8-1-2024

Me, not-me: Voice note use predicts self-voice recognition and liking

Justin Thomas
King Abdulaziz Center for World Culture (Ithra)

Jigar Jogia
Zayed University, jigar.jogia@zu.ac.ae

Mariapaola Barbato
Zayed University

Richard Bentall
The University of Sheffield

Computers in Human Behavior Reports 15 (2024) 100446

Contents lists available at ScienceDirect

Computers in Human Behavior Reports

journal homepage: www.sciencedirect.com/journal/computers-in-human-behavior-reports

Me, not-me: Voice note use predicts self-voice recognition and liking

Justin Thomas [a,*], Jigar Jogia [b], Mariapaola Barbato [b], Richard Bentall [c]

[a] Digital Wellbeing Program (Sync), King Abdulaziz Center for World Culture (Ithra), Saudi Arabia
[b] College of Natural and Health Sciences, Department of Psychology, Zayed University, United Arab Emirates
[c] Department of Psychology, University of Sheffield, UK

ARTICLE INFO

Keywords:
Voice notes
Voice memos
Self-voice
Voice recognition

ABSTRACT

Voice notes, spoken messages recorded and sent via smartphones, have become a widespread means of communication. A likely consequence of this situation is that voice note users become more frequently exposed to recordings of their own voices (self-voices). This correlational study examined if frequent exposure to recordings of the self-voice via voice note replay was associated with improved self-voice recognition (accuracy and response latency) and self-voice liking. Participants (N = 128), regular voice note users, reported voice note replay frequency. They also reported self-voice satisfaction/liking. Finally, participants completed a novel self-voice recognition task, where, across 20 trials, they identified whether a recording was them (self-voice) or not (non-self-voice). The tendency to frequently replay voice notes was positively correlated with self-voice liking and recognition accuracy. These findings may have implications for the treatment of social anxiety disorder and auditory verbal hallucinations.

Voice notes are spoken messages typically recorded on smartphones (Collins English Dictionary, 2022). The use of voice notes as an alternative to text messages and real-time voice calls has increased rapidly in recent years (Singh, 2022). Voice note users can, and frequently do, replay their voice notes before and after sending them. This presents a novel research question: is frequent voice note replay associated with a heightened ability to recognise one's own voice (self-voice recognition accuracy) and with more favourable evaluations of the sound of one's own voice (self-voice liking)? Answering such questions may eventually contribute to informing interventions for individuals who express extreme dislike for the sound of their own voices, as is the frequently the case in social anxiety disorder (Lundh et al., 2002). Similarly, it may also have implications for those who experience more subtle difficulties with self-voice processing, as hypothesized to be the case in auditory verbal hallucinations (Conde et al., 2016) (see Fig. 1).

The notion that digital technologies can have unintended consequences for psychological functioning has received increasing attention in recent decades. Previous studies exploring the behavioural and neurocognitive implications of digital technology have identified both positive and negative impacts (Korte, 2020; Small et al., 2020). The potential harms include phenomena such as behavioural addiction (Griffiths, 2000), attentional deficits (Ra et al., 2018), poorer working memory (Moisala et al., 2016), and putative links with elevated rates of depression (Twenge et al., 2018). In terms of beneficial impacts, frequent digital technology use has been associated with reduced cognitive decline in older adults (Small et al., 2020). Similarly, individuals who regularly play action video games are more adept at learning novel, real-world, sensorimotor tasks compared to their non-gaming counterparts (Gozli et al., 2014).

One conceptual framework that can help explain some of the positive/beneficial impacts of digital technology is neural exercise theory (Korte, 2020; Small et al., 2020). Within this framework, repeated technology use, results in lower neural activity associated with the task being performed, reflecting the acquisition of greater cognitive efficiency (Small et al., 2020). Similarly, intense and protracted technology use can also result in neurological changes at the cellular level - cortical plasticity (Gindrat et al., 2015). Regularly hearing a digital recording of one's own voice may, over time, also lead to an associated cognitive efficiency and neurological change. Furthermore, using a non-associative learning framework (Ioannou & Anastassiou-Hadjicharalambous, 2018), we might also predict that repeated exposure to recordings of one's own voice (for many people an aversive stimulus) will lead to habituation, as indicated by attenuated negative emotional reactivity (Ardiel et al., 2017).

It is well established that a significant proportion of people dislike the sound of their own voices. In one study, 31% of participants felt

* Corresponding author.
E-mail address: justin.thomas@aramco.com (J. Thomas).

https://doi.org/10.1016/j.chbr.2024.100446
Received 14 December 2023; Received in revised form 19 May 2024; Accepted 22 June 2024
Available online 24 June 2024

negatively about their voices, while 24% were neutral and 44% positive (Chong et al., 2022). Furthermore, when exposed to recordings of their own voices, many people suggest that the recorded voice sounds alien, somehow not them (Shuster & Durrant, 2003). This frequent perception of the recorded voice sounding odd relates to the way the sound of an utterance is conducted. A person's recorded voice is conducted through the air, unaffected by tissue and bone. Conversely, live/unrecorded utterances are conducted through bone, tissue and air, with approximately equal density (Tonndorf, 1970;). Resultantly, speakers rarely perceive their recorded self-voice as having the same fullness or depth as their live voice (Lee, Drinnan, & Carding, 2005; Shuster & Durrant, 2003). However, despite the discrepancy between live versus recorded utterances, individuals are generally adept at identifying their own pre-recorded voices - the discrepancy is not so great as to render the self-voice unrecognisable (Rosa et al., 2008).

Such self-voice processing abilities have mental health implications. For instance, the ability to discriminate self- and non-self-vocalization is an essential element of self-awareness, subserving self-monitoring during verbal communication (Conde et al., 2018). Difficulties with such self-voice processing are implicated in auditory verbal hallucinations (Conde et al., 2016). Similarly, self-voice dysphoria (extreme dislike) may be implicated in social anxiety (Lundh et al., 2002).

Frequent exposure to one's own voice, due to regularly communicating via voice notes, may lead to greater self-voice familiarity and liking. The objective of the present study is to explore the relationships between frequency of voice note use, self-voice recognition accuracy and self-voice liking. To the best of our knowledge, this is the first study to explore the possible implications of frequent voice note use for voice recognition and liking. As a preliminary exploration of this idea, we propose the following correlational hypotheses: (1) that frequent voice note replay (self-voice exposure) will be associated with greater self-voice liking, along with (2) more accurate and (3) faster self-voice recognition.

1. Methods

1.1. Participants

A convenience/opportunity sample of Arabic-speaking college women ($N = 128$) with no history of speech or hearing problems participated. The correlational nature of the study benefited from a sample of over 100 to ensure that sample correlation coefficients adequately reflected of the population correlation coefficients (Hole, 2023). The mean age for the participants was 20.90 ($SD = 2.31$) and the study was approved by the participating institution's internal review board (ZU19049F).

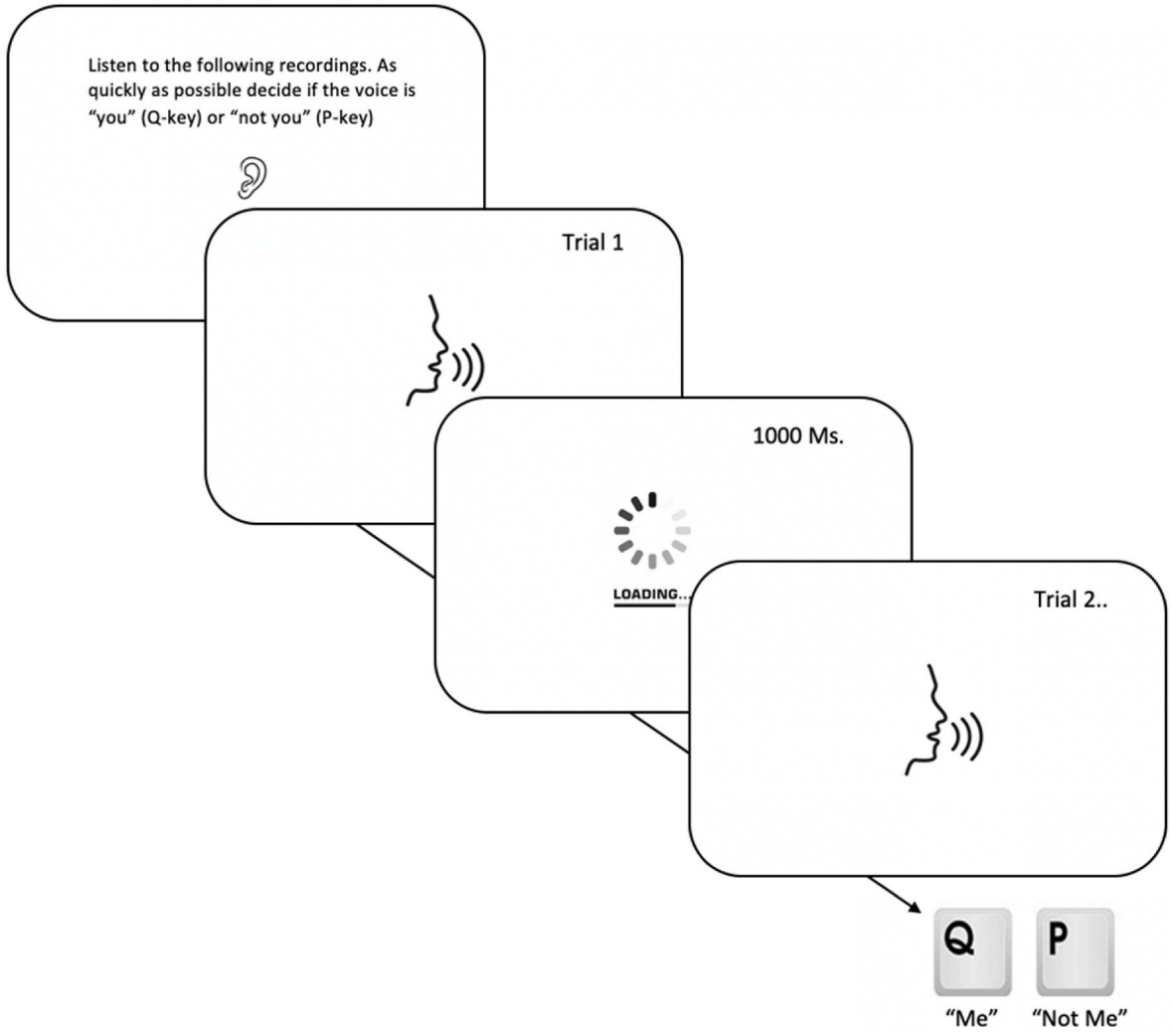

Fig. 1. A depiction of the steps involved in the SVR task. The target voice was either the participants own voice or that of another person (age-matched female). Participants categorized target voices as "own" or "other" by pressing the Q or P key respectively.

1.2. Materials

1.2.1. Self-voice recognition (SVR) task

The SVR task (developed by the first author) allowed participants to digitally record their own voices. For the current task, individuals were instructed to record the Arabic phrase: "marhaban bil 'alam" (hello world). The initial voice recording was performed prior to participants completing demographics and other questions, leaving approximately a 5-min gap between recording the phrase and being asked to perform the self-voice recognition trials.

The SVR element of the task required participants to successfully identify their previously recorded phrase (self-voice) across 20 trials. In 10 of the trials, the recorded voice was the participant's own. However, in the remaining 10 trials, the recorded phrase was voiced by other people (non-self voices), specifically, age-matched college women from the same city. All participants were instructed to decide, as quickly as possible, if the voice being played was them ("me") or not them ("not me"). Responses were indicated by pressing either the "Q" or "P" keys, respectively. There was a 1-s (1000 ms.) break between each trial, and the presentation order of trials (self vs non-self voice) was uniquely randomised for each participant. Accuracy and response times were recorded for each trial. When errors were made, they were recorded as either false positives (erroneously endorsing another person's voice as your own) or false negatives (erroneously dismissing one's own voice as being that of another person). Figure one details the sequence of the SVR trials.

A key strength of this approach is that it is an objective task-based measure of self-voice recognition. Additionally, once the task is understood, it can be performed alone without the potentially off-putting presence of the experimenter. However, one limitation of the SVR task is that such an assessment was new to many participants and therefore prone to being misunderstood. However, to ensure all participants were clear on how to perform the task, they performed an observed dummy run prior to performing the live trials. Furthermore, prior to conducting the study, the SVR task was validated/piloted with 20 students. The initial validation aimed to assess if the task was prone to floor or ceiling effects. Based on the performance data from the pilot, and the participant's feedback, the SVR task was adequately challenging and task performance was approximately normally distributed for response times and self-voice recognition accuracy.

1.2.2. Self-voice liking

Self-voice liking was assessed by a single-item self-report: "In general, how do you feel about the sound of your own voice?". The response scale was from 1 to 10, where 1 equalled "hate my voice" and 10 equalled "love my voice".

1.2.3. Voice note replay frequency

Voice note replay frequency was assessed by a single-item self-report: "How frequently do you replay your own voice notes?". The responses were 0–4: never, rarely, sometimes, often and always.

1.3. Procedure

Participants were tested individually in a soundproof laboratory. They were first tasked with making a digital recording of themselves uttering the Arabic phrase: "marhaban bil 'alam" (hello world). After this, participants completed the self-report items, followed by the SVR task. On average, approximately 5 min passed between the voice recording and the voice identification task.

2. Results

2.1. Self-voice recognition accuracy

Accurate self-voice recognition was defined as correctly identifying each instance of one's own voice (true positives), as well as correctly identifying alien voices (true negatives). In short, accurate self-voice recognition was indicated by making zero errors of any type (false positives or false negatives). Such accurate self-voice recognition, across all 20 trials, was obtained by 98 (76.56%) of participants. The error rate for inaccurate participants (23.44%) ranged from 1 to 7, with a mean error rate of 0.46 (SD = 1.16). False negatives were marginally more common (M = 0.24, SD = 0.70) than false positives (M = 0.21, SD = 0.65). Given that there were only two-types of error that participants could make (false positives and negatives) we explored differences using a paired-samples t-test. The error type differences were not statistically significant. Table one details the study's descriptive statistics.

2.2. Voice note replay

Participants were also asked about how frequently they replayed their own voice note recordings. This was scored on a 5-point scale the frequencies were as follows: never 17 (13.28%), rarely 35 (27.24%), sometimes 40 (31.25%), often 24 (18.75%), always 12 (9.37%) (see Table 1).

2.3. Correlations

Based on the continuous (interval and ratio level) data being analyzed, we explored all correlations using Pearson's product moment. As hypothesized, frequency of replaying one's own voice notes was correlated with self-voice recognition accuracy and with self-voice liking. Table 2 details the correlational analysis.

Additional analysis explored the relationship between voice liking and recognition error-types (false positives vs. false negatives). Both false positives (misidentifying another person's voice as one's own) and negatives (failure to correctly identify one own voice) were inversely correlated with voice liking. However, only the association with false positives reached statistical significance, r (126) = −0.249, p = 0.002. Similarly, only false positives were associated with less frequent voice note use, r (126) = −0.160, p = 0.035. There were no significant associations between error type and recognition response latencies.

3. Discussion

As in previous studies, self-voice dislike (35%) was common (Chong et al., 2022). Similarly, in line with past research (Rosa et al., 2008), participants were able to identify the pre-recorded self-voice, discriminating it from a non-self voice at rates far above chance.

In line with study hypotheses 1 and 2, voice note replay was positively correlated with self-voice recognition (accuracy) and self-voice liking. Voice note replay was correlated with self-voice recognition response latency (hypothesis 3), that is, frequent voice note users performed faster on the SVR task. However, this association failed to reach statistical significance, perhaps due to the relatively small number of trials performed. For hypothesis one, the observed relationship between frequent voice note use and improved self-voice recognition accuracy is with aligned with neural exercise theory (Korte, 2020; Small et al., 2020). Within this framework, frequent performance of a task (e.g.

Table 1
Descriptive statistics for continuous variables.

	Mean	*SD*	Median	*IQR*
Age	20.90	2.31	21	19–21
Self-voice liking	5.43	2.33	5.5	4–7
Recognition RT	120.38	37.75	113	97.75–144.25

Note: Recognition RT = self-voice recognition response time (milliseconds). *IQR* = inter quartile range, an indication of data variability. The number of participants in the analysis ranged from 119 to 128 due to occasional missing data for age.

Table 2
Correlations between key study variables.

	Age	Voice Liking	Recognition Accuracy	Recognition RT
Voice Note Replay	0.006	0.378**	0.152*	−0.110
Age		−0.064	−0.107	−0.053
Voice Liking			0.225**	−0.020
Recognition Accuracy				−0.065

** = $p < 00.01$ * = $p < 00.05$.
Note: Recognition RT = self-voice recognition response time (milliseconds). The number of participants in the analysis ranged from 119 to 128 due to occasional missing data for age.

listening one own voice) can lead to lower neural activity associated with the task in question, arguably reflecting an acquired cognitive efficiency. Similarly, the association between self-voice liking and frequent voice note use can be viewed through a non-associative learning lens (Ioannou & Anastassiou-Hadjicharalambous, 2018), where repeated exposure to an aversive stimulus (for some individuals, the sound of one's own voice) leads to habituation and the attenuation of negative emotional reactivity (Ardiel et al., 2017). However, it is also possible that people who like their own voices to begin with are more inclined toward listening to their own voice notes. It is also possible that a positive feedback loop is established, where self-voice liking leads to more voice note replays, which, in turn, fuels increases in self-voice liking. Ultimately, prospective longitudinal studies and experimental designs will be required to uncover possible causal relationships.

This was a preliminary study, and as such, it has several important limitations. Firstly, the correlational design precludes causal or temporal inferences, while the all-female student sample limits generalizability. Further limitations include the reliance on self-reports of voice note replay frequency and a relatively small number of trials (N= 20) on the self-voice recognition task. Future studies might use prospective longitudinal designs and behavioural measures to capture actual voice note replay frequency. Additionally, doubling the number of trials on the self-voice recognition task might allow for a more sensitive assessment of self-voice recognition response latencies. Limitations aside, we view the present study as a preliminary step towards experimental designs. Future controlled studies might prescribe voice note replay to examine its impact on self-voice liking, and recognition accuracy and latency.

This study represents a preliminary investigation of a potentially important phenomenon, especially given our increasing use of voice notes. Furthermore, the links between self-voice liking/recognition, social anxiety disorder (Lundh et al., 2002) and auditory verbal hallucinations (Conde et al., 2016) offer potential areas for therapeutic application. Perhaps using voice note replay as a form of self-voice exposure, with the aim of decreasing self-voice dislike and enhancing self-voice recognition. The present study, although correlational, provides an indication that further experimental research is merited. Such explorations may lead to new interventions for social anxiety symptomatology and auditory verbal hallucinations. Furthermore, this computerised, performance-based assessment of self-voice recognition may prove useful in future investigations targeting self-voice processing capabilities.

CRediT authorship contribution statement

Justin Thomas: Writing – review & editing, Writing – original draft, Methodology, Formal analysis, Conceptualization. **Jigar Jogia:** Formal analysis, Data curation. **Mariapaola Barbato:** Formal analysis, Data curation. **Richard Bentall:** Formal analysis, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

References

Ardiel, E. L., Yu, A. J., Giles, A. C., & Rankin, C. H. (2017). Habituation as an adaptive shift in response strategy mediated by neuropeptides. *Npj Science of Learning, 2*(1), 9. https://doi.org/10.1038/s41539-017-0011-8

Chong, H. J., Choi, J. H., & Lee, S. S. (2022). Does the perception of own voice affect our behavior? *Journal of Voice*. https://doi.org/10.1016/j.jvoice.2022.02.003

Collins English Dictionary. (2022). *Voice note*. In https://www.collinsdictionary.com/submission/20895/voice+note.

Conde, T., Gonçalves, O. F., & Pinheiro, A. P. (2016). A cognitive neuroscience view of voice-processing abnormalities in schizophrenia: A window into auditory verbal hallucinations? *Harvard Review of Psychiatry, 24*(2), 148–163. https://doi.org/10.1097/hrp.0000000000000082

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

Gindrat, A. D., Chytiris, M., Balerna, M., Rouiller, E. M., & Ghosh, A. (2015). [Smartphone use shapes cortical tactile sensory processing from the fingertips]. *Medical Science, 31*, 363–366. https://doi.org/10.1051/medsci/20153104006 (L'utilisation de smartphones façonne le traitement cortical de l'information sensorielle tactile provenant de l'extrémité des doigts.).

Gozli, D. G., Bavelier, D., & Pratt, J. (2014). The effect of action video game playing on sensorimotor learning: Evidence from a movement tracking task. *Human Movement Science, 38*, 152–162. https://doi.org/10.1016/j.humov.2014.09.004

Griffiths, M. (2000). Internet addiction - Time to be taken seriously? *Addiction Research, 8*(5), 413–418. https://doi.org/10.3109/16066350009005587

Hole, G. (2023). *Research skills one correlation interpretation*. http://users.sussex.ac.uk/~grahamh/RM1web/Eight/20things/20you/20need/20to/20know/20about/20interpreting/20correlations.pdf.

Ioannou, A., & Anastassiou-Hadjicharalambous, X. (2018). Non-associative learning. In T. K. Shackelford, & V. A. Weekes-Shackelford (Eds.), *Encyclopedia of evolutionary psychological science* (pp. 1–13). Springer International Publishing. https://doi.org/10.1007/978-3-319-16999-6_1027-1.

Korte, M. (2020). The impact of the digital revolution on human brain and behavior: Where do we stand? *Dialogues in Clinical Neuroscience, 22*(2), 101–111. https://doi.org/10.31887/DCNS.2020.22.2/mkorte

Lundh, L.-G., Berg, B., Johansson, H., Nilsson, L. K., Sandberg, J., & Segerstedt, A. (2002). Social anxiety is associated with a negatively distorted perception of one's own voice. *Cognitive Behaviour Therapy, 31*(1), 25–30. https://doi.org/10.1080/16506070252823634

Moisala, M., Salmela, V., Hietajärvi, L., Salo, E., Carlson, S., Salonen, O., … Alho, K. (2016). Media multitasking is associated with distractibility and increased prefrontal activity in adolescents and young adults. *NeuroImage, 134*, 113–121. https://doi.org/10.1016/j.neuroimage.2016.04.011

Ra, C. K., Cho, J., Stone, M. D., De La Cerda, J., Goldenson, N. I., Moroney, E., … Leventhal, A. M. (2018). Association of digital media use with subsequent symptoms of attention-deficit/hyperactivity disorder among adolescents. *JAMA, 320*(3), 255–263. https://doi.org/10.1001/jama.2018.8931

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

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

Singh, M. (2022). *People are sending seven billion voice messages on WhatsApp every day*. TechCrunch. March/30/2022 https://techcrunch.com/2022/03/30/people-are-sending-7-billion-voice-messages-on-whatsapp-every-day/.

Small, G. W., Lee, J., Kaufman, A., Jalil, J., Siddarth, P., Gaddipati, H., … Bookheimer, S. Y. (2020). Brain health consequences of digital technology use. *Dialogues in Clinical Neuroscience, 22*(2), 179–187. https://doi.org/10.31887/DCNS.2020.22.2/gsmall

Twenge, J. M., Joiner, T. E., Rogers, M. L., & Martin, G. N. (2018). Increases in depressive symptoms, suicide-related outcomes, and suicide rates among U.S. adolescents after 2010 and links to increased new media screen time. *Clinical Psychological Science, 6*(1), 3–17. https://doi.org/10.1177/2167702617723376