Self-Recognition Shapes Evaluations of (Self-) Voice Attractiveness

Victor Rosi[1], Nadine Lavan[2]*, Carolyn McGettigan[1]*

*These authors contributed equally as senior authors.

1 Department of Speech, Hearing and Phonetic Sciences, University College London, UK

2 Department of Biological and Experimental Psychology, School of Biological and Behavioural Sciences, Queen Mary University of London, UK

Corresponding author: Victor Rosi (v.rosi@ucl.ac.uk)

This preprint has not yet been peer-reviewed

Abstract

Hearing our own voice is commonly assumed to be an unpleasant experience. In contrast, some studies of voice perception have reported that we actually find our own voice to be more attractive than other voices. It remains however unclear whether previously observed enhancements of self-voice attractiveness were mediated by the recognition of the self-voice during the task, because recognition was never explicitly and rigorously examined in previous studies. We therefore conducted a voice attractiveness rating task under conditions that limited the likelihood of self-voice recognition, in order to directly compare the perceived attractiveness of one's own voice when recognised versus when not. We found that both participants who did (N=59) and did not (N=80) recognise their own voice rated their own voice differently from other voices. However, crucially, the direction of this difference depended on whether the self-voice was recognised during the task or not: listeners who spontaneously recognised their voice rated it as more attractive than other voices, whereas listeners who did not recognise their voice rated it as less attractive. The findings help to explain one source of contradictory perspectives in the voice perception literature and beyond. highlighting the intricate interplay between self-perception, first impressions, and identity perception.

Keywords: self-voice perception; attractiveness; voice recognition

1 Introduction

People often report disliking the sound of their own recorded voice, such that we often hear statements like "*I hate listening to recordings of myself. It sounds so different from what I expect*" or "*It doesn't sound like me at all.*" These apparent unfamiliarity and unpleasantness brought by recordings of the self-voice may be explained by the fact that the self-voice is usually experienced multimodally through the combined input of bone conduction, vibrotactile stimulation, and air conduction (Maurer & Landis, 1990; Orepic et al., 2023). Thus, hearing a recorded version of one's own voice, which differs from the way we typically experience the self-voice, may explain why people often find it to sound alien and unpleasant.

Yet, it appears that under certain conditions, humans *do* like the sound of their recorded voice. Specifically, studies have reported that individuals perceive their own voice as more attractive than others' voices (Hughes & Harrison, 2013; Peng et al., 2019, 2020). Hughes and Harrison (2013) asked participants to rate the attractiveness of recordings of their own voice and of other participants' voices counting from 1 to 9, without revealing to the participants that their own voice was included in the set of stimuli. They found that participants rated their own voice as more attractive than the voices of others. Moreover, participants rated their own voice as more attractive than it was rated by others. The authors proposed that this result may be explained either by the exceptional familiarity of the self-voice or by *implicit egotism* – the tendency for positive self-evaluations to extend to self-related stimuli, triggering unconscious self-associations (Jones et al., 2004). More recently, Peng and colleagues used a similar paradigm and investigated if attractiveness ratings would be influenced by the gender of participants, stimulus type (i.e., vowel, number, word) or acoustic manipulations (i.e., fundamental frequency and loudness) of the self-voice (Peng et al., 2019). They reported that regardless of the gender group (i.e., male group, female group) or the stimulus type, the self-voice was consistently rated as more attractive than other voices, while acoustic manipulations of the fundamental frequency of the self-voice led to a decrease in perceived attractiveness. In a subsequent study, the authors investigated how familiarity towards a voice and self-positivity bias impact the self-enhancement effect of voice attractiveness. To do so, they introduced another familiar voice - the voice of a friend - alongside the self-voice and unfamiliar voices, and used a self-reference valence task where participants judged whether positive and negative traits applied to themselves (Cai et al., 2016) to quantify the self-positivity bias for each participant. They found that both familiarity and self-positivity bias positively influenced the attractiveness ratings of the participants' own voice and their friends' voices.

Although the studies reviewed above are consistent in reporting self-voice enhancements for perceived attractiveness, there are nonetheless some inconsistencies in methodology that limit the conclusiveness of the findings for theory development. One key factor lacking clarity is the extent to which listeners in voice attractiveness ratings tasks were consciously aware of the presence of their own voice in the task, and the extent to which this awareness or recognition might have affected the perceptual evaluation of attractiveness. Hughes and Harrison (2013) intended to investigate the *unconscious* perception of self-voice attractiveness, such that they attempted to minimise the likelihood of participants consciously recognising their voice during the attractiveness rating task (Hughes & Harrison, 2013). To do this, they included distractor tasks between recording the participants' voices and the attractiveness rating task, in order that participants would be less likely to anticipate hearing their own voice. At the end of the study, they informed participants that they had heard their own voice during the attractiveness rating task, reporting that "most participants conveyed surprise" (p. 947) at this information while only a few participants suspected their voice had been included. However, the exact number of participants who did and did not recognise their own voice was not reported, nor did the authors directly account for this factor in their analysis. It is therefore not possible to conclusively attribute the enhanced self-voice attractiveness ratings to a process that occurs even without recognition of the self-voice. Unlike Hughes and Harrison, Peng and colleagues did not expressly investigate the unconscious evaluation of the self-voice and did not use distractors (Peng et al., 2019). However, the authors used an explicit voice recognition task that followed the main attractiveness task in their experiment and showed that 93% of participants were able to recognise their own voice among those of other speakers. They therefore showed that only a minority of people failed to recognise their own voice in an explicit recognition task, which may challenge the assumption that most participants could have remained naive to the presence of their own voice in Hughes and Harrison's study. On the other hand, Peng et al.'s study still did not directly quantify whether participants recognised their voice *while rating attractiveness*. It is therefore again difficult to determine whether the mechanism driving higher self-voice attractiveness ratings is tied to being able to recognise your own voice. In their subsequent study, Peng and colleagues more directly addressed the impact of self-recognition on attractiveness judgements by asking participants immediately after the attractiveness rating task if they had recognised their own voice or the voice of their friend during the rating task (Peng et al., 2020). For the self-voice, only 8 participants out of 62 reported that they did not recognise their own voice. Despite this very small sample, the authors nonetheless reported a similar effect of enhanced attractiveness for the self-voice when comparing sub-groups of participants who did versus did not recognise their own voice. However, analyses based on only 8 participants do not allow for statistically reliable conclusions. This is crucial because, with a more balanced sample size for friend voices, recognition appeared to be a key factor: participants who recognised their friend's voice (N = 42) rated it as more attractive, but those who did not (N = 20) showed no such effect. Although this led the authors to conclude that familiarity alone is insufficient to guarantee higher voice attractiveness, it is possible to speculate that recognition itself drives the way participants evaluate the attractiveness of familiar voices, including the self-voice. Thus, establishing a firmer evidence base on how self-recognition shapes self-voice attractiveness requires a study with sufficiently large sub-samples of participants who do and do not recognise their own voice during the attractiveness rating task. Such a design should ensure adequate statistical power for robust analysis.

We therefore investigated whether explicitly recognising one's own voice influences the evaluation of its attractiveness, to examine the mechanisms underpinning the previously observed self-enhancement in the context of attractiveness ratings. To do this, participants rated the attractiveness of a set of voice recordings that included both their own voice and the voice of others (Hughes & Harrison, 2013; Peng et al., 2019, 2020). We first conducted a pilot study to establish an experimental design that would yield sufficiently large samples of both participants who did and did not recognise their own voice during the attractiveness rating task. The speech stimuli consisted of short vowel sounds, which together with distractor tasks ensured enough participants who did and did not recognise their own voice. We predicted that self-voice attractiveness ratings would differ depending on whether participants spontaneously recognised their own voice or not during the ratings task. Specifically, if the self-enhancement in attractiveness ratings is independent of self-recognition, we should observe self-enhancement in all participants. Although if self-recognition actually mediates self-enhancement, we should see this effect only in the participants who recognised themselves while rating attractiveness.

2 Pilot Study

Previous studies reporting enhanced self-voice attractiveness either did not rigorously control for participants' recognition of their own voice (Hughes & Harrison, 2013; Peng et al., 2019) or lacked sufficient numbers of participants who did not recognise their voice (Peng et al., 2020), thus limiting conclusions about the nature of the self-voice enhancement. The present pilot study was therefore designed to ensure an adequate number of participants who did and did not recognise their own voice during the attractiveness rating task. Because participants tended to easily recognise their voice in previous studies (Peng et al., 2019, 2020), the main challenge was therefore to design an experiment that would guarantee enough participants who *did not* recognise themselves. Similarly to previous studies, participants rated the attractiveness of recordings of their own voice and those of other participants. The experiment consisted of two online sessions: a first session in which participants recorded their voice and a second session where they completed the attractiveness rating task. Following Hughes and Harrison (2013), both sessions included distractor tasks to reduce participants' expectation of hearing their own voice in the second session: namely, we added a distractor attractiveness rating task in the first session and a decoy recording task in the second session, such that both sessions were identical in their structure (recording followed by rating). Only the attractiveness ratings from the second session, on the recordings from the first session, were analysed. We also evaluated the influence of speech content on self-voice recognition by having participants record and rate both vowels and digits. The results of this pilot study both informed the experimental design and the sample size for the main study.

2.1 Methods

2.1.1 Participants

39 participants (Mean age = 31, Age range = 21–40, 20 females, 19 males) were recruited online through Prolific (prolific.co). They were first-language speakers of English who were both born and living in the United States.

Ethical approval was obtained from the local ethics committee (SHaPS-2023-CM-038), and all participants gave their informed consent prior to the testing. Participants were rewarded at a rate of £10/hour for each session.

2.1.2 Materials & Procedure

The study was conducted online using the Gorilla Experiment Builder testing platform (Anwyl-Irvine et al., 2020) and comprised two sessions completed by the same group of participants. As mentioned above, the first session aimed to collect participants' voice recordings, while the second session mainly consisted of an attractiveness rating task on the recordings obtained in the first session. With the addition of distractor tasks, both sessions consisted of a voice recording task and an attractiveness rating task. Figure 1 present an overview of the experimental design.

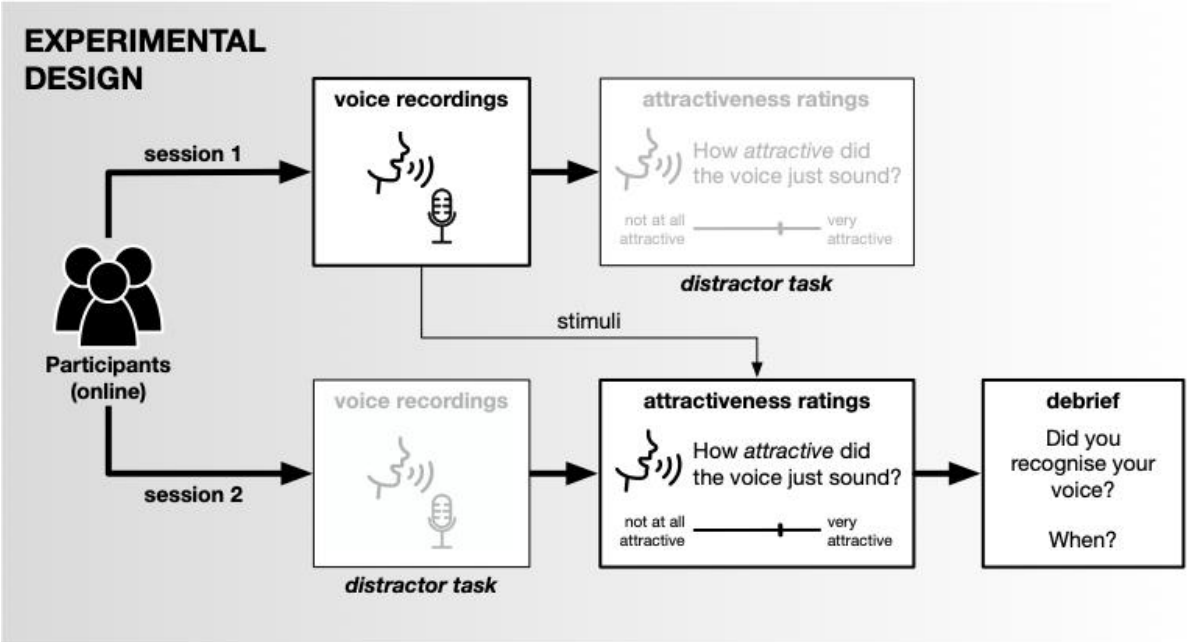

Figure 1 - Overview of the experimental design

2.1.2.1 Recording task (first session)

During the recording task, we collected voice recordings that would be included in the attractiveness rating task of the second session. First, participants were asked to make sure to be in a quiet environment with a working microphone. Then, they were required to read aloud four vowels (i.e., /aː/, /iː/, /uː/, and /æ/), digits from one to ten, and seven colour words (i.e., blue, green, red, black, white, grey, pink). Before each individual sound was recorded, the content to be read was displayed on screen for two seconds. Then, participants were asked to read the vowel, word, or colour within three seconds. We required participants to produce sustained vowels, lasting three seconds. Each type of sound was recorded twice, yielding 42 stimuli per participant. Recordings of colours served as distractors and were not included in the attractiveness ratings. To create the final stimulus sample, for each speaker, we kept one recording each of the vowels /aː/ and /iː/, and of the digits "six" and "eight", thus yielding to 156 stimuli in total (4 stimuli per participant * 39 participants) that would be evaluated in the attractiveness rating task of the second session (see below). Vowels were trimmed to 400 ms long vowels, with 50 ms ramps (i.e., half Hanning window) at the onset and offset of the audio signal. All stimuli were resampled to 22050Hz and RMS-normalised using the Python library *ffmpeg-normalize* (https://github.com/slhck/ffmpeg-normalize).

2.1.2.2 Attractiveness rating task (second session)

In the main attractiveness rating task, participants rated the attractiveness of 39 voices, including samples of their own voice and the voice of other participants who completed the first session. During each trial they listened to a recorded voice sample and rated the attractiveness of the voice using a slider ranging from 0 (not at all attractive) to 100 (very attractive). Participants rated the attractiveness of vowels (78 trials) in block one, and of digits (78 trials) in block two. Stimuli were presented in a randomised order in both blocks. Each participant provided four attractiveness ratings for their own voice, and 152 ratings for other voices. We also presented six vigilance trials at random locations within the task, where participants were asked to select a number between 1 and 6 that was announced to them beforehand.

2.1.2.3 Distractor tasks (first and second sessions)

Distractor tasks were included as such that both sessions had the same structure: A voice recording task, and an attractiveness rating task. Thus, in the first session, the distractor task consisted of an attractiveness rating task similar to the main one presented after the main voice recording task. In this task, participants were presented with 80 voice recordings of colour words (i.e., red and blue) and digits (one and three) collected in a previous study (Rosi, Payne, et al., 2025) and did not include the voice of participants. In the second session, the distractor task was a voice recording task identical to the one from the first session, that was presented before the main attractiveness rating task.

2.1.2.4 Debrief session (second session)

In a debrief session at the end of the second session, participants were asked if they heard their voice in the listening task, and if so, when, using a slider ranging from 0 (start) to 100 (end). On the slider, 50 corresponded to the end of block one and the start of block two. To estimate the potential influence of participants' evaluation of the overall attractiveness of their own voice on attractiveness ratings they provided on their voice, they were also asked to rate the overall attractiveness of their own voice (in the absence of hearing a recording of it) using a slider from 0 (not at all attractive) to 100 (very attractive). These ratings were analysed separately from the main analysis to examine potential differences between participants who did and did not recognise their own voice (see below).

2.1.3 Analysis

2.1.3.1 Main Analysis

The analysis aimed to investigate whether recognising their own voice during the attractiveness rating task modulates participants' perception of their voice's attractiveness relative to how other participant perceived it. To do this, we assessed attractiveness ratings with a linear mixed model using *afex* in R (version 4.4.2). The fixed factors of *Self-Recognition* – whether participants recognised their voice or not in the attractiveness rating task – and *Voice Identity* – whether the voice was the participant's own voice (self) or another participant's voice (other) – were included in our linear mixed model. For each trial, the factor *Self-Recognition* was coded 'yes' if the speaker of that trial's stimulus recognised their voice in the listening task 'no' if they did not. *Participant* and *Stimulus* were added as random intercepts. The final model took the following form:

$$Attractiveness \sim Self\ Recognition * Voice\ Identity + (1|Participant) + (1|Stimulus)$$

We tested the significance of the interaction by performing a likelihood-ratio test. Depending on the significance of the factors and their interaction, we ran post hoc pairwise comparisons using *emmeans* with a Bonferroni correction.

The analysis was conducted separately for vowels (block one) and digits (block two). For vowels, the factor *Self-Recognition* was coded as 'yes' if participants indicated that they heard their voice before the end of block one, and 'no' if not. For digits, the factor *Self Recognition* was coded as 'yes' if participants indicated that they heard their voice at any time during the listening task, and 'no' if not.

To investigate if attractiveness ratings on self-voice recordings would be influenced for each group by participants' broad perception of the attractiveness of their voice, we performed a two-sample t-test in *R* comparing attractiveness ratings obtained in the debrief between participants who recognise their voice in the main attractiveness rating task and participants who did not.

Similarly to previous studies, we performed a complementary analysis, where the factor of *Self-Recognition* is this time coded according to the listener on each trial, i.e., if the listener recognised their own voice in the main attractiveness rating task, then every trial completed by that listener was coded with 'yes' for *Self-Recognition*, and if they did not recognise themselves every trial they completed was coded with 'no' for *Self-Recognition*. This analysis yielded results similar to those presented in the article and is reported in the supplementary information (https://osf.io/q8drs/).

2.1.3.2 Interrater agreement

Interrater agreement within each group of participants (those who recognised their own voice, and those who did not) for each block (vowels, digits) was assessed by computing a two-way random intraclass correlation coefficient (ICC2k) on attractiveness ratings using the *psych* package.

2.2 Results

On average the first session lasted 15 minutes and the second session lasted 20 minutes.

Out of the initial number of 39 participants who signed up for the study, 36 completed the two sessions (see details on each group in Table 1). All data from participants who did not complete the two sessions were excluded from the analysis. Participants who responded incorrectly to more than 20% of the vigilance test in the attractiveness rating task would have been also excluded. However, all participants who completed the two sessions performed above this threshold, and thus none were removed from the analysis.

2.2.1 Self-Recognition and Interrater agreement

For both vowels and digits, participants were divided based on whether they recognised their own voice in each of the blocks of the attractiveness rating task. 10 participants recognised their voice and 26 did not in block one (vowels), 22 participants recognised their voice in block two (digits), and 10 did not recognise their voice at all until the debrief at the end of the experiment. Agreement was moderate to good for both participants who did and did not recognise themselves during block one (see Table 1). In block two, agreement was high for participants who recognised their voice, but moderate for participants who did not recognise their voice (see Table 1).

Table 1 – Sample size and interrater agreement within each group of participants (Pilot study). ***: p<.001

Block	**Self-Recognition**			
	yes		**no**	
	N	*ICC2k*	*N*	*ICC2k*
Block one (vowels)	10	.72***	26	.71***
Block two (digits)	22	.84***	14	.55***

2.2.2 Mixed-model analysis

For both blocks separately, we investigated whether attractiveness ratings differed depending on whether a voice sample was evaluated by its owner (i.e., the participant who produced it) or by another participant, and on whether participants recognised their own voice during the task. Figures 2(A) and 2(B) illustrate the attractiveness ratings on vowels and digits recordings respectively.

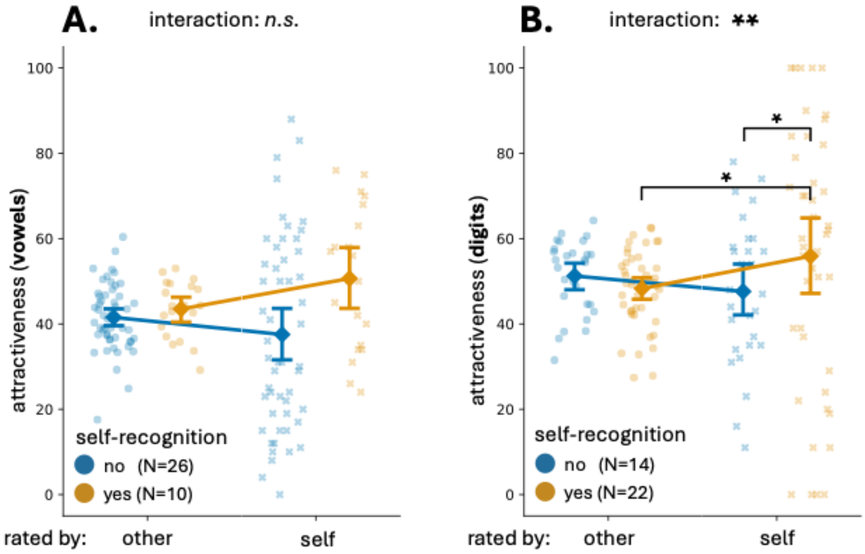

Figure 2 - Attractiveness ratings on voice recordings from the pilot study as a function of *Voice Identity* (other, self) and *Self-Recognition* (no, yes) for vowels (**A.**) and digits (**B.**). **Blue '•':** Mean ratings on voice recordings of participants who did not recognise their own voice, by other participants. **Gold '•':** Mean ratings on voice recordings of participants who recognised their own voice, by other participants. **Blue 'x':** Mean ratings on voice recordings of participants who did not recognise their own voice, by themselves. **Gold 'x':** Mean ratings on voice recordings of participants who recognised their own voice, by themselves. Error bars indicate 95% confidence intervals of the means. *:p<.05, **:p<.01.

In block one consisting of vowels, there was no significant interaction between *Self-Recognition* and *Voice Identity*, $\chi^2(1) = 2.59$, p = .107. However, we observed a main effect of *Self-Recognition*, $\chi^2(1) = 3.87$, p = .049, with increased attractiveness ratings on voices from speakers who recognised their voice (*M* = 46.0, *95% CI* [38.3,53.6]) compared to ratings on voices of participants who did not recognise their voice (*M* = 40.0, *95% CI* [340,45.9].

In block two consisting of digits, we observed a significant interaction between *Self-Recognition* and *Voice Identity*, $\chi^2(1) = 9.07$, *p* = .003. Post hoc comparisons revealed that participants who recognised their voice during the task rated their own voice significantly more attractive than other participants rating their voice (*p* = .004 ; M_{self} = 57.4, *95%* CI_{self} [47.6,67.3] ; M_{other} = 48.3, *95%* CI_{other} [42.6,53.9]). Participants who did not recognise their own voice showed no difference between self-voice rating and others rating their voice (*p* = .120 ; M_{self} = 45.2, *95%* CI_{self} [33.3,57.1] ; M_{other} = 51.3, *95%* CI_{other} [45.1,57.6]). When comparing voices from speakers who recognised their own voice with those from speakers who did not, we observed that participants who recognised their voice rated it as significantly more attractive than participants who did not recognised their voice (*p* = .023). In contrast, there was no significant difference in how attractive they rated other voices (*p* = .156).

2.2.3 Evaluation of overall self-voice attractiveness (debrief)

On average, participants' post-task rating of the overall attractiveness of their own voice did not differ between participants who recognised their voice (*M* = 53.7, *95% CI* [38.9,68.6]) and participants who did not (*M* = 38.8, *95% CI* [27.4,50.15]), *t*(34) = -1.68, *p* = .102, *Cohen's d* = 0.53. These results indicate that participants' ratings of their own voice's attractiveness in the context of other voices in the listening task are independent of their overall evaluation of their self-voice.

2.3 Discussion

The purpose of this pilot study was to test and fine-tune an experimental design to investigate whether recognising one's own voice in an attractiveness rating task influences participants' rating on their own voice. – Critically, we were able to gauge the likelihood of self-recognition when listening to spoken vowels versus spoken digits. Specifically, more than two thirds of participants did not recognise themselves from vowels, while only around a third of participants failed to recognise their own voice during the block containing spoken digits. However, we observed common trends in attractiveness ratings for both types of voice stimuli: participants who recognised their own voice tended to rate it as more attractive than did others, whereas participants who did not recognise their voice showed no such pattern, and in fact tended to unconsciously evaluate their own voice as less attractive than others. Given no apparent impact of speech type and considering the likelihood of increased self-voice recognition rates across a longer task, we opted to use only vowel stimuli in the main experiment.

In terms of the trends in the data, our findings differ from previous studies that, taken together, showed that the self-voice is always rated as more attractive, regardless of being recognised (Hughes & Harrison, 2013; Peng et al., 2019, 2020). Although preliminary and non-significant, our observed trends suggest that the enhanced attractiveness ratings of the self-voice reported in previous studies may be mediated by self-recognition during the task. In contrast, the unconscious evaluation of the self-voice seems to yield lower attractiveness ratings.

3 Main Study

In the main study, we further investigated whether self-recognition influences attractiveness ratings of the self-voice. To build on the trends observed in the pilot study and achieve statistical significance, the main study adopted a similar design. Thus, it consisted of two online sessions: participants completed a voice recording task in the first session and an attractiveness rating task in the second session on the voice recordings collected. In light of findings from the pilot study, where participants recognised themselves frequently when hearing digits in the second block of attractiveness ratings, we only used vowels in the main study across two blocks of attractiveness ratings. This approach not only stripped away linguistic information that might assist recognition but also doubled the number of data points for vowels in our analysis. The required sample size for the main experiment was determined based on effect sizes estimated from t-tests comparing self-ratings and other-ratings in the vowel block of the pilot experiment. See 3.1.1. for the details of the sample size calculation.

3.1 Methods

3.1.1 Sample size estimation

A power analysis based on the pilot study was conducted using G*Power (Faul et al., 2009). To replicate the pilot findings while optimising the number of participants who do not recognise their voice during the task, we based our power calculation on effect sizes from the ratings on vowels. Specifically, we considered the effect sizes from two paired t-tests, i.e., the difference between rating-by-self and rating-by-others of stimuli from participants who recognised their own voice on one side, and on stimuli from participants who did not recognise their own voice on the other side. For stimuli from participants who did not recognise themselves, average self-voice ratings were numerically lower than ratings from others, $t(37)=-0.92$, $p=.365$, *Cohen's d*= -0.26. For stimuli from participants who recognised themselves, average self-ratings were numerically higher than ratings from others, $t(37)=1.26$, $p=.233$, *Cohen's d*= 0.54. Because the effect size for participants who did not recognise themselves was too small to make the required sample size feasible (~350 participants in total), we based the power calculation on the evaluation of participants who did recognise themselves. Thus, with Cohen's $d=0.54$, a sample size of at least 39 participants was estimated to achieve 95% power at an α level of .05 (no nonsphericity correction was needed). Maintaining the "self-recognised/not self-recognised" proportions of participants observed in the pilot (i.e. roughly 1:2) yielded an estimated total sample size of about 120 participants. To account for the fact that we added a second block of attractiveness ratings on vowels, for potential exclusions and/or withdrawals, and for the stochasticity of participants recognising their voice during the experiment, the final recruitment target was 150 participants.

3.1.2 Participants

150 participants (Mean age = 40, Age range = 18–65, 75 females, 75 males) were recruited online through Prolific (prolific.co). Participant were first-language speakers of English, and were born and living in South-East England.

Ethical approval was obtained from the local ethics committee (SHaPS-2023-CM-038), and all participants gave their informed consent prior to the testing. Participants were rewarded at a rate of £10/hour for each session.

3.1.3 Materials & Procedure

The design of the main study was very similar to the pilot experiment. The study was conducted online using the Gorilla Experiment Builder testing platform (Anwyl-Irvine et al., 2020), and consisted of two sessions. The experimental design followed the same structure as for the pilot study (see Figure 1). To limit the duration of the main listening task in the second session, the 150 participants were divided into three groups of 50 that each completed the two sessions separately. Unless indicated below, methods for the pilot experiment were replicated for the main experiment.

3.1.3.1 Recording task (first session)

In the main voice recording task, participants provided three 3-second-long recordings of the sustained vowels /aː/, /iː/, /uː/. For the set of stimuli of the main attractiveness rating task, we kept one recording of the vowel /aː/ and one recording of the vowel /iː/ for each speaker yielding 100 stimuli per group of 50 participants. Similarly to the pilot experiment, vowels were cut to 400 ms, resampled to 22050 Hz and RMS-normalised.

3.1.3.2 Attractiveness rating task (second session)

In the main attractiveness rating task, participants rated the attractiveness of the 100 voice stimuli, including their own voice and the voices of the other participants from the same group. The whole set of stimuli was presented twice in a randomised order in two distinct blocks. Thus, each participant provided four attractiveness ratings for their own voice, and 196 ratings for other voices. We also presented six vigilance tests at random locations.

3.1.3.3 Distractor tasks (first and second sessions)

For the distractor task of the first session, participants completed an attractiveness rating task on 20 voice recordings of the vowels /aː/ and /iː/ after the main voice recording task. Voice samples were collected from a previous study (Rosi, Soopramanien, et al., 2025) and did not include the participants' voices. The distractor task in the second session was a recording task identical to the main recording task from the first session, presented before the main attractiveness rating task.

3.1.3.4 Debrief session (second session)

Similarly to the pilot study, participants were asked if they heard their voice in the main attractiveness rating task, and if so, when. They were also asked to rate the overall attractiveness of their own voice (without listening to a recording of their voice).

3.1.4 Analysis

3.1.4.1 Main Analysis

The main analysis aimed to investigate whether self-recognition during the attractiveness rating task influenced the evaluation of the self-voice. Therefore, like in the pilot study, we used a linear mixed model on attractiveness ratings with *Self-Recognition* (yes/no) and *Voice Identity* (self/other) as fixed factors. As participants were split into three distinct groups, we attempted to model group variability by including *Participant* nested in *Group* as a random intercept (i.e., '*1|Group/Participant)*'). But this led to a singular fit, where *Group* accounted for very little variance, so *Group* was excluded from the model. Random factors therefore consisted of *Participant* and *Stimulus*. As a result, the final model was identical to the one used in the pilot study took. Since we used all data from both blocks for the analysis, *Self-Recognition* was coded as 'yes' whenever a participant recognised their voice, and 'no' if not.

We also conducted a complementary analysis where the factor of *Self-Recognition* is this time coded according to the listener and not the speaker on each trial. This analysis yielded results similar to those presented in the main manuscript and is reported in the supplementary information (https://osf.io/q8drs/).

3.1.4.2 Interrater agreement

The interrater agreement was performed separately for groups of participants who recognised themselves and those who did not, within each group of 50 participants by computing a two-way random intraclass correlation coefficient (ICC2k) on the attractiveness ratings.

3.1.5 Transparency and Openness

The sample size for the main experiment was determined through a power analysis informed by the pilot study results (see 2.1.1). All experimental data and analysis scripts can be found as additional online materials at https://osf.io/q8drs/. Data were analysed using R (4.4.2) and the packages *afex*, *lme4*, and *emmeans*. This study was not preregistered.

3.2 Results

On average the first session lasted 8 minutes and the second session lasted 18 minutes. Out of the initial number of 150 participants who signed up for the study, 139 completed the two sessions (see details on each group in Table 1). No participant was excluded due to failing more than 20% of vigilance trials.

3.2.1 Self-Recognition and Interrater agreement

Within each of the three groups, participants were divided based on whether they recognised their own voice in the attractiveness rating task. Overall, 59 participants recognised their voice and 80 did not. Interrater agreement was good for all groups (see Table 2).

Table 2 – Sample size and interrater agreement within each group of participants (Main study). ***:p<.001

Group	N total	Self-Recognition			
		yes		no	
		N	*ICC2k*	*N*	*ICC2k*
1	44	18	.84***	26	.86***
2	48	22	.86***	26	.85***
3	47	19	.78***	28	.89***

3.2.2 Mixed Model Analysis

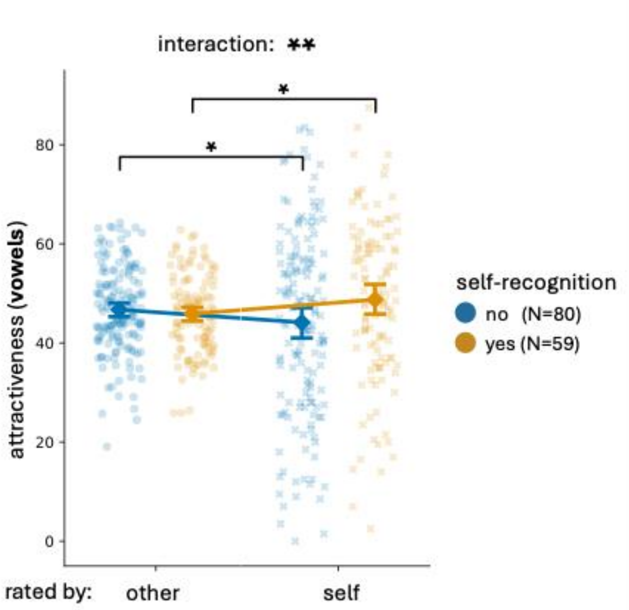

Figure 3 - Attractiveness ratings on voice recordings of vowels from the main study as a function of *Voice Identity* (other, self) and *Self-Recognition* (no, yes). **Blue '•':** Mean ratings on voice recordings of participants who did not recognise their own voice, by other participants. **Gold '•':** Mean ratings on voice recordings of participants who recognised their own voice, by other participants. **Blue 'x':** Mean ratings on voice recordings of participants who did not recognise their

own voice, by themselves. **Gold 'x':** Mean ratings on voice recordings of participants who recognised their own voice, by themselves. Error bars indicate 95% confidence intervals of the means. *:p<.05, **:p<.01.

Figure 3 illustrates the attractiveness ratings on voice recordings collected during the main attractiveness rating task. We observed a significant interaction between *Self-Recognition* and *Voice Identity*, $\chi^2(1) = 8.29$, $p = .004$. Post hoc comparisons indicated that participants who recognised their voice during the listening task rated the self-voice as more attractive than other participants' ratings on that same voice (p=.049 ; M_{self} = 48.1, *95% CI*$_{self}$ [45.0,51.1] ; M_{other} = 45.8, *95% CI*$_{other}$ [43.7,48.0]). Conversely, participants who did not recognise their own voice rated the self-voice as less attractive than others rated their voice (p = .031 ; M_{self} = 44.7, *95% CI*$_{self}$ [41.9,47.4] ; M_{other} = 46.8, *95% CI*$_{other}$ [44.7,48.8]). When comparing voices from speakers who recognised their own voice with those from speakers who did not, there were no significant differences in how attractive they rated their own voice (p = .056) or how attractive they rated other voices (p = .358).

3.2.3 Evaluation of overall self-voice attractiveness (debrief)

On average, participants' post-task ratings of the overall attractiveness of their own voice did not differ between those who did not recognise their voice (M = 41.2, *95% CI* [36.5,45.8]) and those who did (M = 39.1, *95% CI* [33.6,44.6]), $t(124) = -0.57$, p = .569, *Cohen's d* = -0.1. Similarly to the pilot experiment, these results indicate that participants' ratings of their own voice's attractiveness in the context of other voices are independent of their overall evaluation of their self-voice.

4 Discussion

The present study aimed to investigate the role of self-recognition in the evaluation of the attractiveness of the recorded self-voice, to help clarify the potential factors that may underpin the enhanced self-voice attractiveness ratings observed in previous studies. To do this, we tested whether self-recognition was present during the attractiveness ratings, which was not done in previous studies (Hughes & Harrison, 2013; Peng et al., 2019, 2020). We indeed found that all participants rated their own voice differently from other voices. However, more importantly, self-recognition affected attractiveness evaluations of the self-voice, such that participants who recognised their own voice rated it as more attractive than others, whereas those who did not recognise it rated it as less attractive.

We thus partially replicated previous work by showing that participants who recognised their own voice rated it as more attractive than others (Peng et al., 2019, 2020). Previous investigations have highlighted the human ability to easily recognise self-voice stimuli (Hughes & Nicholson, 2010; Rosa et al., 2008). Based on this evidence and our findings, we suggest that the higher perceived attractiveness of the self-voice that was observed in previous studies, including Harrison and Hughes (2013)'s study, was likely underpinned by most participants spontaneously recognising their own voice. Consequently, self-recognition may then trigger mechanisms or processes that result in a self-enhancement effect, such as self-positivity biases, or implicit egotism. Alternatively, the observed self-enhancement effect may simply reflect a reward response dependent on participants "finding" their own voice as a familiar stimulus among unfamiliar stimuli. Indeed, this explanation would be in line with previous research that suggests that the brain's reward system is engaged when (anticipating and) experiencing a personally-relevant and valued voice (Kanber et al., 2025). Finally, the enhanced evaluation of the self-voice upon recognition also aligns with previous research showing that individuals tend to perceive voices that are perceptually or acoustically similar to their own as more attractive than those of others (Peng et al., 2019; Rosi, Soopramanien, et al., 2025).

In contrast, we found that participants who did not recognise their own voice provided lower attractiveness ratings on the self-voice than others. A possible explanation for why these participants may have failed to recognise their own voice and consequently rated it lower in attractiveness could have been due to some voice recordings having poor recording quality, given that they were collected online without control over the recording equipment. Poor recording quality would have rendered these voices both less recognisable and perceptually unattractive. However, this explanation is unlikely, as our main analysis showed that participants rated the voices of others who did vs. did not recognise themselves to be equally attractive (see 3.2.2) - an audio quality explanation would predict lower perceived attractiveness for unrecognised voices. On the other hand, as noted in the Introduction, humans typically experience their own voice multimodally (Maurer & Landis, 1990; Orepic et al., 2023) rather than as unimodal audio stimuli as in the current study. It is therefore most likely that, in the context of our online experiment, self-voice recordings sounded like a less natural and less self-similar version of the embodied self-voice. Thus, participants perceiving a sample of their own recorded voice as unnatural and less familiar potentially caused some degree of discomfort and eeriness that led to lower attractiveness ratings. In other words, rather than reflecting unconditional self-enhancement when evaluating the self-voice without recognition, the observed lower attractiveness ratings of the self-voice might signal the implicit detection of a perceptual incongruence with the self. On another note, Peng and colleagues also found that recognition had an impact on attractiveness ratings of familiar voices, which, when not recognised by listeners, were rated similarly to other voices. In contrast, the lower ratings for the self-voice that we observed when it is not recognised further highlight its distinct processing compared to other voices, including familiar voices (Graux et al., 2015; Kaplan et al., 2008).

To limit the likelihood of self-voice recognition, we used use stimuli and an experimental setup that differ substantially from more ecologically valid conditions for listening to recorded voices. Future research could explore ways to generalise these findings to different kinds of voice stimuli and to less controlled experimental conditions. Based on the attractiveness rating task on digits from our pilot study, we hypothesise that using longer speech content would make it more challenging to mitigate against all participants recognising their own voice. Nonetheless, previous studies have shown the generalisability of social evaluation patterns on voice identities across different recording contents or durations (Groyecka-Bernard et al., 2022; Mahrholz et al., 2018; Mileva & Lavan, 2023). One potential avenue to explore could be to exploit the phenomenon of speaker "change deafness" - where a sudden change in speaker during a perceptual task can go unnoticed (Fenn et al., 2011) - to insert the self-voice within the task. Alternatively, self-recognition during the task could be manipulated via the use of voice morphs as stimuli, where the voice-identity of a participant and another speaker are merged (Belin & Kawahara, 2024; Nussbaum et al., 2023; Orepic et al., 2023). Finally, our study design did not address demographic factors about participants that could influence voice attractiveness ratings such as gender, age, and cultural background (Chun et al., 2024; Peng et al., 2021; Weiss et al., 2021). A follow-up study addressing these questions would be valuable.

In conclusion, this work shows that self-recognition significantly modulates the perceived attractiveness of the self-voice. Participants who recognised their own voice rated it more positively, whereas those who did not tended to rate it less positively – challenging both the common assumption that people always dislike hearing their own voice and the claim that self-voices are consistently evaluated more positively. Thus, while still being mediated by familiarity and self-relevance, previously reported self-enhancement effects may be contingent on explicit self-recognition.

Data availability

Raw datasets, R scripts, and supplementary information are available at https://osf.io/q8drs/.

Funding sources

This work was supported by a Leverhulme Research Leadership Award (RL-2016-013) and a British Academy Mid-Career Fellowship (MCFSS23\230112) awarded to C McGettigan, a Henry Wellcome Fellowship from Wellcome (220448/Z/20/Z) awarded to N Lavan, UKRI grants from the Marie Skłodowska-Curie Actions Doctoral Network "Voice

Communication Sciences" (action 101168998) awarded to C McGettigan (EP/Z535412/1) and N Lavan (EP/Z535345/1), and a British Academy Postdoctoral Fellowship (PFSS24\24003) awarded to V Rosi.

References

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

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

Cai, H., Wu, L., Shi, Y., Gu, R., & Sedikides, C. (2016). Self-enhancement among Westerners and Easterners: A cultural neuroscience approach. *Social Cognitive and Affective Neuroscience*, *11*(10), 1569–1578. https://doi.org/10.1093/scan/nsw072

Chun, E., Jeong, Y., Kim, H., Kim, N., Kim, S., & Lee, Y. (2024). Perception of Voice Attractiveness: Effects of Speaking Rate, Gender, and Age. *Communication Sciences & Disorders*, *29*(2), 462–472. https://doi.org/10.12963/csd.240029

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

Fenn, K. M., Shintel, H., Atkins, A. S., Skipper, J. I., Bond, V. C., & Nusbaum, H. C. (2011). When Less is Heard than Meets the Ear: Change Deafness in a Telephone Conversation. *Quarterly Journal of Experimental Psychology*, *64*(7), 1442–1456. https://doi.org/10.1080/17470218.2011.570353

Graux, J., Gomot, M., Roux, S., Bonnet-Brilhault, F., & Bruneau, N. (2015). Is my voice just a familiar voice? An electrophysiological study. *Social Cognitive and Affective Neuroscience*, *10*(1), 101–105. https://doi.org/10.1093/scan/nsu031

Groyecka-Bernard, A., Pisanski, K., Frąckowiak, T., Kobylarek, A., Kupczyk, P., Oleszkiewicz, A., Sabiniewicz, A., Wróbel, M., & Sorokowski, P. (2022). Do Voice-Based Judgments of Socially Relevant Speaker Traits Differ Across Speech Types? *Journal of Speech, Language, and Hearing Research*, *65*(10), 3674–3694. https://doi.org/10.1044/2022_JSLHR-21-00690

Hughes, S. M., & Harrison, M. A. (2013). I like My Voice Better: Self-Enhancement Bias in Perceptions of Voice Attractiveness. *Perception*, *42*(9), 941–949. https://doi.org/10.1068/p7526

Hughes, S. M., & Nicholson, S. E. (2010). The processing of auditory and visual recognition of self-stimuli. *Consciousness and Cognition*, *19*(4), 1124–1134. https://doi.org/10.1016/j.concog.2010.03.001

Jones, J. T., Pelham, B. W., Carvallo, M., & Mirenberg, M. C. (2004). How Do I Love Thee? Let Me Count the Js: Implicit Egotism and Interpersonal Attraction. *Journal of Personality and Social Psychology*, *87*(5), 665–683. https://doi.org/10.1037/0022-3514.87.5.665

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

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

Mahrholz, G., Belin, P., & McAleer, P. (2018). Judgements of a speaker's personality are correlated across differing content and stimulus type. *PLOS ONE*, *13*(10), e0204991. https://doi.org/10.1371/journal.pone.0204991

Maurer, D., & Landis, T. (1990). Role of Bone Conduction in the Self-Perception of Speech. *Folia Phoniatrica et Logopaedica*, *42*(5), 226–229. https://doi.org/10.1159/000266070

Mileva, M., & Lavan, N. (2023). Trait impressions from voices are formed rapidly within 400 ms of exposure. *Journal of Experimental Psychology: General*, *152*(6), 1539–1550. https://doi.org/10.1037/xge0001325

Nussbaum, C., Pöhlmann, M., Kreysa, H., & Schweinberger, S. R. (2023). Perceived naturalness of emotional voice morphs. *Cognition and Emotion*, *37*(4), 731–747. https://doi.org/10.1080/02699931.2023.2200920

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

Payne, B., Addlesee, A., Rieser, V., & McGettigan, C. (2024). Self-ownership, not self-production, modulates bias and agency over a synthesised voice. *Cognition*, *248*, 105804.

Peng, Z., Hu, Z., Wang, X., Jiao, T., Li, H., & Liu, H. (2021). Gender context modulation on the self-enhancement effect of vocal attractiveness evaluation. *PsyCh Journal*, *10*(6), 858–867. https://doi.org/10.1002/pchj.472

Peng, Z., Hu, Z., Wang, X., & Liu, H. (2020). Mechanism underlying the self-enhancement effect of voice attractiveness evaluation: Self-positivity bias and familiarity effect. *Scandinavian Journal of Psychology*, *61*(5), 690–697. https://doi.org/10.1111/sjop.12643

Peng, Z., Wang, Y., Meng, L., Liu, H., & Hu, Z. (2019). One's own and similar voices are more attractive than other voices. *Australian Journal of Psychology*, *71*(3), 212–222. https://doi.org/10.1111/ajpy.12235

Pinheiro, A. P., Sarzedas, J., Roberto, M. S., & Kotz, S. A. (2023). Attention and emotion shape self-voice prioritization in speech processing. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, *158*, 83–95. https://doi.org/10.1016/j.cortex.2022.10.006

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

Rosi, V., Payne, B., & McGettigan, C. (2025). Effects of self-similarity and self-generation on the perceptual prioritization of voices. *Journal of Experimental Psychology: Human Perception and Performance*, *51*(8), 996–1007. https://doi.org/10.1037/xhp0001325

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

Weiss, B., Trouvain, J., Barkat-Defradas, M., & Ohala, J. J. (Eds.). (2021). *Voice Attractiveness: Studies on Sexy, Likable, and Charismatic Speakers*. Springer. https://doi.org/10.1007/978-981-15-6627-1