Effects of Self-similarity and Self-Generation on the Perceptual Prioritisation of Voices

Victor Rosi[1], Bryony Payne[1,2], and Carolyn McGettigan[1]

[1] Department of Speech Hearing and Phonetic Sciences, University College London, United Kingdom

[2] Department of Psychology, King's College London, United Kingdom

Word count: 7,168

Author note

Experimental data and analysis scripts in R for the three experiments are available at https://osf.io/93xq4/.

Abstract

The Self-Prioritisation Effect (SPE) reflects the ability to efficiently discern self-relevant information. The self-voice emerges as a crucial identity marker, due to its inherent self-relevance, and previous work has demonstrated the perceptual and cognitive advantages of the self-voice over other voices. Yet, the extent to which humans prioritise their self-voice when they hear it is because it is both self-similar ("That sounds like my voice") and self-generated ("I said that") remains understudied. Here, we examined impacts of self-similarity and self-generation on the SPE through 3 experiments. In each experiment, participants learned associations between three voices and three identities (self, friend, other), then performed a task requiring them to perceptually match the heard voices with visual labels ("you", "friend", "stranger"). Experiment 1 revealed an augmented SPE when the self-associated voice in the task was the participant's own self-similar and self-generated voice. In Experiment 2, the SPE was diminished when the self-voice was associated with the "stranger" label – here, the other-associated, but self-similar and self-generated, voice was similarly prioritised to a self-associated but unfamiliar voice. In Experiment 3 we investigated the role of self-generation, by associating the self with a self-similar but machine-generated audio clone of the participant. The SPE was again enhanced. In sum, we demonstrate that listeners show flexibility in their mental representation of self, where multiple sources of self-related information in the voice can be jointly and severally prioritised, independently of self-generation. These findings have implications for the application of self-voice cloning within voice-mediated technologies.

Keywords: Voice Identity, Self-Prioritisation Effect, Perceptual Matching, Voice Cloning

Public Significance Statement

The human voice is a highly self-relevant stimulus and an important expression of self-identity. Here we show that voices whose self-relevance is enhanced by being acoustically similar to the self, or by being associated with the self in a task, are perceived faster and more accurately. We also harness state-of-the-art voice cloning technology to show that self-similar voices are prioritised even when they are machine generated.

Introduction

Humans possess the ability to quickly and efficiently recognise information relevant to their own identity. Crucially, the self-voice emerges as one of the most significant identity markers due to its high perceptual familiarity and its self-generated nature (McGettigan, 2015) - for a voice to exist, an action needs to occur, thus tightly linking this familiar auditory marker of self identity to speaker's agency over their vocal anatomy. Therefore, beyond the perceptual biases brought by bone conduction during speech production (Maurer & Landis, 2009; Orepic et al., 2023), the memory of self-generated vocal information is also an important factor of self-recognition as a proof of our identity (Mulligan & Lozito, 2004). Numerous experiments have indicated that both perceptual and cognitive processing of the self-voice benefits from its self-relevance in comparison to other voices (Candini et al., 2014; Hughes & Harrison, 2013; Pinheiro et al., 2019; Xu et al., 2013). However, the extent to which people perceptually prioritise their own voice, precisely because it is self-relevant and self-generated, needs further investigation.

A wealth of studies has shed light on how previously unfamiliar information can become self-relevant and be subsequently prioritised, an effect known as the Self-Prioritisation Effect (SPE)(Golubickis et al., 2020; B. Payne et al., 2021; S. Payne et al., 2017; Schäfer et al., 2016; Sui et al., 2012). In their seminal study, Sui et al., (2012) employed a perceptual matching paradigm to investigate self-prioritisation. Specifically, they explored the degree to which a geometric shape was perceptually prioritised when associated with the self. To do this, participants were taught three pairings between shapes and the labels "YOU", "FRIEND", and "STRANGER". Subsequently, participants were asked to decide as quickly as possible whether a shape presented onscreen matched the pre-associated label. Participants demonstrated quicker and more accurate recognition of the shape associated to the self (i.e. "YOU") relative to the other shapes. These results unveiled how ownership over a stimulus yields perceptual prioritisation. This finding has been replicated with different stimuli in numerous studies, demonstrating the ability to broaden the concept of self-relevance to unfamiliar information (Golubickis et al., 2020; Schäfer et al., 2016; Sui et al., 2014; Sui & Humphreys, 2017).

Using this perceptual matching paradigm, B. Payne et al. (2021) provided evidence that even unfamiliar voices – which are, inherently, signals relating to others (and their actions) and not to oneself – can become self-relevant and prioritised via perceptual matching. Until recently, the perceptual matching paradigm has, purposefully, only explored the degree of prioritisation afforded to previously unfamiliar stimuli. Yet, Kirk & Cunningham (2024) have newly demonstrated that when it is not an unfamiliar voice that is self-associated but, rather, one's own recorded voice, the degree of prioritisation afforded to that voice is greater than that applied to an unfamiliar and newly self-associated voice. This may not be surprising, but it highlights that the "true" self-voice, when heard as a recording, garners an enhanced degree of prioritisation, likely because it is inherently self-relevant, due to numerous factors including it being both self-similar ("That sounds like my voice") and self-generated ("I said that").

As a cognitive construct, the mental representation of self is regarded as flexible through time and composed of multiple parameters relative to a social context (Oyserman et al., 2012; Sforza et al., 2010). For example, in a study using perceptual matching with faces,

S. Payne et al. (2017) demonstrated that the self-relevance of newly self-associated information may be neither complete, nor long-lasting. Before and after the perceptual matching task, participants assessed the similarity between their self-face and morphed images including both the newly self-associated face and their self-face. While the perceptual matching task showed an SPE, there was no consequent increase in the perceived similarity between the newly associated face and the participant's self-face.

The flexibility of the self has also been observed for voice perception. Using perceptual matching, Kirk & Cunningham (2024) proposed to swap the ownership of voices by associating participants' own recorded voice to other identities (i.e., "FRIEND" or "STRANGER") and an unfamiliar voice to the self. As a result, they consistently observed the perceptual prioritisation of participants' self-voice, regardless of the identity with which the self-voice was temporarily associated in the task – although the absence of data from control listeners meant that they were unable to account for self-prioritisation independent of stimulus effects (as their three independent participant groups each contributed, by design, a unique set of self-voice recordings). An observed consistency in self-voice prioritisation could be explained by the attentional capture ability of the self-voice (Conde et al., 2018; Daryadar & Raghibi, 2015; Pinheiro et al., 2023) and self-related information in general (Alexopoulos et al., 2012; Brédart et al., 2006; Ocampo & Kahan, 2016). The finding that the self-voice was prioritised, regardless of the in-task association, suggests that prioritisation may be hierarchical based on stimulus self-relevance. Specifically, one's self-voice, as a stimulus both resemblant of and generated by the self, is prioritised more highly than a new voice, which has not been self-generated, nor (at least by design) resembles the self, and is only temporarily self-relevant.

Previous findings showing dissociable effects of voice ownership and agency on early neural responses to heard speech stimuli (Hubl et al., 2014) suggest that the prioritisation of voices may be separably influenced by factors related to self-similarity and self-generation. To address the relationship between self-relevance and voice prioritisation we propose above, we therefore need to better understand the contributions of self-similarity and self-generation as components of self-relevance for voice perception. Specifically, we can ask whether self-similarity and self-generation are independently apt to enhance voice prioritisation, or whether there are contingencies: For example, given the nature of the self-voice as intrinsically linked to bodily agency, any enhanced prioritisation of one's own self-voice relative to a newly self-associated voice (see Kirk & Cunningham, 2024) may be dependent on having produced the voice recording oneself. This is important, because advanced deep-learning tools have recently enabled the highly natural synthesis of human voice identities. Known as voice cloning, this process commonly relies on generative models that capture key identity markers in a voice, such as timbre and fundamental frequency, to automatically produce speech through text-to-speech synthesis (Arık et al., 2018). Thus, these models can generate speech with high acoustic and perceptual similarity to a speaker's own self-voice, even if the resulting content was never self-generated. Importantly, if a person's cloned voice is as self-*relevant* for voice perception as their own self-voice, it could offer great benefits to individuals with motor neurone diseases who are losing capacity for speech motor control (Cave & Bloch, 2021). By leveraging voice cloning, it would be possible to investigate the perceptual prioritisation that individuals may assign to a self-similar but not self-generated artificial voice. Recent evidence that a sense of ownership over a *non*-self-generated, *non*-self-similar synthesised voice identity was alone sufficient for it to become perceptually prioritised (Payne et al., 2024) suggests that synthetic voices indeed offer a suitable medium for investigating self-voice prioritisation.

In the present study, we therefore investigated how, as aspects of overall self-relevance, the self-similarity and self-generation of the true self-voice influence the self-prioritisation effect (SPE). To do this, we conducted three experiments, each employing a perceptual matching task with voice stimuli (B. Payne et al., 2021). In Experiment 1, we aimed to quantify the impact of the self-relevance of the true self-voice on the SPE. Participants recorded their own self-voice, then in a perceptual matching paradigm heard their voice congruently paired with the label “YOU”, while two unfamiliar voices were associated with “FRIEND” and “STRANGER”. In a control group, participants heard only unfamiliar voices associated with the three identity labels. The inclusion of a control group crucially allowed us to determine the effects of self-relevance independent of stimulus-related effects (cf Kirk & Cunningham, 2024). If perceptual prioritisation of voices is affected by stimulus self-relevance, then the greater self-relevance of the self-voice (as both self-similar and self-generated) should generate an enhanced SPE for participants who heard their own self-voice, relative to participants for whom the very same voice was heard as unfamiliar and arbitrarily self-associated. If prioritisation is unaffected by the self-relevance of the stimulus, we should instead see a similar SPE across the two participant groups.

In Experiment 2, we investigated how the SPE is affected when the self-voice is associated with an incongruent identity: specifically, by associating the self-voice with “STRANGER” and an unfamiliar voice with "YOU" in an experimental group, and again comparing the experimental group's performance with a group of control participants hearing the same voices and associations. Here, if perceptual prioritisation of voices is affected by stimulus self-relevance, then the SPE should be diminished by the competing association between the participant's self-voice and a non-self identity association, relative to a group of control participants. If prioritisation is not affected by stimulus self-relevance, there should be no difference between the two participant groups in the perceptual matching task.

Experiments 1 and 2 thus allowed us to test whether the greater self-relevance of the self-voice, as a stimulus that is both self-similar and self-generated, modulates the SPE, or whether the SPE is instead immutable and independent of stimulus self-relevance. However, neither of these experiments, nor those of Kirk & Cunningham (2024), can determine which component of stimulus self-relevance might underpin its impacts on the SPE: its self-similarity or its self-relevance. In Experiment 3 we dissociated these components by examining the SPE for self-associated voice clones, and comparing the performance of participants who heard their own voice clones with control participants who heard the clones as unfamiliar self-associated identities. Self-prioritisation of unfamiliar self-associated synthetic voices has been demonstrated before (Payne et al., 2024). However, voice cloning allows us to create specific self-associated voice stimuli that have high acoustic similarity to the self-voice but which are, crucially, artificially (and therefore externally) generated. If self-relevance-related enhancement of the SPE is independent of the self-generation of the stimulus, we should observe an enhanced SPE for participants hearing a self-associated self-voice clone, compared with control participants listening to self-associated clones of unfamiliar voices. Conversely, if self-generation is a critical component of self-relevance for self-prioritisation, we should see an equivalent SPE in the two participant groups.

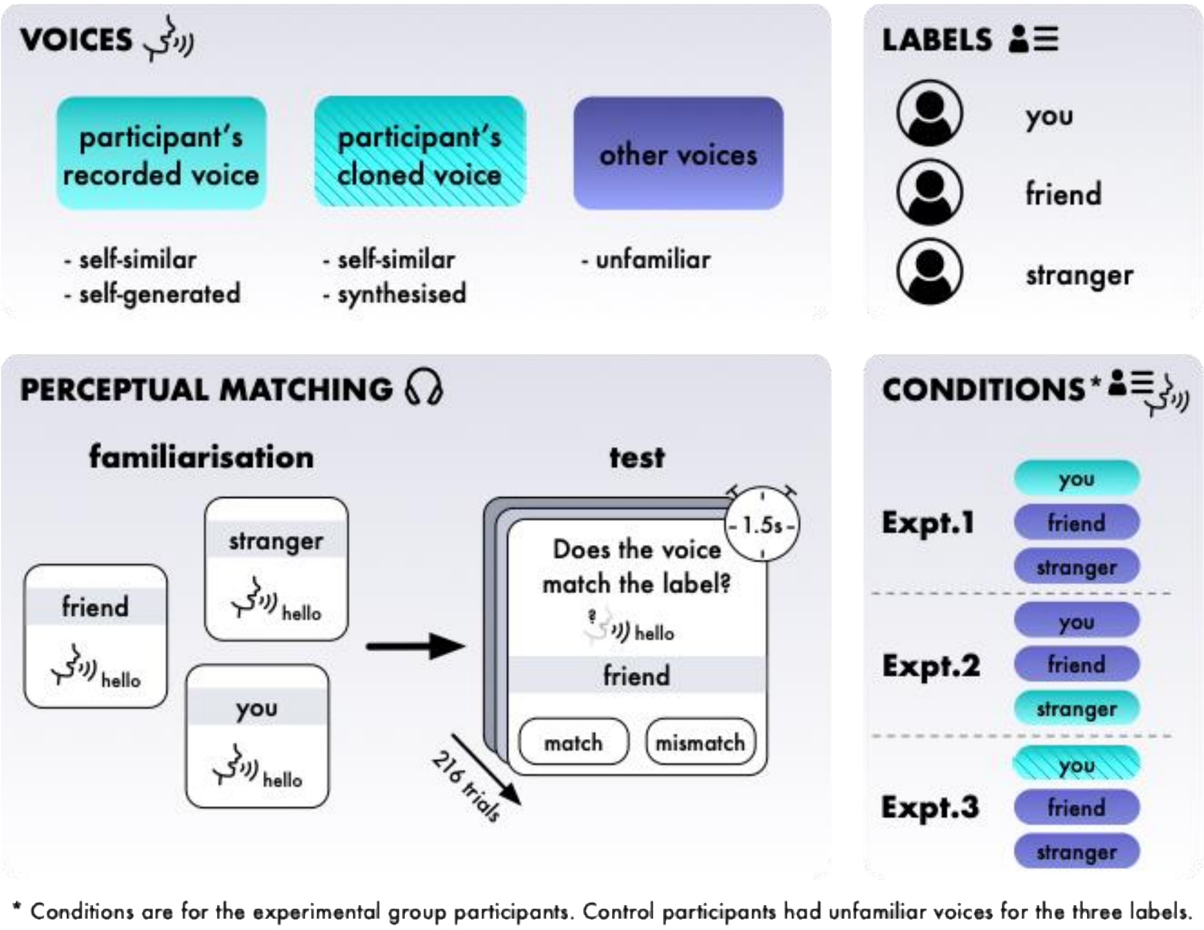

Figure 1 - Overview of the study. The labels "you" and "stranger" correspond respectively to the identities "self" and "other".

Experiment 1

In this first experiment, we investigated the impact of pairing the self-associated voice with the participant's own recorded voice on the SPE.

Methods

Participants

According to an *a priori* sample size computation using G*Power (Faul et al., 2009) based on a pilot experiment with the same paradigm, a sample size of at least 56 participants was sufficient to observe an effect of moderate size for the interaction (η^2=0.035) with α set to 0.05 1-β to 0.80 and the non-sphericity correction to 0.75. Due to requirements from other similar experiments being conducted in the lab (not presented in this paper), a total of 96 participants completed the task. They were split into two groups of 48 experimental participants (mean age= 27.2 years, age range=18-40, 28 females, 20 males) and 48 control participants (mean age=25.6 years, age range=18-40, 28 females, 20 males). Participants were recruited online via Prolific (www.prolific.com) as British monolingual speakers of English with a standard Southern British accent and no hearing impairment or difficulties. Eligible participants had 90% approval rate on Prolific. Ethical approval was obtained (SHaPS-2019-CM-030), and all participants gave their informed consent prior to the testing. Participants were paid for their participation at a rate of £7.50 per hour.

Stimuli

All participants encountered three voice identities, each associated with one of the labels “SELF”, “FRIEND”, and “STRANGER”. Each voice identity provided one sample of the word “hello”. For experimental participants, the self-associated voice was their own recorded self-voice, while the friend-associated and other-associated voices were unfamiliar voices. For each experimental participant, a matched control participant encountered the same trio of voice identities. In this way, we could compare prioritisation effects across groups to investigate the manipulation of interest, while also controlling for stimulus variance. All voices were gender matched to the participants' stated gender.

Self-associated stimuli

Experimental participants were invited to a recording session via the online testing platform Gorilla (www.gorilla.sc; Anwyl-Irvine et al., 2020) prior to the main task. Before starting the recording session, participants were asked to make sure to be in a quiet environment with a working microphone on a computer, and to do a microphone check. To collect natural recordings of the participants saying the target word "Hello" during the recording session, it was first embedded in various sentential contexts, which participants were asked to read aloud across 9 trials. Before each trial, the text was displayed on the screen for 20 seconds. Then, after a countdown, participants were asked to start reading the text. Trials 1 to 9 aimed at accustoming participants to recording their voice, such that they sound more natural afterwards. Following this, participants completed three additional trials where they repeated the word "Hello" ten times per trial. A sample of the word "hello" was then extracted from these last two trials for each participant and paired with the label "YOU" in the experiment.

Friend-associated and Other-associated stimuli

Four participants (mean age=31.8, age range=29-35, 2 females, 2 males) not taking part in the main task were recruited to record their voice. They were British monolingual speakers of English with a Standard Southern British English accent. Recordings of

“Hello” from these voices were used in the main task as either the “FRIEND” or “STRANGER” voice. For both genders, the pairing of the two voices with the labels “FRIEND” and “STRANGER” was counterbalanced across participants.

All stimuli were RMS-normalised using PRAAT (Boersma & van Heuven, 2001). See also the Supplemental Materials for details of stimulus durations across all experiments in the current study.

Procedure

Participants were invited to do the task online via Gorilla on a computer. They were asked to ensure that they were in a quiet environment with minimal background noise. They were also asked to wear headphones and to do a sound check prior to the task. During the task, they used the keyboard to submit responses.

The perceptual matching task consisted of two phases, a familiarisation phase and a test phase (see *Perceptual Matching* in Fig. 1).

Familiarisation phase

Before the familiarisation phase, participants were told that they will be presented three voices: “one that belongs to you”, “one that belongs to a friend”, and “one that belongs to a stranger”. For the experimental participants of this study, the self-associated voice was introduced as “your own voice (previously recorded) which belongs to you”. Then, each voice-label pair was presented four times by simultaneously playing the voice recording through the headphones and displaying the label on the screen. Each label remained on-screen for 3000ms and the voice sample was played 500ms after the label's onset. A fixation cross was displayed for 500ms between presentations of voice-label pairs. The order of presentations was randomised. The overall familiarisation phase lasted one minute.

Test phase

Participants trained for the task in 12 practice trials before the main test phase. The test phase consisted of three blocks of 72 trials. For both stages, the order of the trials was randomised. Each trial consisted of a fixation cross displayed on screen for 500ms, followed by a voice stimulus. Right after the voice sample, a label (i.e., "YOU", "FRIEND", or "STRANGER") was displayed on screen for 1500ms. Participants were then requested to answer as fast as they could whether the voice and the label were a match by pressing either the left-arrow key (MATCH) or the right-arrow key (MISMATCH) on their keyboard. Visual feedback on accuracy was given after each trial. If participants did not answer before the 1500ms limit of presentation of the question, the displayed feedback was "too slow". During the main test phase, an indication of the overall accuracy was given at the end of each block. The experiment lasted 25 minutes on average. There was an equal number of match and mismatch trials and a balanced distribution of "label-voice identity" combinations for the mismatch trials. The order of trials was randomised.

Analysis

To quantify the SPE, we assessed three dependent variables based on participants' responses: trialwise reaction time (in milliseconds), trialwise accuracy (correct/incorrect), and sensitivity (d') per voice. For the analysis of our results, the design of our model comprised *voice identity* (self, friend, other) as the first fixed factor. As our question was "how does the self-voice impact the SPE when it is self-associated?", our second factor was the *condition* (experimental: own recorded self-voice as self; control: new, unfamiliar voice as self). In accordance with previous perceptual matching studies (B. Payne et al., 2021; Sui et al., 2012), we excluded trials with responses quicker than 200ms and longer than 1500ms, and analysed only the match trials, i.e., trials where the voice identity matched the label, as the SPE does not affect mismatch trials. Further, only reaction times for match trials with a correct response were included in the analysis.

Both reaction time data and sensitivity were assessed with linear mixed models (LMM) using *lme4* in the R environment. They included an interaction between *voice identity* and *condition*, and *participant* as a random factor. For example, for reaction time, the model was:

$$reaction\ time \sim condition * voice\ identity + (1|participant)$$

To assess accuracy, we used a binomial generalised linear mixed model (GLMM), with trialwise accuracy coded as 1 (correct) or 0 (incorrect) and including the same fixed factors and random factor as for the assessment of reaction time and sensitivity.

For each model we tested the significance of the interactions and the effects by performing likelihood ratio tests. Depending on the significance of the factors *voice identity* and *condition* and their interaction, we ran *post-hoc* pairwise comparisons using *emmeans* with a Bonferroni correction for 9 pairwise comparisons.

Transparency and Openness

Sample size for the three experiments was consistent and chosen to be larger than the sample size used in the reference study (B. Payne et al., 2021). All experimental data and analysis scripts in R for the three experiments are available at https://osf.io/93xq4/. Data were analysed using R (4.4.0) and the packages *lme4* (1.1-35.3) and *emmeans* (1.10.1).

Results & Discussion

Descriptive statistics for reaction time, accuracy and sensitivity (d') across conditions and voice identities are given in Table 1. See Supplemental Materials for a full report on the statistical tests (https://osf.io/93xq4/).

Table 1. Mean RT, accuracy and sensitivity (d') in Experiment 1 (match trials)

Condition	**Voice Identity**	**Mean RT (ms)**	**Accuracy**	**d'**
Experimental	Self	460(149)	0.98(0.12)	4.09(0.31)
	Friend	537(208)	0.97(0.18)	3.75(0.58)

(participant's voice as 'YOU')	Other	582(227)	0.92(0.26)	3.36(0.76)
Control	Self	518(200)	0.94(0.24)	3.26(0.96)
(three unfamiliar	Friend	555(204)	0.93(0.25)	3.13(0.94)
voices)	Other	582(219)	0.90(0.30)	2.85(0.96)

Note. RT: reaction time; Accuracy: proportion correct. Standard deviations appear within parenthesis.

Reaction time

Figure 2 presents reaction times for correctly assessed match trials. There was a significant interaction between *voice identity* and *condition* ($\chi^2(2)=34.144$, $p<.001$). For both conditions we observe an SPE with the self-associated voice as the quickest recognised voice (ps<.001). Further, reaction times were quicker to the friend-associated voice relative to the other-associated voice in both conditions (ps<.001). Importantly, participants' reaction times to the self-associated voice were significantly faster when it was their own recorded voice relative to a newly self-associated voice ($p=.049$). In comparison, the reaction times for the friend-associated and the other-associated voices did not differ between conditions (ps=1). This demonstrates that having the participants' self-voice as the self-associated voice enhanced the SPE.

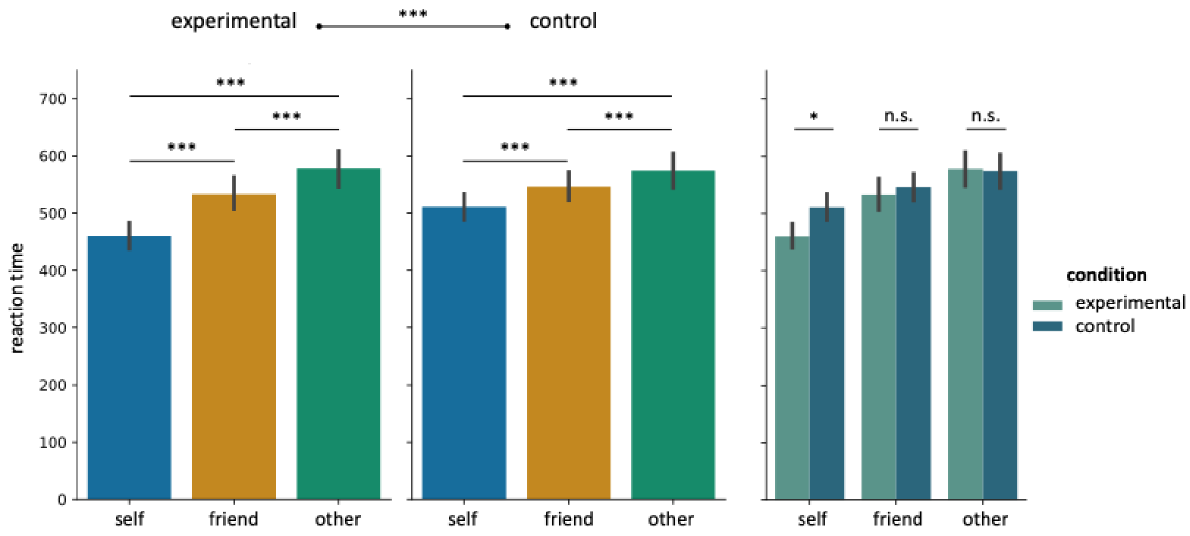

*Figure 2 - Experiment 1: Mean reaction times (RT) as a function of the voice identity (self, friend, or other) and voice type (true self-voice vs. new self- voice). Graph models for correct match trials only. Error bars indicate the 95% confidence intervals of the means. The top bar indicates the significance of interaction between the factors 'voice identity' and 'condition'. *: p<.05 ; **: p<.01 ; ***: p<.001 ; n.s.: non significant.*

Sensitivity

There was no interaction between *voice identity* and *condition* ($\chi^2(2)=3.647$, $p=.161$). However, we observed a significant main effect of *voice identity* ($\chi^2(2)=40.114$, $p<.001$), with an increased sensitivity to the self-associated voice relative to the friend-associated voice ($p=.017$) and other-associated voice ($p<.001$). Thus, the self-voice was prioritised regardless of the condition. There was also a main effect of *condition* ($\chi^2(2)=22.340$, $p<.001$) with the participants of the experimental group having an increased sensitivity in comparison to the participants of the control group. Overall, this result demonstrates an SPE in both conditions, with experimental participants showing a higher sensitivity. This distinction might be due to experimental participants being overall better at discriminating their real voice from other voices.

Accuracy

There was a significant interaction between *voice identity* and *condition* ($\chi^2(2)=112.040$, $p<.001$). Post hoc pairwise t-tests revealed that experimental participants were significantly more accurate than the control participants at recognising the self-associated-voice and the friend-associated voice ($ps<.01$) but not the other-associated voice ($p=.825$). On one hand, accuracy scores for the experimental group demonstrate an SPE, with the self-associated voice more accurately recognised than the friend-associated ($p=.01$) voice and the other-associated voice ($p<.001$). On the other hand, for the control group, the self-associated voice was more accurately recognised than the other-associated voice ($p<.001$), but not the friend-associated voice ($p=1$).

In sum, we observe an enhancement of the SPE when the self-associated voice is the participant's own recorded self-voice, as shown by reduced reaction times. This suggests that the SPE is indeed affected by self-relevance of the self-associated stimulus. Further, sensitivity and accuracy results suggest that encountering the self-voice as a self-associated voice improved the overall performance on the perceptual matching task: this finding might be due to overall reduced cognitive load, given the experimental participants were likely better able to recognise what was – or was not – the truly self-associated voice amongst the voice identities. In Experiment 2, we evaluated how performance is impacted when both the self-associated voice and the other-associated voice are self-relevant.

Experiment 2

In Experiment 1 we showed that the SPE was enhanced when the voice stimulus and the identity label were congruent (i.e., self-voice associated with the label 'YOU') in a perceptual matching task. In Experiment 2, we investigated the extent to which having an incongruent pairing of the self-voice and the task-associated identity interfered with the SPE, this time assigning the self-voice to the label 'STRANGER'. Thus, from the perspective of the listener, the voice associated to the self being unfamiliar is inherently neither self-similar nor self-generated, whereas the voice associated to 'STRANGER' is both self-similar and self-generated.

Methods

Unless indicated below, methods for Experiment 1 were replicated for Experiment 2.

Participants

96 participants completed the study. They were split into two groups of 48 experimental participants (mean age=27.3 years, age range=18-40, 26 females, 22 males) and 48 control participants (mean age=26.3 years, age range=18-40, 26 females, 22 males). Participants were recruited online via Prolific (www.prolific.com) as British monolingual speakers of English with a standard Southern British accent and no hearing impairment or difficulties. Ethical approval was obtained (SHaPS-2019-CM-030), and all participants gave their informed consent prior to the testing. Eligible participants had 90% approval rate on Prolific. Participants were paid for their participation at a rate of £7.50 per hour.

Stimuli

The 48 experimental participants were invited to complete the same voice-recording task as in Experiment 1. Again, the target word 'Hello' was extracted from each participant's voice recordings. The voice stimuli used in Experiment 1 as the friend-associated voice and the other-associated voice were included again. However, in Experiment 2 these were either the self-associated voice or the friend-associated voice while the other-associated voice was the participant's voice recordings.

Procedure

Unlike Experiment 1, participants were advised that their recorded voice may not be self-associated within the task.

Results & Discussion

Descriptive statistics for reaction time, accuracy and sensitivity (d') across conditions and voice identities are given in Table 2. See Supplemental Materials for a full report on the statistical tests (https://osf.io/93xq4/).

Table 2. Mean RTs, accuracy and sensitivity (d') in Experiment 2 (match trials)

Condition	Voice Identity	Mean RT (ms)	Accuracy	d'
Experimental	Self	533(219)	0.94(0.23)	3.37(0.77)
(participant's voice	Friend	550(208)	0.96(0.19)	3.46(0.64)
as 'STRANGER')	Other	545(202)	0.92(0.27)	3.22(0.70)

Control	Self	524(205)	0.96(0.20)	3.41(0.75)
(three unfamiliar	Friend	555(212)	0.96(0.20)	3.44(0.73)
voices)	Other	567(222)	0.92(0.28)	3.28(0.94)

Note. RT: reaction time; Accuracy: proportion correct. Standard deviations appear within parenthesis.

Reaction time

Figure 3 presents reaction times for correctly assessed match trials. There was a significant interaction between *voice identity* and *condition* ($\chi^2(2)=7.715$, p=.021). In the control group, wherein all three voices were unfamiliar, participants' reaction times were fastest for the unfamiliar self-associated voice relative to the other two voices (ps<.001), showing the classic SPE. In the experimental group, the self-associated voice was recognised quicker than the friend-associated voice (p=.012). In contrast, the differences in reaction time between the self-associated voice and the other-associated voice, as well as between the friend-associated voice and the other-associated voice, were non-significant (self-other: p=.367; friend-other: p=1). These results demonstrate that, for the experimental group, the degree of prioritisation for the new unfamiliar voice associated with the self is similar to the degree of prioritisation afforded to the self-voice when it has been associated with another. Thus, the self-relevance of the self-voice interferes with the classic SPE, allowing that the unfamiliar self-associated voice only be prioritised relative to the friend-voice which, newly, becomes the voice with the least relevance to self.

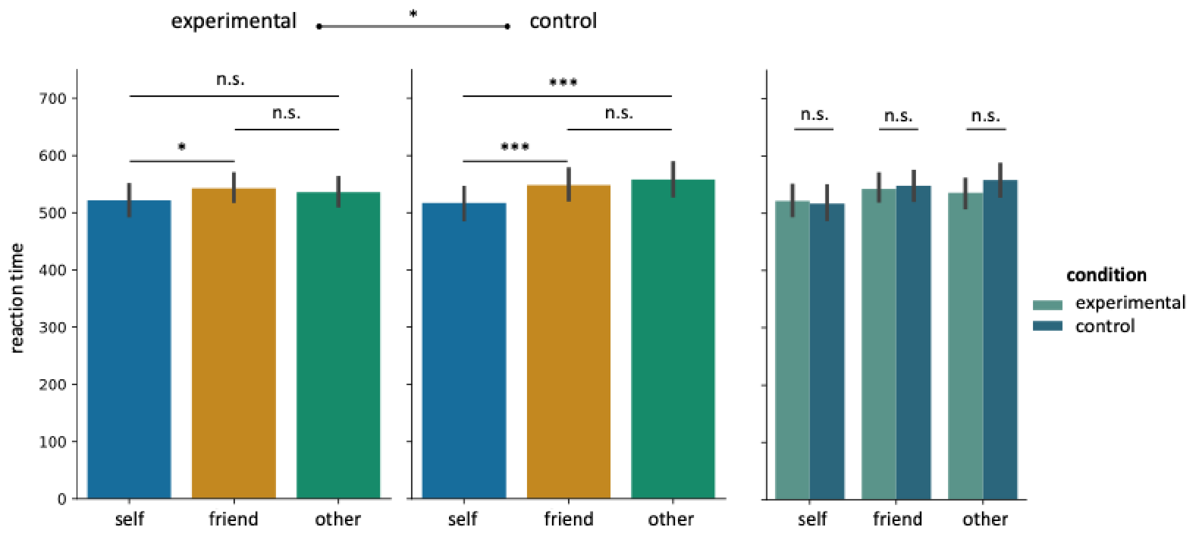

*Figure 3 - Experiment 2: Mean reaction times (RT) as a function of the voice identity (self, friend, or other) and voice type (true self-voice vs. new self- voice). Graph models for correct match trials only. Error bars indicate the 95% confidence intervals of the means. The top bar indicates the significance of interaction between the factors 'voice identity' and 'condition'. *: p<.05 ; **: p<.01 ; ***: p<.001 ; n.s.: non significant.*

Sensitivity

The interactive effect of *condition* and *voice identity* on sensitivity was not significant ($\chi^2(2)=0.209$, p=.900). In addition, there was no significant main effect of *voice identity* ($\chi^2(2)=5.722$, p=.057) or *condition* ($\chi^2(2)=0.038$, p=.846). In sum, from the perspective of sensitivity, there was no significant difference between groups, with no evidence of an SPE.

Accuracy

There was no significant interaction between *condition* and *voice identity* ($\chi^2(2)=3.775$, p=.151). There was a significant main effect observed for *voice identity* ($\chi^2(2)=62.17$ p<.001), but not for *condition* ($\chi^2(2)=0.345$, p=.557). Post hoc tests revealed that accuracy for the self-owned voice and the friend-owned voice did not significantly differ (p=291), but accuracy for both the self-associated and friend-associated was greater than that for the other-associated voice (ps<.001). The finding that accuracy for the other-associated voice was similarly reduced for both groups can be interpreted as a remnant of the SPE in the experimental group, hence supporting the idea that the effect was not entirely disrupted.

In sum, according to the reaction time results, we observed a partial disruption of the SPE with the self-associated voice being prioritised over the friend-associated in both groups. However, it was not prioritised over the other-associated voice when that voice corresponds to the participant's own recorded voice. The accuracy results further support that a temporarily incongruent presentation of a self-similar voice impairs the SPE but does not make it disappear entirely. This suggests that, within this experiment, the temporarily self-associated voice and the true self-voice were both sources of self-relevance. As a result, we observed a certain perceptual flexibility of the self where, according to our findings, the self-associated voice is no more prioritised than the self-voice, when the self-voice is associated with someone else.

Experiment 3

In Experiment 1 and Experiment 2 we showed that by manipulating self-relevance of the voices associated with the self- and other-associated voices, we could enhance or diminish the SPE. As mentioned in the Introduction, the human self-voice exists as a result of the speaker's actions, making the experience of voice generation a fundamental aspect of the voice's self-relevance. For both theoretical and applied reasons, it is important to understand how this aspect of self-relevance influences voice prioritisation. In Experiment 3, we investigated how a self-similar but not self-generated voice impacts the prioritisation of voices. To do this, we replicated the conditions of Experiment 1, here pairing the self either with a synthesised, cloned version of the participant's self-voice (Experimental group), or instead with the cloned version of an unfamiliar other's voice (Control group).

Methods

Unless indicated below, methods for Experiment 1 were replicated for Experiment 3.

Participants

96 participants completed the study. They were split into two groups of 48 experimental participants (mean age=31.8 years, age range=18-40, 24 females, 24 males) and 48 control participants (mean age=30.9 years, age range=18-40, 24 females, 24 males). Participants were recruited online via Prolific (www.prolific.com) as residents of the United States, with US English as first language, and with no hearing impairment or difficulties. Eligible participants had 90% approval rate on Prolific. Ethical approval was obtained (SHaPS-2023-CM-038), and all participants gave their informed consent prior to the testing. Participants were paid for their participation at a rate of £9.50 per hour.

Stimuli

In this experiment, all stimuli consisted of artificial voice clones synthesised based on the recordings of real voices. To do this, we used the *Instant Voice Cloning* tool within the Python API of the Text-to-Speech software *Elevenlabs* (https://elevenlabs.io/).

Self-associated stimuli

Experimental participants were invited to a recording session on the online testing platform Gorilla prior to the main task. The purpose of this session was to obtain high quality recordings of each participant saying a variety of sentences as naturally as possible. There were 45 sentences to record from the list of Harvard sentences the *Rainbow passage* and *Arthur the Rat*. Importantly, none of the sentences contained the target word "Hello". Before starting the recording session, participants were asked to make sure to be in a quiet environment with a working microphone on a computer, and to do a microphone check. At each trial, a sentence was displayed on screen for five seconds, ensuring participants had time to comprehend and read aloud the required sentence. Then, after a three-second countdown, participants were asked to read the text aloud. For each participant, the recordings were concatenated in one single audio that was used to create the voice clones using *Elevenlabs* and generate two speech samples of the word "Hello" that we used in the experiment. All stimuli were RMS-normalised using the Python library *ffmpeg-normalize*.

Friend-associated and Other-associated stimuli

A further four participants (mean age = 32.6 years, range = 18-40, 2 female, 2 male) were recruited to clone their voice. They were residents of the United States, with US English as their first language. Samples of "Hello" from these four cloned voices were used within the main experiment's tasks as either the 'friend' or 'other' voice. These participants were part of the participant pool of a pilot experiment and did not participate in the main experiment.

Acoustic similarity

To validate the acoustic similarity between a cloned sample and the recordings obtained in the first session, we leveraged a state-of-the-art, pre-trained x-vector model (Ravanelli et al., 2021). In the field of automatic speaker recognition tasks, an x-vector is an embedding extracted from a speech sample used to summarise the acoustic information of speaker identity in a multidimensional speaker space. For a given speech sample, the model takes its spectro-temporal characteristics (i.e., MFCCs) as input and outputs an x-vector. Then, the cosine similarity between speech samples or clusters' centroids allows to assess the similarity between speakers. For each experimental participant, we computed the cosine similarity between the x-vector of a clone sample and the centroid of the x-vectors of the recordings used to synthesise the clones. We then compared it with the average cosine similarity between the centroid of a participant's recordings and the friend-associated and other-associated clones. As a result, the average acoustic similarity between clones and recordings of experimental participants' voices was significantly higher than the average acoustic similarity between the friend-associated and other-associated clones with participants' recordings ($T(94)=2.96$, $p=.004$). See also the Supplemental Materials for a similar analysis based on psycho-acoustically relevant measures of the stimuli.

Procedure

Participants were explicitly told that a synthesised version of their voice – that was not their real voice – was associated with the self.

Results & Discussion

Descriptive statistics for reaction time, accuracy and sensitivity (d') across conditions and voice identities are given in Table 3. See Supplemental Materials for a full report on the statistical tests (https://osf.io/93xq4/).

Table 3. Mean RTs, accuracy and sensitivity (d') in Experiment 3 (match trials)

Condition	Voice Identity	Mean RT (ms)	Accuracy	d'
Experimental *(participant's cloned voice as 'YOU')*	Self	509(170)	0.92(0.27)	3.06(0.97)
	Friend	632(224)	0.82(0.38)	1.99(1.06)
	Other	658(230)	0.74(0.44)	1.57(1.0)
Control *(three unfamiliar cloned voices)*	Self	567(210)	0.89(0.31)	2.46(1.01)
	Friend	624(216)	0.84(0.37)	2.03(1.04)
	Other	653(232)	0.76(0.43)	1.61(1.02)

Note. RT: reaction time; Accuracy: proportion correct. Standard deviations appear within parenthesis.

Reaction time

Figure 4 presents reaction times for correctly assessed match trials. Overall, results mirrored those of Experiment 1. Specifically, there was a significant interaction between *voice identity* and *condition* ($\chi^2(2)=55.251$, $p<.001$). We observed an SPE for both conditions with the self-associated voice being the quickest recognised voice (ps<.001). Further, reaction times were quicker to the friend-associated voice relative to the other-associated voice in both conditions (ps<.01). Importantly, participants' reaction times to the self-associated voice were significantly faster when it was a synthesised version of their own voice relative to the synthesised voice of another ($p=.034$). In comparison, the reaction times for the friend-associated and the other-associated voices did not differ between conditions (ps=1). This demonstrates that having a cloned version of the self-voice enhances the SPE, likely because it is acoustically similar to the true self-voice and so more self-relevant.

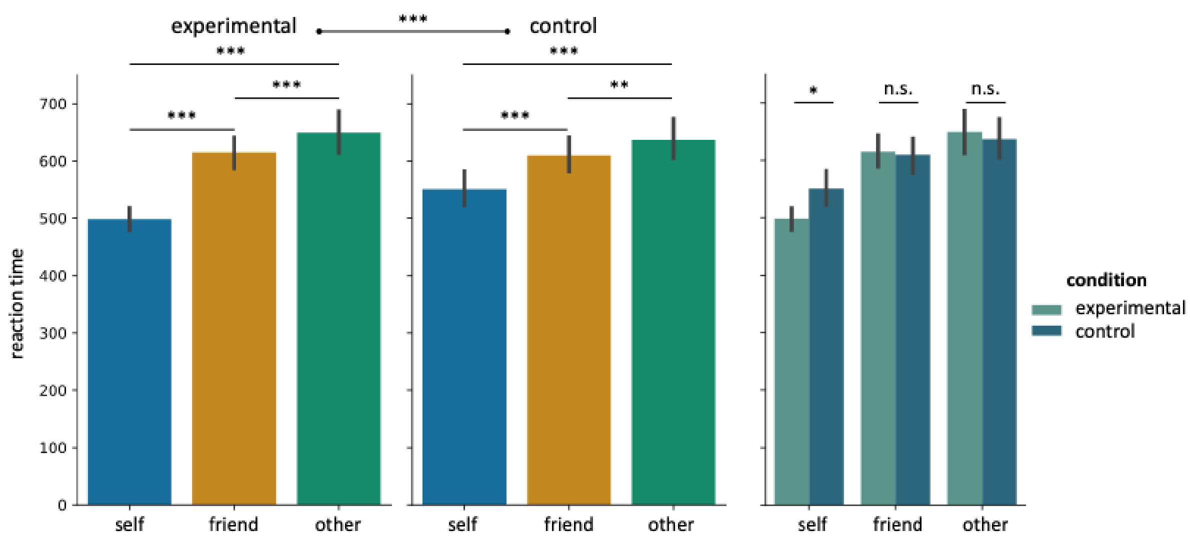

*Figure 4 - Experiment 3: Mean reaction times (RT) as a function of the voice identity (self, friend, or other) and voice type (true self-voice vs. new self- voice). Graph models for correct match trials only. Error bars indicate the 95% confidence intervals of the means. The top bar indicates the significance of interaction between the factors 'voice identity' and 'condition'. *: p<.05 ; **: p<.01 ; ***: p<.001 ; n.s.: non significant.*

Sensitivity

There was a significant interaction between *voice identity* and *condition* ($\chi^2(2)=9.95$, $p=.007$). Experimental participants exhibited an SPE with highest sensitivity for the self-associated voice ($ps<.001$). Similarly, control participants exhibited a significantly greater sensitivity to the self-associated voice compared to the other-associated voice ($p<.001$), but not compared to the friend-associated voice ($p=.086$). Importantly, sensitivity to the self-associated voice was higher in the experimental group than the control group ($p=.035$), further supporting an enhancement of the SPE in the experimental group.

Accuracy

There was a significant interaction between *voice identity* and *condition* ($\chi^2(2)=15.43$, $p<.001$). There was nonetheless an SPE present for both groups ($ps<.001$) with the self-associated voice being the most accurately assessed voice ($ps<.001$). However, post hoc tests did not reveal any significant difference between the two groups that would support the theory of an enhanced SPE for the experimental group ($p=.244$).

In Experiment 3, we show for the first time that a self-associated cloned voice that is acoustically similar to the self-voice enhances the SPE over other synthesised voices, by reducing reaction times and increasing sensitivity. This finding suggests that while self-generation is a crucial day-to-day aspect of the self-voice, the self-relevance of a self-associated voice can still be enhanced via increasing its acoustic similarity to the self-voice.

General Discussion

In this study we investigated the various influences of self-similarity and self-generation of a voice on the self-prioritisation effect. It has previously been shown that having a sense of ownership over a voice induces its perceptual prioritisation (B. Payne et al., 2021). Moreover, it has been demonstrated that using the self-voice of participants in perceptual matching experiments can modulate the SPE (Kirk & Cunningham, 2024). Here we replicated aspects of these previous results and extended them. First, in Experiment 1, we showed that when the self-voice is associated with a congruent identity (i.e., self), the SPE is significantly enhanced in comparison to the classic SPE observed with unfamiliar voices. Second, in Experiment 2, we showed that by assigning the self-voice to the incongruent identity of a stranger, the SPE is partially disrupted. Finally, in Experiment 3, we showed that hearing a self-associated clone of the self-voice enhanced the SPE compared with hearing the self-associated cloned voice of an unfamiliar other.

Previous research has demonstrated that the mental representation of the self can be flexible depending on the context (S. Payne et al., 2017; Sforza et al., 2010). We extended these findings by showing how manipulating the in-task association (i.e., self, friend, other) of the self-voice can either reinforce or weaken task-relevant self-prioritisation. Thus, we observed that when the participant's own voice was other-associated (i.e. with the "STRANGER" label), the prioritisation afforded to it was similar to the prioritisation afforded to the newly self-associated unfamiliar voice. We note that this outcome is different from Kirk & Cunningham (2024): Using perceptual matching, they showed that the self-voice is always prioritised, regardless of its in-task identity association (i.e., self, friend, other). This discrepancy may be explained by the absence of a control group in Kirk and Cunningham's design, which is necessary to rule out the influence of stimulus properties (e.g. by accounting for any baseline differences in the perceptual distinctiveness of the three task voices). Our findings instead suggest flexibility in the classification of self-relevant information, where listeners can manage and prioritise multiple different sources of self-relevance within the task.

In line with previous studies demonstrating an SPE enhancement for more self-related stimuli (Golubickis et al., 2020), we too observed an enhanced SPE when participants' self-voice (Expt. 1) or an acoustically similar version (Expt. 3) served as the self-associated voice. This supports the idea that stimuli that have more self-relevance are more highly prioritised. Here, we newly show that a voice stimulus may be perceptually prioritised by a listener by tuning its acoustic characteristics to the listener's self-voice. However, in a previous study using a perceptual matching task with unfamiliar voices (B. Payne et al., 2021), there was no SPE enhancement when the self-relevance of the voice was increased by gender-matching the self-associated voice to that of the participant, in contrast to mismatched gender conditions. As part of the identity of a speaker, voice gender (in this case, masculine or feminine) has been demonstrated to be consistently related to specific acoustic features including the fundamental frequency and formant spacing (Pépiot, 2015; Skuk & Schweinberger, 2014). However, in this case, matching acoustic properties at a coarse level may have contributed insufficient additional self-relevance for SPE enhancement. Taken together, our current finding and previous work suggests that manipulating the acoustic similarity of a voice to the self-voice leads to a diversity of effect sizes of enhanced SPE. Using voice morphing or by parametrically manipulating voice features that are relevant for speaker recognition (Xu et al., 2013), future studies should explore how the SPE varies with the acoustic (and perceptual) similarity between the self-associated voice and the true self-voice. This approach would provide insights on the weighted importance of certain acoustic features in driving the mechanisms of self-prioritisation and allow the manipulation of self-similarity while maintaining self-generation.

A self-similar artificial voice enhancing the SPE suggests that the self-relevance of a voice can be increased independently of self-generation. Despite the cloned voice being explicitly presented as artificially synthesised, the high acoustic similarity of the self-associated cloned voices was sufficient to induce an enhancement of the SPE. One possible explanation is that while a non-self-generated cloned voice may weaken its relevance to the self, the auditory familiarity of that voice, persisting through the cloned voice timbre, still enables an enhancement of perceptual prioritisation. This is promising for individuals with motor neuron disease seeking to use a bespoke synthesised voice for social interaction (Cave & Bloch, 2021). Future research should further investigate the impact of self-generation by directly comparing the SPEs for recorded and cloned self-voices.

There is a debate regarding whether familiarity or self-relevance underpins the perceptual prioritisation of self-related stimuli (Maire, 2021). Previous research has demonstrated that familiar voices receive special treatment at a cognitive and neurological level (Kanber et al., 2022; Plante-Hébert et al., 2021). Thus, some consider that high familiarity mediates self-biases (Amodeo et al., 2023). Conversely, others argue that cognitive processes involving self-related information are distinct from those related to familiarity (Bortolon & Raffard, 2018). Considering our use of participants' recorded and cloned self-voices in our study, we could not separate self-relevance from familiarity – participants were furthermore made explicitly aware of the presence of their self-voice in the experiments, thus minimising variation in self-voice recognition accuracy as a plausible index of familiarity within our study. However, we acknowledge that opportunities remain to further disentangle self-relevance from aspects of familiarity. For example, the stimulus used through the three experiment of this study was a very frequently used, and highly familiar, word ("Hello"). To further dissociate self-relevance (in its various forms) from familiarity, further experiments could be conducted using voice cloning and/or manipulated self-voice recordings to generate linguistically less familiar but acoustically (and perceptually) self-similar test stimuli (e.g. pseudowords, words in other languages). For example, previous research in the visual domain has shown that valence affects the self-prioritisation effect (Constable et al., 2021; Jalalian et al., 2024). Similarly, varying the semantic content of voice clone stimuli along a valence dimension may reveal effects on the self-prioritisation of a self-similar voice. Future studies could use voice cloning to explore how alignment with one's self-concept impacts self-relevance, for example, by having a voice clone express opinions that the participant agrees with or disagrees with.

Although the current study was specifically concerned with the perception of the self-voice as an audio stimulus – modelling how listeners typically experience recordings of their voice in everyday life, and how users of Augmentative and Alternative Communication (AAC) devices typically experience a synthesised self-voice – it should be noted that the self-voice is most commonly experienced as a multimodal stimulus, via both audio and bone/tissue conduction, during speaking (see Orepic et al., 2023). Future work could therefore investigate self-voice prioritisation under multimodal presentation conditions that better approximate the perceptual correlates of self-speech, for example by presenting audio recordings with simultaneous cutaneous stimulation to the throat. Given that people experience their voice during self-speech more often than during audio-only self-voice listening, we speculate that modulating self-relevance within more embodied task paradigms could have larger effects on the SPE than reported here.

In conclusion, our study yields key findings on how different aspects of self-relevance impact perceptual prioritisation. First, we demonstrated that the self-voice, an inherently self-relevant stimulus, significantly modulates the SPE (Expts. 1 and 2). Second, we established the flexibility of the mental representation of the self through the voice, revealing one's ability to deal with multiple sources of self-relevant information (Expt. 2). Finally, we showed for the first time that a non-self-generated, yet self-similar, synthesised voice enhances the SPE over other synthesised voices (Expt. 3). This suggests that voice cloning technology applications could be used to harness the perceptual benefits of self-relevance without the need for self-generation, for example in personalised voice alerts and reminders.

Acknowledgments

This work was supported by a Leverhulme Research Leadership Award (RL-2016-013) and a British Academy Mid-Career Fellowship (MCFSS23\230112), both awarded to C McGettigan.

Author Contribution Statement

Conceptualisation: C.M., B.P., V.R., Methodology: C.M., B.P., V.R. Investigation, data collection and analysis: B.P., V.R. Manuscript draft: V.R. Manuscript review and editing: C.M., B.P., V.R. Supervision and funding acquisition: C.M.

REFERENCES

Alexopoulos, T., Muller, D., Ric, F., & Marendaz, C. (2012). I, me, mine: Automatic attentional capture by self-related stimuli. *European Journal of Social Psychology*, *42*(6), 770–779. https://doi.org/10.1002/ejsp.1882

Amodeo, L., Nijhof, A. D., Brass, M., & Wiersema, J. R. (2023). The relevance of familiarity in the context of self-related information processing. *Quarterly Journal of Experimental Psychology*, *76*(12), 2823–2836. https://doi.org/10.1177/17470218231154884

Anwyl-Irvine, A. L., Massonnié, J., Flitton, A., Kirkham, N., & Evershed, J. K. (2020). Gorilla in our midst: An online behavioral experiment builder. *Behavior Research Methods*, *52*(1), 388–407. https://doi.org/10.3758/s13428-019-01237-x

Arık, S. Ö., Chen, J., Peng, K., Ping, W., & Zhou, Y. (2018). Neural voice cloning with a few samples. *Proceedings of the 32nd International Conference on Neural Information Processing Systems*, 10040–10050.

Boersma, P., & van Heuven, V. (2001). *Speak and unSpeak with PRAAT*. *5*(9).

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

Brédart, S., Delchambre, M., & Laureys, S. (2006). Short article: One's own face is hard to ignore. *Quarterly Journal of Experimental Psychology*, *59*(1), 46–52. https://doi.org/10.1080/17470210500343678

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

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

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: Official Journal of the International Organization of Psychophysiology*, *133*, 66–78. https://doi.org/10.1016/j.ijpsycho.2018.08.007

Constable, M. D., Becker, M. L., Oh, Y.-I., & Knoblich, G. (2021). Affective compatibility with the self modulates the self-prioritisation effect. *Cognition and Emotion*, *35*(2), 291–304. https://doi.org/10.1080/02699931.2020.1839383

Daryadar, M., & Raghibi, M. (2015). The Effect of Listening to Recordings of One's Voice on Attentional Bias and Auditory Verbal Learning. *International Journal of Psychological Studies*, *7*(2), Article 2. https://doi.org/10.5539/ijps.v7n2p155

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*(4), 1149–1160. https://doi.org/10.3758/BRM.41.4.1149

Golubickis, M., Falbén, J. K., Ho, N. S. P., Sui, J., Cunningham, W. A., & Neil Macrae, C. (2020). Parts of me: Identity-relevance moderates self-prioritization. *Consciousness and Cognition*, *77*, 102848. https://doi.org/10.1016/j.concog.2019.102848

Hubl, D., Schneider, R. C., Kottlow, M., Kindler, J., Strik, W., Dierks, T., & Koenig, T. (2014). Agency and Ownership are Independent Components of 'Sensing the Self' in the Auditory-Verbal Domain. *Brain Topography*, *27*(5), 672–682. https://doi.org/10.1007/s10548-014-0351-0

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

Jalalian, P., Svensson, S., Golubickis, M., Sharma, Y., & Macrae, C. N. (2024). Stimulus valence moderates self-learning. *Cognition and Emotion*, *38*(6), 884–897. https://doi.org/10.1080/02699931.2024.2331817

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

Kirk, N. W., & Cunningham, S. J. (2024). Listen to yourself! Prioritization of self-associated and own voice cues. *British Journal of Psychology*, *n/a*(n/a), 1018. https://doi.org/10.1111/bjop.12741

Maire, H. (2021). « Moi d'abord ! » ou l'égocentrisme ordinaire: Une revue critique de l'effet de priorité au Soi. ["Me first!" or ordinary self-centeredness: A critical review of the priority effect on the self.]. *Canadian Journal of Experimental Psychology / Revue Canadienne de Psychologie Expérimentale*, *75*(3), 307–325. https://doi.org/10.1037/cep0000238

Maurer, D., & Landis, T. (2009). Role of Bone Conduction in the Self-Perception of Speech. *Folia Phoniatrica et Logopaedica*, *42*(5), 226–229. https://doi.org/10.1159/000266070

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

Mulligan, N. W., & Lozito, J. P. (2004). Self-generation and memory. In *The psychology of learning and motivation: Advances in research and theory, Vol 45* (pp. 175–214). Elsevier Academic Press.

Ocampo, B., & Kahan, T. A. (2016). Evaluating automatic attentional capture by self-relevant information. *Cognitive Neuroscience*, *7*(1–4), 22–23. https://doi.org/10.1080/17588928.2015.1075486

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

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

Oyserman, D., Elmore, K., & Smith, G. (2012). *Oyserman, D., Elmore, K., & Smith, G. (2012). Self, self-concept, and identity. Handbook of self and identity, 2, 69-104.* (pp. 69-104.).

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

Payne, S., Tsakiris, M., & Maister, L. (2017). Can the Self Become Another? Investigating the Effects of Self-Association with a New Facial Identity. *Quarterly Journal of Experimental Psychology*, *70*(6), 1085–1097. https://doi.org/10.1080/17470218.2015.1137329

Pépiot, E. (2015). Voice, speech and gender: *Corela. Cognition, représentation, langage*, *HS-16*, Article HS-16. https://doi.org/10.4000/corela.3783

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., 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

Plante-Hébert, J., Boucher, V. J., & Jemel, B. (2021). The processing of intimately familiar and unfamiliar voices: Specific neural responses of speaker recognition and identification. *PLOS ONE*, *16*(4), e0250214. https://doi.org/10.1371/journal.pone.0250214

Ravanelli, M., Parcollet, T., Plantinga, P., Rouhe, A., Cornell, S., Lugosch, L., Subakan, C., Dawalatabad, N., Heba, A., Zhong, J., Chou, J.-C., Yeh, S.-L., Fu, S.-W., Liao, C.-F., Rastorgueva, E., Grondin, F., Aris, W., Na, H., Gao, Y., … Bengio, Y. (2021). *SpeechBrain: A General-Purpose Speech Toolkit* (No. arXiv:2106.04624). arXiv. https://doi.org/10.48550/arXiv.2106.04624

Schäfer, S., Wesslein, A.-K., Spence, C., Wentura, D., & Frings, C. (2016). Self-prioritization in vision, audition, and touch. *Experimental Brain Research*, *234*(8), 2141–2150. https://doi.org/10.1007/s00221-016-4616-6

Sforza, A., Bufalari, I., Haggard, P., & Aglioti, S. M. (2010). My face in yours: Visuo-tactile facial stimulation influences sense of identity. *Social Neuroscience*, *5*(2), 148–162. https://doi.org/10.1080/17470910903205503

Skuk, V. G., & Schweinberger, S. R. (2014). Influences of fundamental frequency, formant frequencies, aperiodicity, and spectrum level on the perception of voice gender. *Journal of Speech, Language, and Hearing Research: JSLHR*, *57*(1), 285–296. https://doi.org/10.1044/1092-4388(2013/12-0314)

Sui, J., He, X., & Humphreys, G. W. (2012). Perceptual effects of social salience: Evidence from self-prioritization effects on perceptual matching. *Journal of Experimental Psychology: Human Perception and Performance*, *38*(5), 1105–1117. https://doi.org/10.1037/a0029792

Sui, J., & Humphreys, G. W. (2017). The ubiquitous self: What the properties of self-bias tell us about the self. *Annals of the New York Academy of Sciences*, *1396*(1), 222–235. https://doi.org/10.1111/nyas.13197

Sui, J., Sun, Y., Peng, K., & Humphreys, G. W. (2014). The automatic and the expected self: Separating self- and familiarity biases effects by manipulating stimulus probability. *Attention, Perception, & Psychophysics*, *76*(4), 1176–1184. https://doi.org/10.3758/s13414-014-0631-5

Xu, M., Homae, F., Hashimoto, R., & Hagiwara, H. (2013). Acoustic cues for the recognition of self-voice and other-voice. *Frontiers in Psychology*, *4*. https://www.frontiersin.org/journals/psychology/articles/10.3389/fpsyg.2013.00735