforce this idea: studies from as early as the 1960s reported stronger galvanic skin responses to self-voice compared to other voices, even when participants did not consciously recognize the voice as their own (Olivos, 1967; Rousey & Holzman, 1967).

Behavioral research likewise points to the special status of the self-voice. In voice identity matching tasks, participants more accurately and rapidly identify their own voice compared to others (Kirk & Cunningham, 2025), and this effect holds even when stimuli are AI-generated self-voice approximations (Rosi et al., 2025). The self-voice also influences various perceptual and cognitive functions differently than other voices, including word

recognition (Cheung & Babel, 2022), distance estimation (Wen et al., 2022), face recognition (Hughes & Nicholson, 2010), and multisensory integration (Aruffo & Shore, 2012). The perception of one's own voice is also linked to real-world outcomes: in a recent survey, participants reported that how they perceive their speaking voice influences self-expression and social behavior (Chong et al., 2024).

These distinct effects may arise from the self-voice's ability to capture attention (Conde et al., 2015; Pinheiro et al., 2023), its enhanced affective salience (Conde et al., 2015), or a perceptual bias driven by a sense of ownership (Payne et al., 2024). Interestingly, the self-voice is often rated as more attractive than other voices (Hughes & Harrison, 2013), even when it is not explicitly recognized as self-related (Douglas & Gibbins, 1983).

Together, these findings highlight the functional uniqueness of the self-voice, reflected in its capacity to engage both low-level sensory and higher-level cognitive processes in ways that differ from the processing of even familiar non-self voices.

The self-voice compared to the self-face

Both the face and the voice serve as key self-identifiers and follow parallel processing hierarchies (Yovel & Belin, 2013). Like the self-voice, the self-face is processed differently from familiar and unfamiliar faces (Bortolon & Raffard, 2018). However, several lines of evidence suggest that the self-voice may hold an even more distinctive status.

First, the self-voice is arguably the most frequent expression of the self in daily life. While individuals may only see their own face in mirrors or digital media, they hear their voice every time they speak. Second, the voice plays a more central role in communication than the face: vocal communication is the primary medium of interaction across cultures (Scott, 2019), and the brain may be tuned to prioritize the modality that directly facilitates this function. Third, unlike the self-face, the self-voice inherently engages the motor system, triggering forward

model predictions during vocal production. This motor involvement enhances the integration of auditory, somatosensory, and vestibular cues, potentially amplifying self-relevance through multisensory convergence (see *Motor* and *Multisensory Integration* blocks).

From a clinical standpoint, the self-voice also appears more central. Auditory hallucinations – predominantly in the form of voices – are significantly more common than visual hallucinations (Bauer et al., 2011). Furthermore, difficulties with self-other distinction have been linked to autistic traits in the auditory, but not visual, domain (Chakraborty & Chakrabarti, 2015) (see *Translational Implications*).

When the self-voice and self-face are perceived together, their interaction appears to depend strongly on context. For instance, Hughes and Nicholson (Hughes & Nicholson, 2010) reported that individuals recognize their own face more readily than their own voice and that presenting both simultaneously may impair recognition, perhaps due to the cognitive cost of binding multiple self-related cues. In contrast, Aruffo and Shore (Aruffo & Shore, 2012) showed that the self-voice – but not the self-face – reduces susceptibility to the McGurk illusion, suggesting that the self-voice carries greater perceptual weight during audiovisual speech integration. There is also evidence that auditory and visual self-recognition are partially dissociable. Chakraborty and Chakrabarti (Chakraborty & Chakrabarti, 2015) observed no correlation between self-recognition performance across modalities, supporting the view that self-identity processing is modality-specific (Young et al., 2020). Although self-voice recognition may rely on less robust or more ambiguous identity cues than the face (Hughes & Nicholson, 2010), it nonetheless engages distinctive cognitive and neural mechanisms compared to other voices. The 'specialness' of the self-voice may therefore manifest in attentional capture, affective salience, or sensorimotor integration rather than in absolute recognition accuracy.

Box 1: Glossary

Auditory-verbal hallucinations: Hearing voices or speech that seem real but are not produced by an external source. One of the most common symptoms of schizophrenia, but also present in the general population.

Bodily self-consciousness: The basic sense of owning and controlling one's body and being located in space. It is supported by multisensory integration of exteroceptive (e.g., vision, touch, sound), interoceptive (e.g., respiration, heartbeat, internal bodily sensations), and proprioceptive signals (information about body position and movement).

BOLD (Blood-Oxygen-Level-Dependent) Signal: The signal measured in fMRI that reflects changes in blood oxygenation linked to neural activity. When a brain region becomes more active, local blood flow increases, producing a change in the BOLD signal.

Bone conduction: Alternative pathway through which the sound of our voice travels to our inner ears. It is based on vibrations of the bones of the skull and other tissues in the head, and it is one of the reasons why we hear our own voice differently from how others hear it.

Corollary discharge: An internal signal the brain sends when starting a movement – such as preparing to speak – to alert sensory areas that a self-generated sound or sensation is about to occur. This helps the brain tell apart sensations caused by oneself from those caused by the outside world.

ECoG (electrocorticography): A brain-recording method that measures electrical activity directly from the surface of the brain.

EEG/MEG (electroencephalography / magnetoencephalography): Techniques that measure the brain's electrical or magnetic activity from outside the head, offering very fast tracking of neural signals.

Efference copy: An internal copy of the motor command the brain issues when performing an action. This copy allows sensory systems to anticipate the specific feedback the action should produce, such as the expected acoustic consequences of one's own voice.

Event-related potential (ERP): A brain response measured with EEG or MEG that is time-locked to a specific event, such as hearing a sound.

fMRI (functional magnetic resonance imaging): A brain-imaging method that tracks changes in blood flow to show which brain areas are active during a task.

Formant frequencies (F1-F5): Resonant frequencies in the voice that shape vowel sounds and help identify speakers. Each person has different formant frequencies due to unique anatomies of vocal and nasal cavities. Typically, five different formants are extracted from the speech signals, and the lower ones change more depending on which sound we are articulating.

Forward model: A computational framework that uses internal motor-related signals to predict the sensory outcome of an action. When speaking, the forward model estimates how the voice should sound and feel, allowing the brain to compare expectations with actual feedback and detect mismatches.

Fundamental frequency (F0): The lowest frequency of a voice, often perceived as its pitch. It is lower in males compared to females.

LPP (Late Positive Potential): A late ERP component (beginning around 300-400 ms) that is larger for emotionally or motivationally significant stimuli. It reflects sustained attention and evaluative processing.

MMN (Mismatch Negativity): An ERP response that occurs when an unexpected sound violates a repeating pattern. It appears automatically (without attention) around 100-250 ms and reflects early auditory change detection.

Motor control (vocal control): How the brain directs and coordinates the muscles needed for producing speech.

N1/M1 suppression: Reduction of early (around 100 ms) brain responses (N1 in EEG/MEG, M1 in MEG) when hearing one's own voice, reflecting the brain's prediction of self-generated sounds.

Norm-based coding of voices: The idea that the brain represents voices by comparing them to an internal "average voice", making distinctive voices easier to recognize.

Predictive coding: A theory proposing that the brain constantly generates predictions about incoming sensory information and updates them based on new evidence.

Prediction error: The difference between what the brain expects to perceive and what it actually perceives.

Readiness potential: A slow buildup of brain activity that occurs before a voluntary movement, including speaking.

Self-concept: A person's understanding of who they are, including their personality traits, abilities, and identity, as well as how they see themselves in relation to others. It reflects both internal self-knowledge and socially shaped aspects of identity, such as roles, group memberships, and how individuals believe they are perceived by others.

Sensory suppression: The brain's reduction of its own sensory responses to self-generated signals (like the sound of one's own voice).

Source-filter model of speech: A model describing how the voice is produced: the vocal cords create a sound source, and the vocal tract shapes it into speech. In the speech signal, the source is typically reflected in the fundamental frequency, and the filter in formant frequencies.

Source localization: Methods used in brain imaging (especially EEG/MEG) to estimate where in the brain electrical activity is coming from.

Box 2: Historical and Disciplinary Context

The self-voice has received comparatively limited attention in psychology and neuroscience relative to other physical expressions of the self, such as the self-face (see Supplementary Material for publication count comparisons with the self-face and the bodily

self). This imbalance likely reflects both historical and methodological factors. Early theories of perception and identity – spanning psychology, neuroscience, and philosophy – treated the face as the primary, stable, and readily observable marker of personhood, cementing its centrality in the neuroscience of identity (Bruce & Young, 1986; Goffman, 1949; Haxby et al., 2000). By contrast, the voice was considered more variable, technically challenging to measure, and primarily a vehicle for social expression rather than a window into self-identity (Hanley et al., 1998; Kreiman, 2024; Laver, 1980). This asymmetry may have created an early, self-reinforcing bias: research on faces generated influential findings that, in turn, attracted additional researchers and funding.

Meanwhile, foundational studies of the self-voice confronted the methodological problem that recorded voices sound different from the voices people hear when speaking, owing to bone-conducted pathways (Gur & Sackeim, 1979; Holzman & Rousey, 1966; Olivos, 1967). This challenge complicated efforts to produce ecologically valid self-voice stimuli and impeded early progress. Moreover, while visual self-recognition could be investigated with static images, studying self-voice processing required dynamic control of auditory feedback and ecologically valid manipulations of vocal identity – technical capacities that have only recently became widely accessible (e.g., Belin & Kawahara, 2025; Kreiman, 2024).

Beyond these historical constraints, research on the self-voice has been fragmented across disciplines. Although much of the empirical foundation comes from cognitive neuroscience, converging evidence from sociophonetics, speech perception, and social psychology highlights the fundamentally interdisciplinary nature of vocal identity. For example, sociophonetics has examined how speakers modulate vocal features to express social identity and group membership (e.g., Bradshaw et al., 2025; Hay, 2018); speech perception research has focused on acoustic and articulatory mechanisms of talker normalization (e.g., Bourguignon et al., 2016); and social psychology has explored how voice cues shape self-

presentation and interpersonal perception (e.g., Klofstad et al., 2012). Yet these traditions rarely interact with the cognitive and neural perspectives that dominate contemporary self-voice studies, and the role of self-voice identity remains underexamined within these adjacent disciplines – for instance, how self- *vs.* other-voice markers shape speech perception processes (for a few examples, see Pinheiro et al., 2016, 2023).

Each of the five building blocks of self-voice processing outlined in our framework has, in fact, deep roots in these complementary fields. The Auditory block resonates with classic work on speech perception and talker normalization, which has examined how listeners accommodate variability in voice acoustics (e.g., Luthra, 2024). The Motor block intersects with phonetics and speech production research, where articulatory dynamics and feedback control are central (e.g., Parrell & Houde, 2019). The Memory block draws on psycholinguistic and associative learning traditions (e.g., Lee & Perrachione, 2022; Perrachione et al., 2011), as well as mean-based coding models of voice identity (e.g., Latinus et al., 2013), which propose that repeated exposure shapes prototypical voice representations centered on a stored "mean" of one's own or others' vocal patterns. The Multisensory integration block connects to embodied cognition research, emphasizing the convergence of auditory, proprioceptive, tactile, and interoceptive signals in action monitoring (e.g., Orepic et al., 2021, 2022; Orepic, Kannape, et al., 2023). Finally, the Self-Concept block integrates insights from social psychology and sociophonetics research, where the voice serves as a marker of identity, personality, and group membership, reflecting an interplay between socially informed vocal patterns and the individual's unique self-expression (Guldner et al., 2024; Ostrand & Chodroff, 2021; Stern et al., 2021). These perspectives complement neuroscientific evidence by showing how self-concept both constrains and is reinforced by vocal self-expression.

Despite these rich disciplinary traditions, they have rarely been synthesized into a unified account. By integrating insights across these domains, the present model aims to reposition the

self-voice as a multidisciplinary construct – one that bridges low-level sensorimotor control with high-level representations of social identity and self-concept.

The Building Blocks of the Self-Voice

The perception of one's own voice is a complex, multi-component process that integrates sensory, motor, and cognitive systems. We propose that this phenomenon can be comprehensively understood through five foundational components or building blocks: Auditory, Motor, Memory, Multisensory Integration, and Self-Concept (Figure 1). These components interact dynamically to support the perception and recognition of the self-voice.

The Auditory block captures the auditory-acoustic aspects of self-voice processing, which can be influenced by interactions with the other blocks. The Motor block refers to the *predictive* mechanisms involved in speech production that transform motor intentions into expected sensory outcomes – for example, through efference copies that anticipate the auditory consequences of one's own speech (Miall & Wolpert, 1996; Wolpert, 1997). The Multisensory Integration block reflects the *convergent* processing of incoming feedback from multiple modalities, including auditory, somatosensory, and proprioceptive cues, which are combined to construct a coherent percept of vocal identity. While the Motor block functions primarily in a forward-modeling capacity, the Multisensory Integration block operates in a *reconstructive* capacity, ensuring consistency between predicted and perceived self-voice signals across sensory domains. The Memory block refers to the internal representation of the self-voice – our sense of "how I think I sound like" – which enables recognition of one's recorded voice even in the absence of concurrent sensorimotor predictions. Finally, the Self-concept block encompasses higher-level aspects of the sense of self, such as personality traits or social identities, which shape and are reinforced by vocal behavior (e.g., speaking in ways that align with the norms of one's social group).

Existing models of speech and voice processing tend to focus primarily on auditory feedback, often underemphasizing the contribution of bone conduction and other bodily signals to natural self-voice perception or addressing only a subset of the relevant processing components. For example, the DIVA (Tourville & Guenther, 2011) and the more recent LaDIVA (Weerathunge et al., 2022) frameworks focus on sensorimotor predictions and incorporate feedforward and feedback loops involving auditory, motor, and somatosensory signals. However, their emphasis remains on externally derived auditory targets rather than on internally generated, multimodal cues that shape the experience of one's own voice. Leading models of voice perception – such as the hierarchical model of voice processing (Belin et al., 2011) and the prototype-based account of voice identity recognition (Lavner et al., 2001a) – similarly focus on acoustic and memory-based mechanisms, largely overlooking the integration of bodily signals during self-generated vocalization. Many other models framed within the predictive coding framework highlight the influence of top-down modulations linked to semantic or contextual information (Caucheteux et al., 2023; Friston et al., 2021; Ralph et al., 2017), yet these frameworks tend to focus on speech and linguistic content rather than on the voice as marker of identity.

In the following sections, we review key empirical findings for each building block, highlight their translational and clinical implications, and outline unresolved questions to guide future research.

Auditory Block

At its core, each voice has a unique acoustic profile that is shaped by an individual's vocal apparatus. While the acoustic parameters that define voice identity remain incompletely understood (Kreiman, 2024), the self-voice introduces additional complexity due to the effects

of bone conduction – an internal transmission pathway absent from our experience of most externally-generated sounds. While the term "bone conduction" suggests that sound is transmitted solely through bones, the process is influenced by the combined properties of all tissues and fluids in the head, including skin, cartilage, and inner ear fluids (Adelman et al., 2015; Hosoi et al., 2019; Stenfelt, 2016).

Bone conduction adds low-frequency energy to the perception of one's own voice, but its acoustic contribution is still not fully characterized (Stenfelt, 2016). Early work (Franke, 1956) and more recent measurements of ear-canal sound pressure (Reinfeldt et al., 2010) have attempted to quantify the frequency ranges where bone conduction dominates air conduction. These studies show frequency-dependent interactions: bone conduction may dominate between 700 Hz and 1.2 kHz (Pörschmann, 2000) or around 1.5-2 kHz (Reinfeldt et al., 2010), with variability across studies. Additionally, speech sounds with greater oral closure (e.g. lip rounding as in /o/) or nasal resonance (e.g., /n/) tend to engage bone conduction more than plosives or open vowels (e.g., /k/, /a/) (Reinfeldt et al., 2010).

Efforts to recreate a more 'natural' self-voice in experimental settings have relied on subjective filtering approaches, where participants adjust or rate the spectral profile of their voice until it matches their internal representation. However, these studies have yielded inconsistent results and high inter-individual variability (Kimura & Yotsumoto, 2018; Shuster & Durrant, 2003; Vurma, 2014), suggesting no universally valid filter exists (Kimura & Yotsumoto, 2018). Despite this, participants often rate filtered self-voices as more natural than unfiltered ones (Shuster & Durrant, 2003), and some neuroimaging studies have even adopted low-pass filtered stimuli to probe brain regions responsive to the self-voice (Kaplan et al., 2008).

A limitation of this spectral filtering approach is its narrow focus on power across frequency bands. A more fruitful strategy may lie in modeling self-voice acoustics using

parameters critical for speaker identity, such as fundamental frequency (F0) and formant dispersion – features rooted in the source-filter model of speech (Baumann & Belin, 2010; Chhabra et al., 2012; Kreiman, 2024). These parameters define a low-dimensional voice space, where perceptually similar voices are positioned closer together, and where acoustic distances correlate with both subjective discriminability (Baumann & Belin, 2010; Chhabra et al., 2012) and neural responses in voice-selective cortex (Latinus et al., 2013). Recent work shows that self-other voice discrimination accuracy scales with acoustic distance in this space (Orepic, Kannape, et al., 2023), indicating that the self-voice relies on the same low-level features that support general voice recognition. While altering these features may not precisely replicate the bone-conduction effect, their perceptual link to voice identity could make them more relevant for approximating the natural perception of one's own voice. These same acoustic dimensions seem to contribute to cognitive biases favoring self-related stimuli (e.g., self-prioritization effect; Kirk & Cunningham, 2025a), highlighting the broader perceptual and cognitive impact of self-voice acoustics.

Higher formants (F3 – F5) appear to carry a disproportionate weight in self-voice perception. Because these formants are more strongly constrained by stable anatomical properties (e.g., laryngeal cavity shape) and less influenced by speech or transient emotional and physiological states (López et al., 2013; Xu et al., 2013), they have been proposed as relatively invariant markers of self-voice identity. However, evidence for this claim, as well as for the precise contribution of any specific acoustic feature to self-voice perception, remains limited. Existing studies typically rely on small samples and show substantial inter-individual variability. Seemingly conflicting findings (e.g., reports of stable higher formants *vs.* unstable spectral filtering) may also stem from methodological differences across studies. For example, some work manipulates self-voice using graphical sliders to approximate bone-conduction acoustics (Kimura & Yotsumoto, 2018; Shuster & Durrant, 2003; Vurma, 2014), whereas other

studies assess behavioral responses to preselected acoustic features (Kirk & Cunningham, 2025; Orepic, Kannape, et al., 2023; Xu et al., 2013; see Supplementary Material for a more detailed discussion).

Motor Block

Unlike static physical features such as the face, the voice must be actively generated (McGettigan, 2015). Its perception is thus intimately tied to motor control. During vocalization, motor signals modulate auditory processing, leading to reduced perceptual salience and attenuated neural responses to self-generated voices compared to external ones (Jennifer & Georgia, 2015; Paraskevoudi & SanMiguel, 2021). This attenuation reflects a core mechanism of voice monitoring, which ensures congruence between intended and produced speech (Pinheiro, Schwartze, & Kotz, 2020).

The dominant theoretical model posits a forward model architecture, in which an efference copy of the motor command predicts the sensory consequences of vocalization such that discrepancies between predicted and actual feedback – prediction errors – trigger corrective motor adjustments (Miall & Wolpert, 1996; Wolpert, 1997). These predictions are thought to be relayed via the cerebellum (Knolle et al., 2012; Pinheiro, Schwartze, & Kotz, 2020; Todorović et al., 2024) to the right anterior superior temporal gyrus/sulcus (Johnson et al., 2021). The ventral precentral gyrus is a potential source of motor-related prediction in vocal production (Khalilian-Gourtani et al., 2024). Electrophysiological signatures of this mechanism include the N1 suppression in EEG (Pinheiro et al., 2018; Pinheiro, Schwartze, Amorim, et al., 2020), M1 suppression in MEG (Houde et al., 2002; Ventura et al., 2009), and reduced gamma-band activity in electrocorticography (Flinker et al., 2010; Towle et al., 2008). Readiness potentials preceding voluntary speech are linked to efference copy generation and

correlated with subsequent sensory suppression (Ford et al., 2014; Pinheiro, Schwartze, Gutiérrez-Domínguez, et al., 2020; Reznik et al., 2018).

Critically, motor-induced suppression is not limited to overt vocalization but occurs also when participants trigger self-voice playback via button press (Johnson et al., 2021; Pinheiro, Schwartze, Amorim, et al., 2020), even in the absence of precise temporal alignment (Orepic et al., 2021). This suggests that self-voice perception recruits broader motor prediction systems, not solely speech-specific mechanisms.

It is important to clarify that, in the context of self-voice perception, feedback mechanisms involve predictive *sensorimotor* processes rather than purely motor signals. Distinguishing between predictive motor signals and incoming sensory feedback highlights the complementary roles of the Motor and Multisensory Integration components: the Motor block reflects top-down predictions of expected vocal outcomes, whereas the Multisensory Integration block captures bottom-up sensory inputs, including tactile, vestibular, and proprioceptive cues, that are integrated to support coherent self-voice perception and vocal agency.

Motor contributions also support self-other voice discrimination. For example, suppression in the right anterior superior temporal cortex occurs only when voice feedback is identified as self-generated (Fu et al., 2006; Johnson et al., 2021). Perturbation studies further show that small, unexpected changes in pitch or intensity elicit compensatory responses, consistent with self-attribution (Burnett et al., 1998; Jones & Munhall, 2002; Natke et al., 2003). In contrast, large deviations often produce following responses, suggesting the voice is attributed to an external source (Burnett et al., 1998; Hain et al., 2000; Larson et al., 2007). Thus, predictive sensorimotor mechanisms help delineate the perceptual boundary between self and other voices.

Self-other distinction is not only theoretically significant but also clinically relevant. Reduced sensory attenuation is associated with anomalous self-experience in both clinical (Beño-Ruiz-de-la-Sierra et al., 2024; van der Weiden et al., 2015) and non-clinical (Duggirala et al., 2024; Pinheiro et al., 2018; Pinheiro, Schwartze, Amorim, et al., 2020) populations, including hallucination-prone individuals and patients with schizophrenia (see *Translational implications*). Such findings underscore the role of motor prediction in maintaining a coherent sense of vocal self.

Memory Block

Voice perception research suggests that speaker identity recognition depends on high-level voice identity representations stored in long-term memory (Andics et al., 2013; Latinus et al., 2013). Similarly, self-voice recognition draws on a memory-based template of one's own voice, shaped through long-term exposure to self-generated auditory feedback and sensorimotor coupling during speech production (Hickok et al., 2011). This internal self-voice representation supports self-other voice discrimination, even in the absence of efference-copy signals, likely via template-matching processes that compare voice input to stored representations. Comparable template-matching mechanisms have also been proposed for other-voice recognition (Lavner et al., 2001b; Maguinness et al., 2018), reinforcing the view that identity judgements arise from comparing incoming voice signals to long-term memory traces.

Electrophysiological evidence supports this notion. During self-other voice discrimination tasks, a self-voice-specific topographic EEG map appears around 350 ms post-voice onset, with hippocampal source localization, possibly reflecting the retrieval of the internal self-voice representation (Iannotti et al., 2022). Similarly, late event-related potentials (ERPs), such as the P3 (~300 ms) and N400 (~400 ms), differ reliably between self- and other-

voice conditions (Conde et al., 2015, 2016, 2018; Graux et al., 2013, 2015; Liu et al., 2019) (see Figure 2B and Table 2).

Neuroimaging studies contrasting responses to self- *vs.* other-voice stimuli are scarce, yet they indicate greater activation for self-voice in regions implicated in voice identity processing. These include the right (Kaplan et al., 2008; Nakamura et al., 2001) and left inferior frontal cortex (Allen et al., 2005), the parainsular cortex (Nakamura et al., 2001), and the right anterior cingulate cortex (Allen et al., 2005) (see Figure 2A and Table 1). Studies examining long-term voice representations have identified the right superior temporal cortex as sensitive to acoustic signal properties related to voice identity (Charest et al., 2013; Latinus et al., 2013). These studies did not directly compare self- with other-voice stimuli but provide complementary evidence for the neural coding of voice identity relevant to self-voice recognition. Consistently, the perceived similarity between a vocal sound and one's natural voice modulates activity in the bilateral superior and middle temporal gyri (Hosaka et al., 2021). Lesions to the right superior temporal gyrus have also been linked to altered self-voice perception (Andrade-Machado et al., 2023).

These findings suggest a functional hierarchy: the right superior temporal cortex encodes detailed acoustic features of voice identity, whereas the right inferior frontal cortex represents more abstract, cognitive-level voice identity representations (Andics et al., 2013; Bestelmeyer & Mühl, 2022). The latter may constitute a modality-independent self-identity hub, as it is activated by both self-voice and self-face stimuli (Kaplan et al., 2008). The right inferior frontal gyrus is also implicated in motor-feedback control (as in DIVA; Tourville & Guenther, 2011) and could serve as a convergence node where sensorimotor predictions both inform and are informed by self-voice identity representations.

However, the mechanisms through which a self-voice representation forms remain unclear. Norm-based coding offers one account, proposing that voice identity is encoded by its

deviation from an internal voice prototype (Latinus et al., 2013; Latinus & Belin, 2011). Adaptation studies support this framework: exposure to male voices biases perception of ambiguous stimuli toward female voices (Schweinberger et al., 2008), and repeated exposure to one identity increases the likelihood of hearing the opposite in morph continua (Latinus & Belin, 2011; Zäske et al., 2010). Voice-selective areas in the superior temporal cortex show greater activation for voices that acoustically deviate from the population average (Latinus et al., 2013). Importantly, the same acoustic features underlying prototype deviation also predict performance in self-other voice discrimination tasks (Orepic, Kannape, et al., 2023) (see *Auditory block*). Thus, individuals may have a central self-voice prototype, formed through long-term experience. Supporting this, anterior portions of the right superior temporal cortex are particularly responsive to deviations in both acoustic and expected properties of the self-voice (Hosaka et al., 2021). Finally, sensory attenuation effects suggest that this self-voice prototype informs forward models in speech production, with prediction errors arising when vocal outputs deviate from the internal template, guiding adjustments and refining self-voice representations. For instance, less prototypical voice outputs elicit reduced sensory attenuation, implying a prediction error from the internal voice template (Niziolek et al., 2013).

Multisensory Integration Block

Under natural speaking conditions, the self-voice is inherently multimodal. In addition to auditory input, self-voice perception is shaped by vibrotactile (Trulsson & Johansson, 2002) and vestibular (Welgampola et al., 2003) signals. Somatosensory feedback from articulator movement is essential for precise vocal control (Tremblay et al., 2003), and somatosensory perturbations can modulate the perception of speech sounds (Ito et al., 2009). Vibrotactile cues, when presented with auditory stimuli, enhance self-other voice discrimination performance (Orepic, Kannape, et al., 2023) and modulate neural responses to the self-voice (Iannotti et al.,

2022). For example, bone-conduction – which provides vibrotactile stimulation – improves self-other voice discrimination relative to air-conduction presentation (Orepic, Kannape, et al., 2023), likely due to added somatosensory cues rather than acoustic differences. However, the integration of sensory modalities during vocalization may vary across individuals. Some speakers rely more on somatosensory than on auditory feedback for articulating certain speech sounds (Lametti et al., 2012).

The multimodal nature of self-voice recognition may be rooted in bodily self-consciousness, the pre-reflective conscious experience of being a self in a body, which depends on the integration of exteroceptive (e.g., visual, auditory, tactile) and interoceptive (e.g., respiration and heartbeat) cues (Blanke et al., 2015; Park & Blanke, 2019). Multisensory conflict paradigms that disrupt bodily self-consciousness – such as mismatches between visual and tactile or interoceptive signals – can alter self-identification and self-location (Blanke et al., 2015). It is plausible, then, that the brain network underlying bodily self-consciousness also contributes to self-voice perception. This network includes the insula, premotor cortex, and cingulate cortex – regions implicated in both bodily self-consciousness (Blanke et al., 2015; Park & Blanke, 2019) and self-voice processing (Iannotti et al., 2022). These overlaps were observed in a recent study showing that multisensory self-voice cues from vocal-tract vibrations shape emotional prosody production and recruit insular, inferior frontal, and cerebellar regions (Selosse et al., 2025).

Behavioral findings provide additional support for the link between bodily self-consciousness and self-voice perception. Sensorimotor conflicts affecting the torso (a bodily self-consciousness hub) alter perceived loudness of self-voice stimuli, whereas similar manipulations on peripheral body parts do not (Orepic et al., 2021). Other experiments show that torso-based bodily self-consciousness manipulations bias participants to falsely attribute voices to self or other in a voice detection task (Orepic et al., 2024). Interoceptive signals

further modulate self-voice perception: self-other voice discrimination performance improves during inhalation relative to exhalation, but is not influenced by cardiac cycle phase (Orepic et al., 2022). This breathing-related modulation is itself amplified by sensorimotor conflicts that disrupt bodily self-consciousness, highlighting a complex interplay between self-voice, interoception, and sensorimotor processing (Orepic et al., 2022).

Self-Concept Block

Self-concept refers to the cognitive and social representation of the self, shaped through development and interpersonal experience (Crone et al., 2022). It comprises both the stability of self-knowledge and the coherence of one's identity over time (Crone et al., 2022). Remarkably, voices convey cues that listeners rapidly use to form impressions of speakers (Lavan et al., 2024) and vocal feedback may likewise shape the speaker's own self-representation over time.

Voice production is influenced by the identity one wishes to project. Speakers adjust vocal characteristics to convey attributes like authority, intelligence, or attractiveness (Guldner et al., 2024; Hughes et al., 2014). Speech patterns are also shaped by social categories and linked to gender roles and sexual identity, as exemplified by pitch modulation in gay men (Gaudio, 1994). When one's voice does not align with one's self-concept, emotional distress can result. For instance, transgender individuals frequently experience dissonance between their experienced gender and their perceived voice (Dacakis et al., 2017b), which can generate profound discomfort and impact mental health (Haskell, 1987) (see *Translational implications*).

Conversely, one's voice can shape one's self-concept. Auditory feedback from one's own voice informs self-perception, including inferences about personality traits (Stern et al., 2021). Persistent changes to the voice – such as those following laryngectomy (Bickford et al., 2013,

2019) or vocal fold paralysis (Francis et al., 2018) – can disrupt established self-representations and prompt compensatory adjustments in self-concept. Similarly, hormone therapy during gender transition is often accompanied by shifts in self-perception (Bultynck et al., 2017), suggesting that hormone-driven changes in vocal anatomy may interact with broader social and psychological transformations to reshape the self-voice and, in turn, the sense of self. However, these effects are difficult to disentangle from concurrent contextual and experiential factors inherent to the transition process.

Recent advances in voice manipulation technologies further underscore the bidirectional link between the acoustic properties of one's voice and self-representation, revealing how even transient alterations to self-voice feedback can modulate emotional, cognitive, and social dimensions of identity. For example, hearing a calmer version of one's voice can reduce anxiety, while deeper voices may enhance one's sense of power (Costa et al., 2018). Even impersonating an elderly voice can reduce negative social evaluations of the elderly (Arakawa et al., 2021). However, it is worth noting that most of these studies – including those using AI-based voice modifications – have tested short-term perceptual or behavioral effects in relatively small samples. Their findings should therefore be interpreted with caution.

The medial prefrontal cortex may serve as a neural substrate for the interaction between self-voice and self-concept. This region is central to self-referential processing (van der Meer et al., 2010) and supports abstract trait-based self-representations (Hu et al., 2016; Levorsen et al., 2023; Marquine et al., 2016). In vocal production, it is more active when speakers express socially relevant information (Guldner et al., 2020) and modulates auditory cortex responses to self-voice feedback (Müller et al., 2015), potentially preparing the auditory system for self-generated stimuli. The medial prefrontal cortex is also more engaged during active vocalization than passive listening, consistent with its role in self-awareness (Jardri et al., 2007) and self-identification (Iannotti et al., 2022).
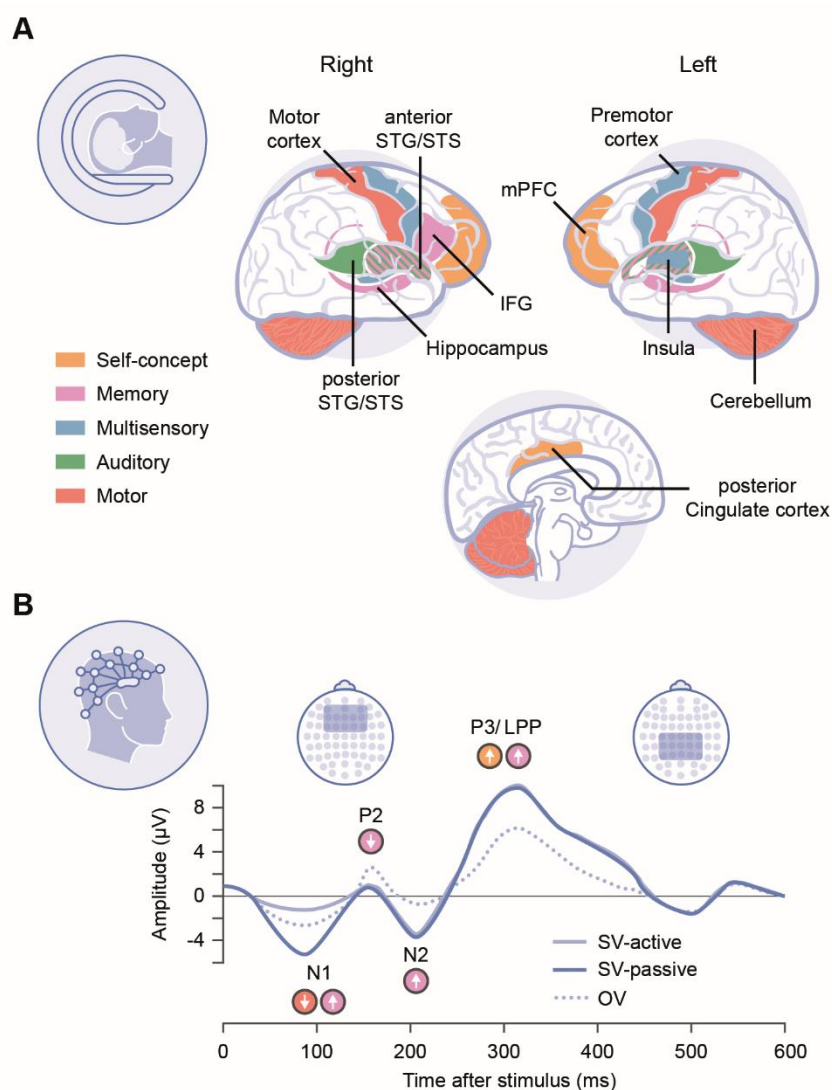
Figure 2. Neural correlates of self-voice processing according to the underlying building blocks, highlighting fMRI (A) and EEG (B) evidence differentiating self-voice (SV) from other voices (OV). Listening to one's own voice preferentially activates right-lateralized regions – including the inferior frontal gyrus, anterior cingulate, and superior temporal cortex – reflecting integration of auditory, motor, and memory signals that underpin self-specific processing. Panel B compares EEG responses to self-voice with explicit attention to identity. ERP studies show that the self-voice elicits early (N1, P2) and later (P3, LPP) neural responses that are stronger than those to familiar or unfamiliar voices. This indicates automatic attentional capture and sustained processing for self-voice stimuli, supporting its prioritized cognitive and affective status. LPP = Late Positive Potential.

Table 1. The Building Blocks of the Self-Voice: Neuroanatomical Correlates

Putative Building Block	Brain region	Functional role	Main Self-Voice Findings
Motor	Motor cortex	• Efference copy formation (Wang et al., 2014) • Control of orofacial movements (Kern et al., 2019) • Sensorimotor adaptation in speech production (Tang et al., 2021)	• Preparatory motor activity predicts sensory attenuation to self-voice feedback (Wang et al., 2014)
	Cerebellum	• Self-monitoring for speech errors (Runnqvist et al., 2016, 2021; Todorović et al., 2024) • Temporal predictions in speech processing (Zoefel et al., 2024)	• Increased activation to unexpected self-voice feedback (Runnqvist et al., 2016, 2021; Todorović et al., 2024)
Multisensory Integration	Insula	• Multisensory integration of exteroceptive and interoceptive signals during speech production facilitating self-monitoring (Kleber et al., 2013, 2017; Kurteff et al., 2024; Selosse et al., 2025; Simmons et al., 2013)	• Increased activation to self-voice feedback in normal vocal production (*vs.* whispering or articulation without acoustic feedback) (Selosse et al., 2025) • Increased activation to self-generated speech *vs.* externally generated speech (Nakamura et al., 2001; Woolnough et al., 2019) • Increased activation to self-dominant voice morphs (Iannotti et al., 2022) • Similar activation to self-generated and passively perceived self-voice (Kurteff et al., 2024)
Memory	Anterior STG/STS	• Segregation of self- and other voices in auditory feedback monitoring (Johnson et al., 2021) • Encoding of long-term acoustic centrality or acoustic representations of	• Reduced activation to self-generated (*vs.* externally generated or unexpectedly shifted self-generated) voice input (e.g., Kurteff et al., 2024))

		voice identity (Andics et al., 2013; Bestelmeyer & Mühl, 2022)	
	Hippocampus	• Autobiographical memory retrieval (Cabeza & St Jacques, 2007; Greenberg et al., 2005) • Self-referential processing (Kurczek et al., 2015)	• Activation specific to self-dominant voice morphs (Iannotti et al., 2022)
	Right IFG	• Multimodal representations of the physical self (Kaplan et al., 2008) • Self-recognition of different appearances of one's self across different periods of life (Apps et al., 2012) • Relay of salient signals to attention networks (Johnson et al., 2021)	• Increased activation to self-voice *vs.* familiar voice (Kaplan et al., 2008; Nakamura et al., 2001)
Self-Concept	Medial PFC	• Higher-order self-referential processes maintaining a cohesive representation of the self (e.g., autobiographical memories, attitudes, emotional evaluations of the self (Feng et al., 2018; Hu et al., 2016; Murray et al., 2015; Sugiura, 2015) • Integration of information about reflective aspects of the self with the processing and evaluation of perceptual representations (Morita et al., 2014) • Integration of psychological and perceptual aspects of the self (Azevedo et al., 2023) • Goal- and outcome-oriented self-referential trait processing in vocal production (Guldner et al., 2020)	• Enhanced communication between left AC and medial PFC (*vs.* baseline) when expecting a self-spoken voice; reduced communication when expecting a played-back voice (Müller et al., 2015)
	Cingulate cortex	• Self-referential processing (Herold et al., 2016; Northoff et al., 2006)	• Increased activation for self- *vs.* other voices (Allen et al., 2005)

Note: AC = auditory cortex; IFG = inferior frontal gyrus; PFC = prefrontal cortex; STG = superior temporal gyrus; STS = superior temporal sulcus.

Table 2. The Building Blocks of the Self-Voice: Electrophysiological Correlates

Putative Building Block	ERP component	Latency (post-stimulus onset)	Functional role	Experimental paradigm	Stimuli	Other-Voice Reference	Main Self-voice Findings
Motor	N1	~100 ms	Comparison of predicted and actual sensory feedback (Knolle et al., 2012)	Speak-listen task / Button-press task	Vowel /a/	Unfamiliar	• SV < OV (amplitude) (Duggirala et al., 2024; Knolle et al., 2019; Pinheiro et al., 2018; Pinheiro, Schwartze, Amorim, et al., 2020)
Memory	N1	~100 ms	Sensory registration of stimulus acoustic properties and automatic attention allocation (Giard et al., 1994)	Self-other voice discrimination	Positive, neutral, and negative adjectives	Unfamiliar	• SV > OV (amplitude) – only neutral (Pinheiro, Rezaii, Nestor, et al., 2016)
				Active oddball	Vowel /a/ and disyllabic words	Unfamiliar	• SV < OV (amplitude – standards) – words and vocalizations (Conde et al., 2018)
Memory	MMN	~170 ms	Pre-attentional detection of auditory deviance (Hofmann-Shen et al., 2020)	Passive oddball	Vowel /a/ and disyllabic words	Unfamiliar	• SV > OV (latency) – vocalizations and words (Conde et al., 2016)
Memory	P2	~200 ms	Early stimulus categorization and automatic attention	Self-other voice discrimination	Positive, neutral, and	Unfamiliar	• SV > OV (amplitude) – only positive (Pinheiro, Rezaii, Nestor, et al., 2016)

			(Crowley & Colrain, 2004); salience detection (Kotz & Paulmann, 2011)		negative adjectives		
	P2	~200 ms	Reduced prediction error (Summerfield et al., 2008)	Active oddball	Vowel /a/ and disyllabic words	Unfamiliar	• SV < OV (amplitude – standards) – words and vocalizations (Conde et al., 2018)
Motor	P2	~200 ms	Conscious detection of a self-initiated sound (Knolle et al., 2012)	Speak-listen task / Button-press task	Vowel /a/	Unfamiliar	• SV < OV (amplitude) (Knolle et al., 2019)
Memory	N2	~250 ms	Early discrimination and categorization of unexpected, deviant stimuli; error awareness (Kotz et al., 2014)	Active oddball	Vowel /a/ and disyllabic words	Unfamiliar	• SV > OV (amplitude) – words and vocalizations (Conde et al., 2018)
				Speak-listen task / Button-press task	Vowel /a/	Unfamiliar	• SV > OV (amplitude – deviants) (Knolle et al., 2019)
Memory	P3a	~300 ms	Involuntary orienting of attention towards unexpected, deviant stimuli (Escera et al., 1998), modulated by their personal	Passive oddball	Vowel /a/	Unfamiliar	• SV < OV (amplitude – deviants) (Graux et al., 2013)
				Passive oddball	Vowel /a/	Familiar	• SV < OV (amplitude) (Graux et al., 2015)
				Passive oddball	Vowel /a/ and	Unfamiliar	• SV < OV (amplitude – deviants) (Conde et al., 2016)

				significance (Roye et al., 2007)		disyllabic words		
Motor	P3a	~300 ms	Involuntary orienting of attention towards unexpected, deviant stimuli (Escera et al., 1998), modulated by their personal significance (Roye et al., 2007)	Button-press task	Vowel /a/	Unfamiliar	• SV > OV (amplitude – deviants) (Knolle et al., 2019)
Memory/ self-concept	P3b	~300 ms	Allocation of higher-order attentional resources to a task-relevant target stimulus after the cognitive evaluation of its significance (e.g., (Polich, 2007))	Active oddball	Vowel /a/ and disyllabic words	Unfamiliar	• SV > OV (amplitude) – only words (Conde et al., 2018)
	N400	~400 ms	Meaning processing and memory retrieval of self-relevant information (Coronel & Federmeier, 2016; Fields & Kuperberg, 2015)	Passive oddball	Names (self name; other name) uttered by SV and OV	Unfamiliar	• SV > OV – other name (Liu et al., 2019)
				Self-other voice discrimination	Positive, neutral, and	Unfamiliar	• SV > OV (amplitude) (Pinheiro et al., 2023)

						negative adjectives	
Memory/ self- concept	LPP	>400 ms	Sustained elaborative processing of self-relevant or emotionally salient stimuli (Hajcak & Foti, 2020)	Self-other voice discrimination	Positive, neutral, and negative adjectives	Unfamiliar	• SV > OV (amplitude) – only emotional (Pinheiro, Rezaii, Nestor, et al., 2016; Pinheiro et al., 2023)

Note: MMN = Mismatch Negativity; LPP = Late Positive Potential; SV = self-voice; OV = other voice.

Dynamic Interactions among the Building Blocks

While the five building blocks are presented separately for conceptual clarity, authentic self-voice perception arises from their dynamic interplay (Figure 3).

The *Auditory* and *Motor* blocks form a predictive loop that continuously monitors the match between intended and perceived vocal outcomes during real-time speech. When mismatches arise, *Motor*-based predictions modulate the *Auditory* block to maintain a stable sensorimotor representation of one's voice (arrow 1 in Figure 3). These predictions originate primarily from the Motor block but are also informed by the *Memory* block, which stores a long-term template of the self-voice (arrow 2). This relationship is reciprocal: persistent *Motor* mismatches can gradually update the *Memory* template, leading to revised predictions over time.

Multisensory Integration mechanisms combine auditory with other exteroceptive and interoceptive feedback to sustain a coherent sense of vocal ownership both during active speech and passive self-voice perception (arrow 4). These rapid multisensory interactions also contribute to *Memory* processes that store and update long-term representations of one's voice (arrow 5). In turn, stored representations shape *Motor*-based predictions during vocal control, forming a bidirectional recalibration loop (arrow 6).

At a higher-representational level, the *Self-Concept* block constrains and interprets these perceptual-motor signals, aligning them with one's beliefs about the self, stable identity traits (e.g., personality), and social identity dimensions (e.g., group membership), thereby maintaining coherence between how one sounds and how one believes oneself to be (arrows 7 and 8). Accordingly, shifts in self-concept – such as those emerging through developmental or clinical processes – can therefore propagate downward to influence perceptual and motor aspects of self-voice processing. Through its influence on *Auditory* and *Motor* components, the

Self-Concept also indirectly shapes the *Memory* block, helping maintain a self-voice template aligned with one's broader identity.

In sum, rather than additive, the five building blocks form a hierarchically organized yet reciprocally connected system that links low-level sensorimotor mechanisms with higher-order self-representations.
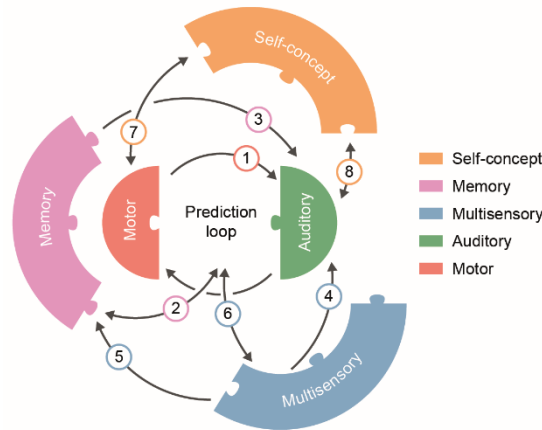

Figure 3. Interactions between the building blocks. 1: Motor-based predictions regulate *Auditory* self-voice perception during speech. 2: *Motor* predictions draw on the stored self-voice *Memory* template, and persistent prediction-perception mismatches gradually update this template. 3: The *Memory* template directly shapes *Auditory* self-voice perception. 4: *Multisensory Integration* influences how *Auditory* signals are combined with other exteroceptive and interoceptive cues, shaping the overall self-voice experience. 5: Sustained *Multisensory* alterations can reshape the *Memory* template. 6: *Multisensory* changes can disrupt the *Motor* prediction loop, and *Motor* predictions can also target non-auditory signals (e.g., somatosensory). 7: Shifts in *Self-Concept* can alter speaking style, and these vocal changes can in turn reinforce aspects of the *Self-Concept*. 8: *Self-Concept* influences self-voice perception, and persistent changes in perceived self-voice can gradually modify one's *Self-Concept*.

Translational Implications of Altered Self-Voice Perception

The self-voice plays a central role in communication, identity, and self-awareness. Disruptions to its perception or processing can therefore have wide-ranging consequences for daily functioning. Given its involvement in a broad spectrum of clinical conditions (see Table 3), a deeper understanding of self-voice perception is of immediate translational relevance. The building blocks framework introduced in this review offers a structured approach to identifying the specific neurocognitive components affected in these conditions, thereby informing more targeted and effective interventions.

Altered self-voice perception is particularly implicated in auditory verbal hallucinations, the experience of hearing voices in the absence of corresponding external auditory input. A prominent theory suggests that auditory verbal hallucinations arise from disruptions in self-monitoring (Frith et al., 1997), particularly in the feedforward sensorimotor mechanisms associated with self-generated speech (see *Motor block*). Supporting evidence includes reduced suppression of the N1 ERP component during vocalization in individuals who experience auditory verbal hallucinations, relative to passive listening to their own voice (Whitford, 2019), as well as altered frontotemporal connectivity (Allen et al., 2008) – both of which may contribute to confusion between self-generated and externally generated voices observed in patients (Pinheiro, Rezaii, Rauber, et al., 2016; Pinheiro et al., 2017; Vukojević et al., 2025; Waters et al., 2012).

Importantly, auditory verbal hallucinations are not confined to individuals with psychiatric or neurological disorders: they also occur in the general population, ranging from benign, transient experiences to clinically significant episodes (Van Os et al., 2009). Hallucination proneness has been linked to altered self-voice processing, including diminished sensory attenuation during vocalization (Mathalon et al., 2019; Pinheiro et al., 2018; Pinheiro, Schwartze, Amorim, et al., 2020) and a tendency to externalize the self-voice when facing perceptual ambiguity (Asai & Tanno, 2013; Pinheiro et al., 2019). These findings underscore the importance of self-voice research for understanding individual vulnerability to altered perceptual experiences and for developing earlier, more personalized intervention strategies.

Self-voice measures also hold promise in neurosurgical contexts, where patients often report concerns about feeling "like a different person" post-surgery. Recent findings suggest that self-other voice discrimination may serve as a sensitive biomarker for detecting pathological alterations in self-processing (Schaller et al., 2021; Voruz et al., 2024). For example, in one patient who developed borderline personality disorder after the resection of a

large meningioma, postoperative assessments revealed a striking inversion in self-other voice discrimination, present in both behavioral and EEG responses – paralleling the patient's disturbed self-concept (Orepic, Iannotti, et al., 2023). These methods may similarly shed light on conditions in which the self-voice changes persistently, such as following total laryngectomy (Bickford et al., 2013, 2019), unilateral vocal fold paralysis (Francis et al., 2018), expressive aphasia (Shadden, 2005), foreign accent syndrome (DiLollo et al., 2014), or gender affirming hormone therapy (Bultynck et al., 2017). Because the self-voice emerges from the interaction of multiple neurocognitive building blocks, these conditions likely involve complex disruptions beyond the auditory modality – affecting sensorimotor integration, memory, and self-concept (see Table 3).

Beyond clinical contexts, the study of the self-voice is increasingly relevant in technology-mediated communication. Emerging technologies now create diverse situations in which the coupling between the natural and digital self-voice is altered – through recorded voices, real-time transformed feedback, and synthetic self-voice clones – each of which engages different components of self-voice processing. Hearing one's recorded voice involves a purely auditory encounter with the self: it activates stored memory representations of one's voice but lacks the motor and somatosensory feedback components typically available during speech, often eliciting feelings of estrangement or surprise (Gur & Sackeim, 1979). In contrast, real-time transformed feedback – such as pitch-shifted or AI-converted voice during speaking – preserves the sensorimotor structure of vocal production yet disrupts the alignment between predicted and perceived auditory outcomes. These manipulations selectively perturb motor-based prediction mechanisms that underlie vocal agency. Finally, synthetic self-voices, including voice clones and deepfakes, present an additional challenge: they reproduce the acoustic signature of the speaker without any bodily origin or motor involvement, thereby challenging the sense of authenticity and ownership, and the anchoring of self-concept in one's

own vocal production. Differentiating among these digital self-voice experiences may help isolate differential effects of emerging technologies on the cognitive and affective boundaries of the self in everyday communication.

Together, these clinical and technological contexts highlight the broad relevance of self-voice perception. As our voices increasingly extend into virtual environments, and as health technologies become more personalized, understanding how the self-voice shapes cognition, emotion, and social behavior becomes both timely and essential (Pinheiro, 2025).

Table 3. Conceptual mapping between self-voice building blocks and clinical contexts

Primary Building Block Affected	Clinical condition	Illustrative findings	Empirical Strength (Single case, Weak Empirical, Moderate Empirical)	Expected Cascade Across Other Building Blocks
Auditory	Hearing loss	• Self-voice dissatisfaction and perceptual changes (Hengen et al., 2020)	WE	Reduced fidelity of auditory input may disrupt sensori**motor** feedback calibration, impair **multisensory integration** between bone- and air-conducted cues, and weaken **memory**-based templates for self-voice identity. Over time, these effects may reshape the **self-concept** via reduced vocal familiarity.
	Cochlear implant adaptation	• Voice recognition impairment (Colby & Orena, 2022) • Impaired voice production quality (dysphonia) (Guastamacchia et al., 2024) • Decreased voice-related quality of life (Bottalico et al., 2023)	ME	
	Dental prosthesis	• Change in voice parameters and in articulation (Hamlet et al., 1976; Seifert et al., 1999)	WE	
	Congenital adrenal hyperplasia and voice endocrinopathies	• Virilization of women's voices with functional impact (Nygren et al., 2009; Stogowska et al., 2022) • Change in subjective and objective voice parameters (Stogowska et al., 2022)	WE	
	Hypogonadotropic hypogonadism	• Changes in mean fundamental frequency after testosterone treatment (Akcam et al., 2004)	WE	
Motor	Laryngectomy (voice loss / prosthetic voice)	• Altered objective and subjective voice outcomes (Botti et al., 2024; Migliorelli et al., 2024; Van Sluis et al., 2018)	WE	Impaired motor predictions degrade **auditory** error monitoring and **multisensory** calibration between motor commands and resulting feedback. **Memory representations** of one's own voice may shift as the vocal signal changes, influencing **self-concept** over time.
	Amyotrophic lateral sclerosis	• Altered voice characteristics (Ramig et al., 1990) • Impaired personalized automatic speech recognition (Cave, 2024)	ME	
	Parkinson's disease	• Deficit in self-voice perception and perceived voice loudness (Contreras-Ruston et al., 2024) • Deficits in perceived loudness and production of vocal emotions (Kwan & Whitehill, 2011)	ME	

		• Differences in mean voice sound level and lower scores on semantic verbal fluency (Steurer et al., 2022)		
	Unilateral vocal fold paralysis (UVFP)	• Differences in acoustic voice features (Oguz et al., 2007) • Affected voice outcome measures (Walton et al., 2019) • Changes in brain adaptation and aerodynamic voice measures (Dedry et al., 2022)	WE	
	Foreign accent syndrome (FAS)	• Changed speech patterns (Dankovičová & Hunt, 2011; Keulen et al., 2016)	SC	
	Functional voice disorder (e.g., muscle tension dysphonia)	• Subjective voice effort, fatigue, and perceived alterations in the self-voice (Chung et al., 2018; Reetz et al., 2019)	WE	
	Auditory-verbal hallucinations	• Impaired sensorimotor predictions in speech production (Allen et al., 2008; Whitford, 2019)	ME	
	Chronic respiratory disorders	• Acoustic voice alterations (Elshebl et al., 2025; Hurtado-Ruzza et al., 2021)	WE	
Multisensory Integration	Functional voice disorder	• Reduced capabilities in processing of body sensations (Smeltzer et al., 2025)	WE	Reduced integration of auditory and somatosensory cues disrupts feedback monitoring and prediction. This may weaken the updating of **memory** traces for self-voice and lead to instability in **self-concept** coherence.
	Dental prosthesis	• Adaptations in voice production due to altered auditory and somatosensory feedback (Barbier et al., 2020; Hamlet & Stone, 1978)	ME	
Memory	Brain lesions, in particular temporal lobe damage	• SOVD impairments (Candini et al., 2018; Orepic, Iannotti, et al., 2023; Voruz et al., 2024)	SC, WE	Degraded voice representations reduce auditory discrimination between self and other voice. They can further impair predictive **motor** control and distort continuity of the **self-concept**.
	Auditory-verbal Hallucinations	• Impaired source monitoring (self *vs.* other) across paradigms, including SOVD (Pinheiro et al., 2017, 2019; Pinheiro, Rezaii, Rauber, et al., 2016; Vukojević et al., 2025; Waters et al., 2012)	ME	
	Autism spectrum disorder	• SOVD impairement (Chakraborty & Chakrabarti, 2015)	WE	
Self-Concept	Gender-affirming hormone therapy	• Changes in self-voice perception (Bultynck et al., 2017)	WE	Changes in psychological self-representation may influence how

	• Associations between self-report of voice femininity and acoustic voice measures (Dacakis et al., 2017a)		**auditory** and **multisensory** cues are weighted during self-voice perception. **Motor** adaptation may occur in attempts to align output with a revised self-concept, reinforcing or destabilizing **memory** representations of one's voice.
Gender dysphoria	• Self-voice as the second most common contributor (Pu et al., 2025)	WE	
Laryngectomy	• Subjective experience of identity change (Bickford et al., 2013, 2019)	WE	
Unilateral vocal fold paralysis	• Affected sense of identity (Francis et al., 2018)	WE	
Foreign accent syndrome (FAS)	• Subjective changes in personal and social identity (DiLollo et al., 2014)	SC	

Note: The table summarizes proposed correspondences between each self-voice building block and selected clinical contexts. Rather than implying isolated one-to-one associations, it points to existing literature, assesses the robustness of evidence, and highlights how disruptions in a primary building block may cascade through others, offering a generative map for hypothesis-driven translational research. Although the cited evidence is representative rather than exhaustive, and mostly limited due to small sample size, it may help guide future clinical studies. SOVD = self-other voice discrimination.

Individual, Cultural, and Developmental Modulators

While the five building blocks provide a useful conceptual structure for understanding self-voice processing, their operation is shaped by contextual, individual, and developmental factors that merit further investigation.

Individual factors such as how self-relevant information is processed (Fenigstein, 1984; Hull et al., 1988) or prior vocal training (Fuchs et al., 2009) may affect the relative weighting and interaction of the components. For example, experienced singers might develop more robust auditory-motor mappings that enhance self-monitoring precision (Jones & Keough, 2008).

Cultural factors are known to affect the sense of self (Chiao et al., 2009, 2010) and likely also shape self-voice processing. In collectivist cultures, where self-concept is primarily relational (Kashima et al., 1995), individuals may rely more on auditory cues reflecting how one's voice is perceived by others. In individualistic cultures, where self-concept emphasizes internal attributes and personal agency (Kashima et al., 1995), memory-based representations of one's own voice may have a greater influence in maintaining a stable sense of one's own voice. Such cultural differences would suggest that the relative weighting of the building blocks can shift depending on social and personal factors.

Developmental dynamics provide another important dimension. Sensitivity to self-voice cues and self-other discrimination evolves across the lifespan, with adolescence representing a particularly salient developmental window (Pinheiro et al., 2024). During puberty, rapid anatomical changes in the larynx and vocal folds substantially alter voice pitch, timbre, and resonance (Hollien et al., 1994), requiring the updating of internal self-voice representations. At the same time, motor control processes adapt to produce stable vocal output despite anatomical changes. These acoustic shifts also engage multisensory integration processes, as

adolescents must reconcile auditory feedback with proprioceptive and bone-conduction signals. Concurrently, heightened social evaluation and identity formation (Crone et al., 2022) amplify the relevance of the self-voice to self-concept. Together, these interactions highlight adolescence as a sensitive period for self-voice processing, in which the auditory, motor, memory, multisensory integration, and self-concept building blocks are dynamically recalibrated (Pinheiro et al., 2024).

Current empirical constraints nevertheless limit the temporal and causal scope of the existing evidence. Most experimental manipulations of the self-voice – including AI-based voice modifications – measure short-term perceptual or behavioral adjustments. Although clinical case reports (e.g., following laryngectomy or voice therapy) suggest longer-term effects, controlled longitudinal data remain scarce. Moreover, disruptions within one building block may cascade across others (see Table 3), yet these interdependencies are rarely tested systematically. Future work should thus examine how individual, cultural, and developmental factors interact to modulate these dynamics, refining the applicability of the proposed framework across contexts and populations. Finally, the multifaceted nature of self-voice processing makes selecting an appropriate experimental approach inherently challenging. Apparent discrepancies across studies often arise from differences in task design (see Supplementary Material). To guide future research, Table 4 summarizes commonly used experimental approaches – with example studies spanning behavioral, EEG, and fMRI/PET methods – organized around three key contrasts: natural *vs.* digital self-voice, active *vs.* passive processing, and implicit *vs.* explicit tasks.

Table 4: Experimental approaches to studying the self-voice

Putative Building Block	Task	Description	Self-voice		Processing		Self-voice reference		Outcome measure			Example studies
			Natural	Digital	Active	Passive	Implicit	Explicit	Behavioral	EEG	fMRI/PET	
Auditory	Same-different task	Participants listen to pairs of voices (including the SV and/or OV) and decide if it is the same or two different voices		X		X	X		Accuracy	--	--	(Candini et al., 2014; Pinheiro et al., 2019)
	Similarity task	Participants rate similarity between two voice recordings		X		X		X	Perceived similarity rating	--	--	(Rosi et al., 2025)
	Self-voice equalization task	Participants move sliders on a GUI equalizer to transform digital into natural self-voice	X	X		X		X	Perceived similarity rating	--	--	(Kimura & Yotsumoto, 2018; Shuster & Durrant, 2003; Vurma, 2014)
Motor	Talk-listen	Participants talk (e.g., vowel /a/) and listen to SV feedback or passively listen to recorded SV	X	X	X	X	X		--	N1 [a]	BOLD contrasts (e.g., auditory cortex) [a]	(Perez et al., 2012)
	Talk-listen with auditory perturbation	Participants talk (e.g., vowel /a/) and listen to SV feedback – in some trials, sensory	X		X		X		Compensatory responses in vocal production	N1 [a]	BOLD contrasts (e.g., auditory cortex) a	(Aucouturier et al., 2016; Franken et al., 2021) (Behroozmand et

		feedback is unexpectedly changed									al., 2015; Fu et al., 2006; Guo et al., 2016; Tourville et al., 2008; Toyomura et al., 2007)
	Button-press	Participants press a button that elicits a vocal sound (e.g., vowel /a/) or passively listen to recorded SV	X	X	X	X		--	N1 [a]	BOLD contrasts (e.g., auditory cortex) [a]	(Duggirala et al., 2024; Pinheiro et al., 2018; Pinheiro, Schwartze, Amorim, et al., 2020)
Memory	Passive oddball	Participants watch a movie while standard and deviant vocal sounds (SV, OV) are presented	X		X	X		--	MMN	--	(Conde et al., 2016; Graux et al., 2013; Rachman et al., 2019)
	Active oddball	Participants count the number of oddball vocal sounds (SV, OV)	X		X	X		--	P3	--	(Conde et al., 2015, 2018)
	SOVD (with or without voice morphing)	Participants assess whether a vocal sound is SV or OV	X		X		X	Accuracy	EEG topographic maps; N1; P2; LPP	--	(Candini et al., 2014; Pinheiro et al., 2019)(Orepic et al., 2023; Rosa et al., 2008) (Iannotti et al., 2022)

Category	Task	Description						Measure			Reference
	Emotion categorization task with SV and OV	Participants assess whether a word is neutral or emotional	X		X	X		--	N1; P2; LPP	--	(Pinheiro et al., 2023)
	Voice-face detection task	Participants assess whether a voice is SV or FV with simultaneous self-face or other face presentation	X		X		X	Accuracy and response time	--	--	(Hughes & Nicholson, 2010)
Multisensory integration	SOVD with concurrent recording of breathing, heartbeat signals, or galvanic shin response	Participants assess whether a vocal sound is SV or OV	X		X		X	Accuracy relative to the respiratory and cardiac phase or the amplitude of the galvanic skin response	--	--	(Orepic et al., 2022) (Gur & Sackeim, 1979)
	SOVD with bone-conducted vibrotactile stimulation	Participants assess whether a vocal sound is SV or OV	X		X		X	Accuracy relative to the type of stimulation	--	--	(Iannotti et al., 2022; Orepic et al 2023)
	Voice detection with concurrent robotic sensorimotor stimulation	Participants report whether they heard a voice embedded in noise, SV and OV are presented at hearing thresholds	X		X	X		False alarms relative to the sensorimotor stimulation	--	--	(Orepic et al., 2024)

Category	Task	Description							Behavioral			Neuroimaging	Reference
	Audio-visual speech integration (McGurk effect)	Participants watch a self-face or other-face while hearing matched or mismatched SV or OV speech sounds		X		X	X		Accuracy	--	--		(Aruffo & Shore, 2012)
	Emotional word production with varying degrees of vocal tract vibration	Participants produce angry, happy, and neutral vocalizations by speaking regularly, whispering or silently articulating	X		X		X		--	--		BOLD contrasts	(Selosse et al., 2025)
Self-concept	Social vocal control task	Participants modulate their own voice to express social traits (e.g., affiliation and competence) or modulate body size	X		X		X		--	--		BOLD contrasts (e.g., medial prefrontal cortex)	(Guldner et al., 2020)
	Trait rating task	Participants rate the intensity of a specific trait (e.g., attractiveness or dominance)		X		X		X	Perceived intensity (e.g., from 1 to 100)	--	--		(Rosi et al., 2025)
	Self-prioritization task	Participants attribute identity labels (self, friend, stranger) to voice stimuli		X		X		X	Response time	--	--		(Kirk & Cunningham, 2025; Payne et al., 2021; Rosi et al., 2025)

Note: BOLD = Blood-Oxygen-Level-Dependent) Signal; LPP = Late Positive Potential; MMN = Mismatch Negativity; SOVD = self-other voice discrimination; SV = self-voice; OV = other voice. [a] Sensory attenuation.

Concluding Remarks

Understanding how individuals develop, represent, and experience a vocal self is central to the broader science of self-consciousness – a topic that has long captivated disciplines ranging from neuroscience and psychology to computational sciences and philosophy. This review proposed a novel framework for deconstructing the building blocks of self-voice perception, encompassing auditory, motor, memory, multisensory integration, and self-concept components. By integrating these foundational processes, the framework offers a transdisciplinary lens through which to explore the cognitive and neural underpinnings of the auditory self.

Importantly, this approach not only synthesizes fragmented findings across domains but also opens new research avenues at the intersection of psychiatry, developmental science, and artificial intelligence (see Table 5). For instance, as voice technologies become increasingly pervasive, future studies should investigate how widespread exposure to synthetic, manipulated, or AI-generated voices may shape internal representations of the self-voice. These questions are especially pressing during sensitive developmental windows, such as puberty, when self-other voice distinction may be particularly malleable (Pinheiro et al., 2024).

In addition to addressing risk and resilience in mental health, the framework may inform the design of interventions and technologies that more closely align with individuals' vocal identity – particularly for those who have lost their natural voice (e.g., due to motor neuron disease, Cave & Bloch, 2021) or who seek voice congruence through gender-affirming therapies.

Ultimately, advancing knowledge of the neurocognitive mechanisms that support the vocal self brings us closer to resolving one of the most intimate paradoxes of human experience: how the voice we produce becomes the voice we hear in the 'mind's ear'. Bridging the gap between natural and digital self-voice is not only a technological challenge – it is a

fundamentally cognitive and emotional one, with wide-reaching implications for identity, agency, and self-awareness in today's digital era.

Table 5. Outstanding Questions

Building block	Future Directions
Auditory	• Which acoustic cues are most effective in making digital self-voice feel natural to the listener? • How stable are those acoustic cues across individuals and different cultures?
Motor	• How do sensorimotor predictions for self-generated vocalizations develop across the lifespan? • Can motor learning interventions enhance self-other voice discrimination in clinical populations?
Memory	• What is the efficacy of self-other voice discrimination as a biomarker for an altered sense of self across different neurological and psychiatric disorders? • How does self-voice representation change across age? • How is self-voice representation affected by sensorimotor prediction errors? • How can we leverage self-voice recognition in clinical contexts to support individuals with disturbed self-other distinction?
Multisensory integration	• What are the neural mechanisms supporting the relationship between self-voice perception and bodily self-consciousness? • What are the underlying neural mechanisms underlying the integration of bone- and air-conducted self-voice components? • How do interoceptive processes occurring during voice production affect one's sense of self? • How do individual and cultural differences affect the weighting of sensory inputs in self-voice integration?

Self-concept	• In what ways do changes in self-voice perception influence social behavior or decision-making in virtual environments?
	• How do voice changes across the lifespan shape short-term and potentially long-term self-concept?
	• What are the ethical implications of manipulating or simulating self-voice in AI-generated content?

Notes

Declaration of Conflicting Interests

The authors declare that there are no conflicts of interest with respect to the authorship or the publication of this article.

References

Adelman, C., Cohen, A., Regev-Cohen, A., Chordekar, S., Fraenkel, R., & Sohmer, H. (2015). Air conduction, bone conduction, and soft tissue conduction audiograms in normal hearing and simulated hearing losses. *Journal of the American Academy of Audiology*, *26*(1), 101–108. https://doi.org/10.3766/jaaa.26.1.11

Akcam, T., Bolu, E., Merati, A. L., Durmus, C., Gerek, M., & Ozkaptan, Y. (2004). Voice changes after androgen therapy for hypogonadotrophic hypogonadism. *The Laryngoscope*, *114*(9), 1587–1591. https://doi.org/10.1097/00005537-200409000-00016

Allen, P., Amaro, E., Fu, C. H. Y., Williams, S. C. R., Brammer, M., Johns, L. C., & McGuire, P. K. (2005). Neural correlates of the misattribution of self-generated speech. *Human Brain Mapping*, *26*(1), 44–53. https://doi.org/10.1002/hbm.20120

Allen, P., Larøi, F., McGuire, P. K., & Aleman, A. (2008). The hallucinating brain: A review of structural and functional neuroimaging studies of hallucinations. *Neuroscience and Biobehavioral Reviews*, *32*(1), 175–191. https://doi.org/10.1016/j.neubiorev.2007.07.012

Andics, A., McQueen, J. M., & Petersson, K. M. (2013). Mean-based neural coding of voices. *NeuroImage*, *79*, 351–360. https://doi.org/10.1016/j.neuroimage.2013.05.002

Andrade-Machado, R., Javarayee, P., Koop, J. I., Farias-Moeller, R., Kim, I., & Lew, S. M. (2023). Neural representations of self-perception of voice: An intracortical evoked potential analysis based on an adolescent with right temporal lobe epilepsy. *Seizure: European Journal of Epilepsy*, *109*, 1–4. https://doi.org/10.1016/j.seizure.2023.04.003

Apps, M. A. J., Tajadura-Jiménez, A., Turley, G., & Tsakiris, M. (2012). The different faces of one's self: An fMRI study into the recognition of current and past self-facial appearances. *NeuroImage*, *63*(3), 1720–1729. https://doi.org/10.1016/j.neuroimage.2012.08.053

Arakawa, R., Kashino, Z., Takamichi, S., Verhulst, A., & Inami, M. (2021). Digital speech makeup: Voice conversion based altered auditory feedback for transforming self-representation. *Proceedings of the 2021 International Conference on Multimodal Interaction*, 159–167. https://doi.org/10.1145/3462244.3479934

Aruffo, C., & Shore, D. I. (2012). Can you McGurk yourself? Self-face and self-voice in audiovisual speech. *Psychonomic Bulletin & Review*, *19*(1), 66–72. https://doi.org/10.3758/s13423-011-0176-8

Asai, T., & Tanno, Y. (2013). Why must we attribute our own action to ourselves? Auditory hallucination like-experiences as the results both from the explicit self-other attribution and implicit regulation in speech. *Psychiatry Research*, *207*(3), 179–188. https://doi.org/10.1016/j.psychres.2012.09.055

Aucouturier, J.-J., Johansson, P., Hall, L., Segnini, R., Mercadié, L., & Watanabe, K. (2016). Covert digital manipulation of vocal emotion alter speakers' emotional states in a congruent direction. *Proceedings of the National Academy of Sciences*, *113*(4), 948–953. https://doi.org/10.1073/pnas.1506552113

Azevedo, R. T., Diaz-Siso, J. R., Alfonso, A. R., Ramly, E. P., Kantar, R. S., Berman, Z. P., Diep, G. K., Rifkin, W. J., Rodriguez, E. D., & Tsakiris, M. (2023). Recognizing the new self: The neurocognitive plasticity of self-processing following facial transplantation. *Proceedings of the National Academy of Sciences*, *120*(14), e2211966120. https://doi.org/10.1073/pnas.2211966120

Barbier, G., Baum, S. R., Ménard, L., & Shiller, D. M. (2020). Sensorimotor adaptation across the speech production workspace in response to a palatal perturbation. *The Journal of the Acoustical Society of America*, *147*(2), 1163–1178. https://doi.org/10.1121/10.0000672

Bauer, S. M., Schanda, H., Karakula, H., Olajossy-Hilkesberger, L., Rudaleviciene, P., Okribelashvili, N., Chaudhry, H. R., Idemudia, S. E., Gscheider, S., Ritter, K., & Stompe, T. (2011). Culture and the prevalence of hallucinations in schizophrenia. *Comprehensive Psychiatry*, *52*(3), 319–325. https://doi.org/10.1016/j.comppsych.2010.06.008

Baumann, O., & Belin, P. (2010). Perceptual scaling of voice identity: Common dimensions for different vowels and speakers. *Psychological Research*, *74*(1), 110–120. https://doi.org/10.1007/s00426-008-0185-z

Belin, P., Bestelmeyer, P. E. G., Latinus, M., & Watson, R. (2011). Understanding voice perception. *British Journal of Psychology*, *102*(4), 711–725. https://doi.org/10.1111/j.2044-8295.2011.02041.x

Belin, P., & Kawahara, H. (2025). STRAIGHTMORPH: A voice morphing tool for research in voice communication sciences. *Open Research Europe*, *4*(154). https://doi.org/10.12688/openreseurope.18055.2

Beño-Ruiz-de-la-Sierra, R. M., Arjona-Valladares, A., Hernández-García, M., Fernández-Linsenbarth, I., Díez, Á., Fondevila Estevez, S., Castaño, C., Muñoz, F., Sanz-Fuentenebro, J., Roig-Herrero, A., & Molina, V. (2024). Corollary discharge dysfunction as a possible substrate of anomalous self-experiences in schizophrenia. *Schizophrenia Bulletin*, *50*(5), 1137–1146. https://doi.org/10.1093/schbul/sbad157

Bestelmeyer, P. E. G., & Mühl, C. (2022). Neural dissociation of the acoustic and cognitive representation of voice identity. *NeuroImage*, *263*, 119647. https://doi.org/10.1016/j.neuroimage.2022.119647

Bickford, J., Coveney, J., Baker, J., & Hersh, D. (2013). Living with the altered self: A qualitative study of life after total laryngectomy. *International Journal of Speech-Language Pathology*, *15*(3), 324–333. https://doi.org/10.3109/17549507.2013.785591

Bickford, J., Coveney, J., Baker, J., & Hersh, D. (2019). Validating the changes to self-identity after total laryngectomy. *Cancer Nursing*, *42*(4), 314–322.

Blanke, O., Slater, M., & Serino, A. (2015). Behavioral, neural, and computational principles of bodily self-consciousness. *Neuron*, *88*(1), 145–166. https://doi.org/10.1016/j.neuron.2015.09.029

Bortolon, C., & Raffard, S. (2018). Self-face advantage over familiar and unfamiliar faces: A three-level meta-analytic approach. *Psychonomic Bulletin & Review*, *25*(4), 1287–1300. https://doi.org/10.3758/s13423-018-1487-9

Bottalico, P., Plachno, A., & Nudelman, C. (2023). Self-reported voice-related quality of life in cochlear implant users. *Revista de Investigación e Innovación En Ciencias de La Salud*, *5*(2), 69–92. https://doi.org/10.46634/riics.232

Botti, C., Lotto, C., Tesauro, P., Guidotti, M., Borghi, A., Molteni, G., Presutti, L., & Fernandez, I. J. (2024). Voice quality after open partial horizontal laryngectomy vs. total laryngectomy with voice prosthesis: A comparative study. *European Archives of Oto-Rhino-Laryngology*, *281*(9), 4897–4902. https://doi.org/10.1007/s00405-024-08735-5

Bourguignon, N. J., Baum, S. R., & Shiller, D. M. (2016). Please say what this word is—Vowel-extrinsic normalization in the sensorimotor control of speech. *Journal of Experimental Psychology: Human Perception and Performance*, *42*(7), 1039–1047. https://doi.org/10.1037/xhp0000209

Bradshaw, A. R., Wheeler, E. D., McGettigan, C., & Lametti, D. R. (2025). Sensorimotor learning during synchronous speech is modulated by the acoustics of the other voice. *Psychonomic Bulletin & Review*, *32*(1), 306–316. https://doi.org/10.3758/s13423-024-02536-x

Bruce, V., & Young, A. (1986). Understanding face recognition. *British Journal of Psychology*, *77*(3), 305–327. https://doi.org/10.1111/j.2044-8295.1986.tb02199.x

Bultynck, C., Pas, C., Defreyne, J., Cosyns, M., den Heijer, M., & T'Sjoen, G. (2017). Self-perception of voice in transgender persons during cross-sex hormone therapy. *The Laryngoscope*, *127*(12), 2796–2804. https://doi.org/10.1002/lary.26716

Burnett, T. A., Freedland, M. B., Larson, C. R., & Hain, T. C. (1998). Voice F0 responses to manipulations in pitch feedback. *The Journal of the Acoustical Society of America*, *103*(6), 3153–3161. https://doi.org/10.1121/1.423073

Cabeza, R., & St Jacques, P. (2007). Functional neuroimaging of autobiographical memory. *Trends in Cognitive Sciences*, *11*(5), 219–227. https://doi.org/10.1016/j.tics.2007.02.005

Candini, M., Avanzi, S., Cantagallo, A., Zangoli, M. G., Benassi, M., Querzani, P., Lotti, E. M., Iachini, T., & Frassinetti, F. (2018). The lost ability to distinguish between self and other voice following a brain lesion. *NeuroImage: Clinical*, *18*, 903–911. https://doi.org/10.1016/j.nicl.2018.03.021

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*, 112–117. https://doi.org/10.1016/j.bandc.2014.10.001

Caucheteux, C., Gramfort, A., & King, J.-R. (2023). Evidence of a predictive coding hierarchy in the human brain listening to speech. *Nature Human Behaviour*, *7*(3), 430–441. https://doi.org/10.1038/s41562-022-01516-2

Cave, R. (2024). How people living with amyotrophic lateral sclerosis use personalized automatic speech recognition technology to support communication. *Journal of Speech, Language, and Hearing Research*, *67*(11), 4186–4202. https://doi.org/10.1044/2024_JSLHR-24-00097

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), 116–129. https://doi.org/10.1111/1460-6984.12588

Chakraborty, A., & Chakrabarti, B. (2015). Is it me? Self-recognition bias across sensory modalities and its relationship to autistic traits. *Molecular Autism*, *6*(1), 20. https://doi.org/10.1186/s13229-015-0016-1

Charest, I., Pernet, C., Latinus, M., Crabbe, F., & Belin, P. (2013). Cerebral processing of voice gender studied using a continuous carryover fMRI design. *Cerebral Cortex*, *23*(4), 958–966. https://doi.org/10.1093/cercor/bhs090

Cheung, S., & Babel, M. (2022). The own-voice benefit for word recognition in early bilinguals. *Frontiers in Psychology*, *13*, 901326.

Chhabra, S., Badcock, J. C., Maybery, M. T., & Leung, D. (2012). Voice identity discrimination in schizophrenia. *Neuropsychologia*, *50*(12), 2730–2735. https://doi.org/10.1016/j.neuropsychologia.2012.08.006

Chiao, J. Y., Harada, T., Komeda, H., Li, Z., Mano, Y., Saito, D., Parrish, T. B., Sadato, N., & Iidaka, T. (2009). Neural basis of individualistic and collectivistic views of self. *Human Brain Mapping*, *30*(9), 2813–2820. https://doi.org/10.1002/hbm.20707

Chiao, J. Y., Harada, T., Komeda, H., Li, Z., Mano, Y., Saito, D., Parrish, T. B., Sadato, N., & Iidaka, T. (2010). Dynamic cultural influences on neural representations of the self. *Journal of Cognitive Neuroscience*, *22*(1), 1–11. https://doi.org/10.1162/jocn.2009.21192

Chong, H. J., Choi, J. H., & Lee, S. S. (2024). Does the perception of own voice affect our behavior? *Journal of Voice*, *38*(5), 1249.e19-1249.e28. https://doi.org/10.1016/j.jvoice.2022.02.003

Chung, D. S., Wettroth, C., Hallett, M., & Maurer, C. W. (2018). Functional speech and voice disorders: Case series and literature review. *Movement Disorders Clinical Practice*, *5*(3), 312–316. https://doi.org/10.1002/mdc3.12609

Colby, S., & Orena, A. J. (2022). Recognizing voices through a cochlear implant: A systematic review of voice perception, talker discrimination, and talker identification. *Journal of Speech, Language, and Hearing Research*, *65*(8), 3165–3194. https://doi.org/10.1044/2022_JSLHR-21-00209

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

Conde, T., Gonçalves, Ó. F., & Pinheiro, A. P. (2016). The effects of stimulus complexity on the preattentive processing of self-generated and nonself voices: An ERP study. *Cognitive, Affective and Behavioral Neuroscience*, *16*, 106–123. https://doi.org/10.3758/s13415-015-0376-1

Conde, T., Gonçalves, Ó. F., & Pinheiro, A. P. (2018). Stimulus complexity matters when you hear your own voice: Attention effects on self-generated voice processing. *International Journal of Psychophysiology*, *133*, 66–78. https://doi.org/10.1016/j.ijpsycho.2018.08.007

Contreras-Ruston, F., Castillo-Allendes, A., Saavedra-Garrido, J., Ochoa-Muñoz, A. F., Hunter, E. J., Kotz, S. A., & Navarra, J. (2024). Voice self-assessment in individuals with Parkinson's Disease as compared to general voice disorders. *Parkinsonism & Related Disorders*, *123*, 106944. https://doi.org/10.1016/j.parkreldis.2024.106944

Coronel, J. C., & Federmeier, K. D. (2016). The N400 reveals how personal semantics is processed: Insights into the nature and organization of self-knowledge. *Neuropsychologia*, *84*, 36–43. https://doi.org/10.1016/j.neuropsychologia.2016.01.029

Costa, J., Jung, M. F., Czerwinski, M., Guimbretière, F., Le, T., & Choudhury, T. (2018). Regulating feelings during interpersonal conflicts by changing voice self-perception. *Conference on Human Factors in Computing Systems - Proceedings*. https://doi.org/10.1145/3173574.3174205

Crone, E. A., Green, K. H., van de Groep, I. H., & van der Cruijsen, R. (2022). A neurocognitive

model of self-concept development in adolescence. *Annual Review of Developmental

Psychology*, *4*(1), 273–295. https://doi.org/10.1146/annurev-devpsych-120920-023842

Crowley, K. E., & Colrain, I. M. (2004). A review of the evidence for P2 being an independent

component process: Age, sleep and modality. *Clinical Neurophysiology*, *115*(4), 732–744.

https://doi.org/10.1016/j.clinph.2003.11.021

Dacakis, G., Oates, J., & Douglas, J. (2017a). Associations between the Transsexual Voice

Questionnaire (TVQ^MtF) and self-report of voice femininity and acoustic voice measures.

International Journal of Language & Communication Disorders, *52*(6), 831–838.

https://doi.org/10.1111/1460-6984.12319

Dacakis, G., Oates, J., & Douglas, J. (2017b). Associations between the Transsexual Voice

Questionnaire (TVQMtF) and self-report of voice femininity and acoustic voice measures.

International Journal of Language & Communication Disorders, *52*(6), 831–838.

https://doi.org/10.1111/1460-6984.12319

Dankovičová, J., & Hunt, C. (2011). Perception of foreign accent syndrome speech and its relation to

segmental characteristics. *Clinical Linguistics & Phonetics*, *25*(2), 85–120.

https://doi.org/10.3109/02699206.2010.513027

Dedry, M., Dricot, L., Van Parys, V., Boucquey, D., Delinte, N., Van Lith-Bijl, J., Szmalec, A., Maryn,

Y., & Desuter, G. (2022). Brain adaptation following various unilateral vocal fold paralysis

treatments: A magnetic resonance imaging based longitudinal case series. *Frontiers in

Neuroscience*, *16*, 947390. https://doi.org/10.3389/fnins.2022.947390

DiLollo, A., Scherz, J., & Neimeyer, R. A. (2014). Psychosocial implications of foreign accent

syndrome: Two case examples. *Journal of Constructivist Psychology*, *27*, 14–30.

https://doi.org/10.1080/10720537.2013.819305

Douglas, W., & Gibbins, K. (1983). Inadequacy of voice recognition as a demonstration of self-

deception. *Journal of Personality and Social Psychology*, *44*(3), 589–592.

https://doi.org/10.1037/0022-3514.44.3.589

Duggirala, S. X., Schwartze, M., Goller, L. K., Linden, D. E. J., Pinheiro, A. P., & Kotz, S. A. (2024). Hallucination proneness alters sensory feedback processing in self-voice production. *Schizophrenia Bulletin*, *50*(5), 1147–1158. https://doi.org/10.1093/schbul/sbae095

Elshebl, O. Z., Mohamed, A. E., & Ahmed, S. A. A. (2025). Study of the impact of chronic obstructive pulmonary disease on vocal function. *International Archives of Otorhinolaryngology*, *29*(04), 001–008. https://doi.org/10.1055/s-0045-1811696

Escera, C., Alho, K., Winkler, I., & Näätänen, R. (1998). Neural mechanisms of involuntary attention to acoustic novelty and change. *Journal of Cognitive Neuroscience*, *10*(5), 590–604. https://doi.org/10.1162/089892998562997

Fang, C. M., Chua, P., Chan, S., Leong, J., Bao, A., & Maes, P. (2024). Leveraging AI-generated emotional self-voice to nudge people towards their ideal selves. *arXiv*. https://arxiv.org/abs/2409.11531

Feng, C., Yan, X., Huang, W., Han, S., & Ma, Y. (2018). Neural representations of the multidimensional self in the cortical midline structures. *NeuroImage*, *183*, 291–299. https://doi.org/10.1016/j.neuroimage.2018.08.018

Fenigstein, A. (1984). Self-consciousness and the overperception of self as a target. *Journal of Personality and Social Psychology*, *47*(4), 860–870. https://doi.org/10.1037/0022-3514.47.4.860

Fields, E. C., & Kuperberg, G. R. (2015). Loving yourself more than your neighbor: ERPs reveal online effects of a self-positivity bias. *Social Cognitive and Affective Neuroscience*, *10*(9), 1202–1209. https://doi.org/10.1093/scan/nsv004

Flinker, A., Chang, E. F., Kirsch, H. E., Barbaro, N. M., Crone, N. E., & Knight, R. T. (2010). Single-trial speech suppression of auditory cortex activity in humans. *Journal of Neuroscience*, *30*(49), 16643–16650. https://doi.org/10.1523/JNEUROSCI.1809-10.2010

Ford, J. M., Palzes, V. A., Roach, B. J., & Mathalon, D. H. (2014). Did I do that? Abnormal predictive processes in schizophrenia when button pressing to deliver a tone. *Schizophrenia Bulletin*, *40*(4), 804–812. https://doi.org/10.1093/schbul/sbt072

Francis, D. O., Sherman, A. E., Hovis, K. L., Bonnet, K., Schlundt, D., Garrett, C. G., & Davies, L. (2018). Life experience of patients with unilateral vocal fold paralysis. *JAMA Otolaryngology–Head & Neck Surgery*, *144*(5), 433–439. https://doi.org/10.1001/jamaoto.2018.0067

Franke, E. K. (1956). Response of the human skull to mechanical vibrations. *The Journal of the Acoustical Society of America*, *28*(6), 1277–1284. https://doi.org/10.1121/1.1908622

Franken, M. K., Hartsuiker, R. J., Johansson, P., Hall, L., & Lind, A. (2021). Speaking with an alien voice: Flexible sense of agency during vocal production. *Journal of Experimental Psychology: Human Perception and Performance*, *47*(4), 479–494. https://doi.org/10.1037/xhp0000799

Friston, K. J., Sajid, N., Quiroga-Martinez, D. R., Parr, T., Price, C. J., & Holmes, E. (2021). Active listening. *Hearing Research*, *399*, 107998. https://doi.org/10.1016/j.heares.2020.107998

Frith, C., Lawrence, A., & Weinberger, D. (1997). The role of the prefrontal cortex in self-consciousness: The case of auditory hallucinations. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *351*(1346), 1505–1512. https://doi.org/10.1098/rstb.1996.0136

Fu, C. H. Y., Vythelingum, G. N., Brammer, M. J., Williams, S. C. R., Amaro Jr, E., Andrew, C. M., Yágüez, L., van Haren, N. E. M., Matsumoto, K., & McGuire, P. K. (2006). An fMRI study of verbal self-monitoring: Neural correlates of auditory verbal feedback. *Cerebral Cortex*, *16*(7), 969–977. https://doi.org/10.1093/cercor/bhj039

Fuchs, M., Meuret, S., Thiel, S., Täschner, R., Dietz, A., & Gelbrich, G. (2009). Influence of singing activity, age, and sex on voice performance parameters, on subjects' perception and use of their voice in childhood and adolescence. *Journal of Voice*, *23*(2), 182–189. https://doi.org/10.1016/j.jvoice.2007.09.007

Gaudio, R. P. (1994). Sounding gay: Pitch properties in the speech of gay and straight men. *American Speech*, *69*(1), 30–57. https://doi.org/10.2307/455948

Giard, M. H., Perrin, F., Echallier, J. F., Thévenet, M., Froment, J. C., & Pernier, J. (1994). Dissociation of temporal and frontal components in the human auditory N1 wave: A scalp

current density and dipole model analysis. *Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section*, *92*(3), 238–252. https://doi.org/10.1016/0168-5597(94)90067-1

Goffman, E. (1949). Presentation of self in everyday life. *American Journal of Sociology*, *55*(1), 6–7.

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), 101–105. https://doi.org/10.1093/scan/nsu031

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

Greenberg, D. L., Rice, H. J., Cooper, J. J., Cabeza, R., Rubin, D. C., & LaBar, K. S. (2005). Co-activation of the amygdala, hippocampus and inferior frontal gyrus during autobiographical memory retrieval. *Neuropsychologia*, *43*(5), 659–674. https://doi.org/10.1016/j.neuropsychologia.2004.09.002

Guastamacchia, A., Albera, A., Puglisi, G. E., Nudelman, C. J., Soleimanifar, S., Astolfi, A., Aronoff, J. M., & Bottalico, P. (2024). Impact of cochlear implants use on voice production and quality. *Scientific Reports*, *14*(1), 12787. https://doi.org/10.1038/s41598-024-63688-3

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), 209–222. https://doi.org/10.3758/s13423-023-02333-y

Guldner, S., Nees, F., & McGettigan, C. (2020). Vocomotor and social brain networks work together to express social traits in voices. *Cerebral Cortex*, 1–17. https://doi.org/10.1093/cercor/bhaa175

Gur, R. C., & Sackeim, H. A. (1979). Self-deception: A concept in search of a phenomenon. *Journal of Personality and Social Psychology*, *37*, 147–169. https://doi.org/10.1037/0022-3514.37.2.147

Hain, T. C., Burnett, T. A., Kiran, S., Larson, C. R., Singh, S., & Kenney, M. K. (2000). Instructing subjects to make a voluntary response reveals the presence of two components to the audio-

vocal reflex. *Experimental Brain Research*, *130*(2), 133–141.

https://doi.org/10.1007/s002219900237

Hajcak, G., & Foti, D. (2020). Significance?... Significance! Empirical, methodological, and

theoretical connections between the late positive potential and P300 as neural responses to

stimulus significance: An integrative review. *Psychophysiology*, *57*(7), e13570.

https://doi.org/10.1111/psyp.13570

Hamlet, S. L., Geoffrey, V. C., & Bartlett, D. M. (1976). Effect of a dental prosthesis on speaker-

specific characteristics of voice. *Journal of Speech and Hearing Research*, *19*(4), 639–650.

https://doi.org/10.1044/jshr.1904.639

Hamlet, S. L., & Stone, M. (1978). Compensatory alveolar consonant production induced by wearing

a dental prosthesis. *Journal of Phonetics*, *6*(3), 227–248. https://doi.org/10.1016/S0095-

4470(19)31155-6

Hanley, J. R., Smith, S. T., & Hadfield, J. (1998). I recognise you but I can't place you: An

investigation of familiar-only experiences during tests of voice and face recognition. *The

Quarterly Journal of Experimental Psychology Section A*, *51*(1), 179–195.

https://doi.org/10.1080/713755751

Haskell, J. A. (1987). Vocal self-perception: The other side of the equation. *Journal of Voice*, *1*(2),

172–179. https://doi.org/10.1016/S0892-1997(87)80042-5

Haxby, J. V., Hoffman, E. A., & Gobbini, M. I. (2000). The distributed human neural system for face

perception. *Trends in Cognitive Sciences*, *4*(6), 223–233. https://doi.org/10.1016/S1364-

6613(00)01482-0

Hay, J. (2018). Sociophonetics: The role of words, the role of context, and the role of words in

context. *Topics in Cognitive Science*, *10*(4), 696–706. https://doi.org/10.1111/tops.12326

Hengen, J., Hammarström, I. L., & Stenfelt, S. (2020). Perception of one's own voice after hearing-

aid fitting for naive hearing-aid users and hearing-aid refitting for experienced hearing-aid

users. *Trends in Hearing*, *24*, 2331216520932467.

https://doi.org/10.1177/2331216520932467

Herold, D., Spengler, S., Sajonz, B., Usnich, T., & Bermpohl, F. (2016). Common and distinct

networks for self-referential and social stimulus processing in the human brain. *Brain*

Structure and Function, *221*(7), 3475–3485. https://doi.org/10.1007/s00429-015-1113-9

Hickok, G., Houde, J., & Rong, F. (2011). Sensorimotor integration in speech processing:

Computational basis and neural organization. *Neuron*, *69*(3), 407–422.

https://doi.org/10.1016/j.neuron.2011.01.019

Hofmann-Shen, C., Vogel, B. O., Kaffes, M., Rudolph, A., Brown, E. C., Tas, C., Brüne, M., &

Neuhaus, A. H. (2020). Mapping adaptation, deviance detection, and prediction error in

auditory processing. *NeuroImage*, *207*, 116432.

https://doi.org/10.1016/j.neuroimage.2019.116432

Hollien, H., Green, R., & Massey, K. (1994). Longitudinal research on adolescent voice change in

males. *The Journal of the Acoustical Society of America*, *96*(5), 2646–2654.

https://doi.org/10.1121/1.411275

Holzman, P. S., & Rousey, C. (1966). The voice as a percept. *Journal of Personality and Social*

Psychology, *4*, 79–86. https://doi.org/10.1037/h0023518

Hosaka, T., Kimura, M., & Yotsumoto, Y. (2021). Neural representations of own-voice in the human

auditory cortex. *Scientific Reports*, *11*(1), 591. https://doi.org/10.1038/s41598-020-80095-6

Hosoi, H., Nishimura, T., Shimokura, R., & Kitahara, T. (2019). Cartilage conduction as the third

pathway for sound transmission. *Auris Nasus Larynx*, *46*(2), 151–159.

https://doi.org/10.1016/j.anl.2019.01.005

Houde, J. F., Nagarajan, S. S., Sekihara, K., & Merzenich, M. M. (2002). Modulation of the auditory

cortex during speech: An MEG study. *Journal of Cognitive Neuroscience*, *14*(8), 1125–1138.

https://doi.org/10.1162/089892902760807140

Hu, C., Di, X., Eickhoff, S. B., Zhang, M., Peng, K., Guo, H., & Sui, J. (2016). Distinct and common

aspects of physical and psychological self-representation in the brain: A meta-analysis of self-

bias in facial and self-referential judgements. *Neuroscience and Biobehavioral Reviews*, *61*,

197–207. https://doi.org/10.1016/j.neubiorev.2015.12.003

Hughes, S. M., & Harrison, M. A. (2013). I like my voice better: Self-enhancement bias in

perceptions of voice attractiveness. *Perception*, *42*(9), 941–949.

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), 107–127.

https://doi.org/10.1007/s10919-013-0163-z

Hughes, S. M., & Nicholson, S. E. (2010). The processing of auditory and visual recognition of self-

stimuli. *Consciousness and Cognition*, *19*(4), 1124–1134.

https://doi.org/10.1016/j.concog.2010.03.001

Hull, J. G., Van Treuren, R. R., Ashford, S. J., Propsom, P., & Andrus, B. W. (1988). Self-

consciousness and the processing of self-relevant information. *Journal of Personality and

Social Psychology*, *54*(3), 452–465. https://doi.org/10.1037/0022-3514.54.3.452

Hurtado-Ruzza, R., Iglesias, Ó. Á.-C., Becerro-de-Bengoa-Vallejo, R., Calvo-Lobo, C., San-Antolín,

M., Losa-Iglesias, M. E., Romero-Morales, C., & López-López, D. (2021). Self-Perceived

Handicap Associated With Dysphonia and Health-Related Quality of Life of Asthma and

Chronic Obstructive Pulmonary Disease Patients: A Case–Control Study. *Journal of Speech,

Language, and Hearing Research*, *64*(2), 433–443. https://doi.org/10.1044/2020_JSLHR-20-

00473

Iannotti, G. R., Orepic, P., Brunet, D., Koenig, T., Alcoba-Banqueri, S., Garin, D. F. A., Schaller, K.,

Blanke, O., & Michel, C. M. (2022). EEG spatiotemporal patterns underlying self-other voice

discrimination. *Cerebral Cortex*, *32*(9), 1978–1992. https://doi.org/10.1093/cercor/bhab329

Ito, T., Tiede, M., & Ostry, D. J. (2009). Somatosensory function in speech perception. *Proceedings of

the National Academy of Sciences*, *106*(4), 1245–1248.

https://doi.org/10.1073/pnas.0810063106

Jardri, R., Pins, D., Bubrovszky, M., Despretz, P., Pruvo, J. P., Steinling, M., & Thomas, P. (2007).

Self awareness and speech processing: An fMRI study. *NeuroImage*, *35*(4), 1645–1653.

https://doi.org/10.1016/j.neuroimage.2007.02.002

Jennifer, O., & Georgia, D. (2015). Transgender voice and communication: Research evidence underpinning voice intervention for male-to-female transsexual women. *Perspectives on Voice and Voice Disorders*, *25*(2), 48–58. https://doi.org/10.1044/vvd25.2.48

Johnson, J., Belyk, M., Schwartze, M., Pinheiro, A. P., & Kotz, S. A. (2021). Expectancy changes the self-monitoring of voice identity. *European Journal of Neuroscience*, *53*, 2681–2695. https://doi.org/10.1111/ejn.15162

Jones, J. A., & Keough, D. (2008). Auditory-motor mapping for pitch control in singers and nonsingers. *Experimental Brain Research*, *190*(3), 279–287. https://doi.org/10.1007/s00221-008-1473-y

Jones, J. A., & Munhall, K. G. (2002). The role of auditory feedback during phonation: Studies of Mandarin tone production. *Journal of Phonetics*, *30*(3), 303–320. https://doi.org/10.1006/jpho.2001.0160

Kaplan, J. T., Aziz-Zadeh, L., Uddin, L. Q., & Iacoboni, M. (2008). The self across the senses: An fMRI study of self-face and self-voice recognition. *Social Cognitive and Affective Neuroscience*, *3*(3), 218–223. https://doi.org/10.1093/scan/nsn014

Kashima, Y., Yamaguchi, S., Kim, U., Choi, S.-C., Gelfand, M. J., & Yuki, M. (1995). Culture, gender, and self: A perspective from individualism-collectivism research. *Journal of Personality and Social Psychology*, *69*(5), 925–937. https://doi.org/10.1037/0022-3514.69.5.925

Kern, M., Bert, S., Glanz, O., Schulze-Bonhage, A., & Ball, T. (2019). Human motor cortex relies on sparse and action-specific activation during laughing, smiling and speech production. *Communications Biology*, *2*(1), 118. https://doi.org/10.1038/s42003-019-0360-3

Keulen, S., Verhoeven, J., Bastiaanse, R., Mariën, P., Jonkers, R., Mavroudakis, N., & Paquier, P. (2016). Perceptual accent rating and attribution in psychogenic FAS: Some further evidence challenging Whitaker's operational definition. *Frontiers in Human Neuroscience*, *10*. https://doi.org/10.3389/fnhum.2016.00062

Khalilian-Gourtani, A., Wang, R., Chen, X., Yu, L., Dugan, P., Friedman, D., Doyle, W., Devinsky, O., Wang, Y., & Flinker, A. (2024). A corollary discharge circuit in human speech. *PNAS*, *121*(50), e2404121121. https://doi.org/10.1073/pnas.2404121121

Kimura, M., & Yotsumoto, Y. (2018). Auditory traits of 'own voice'. *PLOS One*, *13*(6), e0199443.

Kirk, N. W., & Cunningham, S. J. (2025). Listen to yourself! Prioritization of self-associated and own voice cues. *British Journal of Psychology*, *116*(1), 131–148. https://doi.org/10.1111/bjop.12741

Kleber, B., Friberg, A., Zeitouni, A., & Zatorre, R. (2017). Experience-dependent modulation of right anterior insula and sensorimotor regions as a function of noise-masked auditory feedback in singers and nonsingers. *NeuroImage*, *147*, 97–110. https://doi.org/10.1016/j.neuroimage.2016.11.059

Kleber, B., Zeitouni, A. G., Friberg, A., & Zatorre, R. J. (2013). Experience-dependent modulation of feedback integration during singing: Role of the right anterior insula. *The Journal of Neuroscience*, *33*(14), 6070–6080. https://doi.org/10.1523/JNEUROSCI.4418-12.2013

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), 2698–2704. https://doi.org/10.1098/rspb.2012.0311

Knolle, F., Schröger, E., Baess, P., & Kotz, S. A. (2012). The cerebellum generates motor-to-auditory predictions: ERP lesion evidence. *Journal of Cognitive Neuroscience*, *24*(3), 698–706. https://doi.org/10.1162/jocn_a_00167

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

Kotz, S. A., & Paulmann, S. (2011). Emotion, language, and the brain. *Linguistics and Language Compass*, *5*(3), 108–125. https://doi.org/10.1111/j.1749-818X.2010.00267.x

Kotz, S. A., Stockert, A., & Schwartze, M. (2014). Cerebellum, temporal predictability and the updating of a mental model. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *369*, 20130403. https://doi.org/10.1098/rstb.2013.0403

Kreiman, J. (2024). Information conveyed by voice quality. *The Journal of the Acoustical Society of America*, *155*(2), 1264–1271. https://doi.org/10.1121/10.0024609

Kurczek, J., Wechsler, E., Ahuja, S., Jensen, U., Cohen, N. J., Tranel, D., & Duff, M. (2015). Differential contributions of hippocampus and medial prefrontal cortex to self-projection and self-referential processing. *Neuropsychologia*, *73*, 116–126. https://doi.org/10.1016/j.neuropsychologia.2015.05.002

Kurteff, G. L., Field, A. M., Asghar, S., Tyler-Kabara, E. C., Clarke, D., Weiner, H. L., Anderson, A. E., Watrous, A. J., Buchanan, R. J., Modur, P. N., & Hamilton, L. S. (2024). Spatiotemporal mapping of auditory onsets during speech production. *The Journal of Neuroscience*, e1109242024. https://doi.org/10.1523/JNEUROSCI.1109-24.2024

Kwan, L. C., & Whitehill, T. L. (2011). Perception of speech by individuals with Parkinson's Disease: A review. *Parkinson's Disease*, *2011*, 1–11. https://doi.org/10.4061/2011/389767

Lametti, D. R., Nasir, S. M., & Ostry, D. J. (2012). Sensory preference in speech production revealed by simultaneous alteration of auditory and somatosensory feedback. *The Journal of Neuroscience*, *32*(27), 9351 LP – 9358. https://doi.org/10.1523/JNEUROSCI.0404-12.2012

Larson, C. R., Sun, J., & Hain, T. C. (2007). Effects of simultaneous perturbations of voice pitch and loudness feedback on voice F0 and amplitude control. *The Journal of the Acoustical Society of America*, *121*(5), 2862–2872. https://doi.org/10.1121/1.2715657

Latinus, M., & Belin, P. (2011). Anti-voice adaptation suggests prototype-based coding of voice identity. *Frontiers in Psychology*, *2*(JUL). https://doi.org/10.3389/fpsyg.2011.00175

Latinus, M., McAleer, P., Bestelmeyer, P. E. G., & Belin, P. (2013). Norm-based coding of voice identity in human auditory cortex. *Current Biology*, *23*(12), 1075–1080. https://doi.org/10.1016/j.cub.2013.04.055

Lavan, N., Rinke, P., & Scharinger, M. (2024). The time course of person perception from voices in the brain. *Proceedings of the National Academy of Sciences*, *121*(26), e2318361121. https://doi.org/10.1073/pnas.2318361121

Laver, J. (1980). *The phonetic description of voice quality*. Cambridge University Press.

Lavner, Y., Rosenhouse, J., & Gath, I. (2001a). The prototype model in speaker identification by human listeners. *International Journal of Speech Technology*, *4*(1), 63–74. https://doi.org/10.1023/A:1009656816383

Lavner, Y., Rosenhouse, J., & Gath, I. (2001b). The Prototype Model in Speaker Identification by Human Listeners. *International Journal of Speech Technology*, *4*(1), 63–74. https://doi.org/10.1023/A:1009656816383

Lee, J. J., & Perrachione, T. K. (2022). Implicit and explicit learning in talker identification. *Attention, Perception, & Psychophysics*, *84*(6), 2002–2015. https://doi.org/10.3758/s13414-022-02500-8

Levorsen, M., Aoki, R., Matsumoto, K., Sedikides, C., & Izuma, K. (2023). The self-concept is represented in the medial prefrontal cortex in terms of self-importance. *The Journal of Neuroscience*, *43*(20), 3675 LP – 3686. https://doi.org/10.1523/JNEUROSCI.2178-22.2023

Liu, L., Li, W., Li, J., Lou, L., & Chen, J. (2019). Temporal features of psychological and physical self-representation: An ERP study. *Frontiers in Psychology*, *10*, 785.

López, S., Riera, P., Assaneo, M. F., Eguía, M., Sigman, M., & Trevisan, M. A. (2013). Vocal caricatures reveal signatures of speaker identity. *Scientific Reports*, *3*(1), 3407. https://doi.org/10.1038/srep03407

Luthra, S. (2024). Why are listeners hindered by talker variability? *Psychonomic Bulletin & Review*, *31*(1), 104–121. https://doi.org/10.3758/s13423-023-02355-6

Maguinness, C., Roswandowitz, C., & Von Kriegstein, K. (2018). Understanding the mechanisms of familiar voice-identity recognition in the human brain. *Neuropsychologia*, *116*, 179–193. https://doi.org/10.1016/j.neuropsychologia.2018.03.039

Marquine, M. J., Grilli, M. D., Rapcsak, S. Z., Kaszniak, A. W., Ryan, L., Walther, K., & Glisky, E. L. (2016). Impaired personal trait knowledge, but spared other-person trait knowledge, in an individual with bilateral damage to the medial prefrontal cortex. *Neuropsychologia*, *89*, 245–253. https://doi.org/10.1016/j.neuropsychologia.2016.06.021

Mathalon, D. H., Roach, B. J., Ferri, J. M., Loewy, R. L., Stuart, B. K., Perez, V. B., Trujillo, T. H., & Ford, J. M. (2019). Deficient auditory predictive coding during vocalization in the psychosis

risk syndrome and in early illness schizophrenia: The final expanded sample. *Psychological Medicine*, *49*(11), 1897–1904. https://doi.org/DOI:%2520 10.1017/S0033291718002659

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://doi.org/10.3389/fnhum.2015.00129

Miall, R. C., & Wolpert, D. M. (1996). Forward models for physiological motor control. *Neural Networks*, *9*(8), 1265–1279. https://doi.org/10.1016/S0893-6080(96)00035-4

Migliorelli, A., Natale, E., Manuelli, M., Ciorba, A., Bianchini, C., Pelucchi, S., & Stomeo, F. (2024). Analysis of tracheoesophageal voice after total laryngectomy: A single center experience. *Journal of Clinical Medicine*, *13*(15), 4392. https://doi.org/10.3390/jcm13154392

Morita, T., Tanabe, H. C., Sasaki, A. T., Shimada, K., Kakigi, R., & Sadato, N. (2014). The anterior insular and anterior cingulate cortices in emotional processing for self-face recognition. *Social Cognitive and Affective Neuroscience*, *9*(5), 570–579. https://doi.org/10.1093/scan/nst011

Müller, N., Leske, S., Hartmann, T., Szebényi, S., & Weisz, N. (2015). Listen to yourself: The medial prefrontal cortex modulates auditory alpha power during speech preparation. *Cerebral Cortex*, *25*(11), 4029–4037. https://doi.org/10.1093/cercor/bhu117

Murray, R. J., Debbané, M., Fox, P. T., Bzdok, D., & Eickhoff, S. B. (2015). Functional connectivity mapping of regions associated with self- and other-processing. *Human Brain Mapping*, *36*(4), 1304–1324. https://doi.org/10.1002/hbm.22703

Nakamura, K., Kawashima, R., Sugiura, M., Kato, T., Nakamura, A., Hatano, K., Nagumo, S., Kubota, K., Fukuda, H., Ito, K., & Kojima, S. (2001). Neural substrates for recognition of familiar voices: A PET study. *Neuropsychologia*, *39*(10), 1047–1054. https://doi.org/10.1016/S0028-3932(01)00037-9

Natke, U., Donath, T. M., & Kalveram, K. Th. (2003). Control of voice fundamental frequency in speaking versus singing. *The Journal of the Acoustical Society of America*, *113*(3), 1587–1593. https://doi.org/10.1121/1.1543928

Niziolek, C. A., Nagarajan, S. S., & Houde, J. F. (2013). What does motor efference copy represent? Evidence from speech production. *Journal of Neuroscience*, *33*(41), 16110–16116. https://doi.org/10.1523/JNEUROSCI.2137-13.2013

Northoff, G., Heinzel, A., de Greck, M., Bermpohl, F., Dobrowolny, H., & Panksepp, J. (2006). Self-referential processing in our brain-A meta-analysis of imaging studies on the self. *NeuroImage*, *31*(1), 440–457. https://doi.org/10.1016/j.neuroimage.2005.12.002

Nygren, U., Södersten, M., Falhammar, H., Thorén, M., Hagenfeldt, K., & Nordenskjöld, A. (2009). Voice characteristics in women with congenital adrenal hyperplasia due to 21-hydroxylase deficiency. *Clinical Endocrinology*, *70*(1), 18–25. https://doi.org/10.1111/j.1365-2265.2008.03347.x

Oguz, H., Demirci, M., Safak, M. A., Arslan, N., Islam, A., & Kargin, S. (2007). Effects of unilateral vocal cord paralysis on objective voice measures obtained by Praat. *European Archives of Oto-Rhino-Laryngology*, *264*(3), 257–261. https://doi.org/10.1007/s00405-006-0179-7

Olivos, G. (1967). Response delay, psychophysiologic activation, and recognition of one's own voice. *Psychosomatic Medicine*, *29*(5), 433–440.

Orepic, P., Bernasconi, F., Faggella, M., Faivre, N., & Blanke, O. (2024). Robotically-induced auditory-verbal hallucinations: Combining self-monitoring and strong perceptual priors. *Psychological Medicine*, *54*(3), 569–581. https://doi.org/DOI:%252010.1017/S0033291723002222

Orepic, P., Iannotti, G. R., Haemmerli, J., Goga, C., Park, H.-D., Betka, S., Blanke, O., Michel, C. M., Bondolfi, G., & Schaller, K. (2023). Experimentally-evidenced personality alterations following meningioma resection: A case report. *Cortex*, *168*, 157–166. https://doi.org/10.1016/j.cortex.2023.08.006

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

Orepic, P., Park, H.-D., Rognini, G., Faivre, N., & Blanke, O. (2022). Breathing affects self-other voice discrimination in a bodily state associated with somatic passivity. *Psychophysiology*, *59*(7), e14016. https://doi.org/10.1111/psyp.14016

Orepic, P., Rognini, G., Kannape, O. A., Faivre, N., & Blanke, O. (2021). Sensorimotor conflicts induce somatic passivity and louden quiet voices in healthy listeners. *Schizophrenia Research*, *231*, 170–177. https://doi.org/10.1016/j.schres.2021.03.014

Ostrand, R., & Chodroff, E. (2021). It's alignment all the way down, but not all the way up: Speakers align on some features but not others within a dialogue. *Journal of Phonetics*, *88*, 101074. https://doi.org/10.1016/j.wocn.2021.101074

Paraskevoudi, N., & SanMiguel, I. (2021). Self-generation and sound intensity interactively modulate perceptual bias, but not perceptual sensitivity. *Scientific Reports*, *11*(1), 17103. https://doi.org/10.1038/s41598-021-96346-z

Park, H.-D., & Blanke, O. (2019). Coupling inner and outer body for self-consciousness. *Trends in Cognitive Sciences*, *23*(5), 377–388. https://doi.org/10.1016/j.tics.2019.02.002

Parrell, P., & Houde, J. (2019). Modeling the role of sensory feedback in speech motor control and learning. *Journal of Speech, Language, and Hearing Research*, *62*(8S), 2963–2985. https://doi.org/10.1044/2019_JSLHR-S-CSMC7-18-0127

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

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

Perez, V. B., Ford, J. M., Roach, B. J., Loewy, R. L., Stuart, B. K., Vinogradov, S., & Mathalon, D. H. (2012). Auditory cortex responsiveness during talking and listening: Early illness schizophrenia and patients at clinical high-risk for psychosis. *Schizophrenia Bulletin*, *38*(6), 1216–1224. https://doi.org/10.1093/schbul/sbr124

Perrachione, T. K., Del Tufo, S. N., & Gabrieli, J. D. E. (2011). Human voice recognition depends on language ability. *Science*, *333*(6042), 595–595. https://doi.org/10.1126/science.1207327

Pinheiro, A. P. (2025). Beyond acoustics: Self-relevance as a key to voice naturalness (L). *The Journal of the Acoustical Society of America*, *158*(5), 4045–4047. https://doi.org/10.1121/10.0039927

Pinheiro, A. P., Aucouturier, J.-J., & Kotz, S. A. (2024). Neural adaptation to changes in self-voice during puberty. *Trends in Neurosciences*, *47*(10), 777–787. https://doi.org/10.1016/j.tins.2024.08.001

Pinheiro, A. P., Farinha-Fernandes, A., Roberto, M. S., & Kotz, S. A. (2019). Self-voice perception and its relationship with hallucination predisposition. *Cognitive Neuropsychiatry*, *24*(4), 237–255. https://doi.org/10.1080/13546805.2019.1621159

Pinheiro, A. P., Rezaii, N., Nestor, P. G., Rauber, A., Spencer, K. M., & Niznikiewicz, M. (2016). Did you or I say pretty, rude or brief? An ERP study of the effects of speaker's identity on emotional word processing. *Brain and Language*, *153–154*, 38–49. https://doi.org/10.1016/j.bandl.2015.12.003

Pinheiro, A. P., Rezaii, N., Rauber, A., Nestor, P. G., Spencer, K. M., & Niznikiewicz, M. (2017). Emotional self–other voice processing in schizophrenia and its relationship with hallucinations: ERP evidence. *Psychophysiology*, *54*(9), 1252–1265. https://doi.org/10.1111/psyp.12880

Pinheiro, A. P., Rezaii, N., Rauber, A., & Niznikiewicz, M. (2016). Is this my voice or yours? The role of emotion and acoustic quality in self-other voice discrimination in schizophrenia. *Cognitive Neuropsychiatry*, *21*(4), 335–353. https://doi.org/10.1080/13546805.2016.1208611

Pinheiro, A. P., Sarzedas, J., Roberto, M. S., & Kotz, S. A. (2023). Attention and emotion shape self-voice prioritization in speech processing. *Cortex*, *158*, 83–95. https://doi.org/10.1016/j.cortex.2022.10.006

Pinheiro, A. P., Schwartze, M., Amorim, M., Coentre, R., Levy, P., & Kotz, S. A. (2020). Changes in motor preparation affect the sensory consequences of voice production in voice hearers. *Neuropsychologia*, *146*, 107531. https://doi.org/10.1016/j.neuropsychologia.2020.107531

Pinheiro, A. P., Schwartze, M., Gutiérrez-Domínguez, F., & Kotz, S. A. (2020). Real and imagined sensory feedback have comparable effects on action anticipation. *Cortex*, *130*, 290–301. https://doi.org/10.1016/j.cortex.2020.04.030

Pinheiro, A. P., Schwartze, M., & Kotz, S. A. (2018). Voice-selective prediction alterations in nonclinical voice hearers. *Scientific Reports*, *8*(1), 14717. https://doi.org/10.1038/s41598-018-32614-9

Pinheiro, A. P., Schwartze, M., & Kotz, S. A. (2020). Cerebellar circuitry and auditory verbal hallucinations: An integrative synthesis and perspective. *Neuroscience and Biobehavioral Reviews*, *118*, 485–503. https://doi.org/10.1016/j.neubiorev.2020.08.004

Polich, J. (2007). Updating P300: An integrative theory of P3a and P3b. *Clinical Neurophysiology*, *118*(10), 2128–2148. https://doi.org/10.1016/j.clinph.2007.04.019

Pörschmann, C. (2000). Influences of bone conduction and air conduction on the sound of one's own voice. *Acta Acustica*, *86*(6), 1038–1045.

Pu, S., Goldberg, L., Ren, J., Goldberg, A. C., & Courey, M. (2025). Physical features contributing to gender dysphoria: The role of voice. *Otolaryngology–Head and Neck Surgery*, *172*(6), 2018–2025. https://doi.org/10.1002/ohn.1207

Rachman, L., Dubal, S., & Aucouturier, J.-J. (2019). Happy you, happy me: Expressive changes on a stranger's voice recruit faster implicit processes than self-produced expressions. *Social Cognitive and Affective Neuroscience*, *14*(5), 559–568. https://doi.org/10.1093/scan/nsz030

Ralph, M. A. L., Jefferies, E., Patterson, K., & Rogers, T. T. (2017). The neural and computational bases of semantic cognition. *Nature Reviews Neuroscience*, *18*(1), 42–55. https://doi.org/10.1038/nrn.2016.150

Ramig, L. O., Scherer, R. C., Klasner, E. R., Titze, I. R., & Horii, Y. (1990). Acoustic analysis of voice in amyotrophic lateral sclerosis. *Journal of Speech and Hearing Disorders*, *55*(1), 2–14. https://doi.org/10.1044/jshd.5501.02

Reetz, S., Bohlender, J. E., & Brockmann-Bauser, M. (2019). Do standard instrumental acoustic, perceptual, and subjective voice outcomes indicate therapy success in patients with functional dysphonia? *Journal of Voice*, *33*(3), 317–324. https://doi.org/10.1016/j.jvoice.2017.11.014

Reinfeldt, S., Östli, P., Håkansson, B., & Stenfelt, S. (2010). Hearing one's own voice during phoneme vocalization—Transmission by air and bone conduction. *The Journal of the Acoustical Society of America*, *128*(2), 751–762. https://doi.org/10.1121/1.3458855

Reznik, D., Simon, S., & Mukamel, R. (2018). Predicted sensory consequences of voluntary actions modulate amplitude of preceding readiness potentials. *Neuropsychologia*, *119*, 302–307. https://doi.org/10.1016/j.neuropsychologia.2018.08.028

Rosa, C., Lassonde, M., Pinard, C., Keenan, J. P., & Belin, P. (2008). Investigations of hemispheric specialization of self-voice recognition. *Brain and Cognition*, *68*(2), 204–214. https://doi.org/10.1016/j.bandc.2008.04.007

Rosi, V., Payne, B., & McGettigan, C. (2025). Effects of self-similarity and self-generation on the perceptual prioritisation of voices. *Journal of Experimental Psychology: Human Perception and Performance*. https://doi.org/10.1037/xhp0001325

Rousey, C., & Holzman, P. S. (1967). Recognition of one's own voice. *Journal of Personality and Social Psychology*, *6*(4, Pt.1), 464–466. https://doi.org/10.1037/h0024837

Roye, A., Jacobsen, T., & Schröger, E. (2007). Personal significance is encoded automatically by the human brain: An event-related potential study with ringtones. *European Journal of Neuroscience*, *26*(3), 784–790. https://doi.org/10.1111/j.1460-9568.2007.05685.x

Runnqvist, E., Bonnard, M., Gauvin, H. S., Attarian, S., Trébuchon, A., Hartsuiker, R. J., & Alario, F. X. (2016). Internal modeling of upcoming speech: A causal role of the right posterior cerebellum in non-motor aspects of language production. *Cortex*, *81*, 203–214. https://doi.org/10.1016/j.cortex.2016.05.008

Runnqvist, E., Chanoine, V., Strijkers, K., Pattamadilok, C., Bonnard, M., Nazarian, B., Sein, J., Anton, J.-L., Dorokhova, L., Belin, P., & Alario, F.-X. (2021). Cerebellar and cortical correlates of internal and external speech error monitoring. *Cerebral Cortex Communications*, *2*(2), tgab038. https://doi.org/10.1093/texcom/tgab038

Schaller, K., Iannotti, G. R., Orepic, P., Betka, S., Haemmerli, J., Boex, C., Alcoba-Banqueri, S., Garin, D. F. A., Herbelin, B., Park, H.-D., Michel, C. M., & Blanke, O. (2021). The

perspectives of mapping and monitoring of the sense of self in neurosurgical patients. *Acta Neurochirurgica*, *163*(5), 1213–1226. https://doi.org/10.1007/s00701-021-04778-3

Schweinberger, S. R., Casper, C., Hauthal, N., Kaufmann, J. M., Kawahara, H., Kloth, N., Robertson, D. M. C., Simpson, A. P., & Zäske, R. (2008). Auditory adaptation in voice perception. *Current Biology*, *18*(9), 684–688. https://doi.org/10.1016/j.cub.2008.04.015

Scott, S. K. (2019). From speech and talkers to the social world: The neural processing of human spoken language. *Science*, *366*(6461), 58–62. https://doi.org/10.1126/science.aax0288

Seifert, E., Runte, C., Riebandt, M., Lamprecht-Dinnesen, A., & Bollmann, F. (1999). Can dental prostheses influence vocal parameters? *The Journal of Prosthetic Dentistry*, *81*(5), 579–585. https://doi.org/10.1016/S0022-3913(99)70213-1

Selosse, G., Grandjean, D., & Ceravolo, L. (2025). Neural correlates of embodied and vibratory mechanisms associated with emotional prosody production. *Social Cognitive And Affective Neuroscience*, *20*(1), nsaf084. https://doi.org/10.1093/scan/nsaf084

Shadden, B. (2005). Aphasia as identity theft: Theory and practice. *Aphasiology*, *19*(3–5), 211–223. https://doi.org/10.1080/02687930444000697

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. V. L., & Zäske, R. (2021). Who we are. In *The Handbook of Speech Perception* (pp. 365–397). https://doi.org/10.1002/9781119184096.ch14

Simmons, W. K., Avery, J. A., Barcalow, J. C., Bodurka, J., Drevets, W. C., & Bellgowan, P. (2013). Keeping the body in mind: Insula functional organization and functional connectivity integrate interoceptive, exteroceptive, and emotional awareness. *Human Brain Mapping*, *34*(11), 2944–2958. https://doi.org/10.1002/hbm.22113

Smeltzer, J. C., Chiou, S. H., & Shembel, A. C. (2025). Interoception, voice symptom reporting, and voice disorders. *Journal of Voice*, *39*(5), 1236–1245. https://doi.org/10.1016/j.jvoice.2023.03.002

Stenfelt, S. (2016). Model predictions for bone conduction perception in the human. *Hearing Research*, *340*, 135–143. https://doi.org/10.1016/j.heares.2015.10.014

Stern, J., Schild, C., Jones, B. C., DeBruine, L. M., Hahn, A., Puts, D. A., Zettler, I., Kordsmeyer, T. L., Feinberg, D., Zamfir, D., Penke, L., & Arslan, R. C. (2021). Do voices carry valid information about a speaker's personality? *Journal of Research in Personality*, *92*, 104092. https://doi.org/10.1016/j.jrp.2021.104092

Steurer, H., Schalling, E., Franzén, E., & Albrecht, F. (2022). Characterization of mild and moderate dysarthria in Parkinson's Disease: Behavioral measures and neural correlates. *Frontiers in Aging Neuroscience*, *14*, 870998. https://doi.org/10.3389/fnagi.2022.870998

Stogowska, E., Kamiński, K. A., Ziółko, B., & Kowalska, I. (2022). Voice changes in reproductive disorders, thyroid disorders and diabetes: A review. *Endocrine Connections*, *11*(3), e210505. https://doi.org/10.1530/EC-21-0505

Sugiura, M. (2015). Three faces of self-face recognition: Potential for a multi-dimensional diagnostic tool. *Neuroscience Research*, *90*, 56–64. https://doi.org/10.1016/j.neures.2014.10.002

Summerfield, C., Trittschuh, E. H., Monti, J. M., Mesulam, M.-M., & Egner, T. (2008). Neural repetition suppression reflects fulfilled perceptual expectations. *Nature Neuroscience*, *11*(9), 1004–1006. https://doi.org/10.1038/nn.2163

Tang, D.-L., McDaniel, A., & Watkins, K. E. (2021). Disruption of speech motor adaptation with repetitive transcranial magnetic stimulation of the articulatory representation in primary motor cortex. *Cortex*, *145*, 115–130. https://doi.org/10.1016/j.cortex.2021.09.008

Todorović, S., Anton, J.-L., Sein, J., Nazarian, B., Chanoine, V., Rauchbauer, B., Kotz, S. A., & Runnqvist, E. (2024). Cortico-cerebellar monitoring of speech sequence production. *Neurobiology of Language*, *5*(3), 701–721. https://doi.org/10.1162/nol_a_00113

Tourville, J. A., & Guenther, F. H. (2011). The DIVA model: A neural theory of speech acquisition and production. *Language and Cognitive Processes*, *26*(7), 952–981. https://doi.org/10.1080/01690960903498424

Towle, V. L., Yoon, H.-A., Castelle, M., Edgar, J. C., Biassou, N. M., Frim, D. M., Spire, J.-P., & Kohrman, M. H. (2008). ECoG gamma activity during a language task: Differentiating

expressive and receptive speech areas. *Brain*, *131*(8), 2013–2027.

https://doi.org/10.1093/brain/awn147

Tremblay, S., Shiller, D. M., & Ostry, D. J. (2003). Somatosensory basis of speech production.

Nature, *423*(6942), 866–869. https://doi.org/10.1038/nature01710

Trulsson, M., & Johansson, R. S. (2002). Orofacial mechanoreceptors in humans: Encoding

characteristics and responses during natural orofacial behaviors. *Behavioural Brain Research*,

135(1), 27–33. https://doi.org/10.1016/S0166-4328(02)00151-1

van der Meer, L., Costafreda, S., Aleman, A., & David, A. S. (2010). Self-reflection and the brain: A

theoretical review and meta-analysis of neuroimaging studies with implications for

schizophrenia. *Neuroscience and Biobehavioral Reviews*, *34*(6), 935–946.

https://doi.org/10.1016/j.neubiorev.2009.12.004

van der Weiden, A., Prikken, M., & van Haren, N. E. M. (2015). Self-other integration and distinction

in schizophrenia: A theoretical analysis and a review of the evidence. *Neuroscience and*

Biobehavioral Reviews, *57*, 220–237. https://doi.org/10.1016/j.neubiorev.2015.09.004

Van Os, J., Linscott, R. J., Myin-Germeys, I., Delespaul, P., & Krabbendam, L. (2009). A systematic

review and meta-analysis of the psychosis continuum: Evidence for a psychosis proneness-

persistence-impairment model of psychotic disorder. *Psychological Medicine*, *39*(2), 179–

195. https://doi.org/10.1017/S0033291708003814

Van Sluis, K. E., Van Der Molen, L., Van Son, R. J. J. H., Hilgers, F. J. M., Bhairosing, P. A., & Van

Den Brekel, M. W. M. (2018). Objective and subjective voice outcomes after total

laryngectomy: A systematic review. *European Archives of Oto-Rhino-Laryngology*, *275*(1),

11–26. https://doi.org/10.1007/s00405-017-4790-6

Ventura, M. I., Nagarajan, S. S., & Houde, J. F. (2009). Speech target modulates speaking induced

suppression in auditory cortex. *BMC Neuroscience*, *10*, 58. https://doi.org/10.1186/1471-

2202-10-58

Voruz, P., Orepic, P., Coll, S. Y., Haemmerli, J., Blanke, O., Péron, J. A., Schaller, K., & Iannotti, G.

R. (2024). Self-other voice discrimination task: A potential neuropsychological tool for

clinical assessment of self-related deficits. *Heliyon*, *10*(19), e38711.

https://doi.org/10.1016/j.heliyon.2024.e38711

Vukojević, J., Jelić, L., McCormack, K., Sušac, J., Muselimović, I., Bagarić, M., Brečić, P., Dellwo, V., Cifrek, M., Savić, A., & Orepić, P. (2025). *Self-other voice confusion in patients with auditory-verbal hallucinations and nonclinical hallucination proneness*. medRxiv. https://doi.org/10.1101/2025.10.06.25337403

Vurma, A. (2014). The timbre of the voice as perceived by the singer him-/herself. *Logopedics Phoniatrics Vocology*, *39*(1), 1–10. https://doi.org/10.3109/14015439.2013.775334

Walton, C., Carding, P., Conway, E., Flanagan, K., & Blackshaw, H. (2019). Voice outcome measures for adult patients with unilateral vocal fold paralysis: A systematic review. *The Laryngoscope*, *129*(1), 187–197. https://doi.org/10.1002/lary.27434

Wang, J., Mathalon, D. H., Roach, B. J., Reilly, J., Keedy, S. K., Sweeney, J. A., & Ford, J. M. (2014). Action planning and predictive coding when speaking. *NeuroImage*, *91*, 91–98. https://doi.org/10.1016/j.neuroimage.2014.01.003

Waters, F., Woodward, T., Allen, P., Aleman, A., & Sommer, I. (2012). Self-recognition deficits in schizophrenia patients with auditory hallucinations: A meta-analysis of the literature. *Schizophrenia Bulletin*, *38*(4), 741–750. https://doi.org/10.1093/schbul/sbq144

Weerathunge, H. R., Alzamendi, G. A., Cler, G. J., Guenther, F. H., Stepp, C. E., & Zañartu, M. (2022). LaDIVA: A neurocomputational model providing laryngeal motor control for speech acquisition and production. *PLOS Computational Biology*, *18*(6), e1010159. https://doi.org/10.1371/journal.pcbi.1010159

Welgampola, M. S., Rosengren, S. M., Halmagyi, G. M., & Colebatch, J. G. (2003). Vestibular activation by bone conducted sound. *Journal of Neurology, Neurosurgery & Psychiatry*, *74*(6), 771–778. https://doi.org/10.1136/JNNP.74.6.771

Wen, W., Okon, Y., Yamashita, A., & Asama, H. (2022). The over-estimation of distance for self-voice versus other-voice. *Scientific Reports*, *12*(1), 420. https://doi.org/10.1038/s41598-021-04437-8

Whitford, T. J. (2019). Speaking-induced suppression of the auditory cortex in humans and its relevance to schizophrenia. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, *4*(9), 791–804. https://doi.org/10.1016/j.bpsc.2019.05.011

Wolpert, D. M. (1997). Computational approaches to motor control. *Trends in Cognitive Sciences*, *1*(6), 209–216. https://doi.org/10.1016/S1364-6613(97)01070-X

Woolnough, O., Forseth, K. J., Rollo, P. S., & Tandon, N. (2019). Uncovering the functional anatomy of the human insula during speech. *eLife*, *8*, e53086. https://doi.org/10.7554/eLife.53086

Xu, M., Homae, F., Hashimoto, R.-I., & Hagiwara, H. (2013). Acoustic cues for the recognition of self-voice and other-voice. *Frontiers in Psychology*, *4*(735), 1–7. https://doi.org/10.3389/fpsyg.2013.00735

Yee, N., & Bailenson, J. (2007). The proteus effect: The effect of transformed self-representation on behavior. *Human Communication Research*, *36*, 285–312. https://doi.org/10.1111/j.1468-2958.2007.00299.x

Young, A. W., Frühholz, S., & Schweinberger, S. R. (2020). Face and voice perception: Understanding commonalities and differences. *Trends in Cognitive Sciences*, *24*(5), 398–410. https://doi.org/10.1016/j.tics.2020.02.001

Yovel, G., & Belin, P. (2013). A unified coding strategy for processing faces and voices. *Trends in Cognitive Sciences*, *17*(6), 263–271. https://doi.org/10.1016/j.tics.2013.04.004

Zäske, R., Schweinberger, S. R., & Kawahara, H. (2010). Voice aftereffects of adaptation to speaker identity. *Hearing Research*, *268*(1), 38–45. https://doi.org/10.1016/j.heares.2010.04.011

Zoefel, B., Abbasi, O., Gross, J., & Kotz, S. A. (2024). Entrainment echoes in the cerebellum. *Proceedings of the National Academy of Sciences*, *121*(34), e2411167121. https://doi.org/10.1073/pnas.2411167121