Self-Voice Perception: A Multidimensional Process

Helene Charlotte Ortmann

Master Thesis - Applied Cognitive Neuroscience

S4288963
10th April 2025
Department of Psychology
University of Groningen
Examiner/Daily supervisor:
Ana P. Pinheiro
Margarida Marques
Tassos Sarampalis

Abstract

The ability to recognize one's own voice is a unique aspect of auditory perception and essential for understanding altered perceptual experiences, such as auditory hallucinations, where impairments in self-other voice discrimination are frequently observed. However, research on self-other voice discrimination remains limited, particularly regarding the roles of self-other acoustic voice distance, vocal congruence, and emotional valence. This study investigated how these factors influence self-other voice discrimination using a voice categorization task with verbal stimuli. Participants (N = 50) completed a task where they identified along a morphing continuum from other to self-voice whether the voice they heard was "More mine" or " More other" while emotional valence (positive, negative, neutral) was manipulated. Results revealed that self-other voice distance and vocal congruence did not significantly predict discrimination accuracy, contradicting hypotheses derived from prior literature. Emotional valence significantly influenced self-other voice discrimination, with emotionally charged words (positive and negative) requiring lower self-voice content to be categorized as "More mine", suggesting an emotional bias in self-voice discrimination. In addition, negative words, in particular, led to less accurate discrimination compared to neutral and positive words. These findings highlight the role of emotional biases in self-voice recognition and suggest that higher-order cognitive mechanisms, rather than acoustic factors, may play a more prominent role in self-other voice discrimination. The study underscores the need for further research on the neural and cognitive processes underlying self-voice recognition and its clinical and forensic applications.

Keywords: Self-voice recognition, Emotional valence, Vocal congruence, Self-other acoustic voice distance

Self-Voice Perception: A Multidimensional Process

The human voice conveys a wealth of information beyond its linguistic content. Within just milliseconds of speech, listeners are capable of extracting cues related to the speaker's identity, including gender, age, emotional state, and even aspects of their physical and social condition (Kreiman & Sidtis, 2011). Recognizing these characteristics is crucial in social interactions and shapes how we engage with others. Among voices, our own holds a unique place. It is often perceived as more attractive by oneself (Hughes & Harrison, 2013) and also found to capture our attention more rapidly than the voices of friends or strangers (Conde et al., 2018; Kirk & Cunningham, 2024). Specifically, Kirk and Cunningham (2024) showed that self-voice elicits faster reaction times and an attentional bias toward self-relevant auditory stimuli. Complementing this, Pinheiro et al. (2023) provide neural evidence that self-voice engages brain mechanisms that enhance attention and processing efficiency compared to other voices. Conversely, difficulties in recognizing one's own voice and discriminating between self and other voices have been linked to an increased predisposition for auditory hallucinations (Pinheiro et al., 2019). Such insights are critical for advancing our knowledge of auditory perception and hold significant implications for diagnosing and treating psychiatric disorders such as schizophrenia. In particular, this study aims to shed light on three factors that might influence self-other voice discrimination, namely self-other voice acoustic distance, vocal congruence, and emotional valence. To do so, we begin by reviewing relevant literature and defining each of these constructs.

How do we Recognize Voices?

Let us start by mapping out the fundamentals of voice perception. It is a complex cognitive process involving specialized brain pathways. Models have been proposed to understand the flow of information, one of which was by Belin et al. (2011), who introduced the "auditory face model". Here, voices are initially processed in low-level auditory regions,

followed by a structural encoding phase. This phase involves three interacting pathways: a speech pathway (left superior temporal sulcus (STS), prefrontal, and premotor regions), an affective pathway (right temporal-medial regions, amygdala, and insula), and an identity pathway (right anterior STS for familiar voices). During audiovisual integration, these pathways are hypothesized to also interact with face-processing systems. The STS combines auditory and visual inputs, working alongside face-processing regions such as the fusiform face area (FFA) and emotional pathways to form a cohesive perception of voices and faces. This indicates that the processing of voices occurs throughout the brain and involves a network of specialized brain pathways that integrate auditory, emotional, and identity-related information, working together with face-processing systems to create a unified percept. Taken together, this model suggests that voice processing is distributed across the brain, relying on a network of specialized pathways that integrate auditory, emotional, and identity-related information, working alongside face-processing systems to form a unified perceptual experience.

Latinus and Belin (2011) further explored voice perception using "anti-voice" stimuli, where a prototype voice was acoustically inverted. By examining how the anti-voice exposure influenced participants' perception of subsequent voices, they observed that adaptation to anti-voices shifted participants' recognition of identities, revealing that voice identities might be organized in a multidimensional space around an internal prototype. This norm-based coding enables the brain to recognize voices by their deviations from the average, similar to face recognition. This was corroborated by Latinus et al. (2013), who used functional magnetic resonance imaging (fMRI) to examine temporal voice areas (TVA) while participants listened to voices varying in their proximity to male and female prototypes. They identified that higher neural responses were elicited for distinctive voices, suggesting enhanced encoding for deviations from these prototypes. These findings suggest that voice

recognition operates through a norm-based coding system, a sophisticated neural 'compass' that orients perception around internal vocal prototypes and responses scaling when acoustic deviations from this average occur.

Building on this framework, Perrachione et al. (2019) pinpointed key acoustic features, such as pitch, harmonics-to-noise ratio, and speech rate, that determine perceived deviations from vocal prototypes, with pitch standing out as the most influential. This feature-based account of voice perception gained neural grounding when Bestelmeyer and Mühl (2022) revealed a cortical division of labor. While primary voice-sensitive areas responded to acoustic similarity (showing adaptation effects), the anterior temporal lobe (ATL) maintained invariant identity representations, demonstrating the brain's dual processing of physical and abstract voice properties. The hierarchy was further clarified by Staib and Frühholz's (2023) fMRI-based study, where participants listened to voices, non-voice sounds, and artificial sounds. Their results revealed that a specific region in the higher-order auditory cortex, located in the superior temporal area, is specialized for processing voices beyond basic acoustic analysis. Together, these findings suggest that voice perception operates through a hierarchically organized neural system where early auditory regions encode fine-grained acoustic details, while higher-order areas (e.g., ATL, superior temporal cortex) extract stable, identity-relevant information. This dissociation could also explain how humans effortlessly recognize voices despite variations in pitch or speech rate, highlighting the brain’s efficiency in separating sensory input from abstract identity representation. This framework provides a basis for the present study, which investigates how both lower-level acoustic factors and higher-order cognitive factors influence self-other voice discrimination.

How do we Discriminate Our Own Voice from Others’?

Now that we have explored how we perceive voices in general, we now turn to the brain's specialized processing of one's own voice. Before examining the specific factors that may influence self-other voice discrimination, it is essential to understand the distinct neural and cognitive mechanisms involved in recognizing one's own voice. This foundation will help contextualize how features such as self-other voice distance, vocal congruence, and emotional valence may shape this process.

A growing body of research demonstrates that self-voice recognition enjoys unique perceptual priority, as evidenced in faster reaction times (Conde et al., 2018), heightened neural responses (Pinheiro et al., 2019), and even biased attractiveness judgments compared to other voices (Hughes & Harrison, 2013). This self-advantage was also investigated by Kirk and Cunningham (2024). Their findings revealed that when participants' voices were included in a voice-label matching task, they consistently demonstrated faster and more accurate responses compared to other voices, regardless of how the voices were labeled. This highlights the brain's inherent prioritization of self-related vocal cues, emphasizing the unique advantage of self-voice in voice processing. and might suggest the reliance on a robust internal representation to facilitate faster and more efficient recognition compared to other voices. These results provide empirical support for the role of self-specific processing in voice discrimination and suggest that factors such as attentional bias and the strength of internal voice representations may critically influence self-other voice differentiation.

The behavioral advantages of self-voice recognition find their neural counterpart in distinct cortical signatures, as Yankouskaya et al. (2021) demonstrated in their study where heightened activity in the medial prefrontal cortex (mPFC) was measured when individuals heard their own voice compared to others' voices. This indicates that the brain prioritizes self-related auditory cues also on a neural level, emphasizing the unique attention given to one's own voice. This self-prioritization mechanism was further dissected by Iannotti et al.

(2022), who mapped the specific neural circuits governing self-other voice discrimination (SOVD). They identified a self-voice-specific EEG topographic map occurring around 345 milliseconds post-stimulus, activating a network that includes the insula, cingulate cortex, and medial temporal lobe. These regions integrate sensory, emotional, and self-referential signals, contributing to self-voice recognition. While EEG studies have been instrumental in delineating the temporal dynamics of this process, their limited spatial resolution necessitates complementary fMRI evidence. However, it is important to interpret these findings with caution, as EEG studies, while valuable for temporal resolution, may lack the spatial precision of other neuroimaging methods like fMRI. Nevertheless, these findings mirror Bestelmeyer and Mühl's (2022) hierarchical model of voice perception, showing that voice analysis begins with acoustical features but also involves higher-order brain regions involved in self-voice representation.

We will now review relevant research on the factors investigated in our study: self-other voice distance, vocal congruence, and emotional valence, which may influence our ability to distinguish our own voice from others.

The Role of Acoustic Voice Distance in Self-Voice Recognition

On the acoustic level, Orepic et al. (2023) explored the basic features of self-voice recognition by placing participants' own voices within a "voice space" alongside an “other” voice, building on the foundational work of Latinus et al. (2013). They found that the closer a voice was acoustically to the participant's own voice (smaller self-other voice distance), the more challenging it became to distinguish between them. This suggests that self-voice recognition relies on fine-grained acoustic distinctions, aligning with Newham's (1998) conclusion that each voice can be described as a unique "sound-print" that everyone possesses. This difficulty in distinguishing acoustically similar voices raises an intriguing

question: How does the brain ensure accurate self-voice recognition, especially when other voices closely resemble our own?

The Role of Vocal Congruence in Self-Voice Recognition

At the intersection of voice perception and self-awareness lies vocal congruence, the degree to which one's perceived vocal identity aligns with one's internal self-concept (Crow et al., 2021). Research has shown that in populations such as aging individuals (Costa, & Matias, 2005), transgender persons (Pickering, 2015), or those who have undergone procedures, such as laryngectomy (Bickford et al., 2013) the voice no longer aligns with one's internal sense of self, and showed disruption of self-perception, emotional well-being, and social integration. These findings point to a deeper relationship between voice and self-representation, one that underscores why vocal congruence matters.

Crow and colleagues (2021) developed the Vocal Congruence Scale (VCS), a 10-item self-report tool, to measure this construct. Preliminary validation of the VCS in vocally healthy participants revealed a moderately negative correlation with the emotional subscale of the Voice Handicap Index (VHI) but not with the Functional or Physical subscales. This suggests that vocal congruence is more closely tied to emotional and identity-related perceptions of the voice rather than its physical or functional characteristics. Tucker et al. (2024) further underscored this relationship, noting that the voice integrates anatomical, physiological, and psychological aspects of the self, making it a primary marker of identity. When vocal congruence is disrupted, it can lead to a diminished sense of self, as individuals may feel their voice no longer reflects who they are, making its recognition more difficult. Additionally, Chong et al. (2024) discovered that an individual's perception of their speaking voice influences a range of behaviors, from personal expression to social interactions, suggesting that one's attitude toward their voice significantly affects personal, interpersonal, and social presentation. Taken together, these studies show that self-voice recognition is

shaped not only by physical or functional voice characteristics but also by identity-related factors.

The Role of Emotional Valence in Self-Voice Recognition

Emotional cues equally play a significant role in attention capture, as evidenced by their widespread use in advertising strategies to engage or influence audiences. Pinheiro et al. (2023) directly demonstrated this phenomenon in voice perception by showing how attention and emotion interact to enhance self-voice prioritization in speech processing. Participants listened to self-voice and other-voice stimuli under different emotional and neutral conditions, with reaction times and accuracy recorded to assess prioritization. EEG was used to measure event-related potentials (ERPs), focusing on components associated with auditory attention and emotional processing. Their study revealed that both self-relevant (i.e., self-voice) and emotional cues are prioritized in perception. They found that words spoken in one's voice elicited stronger neural responses compared to unfamiliar voices, particularly in early sensory stages. The research also showed that emotional content and attention to speaker identity interactively modulate these neural responses, highlighting the complex interplay between self-relevance, emotion, and attention in shaping how self-voice is processed by individuals.

This work builds on earlier findings by Pinheiro et al. (2016), who laid the groundwork for understanding how speaker identity and emotional valence jointly influence speech processing. In their ERP study, 16 healthy participants listened to 420 prerecorded adjectives that varied in voice identity (self vs. other) and emotional valence (neutral, positive, and negative). Participants were asked to determine whether the speech they heard was their own, someone else's, or if they were unsure. The results showed that self-speech with neutral emotional valence elicited a more negative N1 amplitude, a component of the brain's early response to auditory stimuli. Self-speech with positive valence resulted in a

more positive P2 amplitude, which reflects neural processing associated with the evaluation of emotionally relevant stimuli. Additionally, self-speech with both positive and negative valence led to an increased Late Positive Potential (LPP), which is linked to higher-level cognitive processing and emotional regulation. Convergingly, the study showed that participants were more accurate in processing emotionally charged words (positive or negative) when spoken in their voices than other voices. These findings indicate that emotional valence and speaker identity interactively modulate speech processing at both early and late stages, highlighting the intricate relationship between self-relevance and emotion in voice perception.

Expanding on the role of emotion in voice processing, Xu and Armony (2021) examined how emotional prosody, content, and repetition affect memory recognition of speaker identity. Their study revealed that emotional prosody and content significantly enhance memory for the speaker's identity, with repeated exposure further improving recognition. This suggests that emotional and repetitive aspects of speech play a crucial role in how we remember and identify speakers, complementing the findings of Pinheiro et al. (2016) by demonstrating how emotional cues not only influence immediate speech processing but also shape long-term memory for speaker identity.

The Present study

Despite growing interest in self-other voice discrimination, questions remain unanswered about the underlying mechanisms. The present study aims to investigate three under-explored areas, namely the self-other voice acoustic distance, perceived vocal congruence, and the emotional valence of the spoken content and their influence on one's ability to discriminate between self and other voices. We used a voice categorization task with verbal stimuli and applied the voice morphing technique to create ambiguous stimuli, which is crucial for testing how people perceive their own voices under uncertain conditions

(Belin & Kawahara, 2024). This technique allows us to create voice samples that gradually transition between the participant's own voice and another person's voice, making the distinction between "self" and "other" less clear. By doing so, we can systematically investigate how individuals perceive their own voice when it is acoustically altered, shedding light on self-other voice discrimination under more natural but challenging conditions

First, we examine if the distance between the self and the other voice will influence self-voice recognition on a fundamental level. Unlike previous studies (Orepic et al., 2023; Baumann & Belin, 2010), which used isolated vowel recordings, we extracted vowels directly from spoken words. Based on the findings from Orepic et al. (2023), we hypothesize that a smaller acoustic difference between self and other voices will make identifying one's own voice more challenging, leading to lower accuracy in discriminating between the two (Hypothesis 1). Conversely, we anticipate that a greater self-other voice distance would make differentiating between the voices easier, leading to a higher accuracy in discriminating between them.

Furthermore, we will explore whether a heightened sense of vocal congruence (alignment between one's voice and the sense of self) will lead to an enhanced ability to distinguish between the self and the other voice (H 2) and vice versa (Crow et al., 2021). We expect that a lower perception of vocal congruence will make it more challenging to discriminate one's own voice from that of another voice.

Lastly, we examine whether the emotional valence of the stimuli influences the ability to discriminate between self and other voices. First, drawing upon the aforementioned research by Pinheiro et al. (2016), we posit that, due to the self-positivity bias, participants will more easily identify their own voices when presented with positive stimuli. To assess this, we made use of the point of subjective equality (PSE), which refers to the point at which a participant perceives two stimuli as equally likely to be either 'self' or 'other. We expect a

higher point of subjective equality (PSE) for positive words compared to the PSE for neutral words (H 3a). Conversely, in the context of negative words, we expect participants to be more likely to perceive the heard voice as belonging to someone else, resulting in a lower PSE compared to neutral words (H 3b). Second, we examine how emotion influences task accuracy, where we build on evidence from Pinheiro et al. (2016). Specifically, we expect accuracy to follow the same pattern as the PSE, where discrimination accuracy will be higher for positive words compared to neutral words (H 3c) and lower for negative words compared to neutral words (H 3d).

Methods

Ethical approval was granted by the Ethics Committee of the Faculty of Psychology at the University of Lisbon and conducted following the ethical standards laid down in the Declaration of Helsinki for the Project entitled "The Me and the I: Dissociating Ownership and Agency in Sensorimotor Processing (146/2020 – BIAL Foundation)".

Participants

The participants were students from Lisbon who were native speakers of European Portuguese. For the *Self-Other Voice Distance* and the *Emotional Valence Analyses*, the sample included 50 participants (M_{age}= 22.94, SD_{age} =4.06, N_{male} = 22, N_{female}=28). Six participants were excluded from the *Vocal Congruence Analysis* due to incomplete questionnaire responses, resulting in a final sample of 44 participants (M_{age}= 22.15, *SD* =(3.57), N_{male} = 20, N_{female}=24). All participants had normal hearing and no reported history of psychiatric or neurological disorders. Written consent was provided ahead of the experiment.

Stimuli

Word Selection Procedure

From the 284 bisyllabic words of the Portuguese version of the Affective Norms for English Words (ANEW; Soares et. al, 2012) a total of 220 words were deemed suitable for

selection after excluding those with accents, hyphens, non-noun or noun-adjective word classes, or an atypical number of letters for a bi-syllabic structure (fewer than 4 or more than 6 letters). These words were then categorized into three groups based on semantic valence: neutral (90 words), positive (73 words), and negative (57 words). Words were classified as negative if their valence scores were below 4.0, neutral if their scores ranged from 4.0 to 6.0, and positive if their scores exceeded 6.0.

For the final selection of stimuli, we aimed to control for arousal, dominance, and word frequency across the three valence groups, using values calculated from the full ANEW dataset. Positive and negative words had a generally higher arousal compared to neutral words. Additionally, only words with the same number of letters (either 4, 5, or 6) were retained to ensure phonological consistency in the subsequent morphing process.

However, despite our selection criteria, differences in word frequency persisted across the three valence groups. On average, negative words had the lowest frequency (M = 22.85), followed by neutral words (M = 55.26), while positive words were the most frequently used (M = 98.02). To minimize these differences, we prioritized words within similar frequency ranges for each valence group while maintaining the integrity of the emotional categories.

The experimental task initially aimed to include four words per emotional valence (twelve words in total) to maintain statistical power while ensuring the task duration remained manageable. To account for potential issues with voice recording quality or participants' inability to maintain neutral intonation, six words per valence were recorded. From these, the four words with the highest quality and most neutral intonation were selected based on the judgment of two researchers. The final stimuli consisted of 12 bisyllabic words (e.g., "planta"(plant)), which can be found together with their properties in Appendix A.

Recording of the participant's voice

The audio recordings were captured in a soundproof studio using a Roland R-16 recorder, employing a 44.1 kHz sampling rate and 16-bit quantization in front of a computer. One by one, the words appeared in the center of the screen, and after listening to the "voice model," they were instructed to repeat each word five times. To ensure consistency, participants were asked to adjust the loudness and maintain a neutral prosody, matching the intensity of the voice model before starting their recordings. The use of a voice model, as implemented in the study by Pinheiro et al. (2023), helped minimize variability in speech rate, voice loudness, and pitch across participants. Later, the sound files were downloaded as an MP3 file.

Then, using Audacity® (version 3.7.1; Audacity Team, 2025), background noise was filtered out from the recordings. The words were then segmented and normalized to 65 dB SPL (sound pressure level) at the source using Praat® (Boersma et al., 2025) to ensure consistent intensity across stimuli. This normalization was applied to the recorded audio files before playback through headphones during the experimental task.

To ensure age- and sex-matched “other-voice” stimuli, a 25-year-old female (M_{F0}=255 Hz) and a 27-year-old male (M_{F0}= 95 Hz) recorded the words using the same procedure. For the self-voice condition, the participants’ mean F0 was 195 Hz for females and 120 Hz for males.

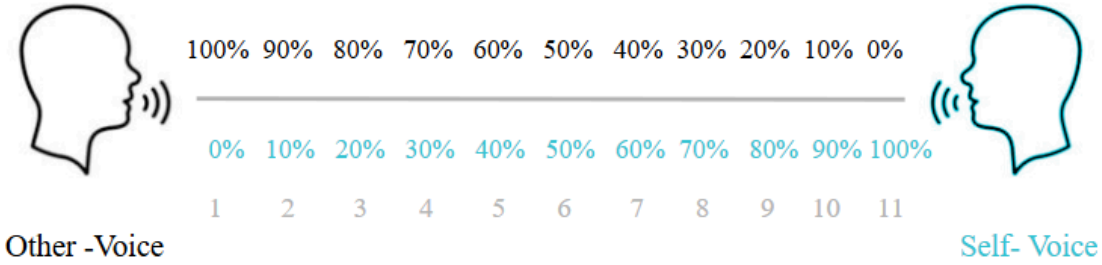
Psychophysical Task

Four words were selected per participant for each emotional category, choosing the most prosodically neutral and clearest token from their recordings. These recordings were then used to create a continuum from "other-voice" to "self-voice" using the TANDEM-STRAIGHT software (Belin & Kawahara, 2024) in MATLAB. Each word was morphed in 11 steps (10% increments), generating 11 stimuli per word, providing a smooth transition between the extremes while ensuring clear perceptual differences. The two extremes represented the unmorphed recordings of the participant’s voice (self-voice) and the

matched “other-voice”, which was selected based on the participant’s sex. The visualization of the experimental stimuli is shown in Figure 1 below.

Figure 1

Visualisation of the Morphing Continuum



Note. The figure illustrates the morphing continuum, beginning with the 'other' voice and progressing in 10% increments toward the 'self' voice. The percentages above the line indicate the proportion of each identity at each step (which are written below). The blue represents the percentage of the “self-voice”, while the black represents that of the “other-voice”.

The task was generated using Qualtrics software (2024) and conducted online. Concurrently, the participants engaged in a *Zoom* (Version 5.0.2) session with the experimenter to ensure their adherence to the established guidelines, including the maintenance of a tranquil environment and the utilization of headphones.

On each trial, participants heard a word from the morphed continuum and responded to “more my voice” or “more other voice” by clicking on one of two buttons labeled “Mais Minha” and “Mais Outra”. Semantic valence (positive, negative, neutral) varied randomly within Qualtrics using its built-in randomization feature across trials. To make sure that participants could identify their own voice, the session started with a training phase, where they were presented with both ambiguous and unambiguous stimuli (i.e., 100% other-voice

and 100% self-voice). Then, in the main task, each stimulus was presented three times, for a total of 396 trials (3 presentations of 12 words, in 11 morph versions).

Questionnaire

After completing the task, participants filled out a questionnaire, first accessing sociodemographic information and answering the question: “How similar do you perceive your own voice to the other voice?” Furthermore, they filled out the Vocal Congruence Scale (Crow et al., 2021), which measures the perceived coherence of one’s voice with one’s self-concept. The scale was translated into Portuguese for this project by two native speakers who were also fluent in English. The translated version was then translated back into English and compared to the original after any inconsistencies were resolved. The two translators reviewed any differences and agreed on the translation into European Portuguese.

Analyses

Self-Other Voice Distance Analysis

To investigate if the self-other voice distance influences the ability to discriminate one's own voice from another (H 1a,b), we performed an acoustic analysis of participants' unmorphed voices and classified them according to the vocal space dimensions as defined by Baumann and Belin (2010).

The acoustic analysis centered on the near-open central vowel [ɐ], a frequent phonological unit in European Portuguese, particularly in unstressed word-final syllables (e.g., ‘culpa’, ‘multa’, ‘porta’). This selection aligns with the natural distribution of [ɐ] in our verbal stimuli, except for ‘natal’, where [ɐ] occurs word-initially. Crucially, this approach diverges from Orepic et al. (2023), who analyzed the open central vowel [a]. Despite its positional constraints, [ɐ] remains acoustically robust for inter-speaker differentiation when extracted via the formant analysis protocol established by Albuquerque et al. (2023), as demonstrated by their validation of vowel formant extraction from naturalistic speech stimuli.

We preserved the methodological compatibility of Orepic et al.(2023) by implementing their coordinate reference system. For each participant, the acoustically clearest [a] token was identified through a quantitative assessment of formant stability and signal-to-noise ratio, ensuring valid cross-study comparisons.

Vocalic parameters were analyzed using Praat © software, with measurements encompassing fundamental frequency (F0) and the first five formant frequencies (F1–F5). Following Orepic et al.'s (2023) biologically grounded framework, we computed sex-specific voice-space coordinates to account for anatomical dimorphism: for male participants, coordinates were derived as x = log(F0) to capture laryngeal source characteristics and y = log(F5 – F4) representing vocal tract filtering properties; for female participants, coordinates followed x = log(F0) and y = log(F1). This systematic approach combines the ecological validity of naturalistic [ɐ] production with the standardized comparative benefits of Orepic et al.'s [a] analysis while maintaining analytical rigor through Albuquerque et al.'s (2023) validated formant extraction protocols for connected speech. The acoustic properties of the vowels can be found in the table below (Table 1).

Table 1

Properties of Extracted Vowels

	Self- Voice	**Other - Voice**
Duration (in ms)	0.112 (0.014)	0.118 (0.014)
F0 (in Hz)	163.06 (51.37)	194.16 (78.47)

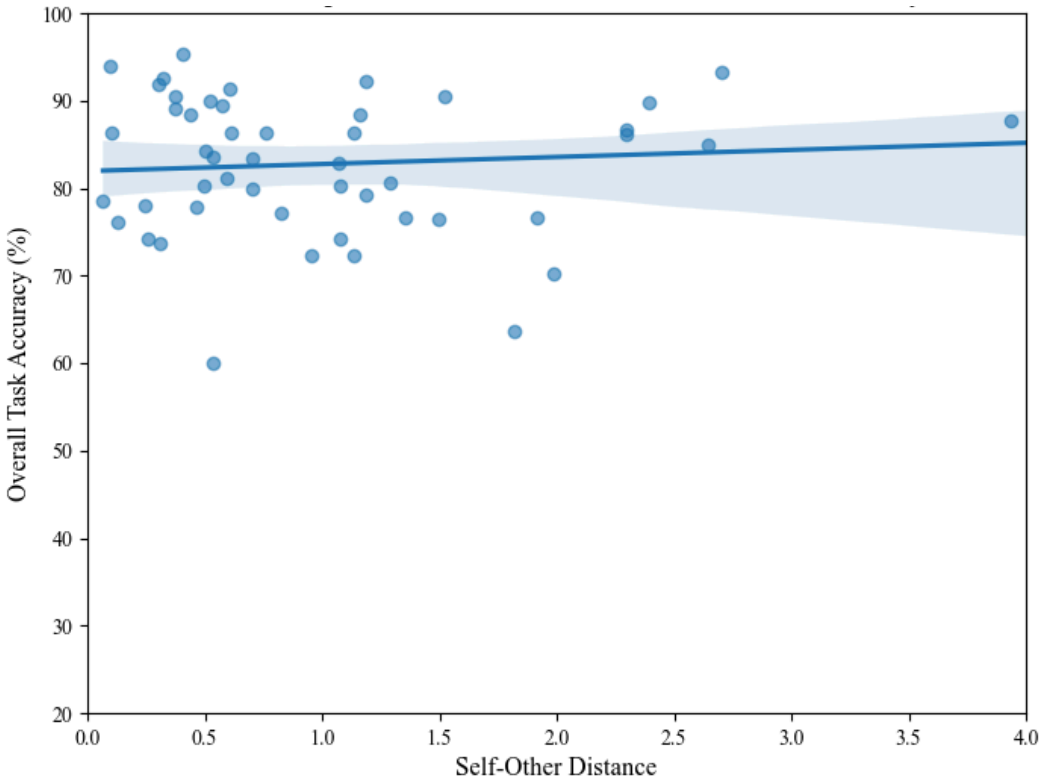
Note. Mean values of acoustic properties (including standard deviation) for the vowel "a" extracted from participants' self-voices versus other voices.

The overall task accuracy was then calculated (M = 82.84%, SD = 7.92, *range* = 60% - 95.28%) to assess the ability of participants to distinguish their own voice from others along the morphing continuum, which varied the amount of self-voice information. The 50% self/other morph step was excluded from these calculations, as the response options ("More mine" or "More other") do not permit a definitively correct answer for this perfectly ambiguous condition. We used task accuracy as it reflects sensitivity to acoustic features and self-voice recognition, aligning with prototype-based models of voice perception (Latinus & Belin, 2011). By capturing deviations from an internal prototype, task accuracy offers insights into the cognitive mechanisms of self-other voice differentiation.

To examine the relationship between self-other voice distance (including both female and male voices) and overall task accuracy, a Pearson correlation was conducted. The results revealed a non-significant positive correlation, r = .111, p = .444, which is inconsistent with H1, visualized below in Figure 2.

Figure 2

Correlation Between Overall Task Accuracy and Self-Other Voice Distance



Note. Scatterplot showing the relationship between overall task accuracy (y-axis) and self-other voice distance (x-axis). Each data point represents one participant. The solid line represents the line of best fit, and the shaded area indicates the 95% confidence interval.

Next, a simple linear regression was conducted to assess whether self-other voice distance predicts self-voice discrimination performance, with overall task accuracy as the dependent variable and self-other voice distance as the independent variable.

The regression model was non-significant, $F(1, 48) = 0.60$, $p = .444$, with voice distance accounting for only 0.12% of the variance in task accuracy ($R^2 = .012$). The coefficient for self-other voice distance was also non-significant ($\beta = 0.110$, $p = .444$).

Given the previously mentioned gender differences and the approach taken by Skuk and Schweinberger (2013) in separating voice space by gender, we also analyzed the correlation between voice distances and task performance separately for male and female participants. The results can be found in the supplementary material.

Exploratory Analysis

As part of our exploratory analysis, we evaluated both the perceived and objective relationships between voice distance and discrimination accuracy. Participants subjectively rated the similarity between their own voice and the other voice via self-report, providing a measure of perceived self-other voice distance. We then computed a Pearson correlation to examine the association between perceived voice distance and task accuracy. This approach enabled a direct comparison of how subjective voice similarity judgments and objective acoustic divergence independently predict self-voice discrimination performance. The results also indicated a nonsignificant negative correlation between perceived self-other voice distance and task accuracy, r = -0.18, p = .209.

Vocal Congruence Analysis

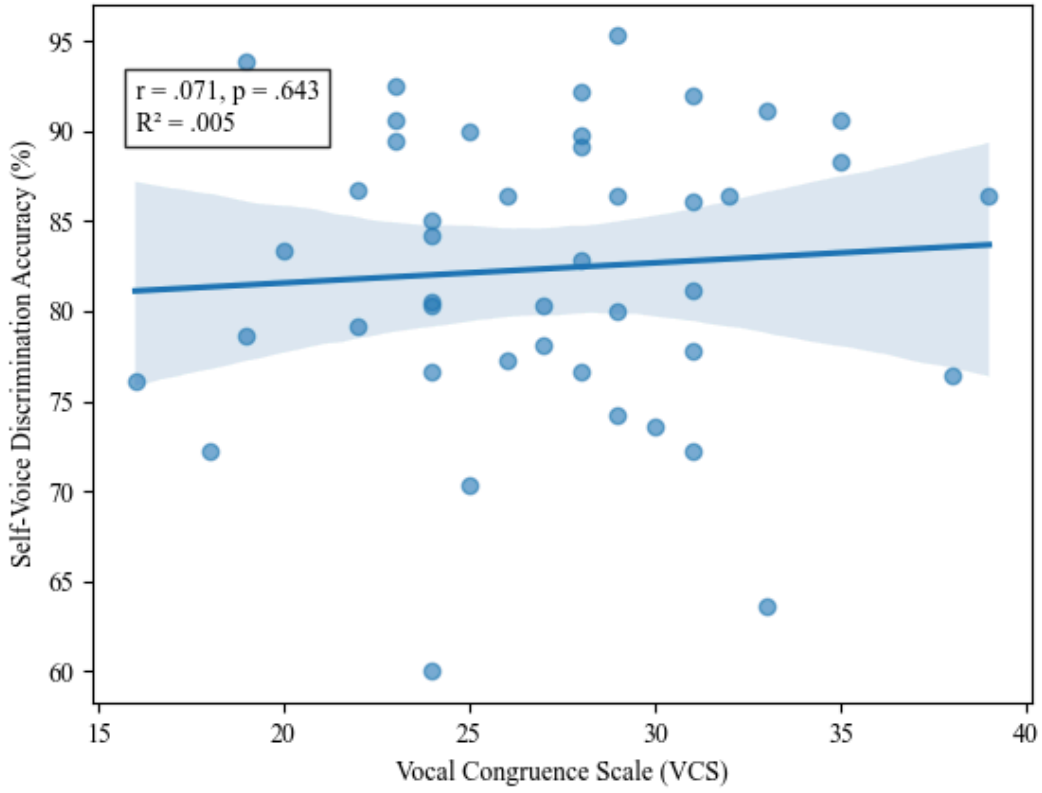
To examine the theorized influence of vocal congruence – conceptualized as the alignment between vocal characteristics and self-representation – upon self-voice discrimination capacity (H2), we derived two principal metrics. First, the aggregated Vocal Congruency Scale (VCS) scores (M = 27.07, SD = 5.15, *range* =16 - 39), which reflect subjective vocal self-perception. Second, the overall task accuracy (M = 82.84%, SD = 7.92%, *range* = 60% - 95.28%), serving as an objective behavioral measure of discrimination performance. Here, the 50% self/other voice morphing step was also excluded from the calculation, as this midpoint offers no objectively correct response, given the binary choice format. Together, these complementary measures allow us to evaluate the interplay between subjective and perceptual dimensions in voice identity processing.

Subsequent analyses were conducted in two phases. First, a bivariate Pearson correlation assessed the relationship between these constructs, revealing a non-significant positive correlation, r =.071, p = .643, which can be seen in the plot below (Figure 3). Next, a simple linear regression model was used to evaluate the predictive validity of VCS scores

(independent variable) on discrimination accuracy (dependent variable). The regression model was non-significant, $F(1, 42) = 0.15$, $p = .643$, with vocal congruence accounting for only 0.5% of the variance in accuracy ($R^2 = .005$).

Figure 3

Relationship between Vocal Congruence and Task accuracy



Note. This scatter plot illustrates the relationship between participants' scores on the Vocal Congruence Scale (VCS) and their accuracy in the task. Each dot represents an individual participant, with VCS scores plotted on the x-axis and task accuracy (in percentage) on the y-axis. The regression line (solid line) shows the trend, while the shaded area represents the 95% confidence interval.

Additional analyses examining the relationship between vocal congruence and task accuracy are provided in the supplementary material. These include (1) task accuracy for 0% and 100% self-voice in relation to vocal congruence and (2) comparisons of task accuracy

between participants with low and high vocal congruence. All analyses yielded non-significant results.

Emotional Valence Analysis

To investigate whether the emotional valence of the stimuli influences self-other voice discriminability, we conducted several analyses.

Effect of Emotional Valence on Self-Other Voice Discrimination

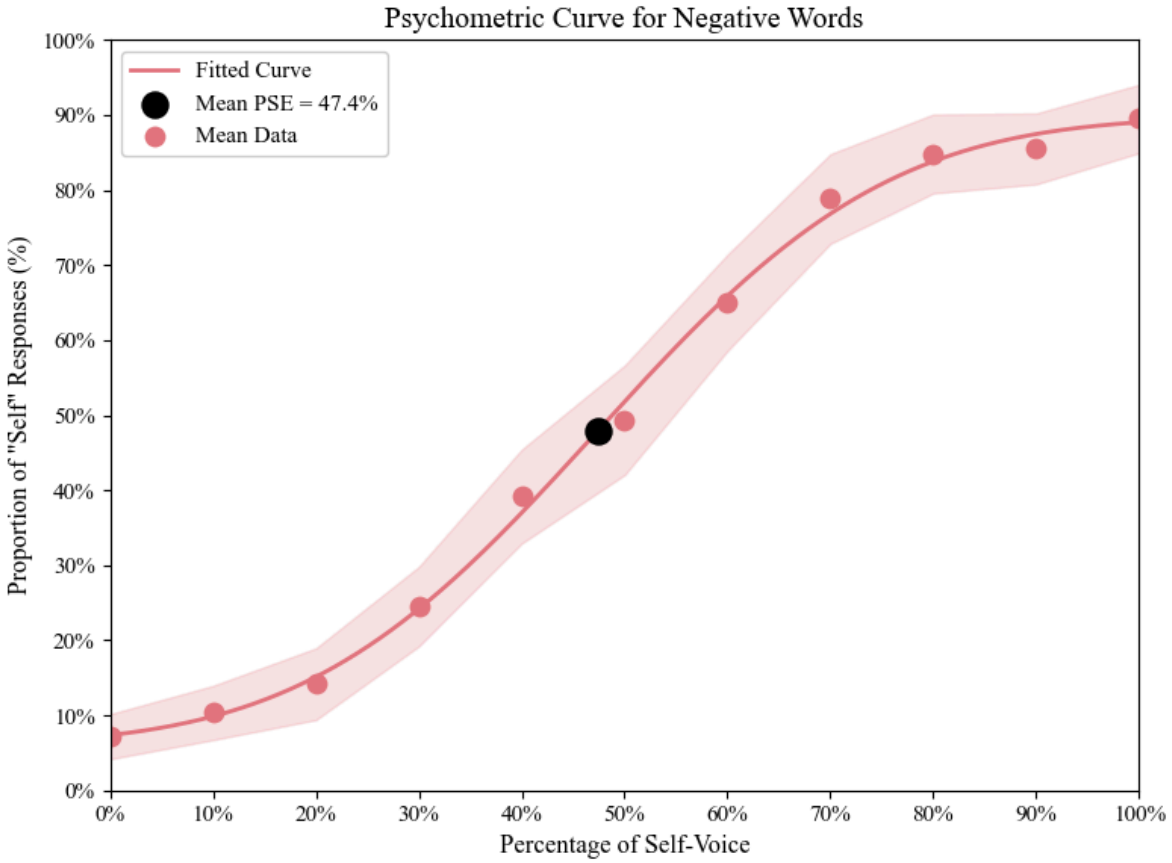
Initial analyses focused on calculating the PSE for each emotional word category (positive, neutral, negative) using the MATLAB function *FitPsycheCurveLogitWH* (Jenkins, 2020). This approach identified the precise morphing continuum threshold where participants perceived voices as equally likely to be self- or other-generated, allowing us to detect emotion-specific variations in voice discrimination sensitivity.

The results showed that the mean PSE was highest for neutral words ($M = 5.32$, $SD = 1.04$), while positive ($M = 4.73$, $SD = 1.10$) and negative words ($M = 4.77$, $SD = 1.62$) resulted in lower PSE values. The corresponding psychometric functions (Figure 4) visually demonstrate these valence-dependent shifts in discrimination sensitivity per emotion.

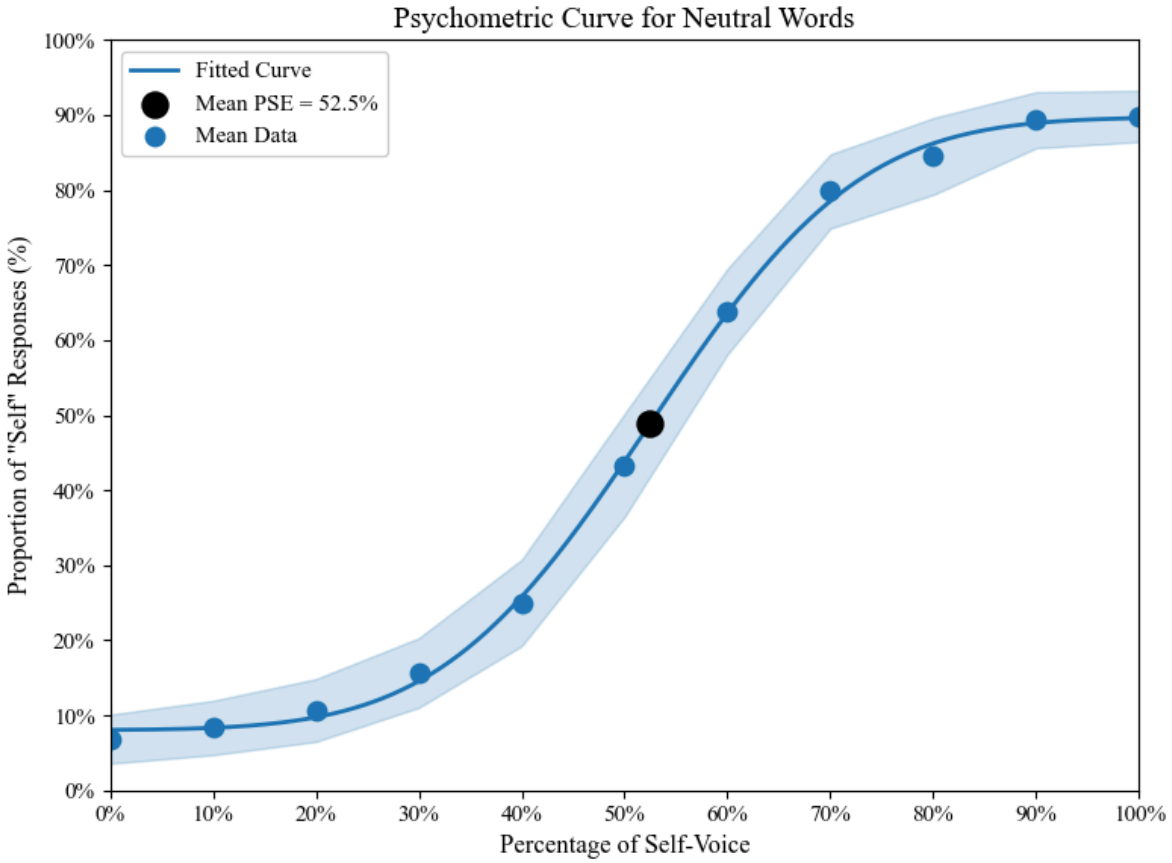
Figure 4

Psychometric curves for self-other voice discrimination across emotional conditions

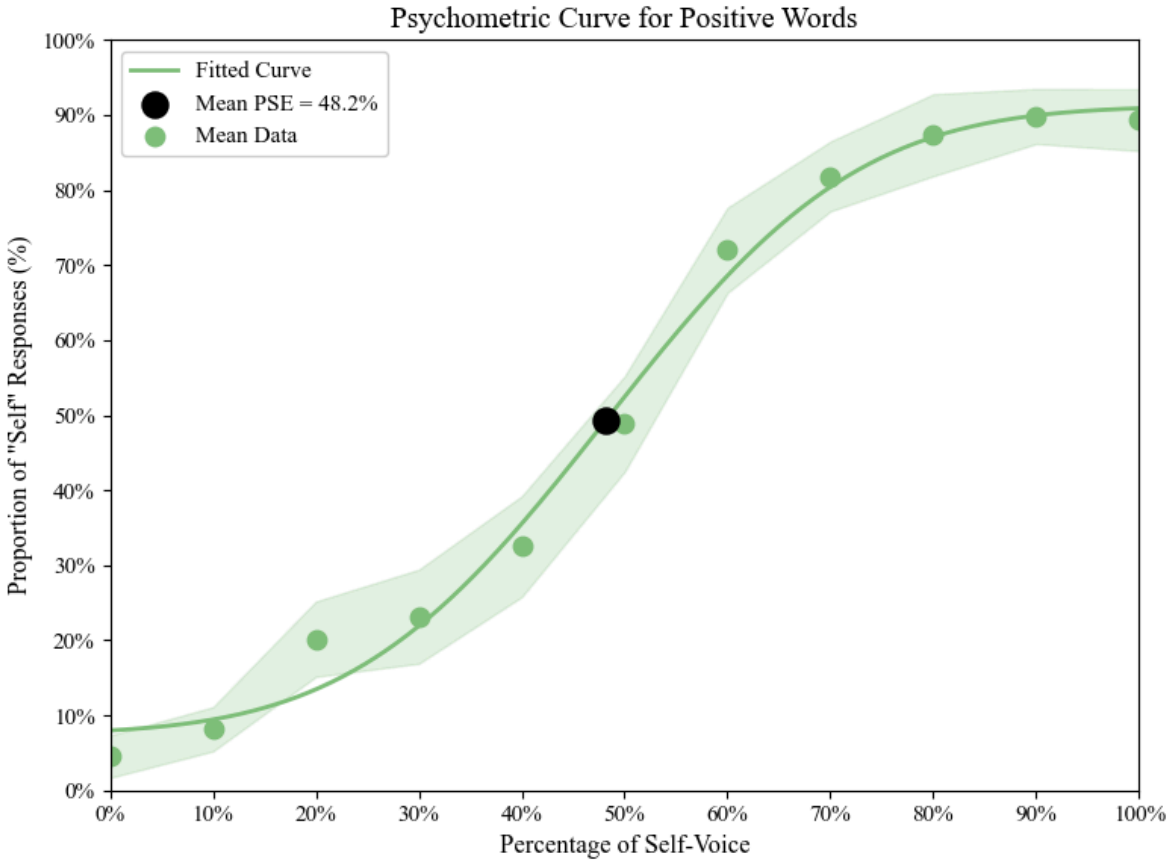
(a)


Psychometric Curve for Negative Words
Fitted Curve
Mean PSE = 47.4%
Mean Data
Proportion of "Self" Responses (%)
Percentage of Self-Voice
0%
10%
20%
30%
40%
50%
60%
70%
80%
90%
100%

(b)


Psychometric Curve for Neutral Words
Fitted Curve
Mean PSE = 52.5%
Mean Data
Proportion of "Self" Responses (%)
Percentage of Self-Voice
0%
10%
20%
30%
40%
50%
60%
70%
80%
90%
100%

(c)



Note. Psychometric curves fitted for different emotions: (a) negative, (b) neutral, and (c) positive during self-other voice discrimination. The x-axis represents the percentage of self-voice in the stimuli, and the y-axis shows the proportion of "self" responses. Shaded areas around each curve represent 95% confidence intervals. The black dot marks the point of subjective equality (PSE).

To formally test emotional valence effects, we conducted a one-way repeated-measures ANOVA with valence as the within-subjects factor and PSE as the dependent variable. The analysis revealed a significant main emotion effect, $F(2, 98) = 4.95$, $p = .009$, $\eta^2 = .092$.

Post-hoc comparisons using Bonferroni correction revealed significant differences in the PSE between neutral and positive words ($p = .017$, $d = -.458$) and between neutral and

negative words (p = .031, d = -.424). No significant difference was found between positive and negative words (p = 1.000, d = -.034).

Effect of Emotional Valence on Task Accuracy

We first computed task accuracy scores separately for each emotional valence condition (positive, negative, neutral) to quantify participants' self-voice discrimination performance across affective contexts objectively and excluded again the 50% self/other voice morphing step. This valence-specific accuracy metric directly tested our hypothesis (H 3c) regarding the emotion's influence on voice identification. The results are shown in the table below (Table 2).

Table 2

Task Accuracy Across Valence Conditions

Condition	Mean Accuracy (%)	SD
Negative	80.75	9.73
Neutral	83.73	8.89
Positive	84.03	8.77

Note. Descriptives of the task accuracy per emotion.

Next, a one-way repeated-measures ANOVA was conducted to examine the effect of emotional valence (positive, negative, neutral) on task accuracy, with emotional valence as

the within-subjects factor. The results indicated a significant main effect of emotional valence, $F(2, 98) = 5.30$, $p = .007$, $\eta^2 = .098$.

Post hoc tests with Bonferroni correction revealed significant differences in accuracy between positive and negative words ($p = .026$, $d = .326$) and between negative and neutral words ($p = .012$, $d = -.359$). No significant difference was found between positive and neutral words ($p = 1.000$, $d = -.033$).

Explorative Analysis

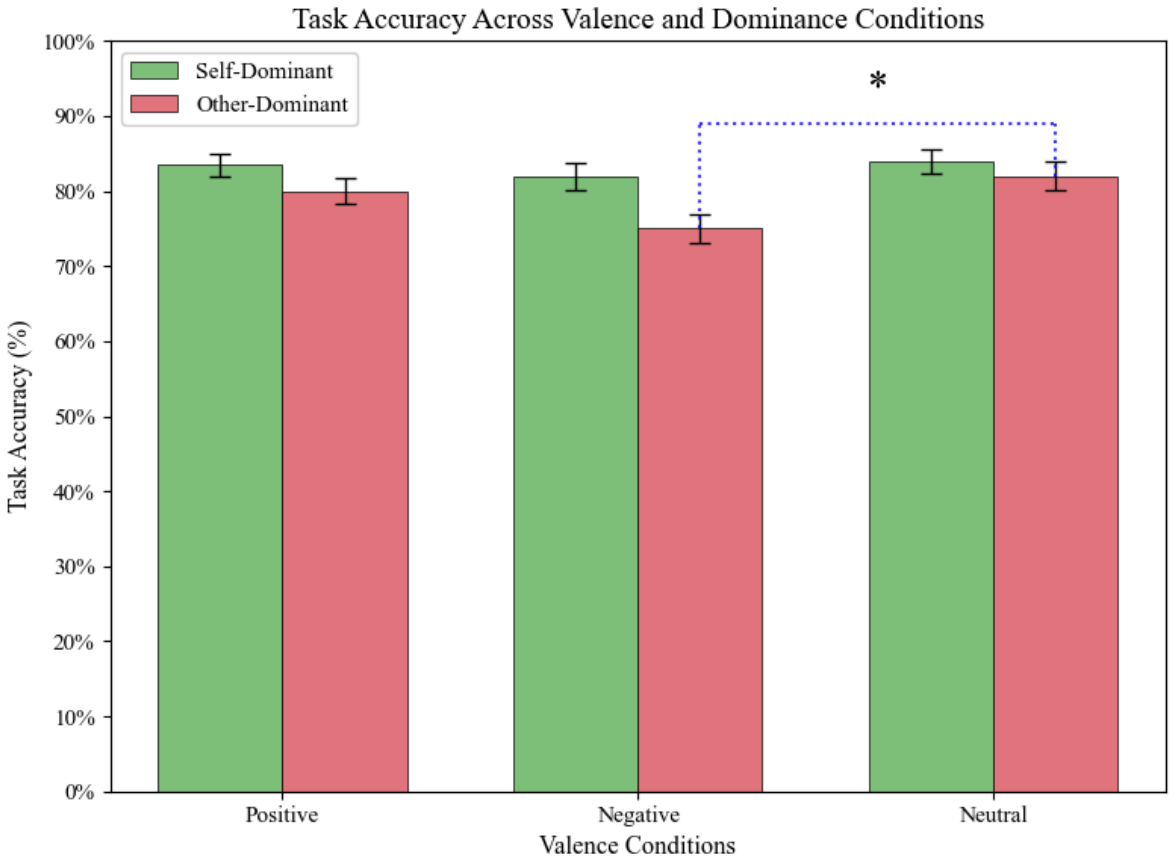
Moving beyond our confirmatory analyses, we investigated how emotional valence interacts with voice dominance. First, task accuracy across valence and identity conditions was calculated. Building on Pinheiro et al. (2016), who reported enhanced accuracy for emotional stimuli dominated by the self-voice, we conducted two separate one-way repeated-measures ANOVAs—each including emotional valence as a within-subjects factor. The analyses were split by voice dominance: one focused on self-dominant trials (60–100% self-voice morphs) and the other on other-dominant trials (0–40% self-voice morphs), followed by Bonferroni-corrected post-hoc comparisons. Visualization of the results can be found below in Figure 5.

Self-Dominant Condition. No significant effect of emotional valence on task accuracy was found, $F(2,98) = 1.88$, $p = .157$, $\eta^2 = .037$.

Other-Dominant Condition. In contrast, the ANOVA revealed a significant effect of emotional valence on accuracy, $F(2,98) = 5.25$, $p = .007$, $\eta^2 = .097$. Post hoc Bonferroni-adjusted comparisons revealed a significant difference between negative and neutral valences ($p = .005$), with lower accuracy for negative stimuli. However, no significant differences were observed between positive and negative ($p = .442$) or positive and neutral valences ($p = .237$).

Figure 5

Task accuracy across valence and dominance condition



Note. This grouped bar plot compares task accuracy for positive, negative, and neutral valence conditions in both Self-Dominant (morphs 60%–100% self-voice) and Other-Dominant (morphs 0%–40% self-voice) conditions. The star indicates the significant difference between neutral and negative words in the other dominant condition.Error bars represent standard error.

To complement these findings, we ran two additional analyses provided in the supplementary materials. First, a 3 (valence: positive, neutral, negative) × 2 (identity: self, other) repeated-measures ANOVA on overall task accuracy was conducted. Second, we examined unambiguous voice trials (0% and 100% self-voice) in two separate one-way

repeated-measures ANOVAs to assess the influence of valence independent of morphing ambiguity. All additional analyses yielded non-significant results.

Discussion

Self-Other Voice Distance Analysis

The first hypothesis (H1) proposed that the acoustic distance between self and other voices would influence the ability to discriminate between them. Specifically, we hypothesized that a smaller self-other voice acoustic distance would lead to lower accuracy in self-voice recognition (H1). Contrary to this prediction, the results did not support the influence of acoustic self-other voice distance on self-voice recognition. The Pearson correlation between self-other voice distance and overall task accuracy revealed a non-significant positive relationship (p= .643), and the regression analysis showed that vocal congruence accounted for only 0.5% of the variance in accuracy. These results suggest that the acoustic similarity between self and other voices does not fundamentally impair or enhance the ability to discriminate one's own voice. This finding contrasts with the predictions of Orepic et al. (2023), which proposed that a greater self-other voice distance would facilitate self-voice recognition, while a smaller distance would hinder it. Looking at our exploratory analysis in which we correlated perceived self-other voice distance, we also did not find a significant result, indicating that self-other voice distance - subjective or objective (acoustical) - does not influence task accuracy. Self-voice discrimination appears to rely on higher-order neural mechanisms, particularly as the stimuli becomes more complex, such as using natural speech instead of isolated vowels as Orepic et al.(2023). Conde et al. (2018) demonstrated that increased stimulus complexity engages additional cognitive and perceptual resources, which likely contribute to self-voice recognition. This is further supported by Bestelmeyer & Mühl (2022), who identified that regions associated with

self-awareness, emotional processing, and interoception—specifically the bilateral anterior insulae and medial frontal gyri—play a key role in differentiating self-voice from other voices. Similarly, our findings suggest that even though we extracted a vowel, the complexity of the word itself exerted a stronger influence on task accuracy, engaging higher-order neural cognitive processes.

Vocal Congruence Analysis

The second hypothesis proposed that vocal congruence - the alignment between one's voice and their self - would influence the ability to discriminate between self and other voices. Specifically, we hypothesized that higher vocal congruence would lead to better discrimination (H2) and vice versa. Contrary to these expectations, the Pearson correlation between VCS scores and overall task accuracy revealed a non-significant relationship (p = .704), and the linear regression showed that vocal congruence only explained 0.3% of the accuracy variance. These results suggest that vocal congruence might not be linked to the ability of self-other voice discrimination. This lack of a significant relationship may reflect the complexity of vocal congruence as a construct. While it is thought to reflect a relationship between one's own voice and one's self-concept, it may not directly translate into improved (or impaired) perceptual or behavioral performance in self-other voice discrimination tasks. Factors like exposure to one's recorded voice and cognitive biases may play a larger role in recognition than congruence alone.

Emotional Valence Analysis

In contrast to the null findings in the previous hypotheses, our results provided strong evidence demonstrating that emotional valence significantly influenced self-other voice discrimination. Following our hypotheses, the results support H 3a, as positive stimuli were associated with lower PSE values than neutral stimuli. Participants identified the stimuli as their own more quickly—that is, they required a smaller percentage of their voice to be

present in the stimuli to perceive it as their own. However, for H 3b, the findings were the opposite of what was predicted; the negative words were also associated with lower PSE values than neutral words. This can be seen in higher Point of Subjective Equality (PSE) values for neutral words compared to both positive and negative words. This aligns with Pinheiro et al. (2023), suggesting an emotional bias in self-voice recognition, which can also be understood as an arousal bias.

We then examined the influence of emotional valence on task accuracy. Our comparison revealed significantly lower accuracy for negative words compared to neutral ones ($p = .012$), supporting Hypothesis 3d. However, no significant difference was found between positive and neutral words, which contradicted H 3c. When considered alongside the PSE findings, these results reveal a dissociation between PSE and accuracy. While emotional stimuli may bias participants toward identifying a voice as their own, even when it contains more 'other-like' features, this bias appears to come at the cost of reduced precision, resulting in increased errors in discrimination.

The additional explorative analysis revealed that in the self-dominant condition (60–100% self-voice), emotional valence had no significant effect on task accuracy, likely reflecting the strong familiarity and automaticity of self-voice processing. Conversely, in the other-dominant condition (0–40% self-voice), accuracy was significantly lower for negative valence compared to neutral. This suggests that negative emotional content amplifies confusion in identity judgments when self-voice cues are sparse, potentially reflecting a self-referential negativity bias—participants misattributed other-dominant voices as their own more frequently with negative words. However, the absence of similar effects for positive valence raises questions: if negative valence disrupts accuracy, why does positive valence not elicit comparable impairments or enhancements? A possibility is that negative words uniquely destabilize perceptual boundaries along the morphing continuum, particularly in

identity-ambiguous (other-dominant) contexts. This aligns with the PSE results, where the emotional bias was driven by other-dominant stimuli, further underscoring the interaction between emotional salience and self-representation in voice perception.

This aligns with the findings of Colliot et al. (2024), who explored the impact of negative emotional stimuli on working memory. Their research indicates that negative stimuli capture attentional resources, ultimately impairing the ability to maintain and manipulate information in working memory. Considering this in the context of Latinus and Belin's (2011) prototype-based model of voice identity, it seems that emotionally loaded stimuli may restrict access to the prototype, leading to lower task accuracy.

Limitations and Future Research

While our findings contribute to understanding self-voice recognition, several limitations should be acknowledged, which also open avenues for future research. One notable limitation is the sample size, which even had to be reduced once (vocal congruence analysis) and remained relatively small, potentially influencing the results. Future studies could aim for a larger and more diverse sample, considering factors such as age, social background, and ethnicity, to increase the robustness and generalizability of findings. Another potential limitation concerns the non-significant results for self-other voice distance, which may be attributed to the methodological approach. Unlike Orepic et al. (2023), who recorded a single vowel, we extracted vowels from words that varied between participants. This variation could have led to a less precise vowel representation, potentially impacting the results. Future research could explore whether using vowels extracted from words versus pure vowel recordings influences self-other voice distance and task performance and whether standardizing the word choice across participants improves precision. In such a design, emotional valence could be conveyed through prosody, allowing greater control over acoustic

features and allowing for an exploration of how emotional prosody may influence self-voice recognition processes.

Similarly, the VSC used in our study was a translated version, and potential discrepancies in translation may have influenced participants' responses due to linguistic and cultural differences in self-perception and voice evaluation.

A more detailed investigation into the VSC, such as identifying specific items that are particularly relevant to self-other voice discrimination, could also provide deeper insights. Furthermore, we relied on normative valence ratings for the words and did not verify their perceived valence with participants. For example, the word “fight” (luta) was classified as neutral based on these norms. Future studies could include participant-based valence ratings to confirm whether the emotional manipulation was effective as intended.

Beyond methodological refinements, future research could explore the response time in the task, as it may offer insights into the cognitive processes behind this emotional bias in self-other voice discrimination. Examining how quickly participants identify voices with different emotional valence could reveal how emotional cues influence the attention and efficiency of voice perception, and shedding light on the automaticity or effort involved in distinguishing self from other voices.

Furthermore, when considering studies related to hallucinations, a different pattern of response to negative words was observed. Individuals prone to hallucinations (Pinheiro et al., 2019) or those diagnosed with hallucinations (Pinheiro et al., 2016) were more likely to identify negative words as belonging to the 'other' voice. In contrast, our study found a tendency to associate negative words more with the self. Since their study did not use morphed stimuli, it would be interesting to explore what response patterns emerge when using morphed stimuli in people diagnosed with auditory hallucinations. Lastly, neurophysiological measures such as EEG or fMRI could help elucidate the underlying

cognitive and neural mechanisms involved in processing self-related auditory stimuli. For instance, looking into the P2, which reflects early processing, and the Late Positive Potential, which captures higher-order emotional and self-referential integration (Pinheiro et al., 2016). This would provide a deeper understanding of how the brain distinguishes self-voice from other voices, particularly in emotionally salient contexts.

Practical Implications

From an applied perspective, these findings have implications for various domains, including clinical research and forensic voice identification.

The emotional bias observed in self-voice discrimination suggests that individuals may be more susceptible to misattributions of voice identity under emotional conditions. This is particularly relevant in clinical contexts, such as schizophrenia, where symptoms like auditory hallucinations often involve self-other voice misattributions (Amorim et al., 2022; Pinheiro et al., 2019). Understanding how emotional valence influences voice recognition may help refine therapeutic interventions targeting voice identity distortions in psychiatric populations.

In forensic settings, where voice recognition plays a role in legal proceedings, our results emphasize the need for caution when relying on witness testimony regarding voice identity (Sherrin, 2015), especially in emotionally charged situations. Since emotional stimuli increase misidentifications, forensic protocols should consider how emotional context might distort recognition accuracy.

Conclusion

While self-other voice distance and vocal congruence did not significantly predict recognition accuracy, emotional valence played a crucial role in modulating self-other voice discrimination. The findings imply that self-voice recognition is a complex, multidimensional process that cannot be fully explained by low-level acoustic features alone. Instead, it relies

on the interplay of sensory input and higher-order cognitive and emotional processing, particularly as stimuli increase in complexity (e.g., whole words versus isolated vowels). These findings lay the groundwork for further exploration of the factors shaping self-other voice discrimination beyond those examined here. They suggest practical implications for various domains, including clinical research and forensic voice identification.

References

Albuquerque, L., Oliveira, C., Teixeira, A., Sa-Couto, P., & Figueiredo, D. (2023). A Comprehensive Analysis of Age and Gender Effects in European Portuguese Oral Vowels. *Journal of Voice*, 37(1), 143. https://doi.org/10.1016/j.jvoice.2020.10.021

Amorim, M., Roberto, M. S., Kotz, S. A., & Pinheiro, A. P. (2022). The perceived salience of vocal emotions is dampened in non-clinical auditory verbal hallucinations. *Cognitive Neuropsychiatry*, 27(2-3), 169–182. https://doi.org/10.1080/13546805.2021.1949972

Audacity Team (2025). Audacity ®: Free Audio Editor and Recorder [Computer program] Version 3.7.1 retrieved from http://audacity.sourceforge.net/

Baumann, O., & Belin, P. (2010). Perceptual scaling of voice identity: common dimensions for different vowels and speakers. Psychological Research PRPF: *An International Journal of Perception, Attention, Memory, and Action*, 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* (London, England: 1953), 102(4), 711–725. https://doi.org/10.1111/j.2044-8295.2011.02041.x

Belin, P., & Kawahara, H. (2024). STRAIGHTMORPH: A Voice Morphing Tool for Research in Voice Communication Sciences. *Open Research Europe*, 4, 154. https://doi.org/10.12688/openreseurope.18055.2

Bestelmeyer, P. E. G., & Mühl, C. (2022). Neural dissociation of the acoustic and cognitive representation of voice identity. *NeuroImage*, 263. 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.

Boersma, Paul & Weenink, David (2025). Praat: doing phonetics by computer [Computer program]. Version 6.4.26, retrieved from http://www.praat.org/

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. https://doi.org/10.1016/j.jvoice.2022.02.003

Colliot, P., Plancher, G., Fournier, H., Labaronne, M., & Chainay, H. (2024). Effect of negative emotional stimuli on working memory: Impact of voluntary and automatic attention. *Psychonomic Bulletin & Review*, 1–9. https://doi.org/10.3758/s13423-024-02593-2

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

Costa, H. Matias, C. (2005). Vocal impact on quality of life of elderly female subjects. *Brazilian Journal of Otorhinolaryngology*, 71(2): 172-178.

Crow, K. M., van Mersbergen, M., & Payne, A. E. (2021). Vocal Congruence: The Voice and the Self Measured by Interoceptive Awareness. *Journal of Voice*, 35(2), 324. https://doi.org/10.1016/j.jvoice.2019.08.027

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.

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

Jenkins, G. (2020). Psychometric curve fitting. GitHub. https://github.com/garethjns/PsychometricCurveFitting

Kirk, N. W., & Cunningham, S. J. (2024). Listen to yourself! Prioritization of self-associated and own voice cues. *British Journal of Psychology* (London, England: 1953). https://doi.org/10.1111/bjop.12741

Kreiman, J., & Sidtis, D. (2011). Foundations of voice studies: An interdisciplinary approach to voice production and perception. *Wiley-Blackwell*. https://search.ebscohost.com/login.aspx?direct=true&scope=site&db=nlebk&db=nlabk&AN=503338

Latinus, M., & Belin, P. (2011). Anti-Voice Adaptation Suggests Prototype-Based Coding of Voice Identity. *Frontiers in Psychology*, 2. 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*: CB, 23(12), 1075–1080. https://doi.org/10.1016/j.cub.2013.04.055

Newham, P. (1998). Therapeutic voicework: Principles and practice for the use of singing as a therapy. *Jessica Kingsley Publishers Ltd*

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

Perrachione, T. K., Furbeck, K. T., & Thurston, E. J. (2019). Acoustic and linguistic factors affecting perceptual dissimilarity judgments of voices. *The Journal of the Acoustical Society of America*, 146(5), 3384. https://doi.org/10.1121/1.5126697

Pickering, J. (2015). Transgender Voice and Communication: Introduction and International Context. *Perspectives on Voice and Voice Disorders*, 25: 25-31.

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., 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., 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

Qualtrics. (2024). Qualtrics survey software [Software]. https://www.qualtrics.com

Skuk, V. G., & Schweinberger, S. R. (2013). Gender differences in familiar voice identification. *Hearing Research*, 296, 131–140. https://doi.org/10.1016/j.heares.2012.11.004

Soares, A. P., Comesaña, M., Pinheiro, A. P., Simões, A., & Frade, C. S. (2012). The adaptation of the Affective Norms for English Words (ANEW) for European Portuguese. *Behavior Research Methods*, 44(1), 256–269. https://doi.org/10.3758/s13428-011-0131-7

Sherrin, C. (2015). Earwitness evidence: The reliability of voice identifications. *Osgoode Hall Law Journal*, 52(3), 819–862.

Staib, M., & Frühholz, S. (2023). Distinct functional levels of human voice processing in the auditory cortex. *Cerebral Cortex* (New York, N.Y. : 1991), 33(4), 1170–1185. https://doi.org/10.1093/cercor/bhac128

Tucker, A. E., Crow, K., Wark, M., Eichorn, N., & van Mersbergen, M. (2024). How Does Our Voice Reflect Who We Are? Connecting the Voice and the Self Using Implicit Association Tests. Journal of Voice: *Official Journal of the Voice Foundation*. https://doi.org/10.1016/j.jvoice.2024.10.018

Xu, H., & Armony, J. L. (2024). Arousal level and exemplar variability of emotional face and voice encoding influence expression-independent identity recognition. Motivation and Emotion, 48(3), 464–483. https://doi.org/10.1007/s11031-024-10066-1

Yankouskaya, A., Sui, J., & Molnar-Szakacs, I. (2021). Self-Positivity or Self-Negativity as a Function of the Medial Prefrontal Cortex. *Brain Sciences*, 11(2). https://doi.org/10.3390/brainsci11020264

Appendix A

Psycholinguistic Properties of The Word Selection

Properties of Negative Words

E - Word	EP - Word	Valence M	Arousal M	Dom M	Frequ	Letters N	Syl N
sad	triste	2.02	5.40	3.34	77.60	6	2
fault	culpa	2.28	6.49	3.28	48.92	5	2
penalty	multa	2.29	6.68	3.81	21.59	5	2
crude	bruto	2.46	6.38	4.53	22.58	5	2
trash	lixo	2.69	5.11	4.28	40.41	4	2
bullet	bala	2.80	6.76	4.62	13.63	4	2
horror	horror	2.88	6.64	3.95	21.71	6	2

Note. E-Word = English word; EP-Word = Equivalent Portuguese word; Dom M = Mean dominance; Frequ = Frequency of the word in usual usage; SylN = Number of syllables.

Properties of Neutral Words

E - Word	EP - Word	Valence M	Arousal M	Dom M	Frequ	Letters N	Syl N
fight	luta	4.05	6.92	4.88	129.73	4	2
hay	feno	4.67	3.58	4.60	3.08	4	2

door	porta	4.98	3.83	4.69	303.57	5	2
saint	santo	5.19	3.07	4.42	68.47	5	2
plant	planta	5.98	3.86	5.50	72.42	6	2
hotel	hotel	5.98	3.63	4.97	73.04	5	2
fight	luta	4.05	6.92	4.88	129.73	4	2

Note. E-Word = English word; EP-Word = Equivalent Portuguese word; Dom M = Mean dominance; Frequ = Frequency of the word in usual usage; SylN = Number of syllables.

Properties of Positive Words

E - Word	**EP - Word**	**Valence M**	**Arousal M**	**Dom M**	**Frequ**	**Letters N**	**Syl N**
party	festa	7.57	6.52	6.03	135.41	5	2
gift	prenda	7.93	5.25	5.89	4.87	6	2
christmas	natal	8.02	5.03	6.89	15.98	5	2
laughter	riso	8.03	6.12	6.67	35.78	4	2
pleasure	prazer	8.18	6.86	7.04	64.96	6	2
cake	bolo	7.55	4.55	5.71	25.48	4	2

Note. E-Word = English word; EP-Word = Equivalent Portuguese word; Dom M = Mean dominance; Frequ = Frequency of the word in usual usage; SylN = Number of syllables.

Supplementary Material: Self-Voice Perception: A Multidimensional Process

Self-Other Voice Distance Analysis

Gender differences

The following are the results for analysing the relation between voice distances and task performance separately for male and female participants.

Female Participants. Descriptive statistics for female participants indicated that the mean self-other distance was $M = 1.02$ ($SD = 0.87$), with scores ranging from 0.15 to 4.10. Pearson's correlation analysis revealed a nonsignificant negative relationship between self-other distance and overall task accuracy, $r = -0.23$, $p = .245$. A simple linear regression was conducted to predict overall task accuracy based on self-other distance. The model was not statistically significant, $F(1, 26) = 1.42$, $p = .245$, and explained only 5.17% of the variance in overall task accuracy ($R^2 = .052$), indicating that self-other distance did not significantly predict overall task accuracy for females.

Male Participants. Descriptive statistics for male participants indicated that the mean self-other distance was $M = 1.21$ ($SD = 0.75$), with scores ranging from 0.21 to 2.72. Pearson's correlation analysis indicated a nonsignificant positive relationship between self-other distance and overall task accuracy, $r = 0.22$, $p = .327$. A simple linear regression was conducted to predict overall task accuracy based on self-other distance. The model was not statistically significant, $F(1, 20) = 1.01$, p = .327, and explained only 4.80% of the variance in overall task accuracy ($R^2 = .048$), indicating that self-other distance did not significantly predict overall task accuracy for males.

Vocal Congruence Analysis

Task accuracy for 0% and 100% self-voice in relation to vocal congruence

0% Self-Voice. Performance on the whole task at 0% self-voice ($N = 50$) had a mean score of $M = 93.72$ ($SD = 8.41$), with scores ranging from 61.11 to 100. Pearson's correlation

analysis revealed a nonsignificant relationship between the score on the VCS and performance on the whole task at 100% self-voice, $r = 0.07$, $p = .630$.

100% Self-Voice. Performance on the whole task at 100% self-voice ($N = 50$) had a mean score of $M = 89.89$ ($SD = 12.15$), with scores ranging from 38.89 to 100, which were lower than at 0% self-voice. Pearson's correlation analysis revealed a nonsignificant relationship between the VCS and performance on the whole task at 100% self-voice, $r = 0.07$, $p = .630$.

Comparisons of task accuracy between participants with low and high vocal congruence

Participants were divided into two groups based on their VCS scores using a median split, with the median score of 27.5 serving as the cutoff. Those with scores greater than 27.5 were classified as having high vocal congruence, while those with scores equal to or less than 27.5 were classified as having low vocal congruence.

Low Vocal Congruence. For participants with low vocal congruence (N = 22), the mean overall task accuracy was $M = 81.43$ ($SD = 7.99$), with scores ranging from 60 to 93.89. The mean VCS score (in percentage) was $M = 57.39$ ($SD = 7.34$), with scores ranging from 40.00 to 67.50. Pearson's correlation analysis revealed a nonsignificant relationship between overall task accuracy and the VCS score, $r = -0.01$, $p = .951$.

High Vocal Congruence. For participants with high vocal congruence (N = 22), the mean overall task accuracy was $M = 83.27$ ($SD = 8.09$), with scores ranging from 63.61 to 95.28. The mean VCS score (in percentage) was M = 77.95 (SD = 8.00), with scores ranging from 70.00 to 97.50. Pearson's correlation analysis revealed a nonsignificant relationship between overall task accuracy and the VCS score, $r = -0.06$, $p = .790$.

Emotional Valence Analysis

Effects of emotion and identity on task performance

A one-way repeated measures ANOVA was conducted to examine the effects of emotion and identity on task performance. The analysis revealed a significant main effect of emotion, $F(2, 98) = 5.30$, $p = .007$, $\eta^2 = .098$, indicating that task performance varied significantly across the three emotion conditions. However, there was no significant main effect of identity, $F(1, 49) = 0.36$, $p = .551$, $\eta^2 = .007$, suggesting that task performance did not differ between the self-dominant and other-dominant conditions. The interaction between emotion and identity was not statistically significant, $F(2, 98) = 2.49$, $p = .088$, $\eta^2 = .048$, indicating that the effect of emotion on task performance did not depend on the identity condition.

Effects of task accuracy at 0% and 100% self-voice for each emotion

0% self-voice. Descriptive statistics for accuracy at 0% self-voice showed a mean accuracy of 95.33 (*SD* = 10.26) for the positive words, 92.67 (*SD* = 10.86) for the negative words, and 93.17 (*SD* = 11.75) for neutral words. A one-way repeated measures ANOVA for accuracy at 0% self-voice revealed no significant effect of emotional condition on task accuracy, $F(2, 98) = 1.35$, $p = 0.26$, indicating that when the voice was not self-relevant, emotional valence did not impact the ability to discriminate the voice.

100% self-voice. For task accuracy at 100% self-voice, the mean accuracy for the positive words was 95.33 (*SD* = 10.26), for the negative words 92.67 (*SD* = 10.86), and for the neutral words 93.17 (*SD* = 11.75). A one-way repeated measures ANOVA was performed to examine accuracy differences across emotional conditions at 100% self-voice. The results indicated a significant effect of emotional condition on accuracy, $F(2, 98) = 3.24$, $p = 0.03$, $\eta^2 < 0.01$, suggesting that the emotional content of the voice influences accuracy in recognizing self-voice.