How Understanding Voice Naturalness Includes Authenticity and Further Aspects of Self-Voice Processing: A Reply to Pinheiro (2025)
--Manuscript Draft--

Manuscript Number:	
Full Title:	How Understanding Voice Naturalness Includes Authenticity and Further Aspects of Self-Voice Processing: A Reply to Pinheiro (2025)
Article Type:	Letters to the editor
Corresponding Author:	Christine Nussbaum Friedrich Schiller University Jena: Friedrich-Schiller-Universitat Jena GERMANY
First Author:	Christine Nussbaum
Order of Authors:	Christine Nussbaum
	Sascha Frühholz
	Stefan R. Schweinberger
Section/Category:	Psychological and Physiological Acoustics
Keywords:	voice; naturalness; self-voice; authenticity
Abstract:	Understanding voice naturalness is of substantial practical importance to very different applied fields, including voice manipulation, synthesis, and pathology. We proposed an inclusive framework to unify manifold research efforts under a concise conceptualization of voice naturalness (Nussbaum et al., 2025). A recent letter by Pinheiro (2025) criticized that our framework (1) overlooks the self-voice and (2) reduces naturalness to considerations of acoustics alone. Both assertions are incorrect. In this response, we resolve these misunderstandings by showing how self-voice research, while facing conceptual and methodological challenges of its own, is easily integrated into our framework and its concept of authenticity.
Additional Information:	
Question	Response
Please read the Transfer of Copyright Agreement and the Instructions. By selecting the appropriate response below, you verify that you have consented to the Transfer of Copyright Agreement (if applicable) and the conditions and representations set forth in the Instructions.	R - Regular

Cover Letter

FRIEDRICH-SCHILLER-UNIVERSITÄT JENA

Institut für Psychologie
Lehrstuhl für Allgemeine Psychologie und kognitive Neurowissenschaften

Univ.-Prof. Dr. Stefan R. Schweinberger

Christine Nussbaum, PhD

Am Steiger 3/Haus 1
07743 Jena

Telefon: 0 36 41 9-45181
Telefax: 0 36 41 9-45182
E-Mail: stefan.schweinberger@uni-jena.de
                christine.nussbaum@uni-jena.de

Universität Jena · Lehrstuhl für Allgemeine Psychologie · Institut für Psychologie · 07737 Jena

To the editors of
Journal of the Acoustical Society of America (JASA)

Jena, 12. Januar 2026

Manuscript submission

Dear Dr. Kreiman, dear Editors of *(JASA)*

attached please find a *letter to the editor* entitled "How *Understanding Voice Naturalness* Includes Authenticity and Further Aspects of Self-Voice Processing: A Reply to Pinheiro (2025)", by Christine Nussbaum, Sascha Frühholz and Stefan R. Schweinberger.

This letter is a response to Pinheiro (2025): "Beyond acoustics: Self-relevance as a key to voice naturalness (L)", which was also published as a letter to the editor in JASA two months ago. In this work, Pinheiro discusses our newly-proposed framework for voice naturalness (published in *Trends in Cognitive Sciences* in May 2025, https://doi.org/10.1016/j.tics.2025.01.010), criticizing that it (1) overlooked the self-voice and (2) reduces naturalness to aspects of voice acoustics alone. We welcome this opportunity to clarify that both assertions by Pinheiro (2025) are incorrect. Quite to the contrary – aspects self-voice processing can be easily integrated in our existing framework and are not at all in conflict with our proposed definitions for voice naturalness. We presume that these misunderstandings arose partly because our article in TiCS discussed self-voice processing only briefly and without much detail, as a natural consequence of the nature of these papers and our focus on a different topic.

In our reply, we resolve these misunderstandings by showing how the self-voice can be connected to aspects of naturalness and authenticity in the context of our framework. Further, we expand on the important points Pinheiro raised, and share our perspective on current challenges around research on self-voice processing.

This response gives us the important opportunity to clarify our conceptual framework of voice naturalness and rectify incorrect claims such as it reduces naturalness to voice acoustics alone. As outlined in the original publication (https://doi.org/10.1016/j.tics.2025.01.010), a concise definitional framework is a crucial prerequisite for a systematic understanding voice naturalness (and – as we hold – for self-voice processing as well). As similar issues also span into the visual domain, our response briefly draws parallels into current controversies that exist for self-face research as well. In short, we believe that controversial debates about proper conceptualizations of timely topics prior to systematic empirical efforts are of high value to researchers within and beyond the auditory domain. We therefore expect this response letter to be of great interest to the wide readership of *JASA*, and look forward to your consideration and evaluation at your earliest convenience.

Yours sincerely,
Christine Nussbaum, Sascha Frühholz and Stefan R. Schweinberger

How *Understanding Voice Naturalness* Includes Authenticity and Further Aspects of Self-Voice Processing: A Reply to Pinheiro (2025)

Christine Nussbaum[1,2], Sascha Frühholz[3,4], and Stefan R. Schweinberger[1,2,5]

[1]Department for General Psychology and Cognitive Neuroscience, Friedrich Schiller University Jena, 07743 Jena, Germany

[2]Voice Research Unit, Friedrich Schiller University, 07743 Jena, Germany

[3]Department of Psychology, University of Oslo, 0371 Oslo, Norway

[4]Cognitive and Affective Neuroscience Unit, University of Zurich, 8050 Zurich, Switzerland

[5]Swiss Center for Affective Sciences, University of Geneva, 1222 Geneva, Switzerland

Correspondence should be addressed to Christine Nussbaum (https://www.allgpsy.uni-jena.de/christine-nussbaum/), Department for General Psychology and Cognitive Neuroscience, Friedrich Schiller University Jena, Am Steiger 3/1, 07743 Jena, Germany. Tel: +49 (0) 3641 945934, E-Mail: christine.nussbaum@uni-jena.de.

Abstract: Understanding voice naturalness is of substantial practical importance to very different applied fields, including voice manipulation, synthesis, and pathology. We proposed an inclusive framework to unify manifold research efforts under a concise conceptualization of voice naturalness (Nussbaum et al., 2025). A recent letter by Pinheiro (2025) criticized that our framework (1) overlooks the self-voice and (2) reduces naturalness to considerations of acoustics alone. Both assertions are incorrect. In this response, we resolve these misunderstandings by showing how self-voice research, while facing conceptual and methodological challenges of its own, is easily integrated into our framework and its concept of authenticity.

What makes a voice sound *natural*? This is a question of high relevance across many domains: in speech-language pathology, across the manifold possibilities of voice modification, and in voice synthesis and cloning, a rapidly developing field that is now invading everyday life from various directions (Lavan et al., 2025). Against this backdrop, it is rather astonishing that an obvious and related question received relatively little attention until now: What do we *mean* by voice naturalness? To address this conceptual gap, we recently published a framework for voice naturalness, which offers a concise definition that is rooted in theoretical models on voice perception and is flexibly applicable to diverse empirical contexts (Nussbaum et al., 2025). In a nutshell, we proposed a taxonomy with two distinct types: deviation-based naturalness and human-likeness-based naturalness. The deviation-based approach assesses naturalness in terms of distance away from a reference, while the human-likeness-based approach assesses naturalness according to its similarity to a specific reference – the human voice. With voice naturalness being a diverse and highly interdisciplinary topic (Pandey et al., 2025), we are pleased to see that our work has already inspired a deeper discussion on its widespread implications.

In a recent letter to the editor, Pinheiro (2025) highlighted the importance of naturalness regarding the perception of one's own voice. A key point in this letter is that impressions of one's own voice (including synthesized versions of it) as sounding "natural to me" are the result of a complex cognitive process that cannot be explained by the acoustics of the voice signal alone. Instead, impressions of self-voice are based on self-relevance, defined by Pinheiro as the alignment between that voice and the listener's internal self-representation. Pinheiro (2025) claims that this is an aspect that our framework "overlooked", and the author proposes self-relevance as a complementary dimension for voice naturalness.

We welcome the opportunity to reply to Pinheiro´s letter. Of course, there is no doubt that the discussion around the perception of one´s own voice touches important issues in voice research, which includes interdisciplinary (e.g., neurocognitive, acoustic, philosophical, and medical) perspectives. Furthermore, we agree that impressions of naturalness in self-voice evaluation deserve

more attention. Above all, however, we want to emphasize that this is not in conflict but in complete accordance with our proposed framework. The points raised by Pinheiro (2025) neither reveal a blind spot nor an incomplete conceptualization of voice naturalness in Nussbaum et al. (2025), and any apparent contradiction stems from misconceptions of our framework. Our response is twofold: in the first part, we resolve these misunderstandings and outline how self-voice perception fits well within our proposed model. In the second part, we expand on the important points Pinheiro raised and share our perspective on current challenges around research on self-voice processing.

Voice naturalness and its links to self-voice processing

Pinheiro (2025) writes that our dimension of "deviation-based naturalness captures how much a voice diverges from typical acoustic patterns" (p. 4045). This is not fully correct. In fact, we refer to the "deviation from a reference that represents maximum naturalness" (Nussbaum et al., 2025, p. 472). Although the difference may seem small, it is crucial because it captures the key point raised by Pinheiro (2025): that impressions of naturalness are influenced by many aspects that go beyond acoustics alone. As we outlined, it is common for listeners to use an internal implicit reference based on their experience and expectations (Nussbaum et al., 2025, p. 472). Logically, the same acoustic signal can result in different subjective impressions, depending on listener and context variables. There is ample evidence for this: For example, listeners judge speech in their own accent as more natural than foreign-accented speech (Kapolowicz et al., 2022). The perception and acceptance of synthetic voices also seems modulated by listeners' personalities (Lee, 2010) and multimodal context, such as the visual appearance of an artificial agent (Mitchell et al., 2011). Moreover, auditory adaptation aftereffects in the perception of acoustically identical voices (Schweinberger et al., 2008) have now been established as a ubiquitous phenomenon. We maintain that a full understanding of voice naturalness requires more systematic research that assesses differences across listeners and contexts, as we outlined in the *outstanding questions* section. Pinheiro (2025) sums this up perfectly: "Naturalness […] must therefore be understood as emerging from the interaction between the acoustic properties of speech and the listener's representational

framework." (page 4046). But in fact, this is the essence of our model (cf. Figure 2, Nussbaum et al. 2025, page 473).

Another point raised by Pinheiro (2025) was that Nussbaum et al. (2025) "overlooked the self-voice". This is incorrect as well. We intentionally decided to only briefly touch on the issue where relevant, in order not to distract from the main points of our paper. What we did consider is an important aspect of self-voice that provides substantial motivation for understanding voice naturalness (that was in turn not mentioned by Pinheiro, 2025): we discussed the seminal work by Yamagishi et al. (2012) and Hyppa-Martin et al. (2024), who evaluate personalized speech-synthesis technology for individuals who lost their biological voice (e.g., due to laryngectomy) and outlined how important it is that these devices sound authentic and like "their own natural voice".

On a related note, while we welcome Pinheiro's call for more scientific attention towards multifaceted aspects that contribute to the impression of a voice as one's own, closer contact with Nussbaum et al. (2025) reveals how this call can be easily integrated into the existing framework. Specifically, our distinction between *naturalness* and *authenticity* represents an obvious entry point (cf. Figure 3, Nussbaum et al. 2025, page 475). Whereas naturalness refers to the holistic impression of a voice (i.e. "does a voice sound natural or unnatural?"), authenticity has been used to refer to specific social signals, such as emotion, gender or identity (i.e. "does this voice sound like an authentic or posed expression of anger?"). Identity authenticity, specifically, is assessed with regard to a specific speaker: "does this voice sound authentically like speaker X?" (e.g.,Roswandowitz et al., 2024). Self-voice assessments can be seen as a special case of identity authenticity: "Does this voice authentically sound *like me*?" Note that this does not imply that self-voice assessment is equivalent to assessment of authenticity for other speakers' identities. Pinheiro (2025) listed several reasons why the self-voice could be a special case, which we will even expand upon below (see also Frühholz & Schweinberger, 2021). We therefore do not object, but we quite simply find that a conceptual clarification linking it to the concepts of voice naturalness and authenticity eradicates the need to declare self-relevance as a complementary dimension.

Current Challenges for Understanding Self-Voice Processing

Here we complement the points raised by Pinheiro by reflecting on current challenges for self-voice research, which are relevant for voice naturalness and beyond. In fact, there are intriguing parallels to voice naturalness research: while highly relevant, previous research on the self-voice has been fragmentary and unsystematic, and hampered by methodological and conceptual challenges. For instance, recordings of the own voice notoriously sound unnatural for a listener, because they lack bone conduction cues (Maurer & Landis, 1990). To become relevant for naturalness research, this unnaturalness of the own voice needs to be considered by presenting stimuli with special devices (i.e., bone-conducting headphones). The vast majority of published papers on the self-voice (including a recent one by Pinheiro et al., 2023) does not consider this issue in their experimental setup, although a few notable exceptions now exist (e.g. Orepic et al., 2023).

Finally, we should not forget that impressions about real people are often elicited by dynamic multisensory (vocal, facial, body motion) cues. The relatively short but intense history of voice perception research (Frühholz & Belin, 2018) suggests that the human brain treats voices and faces as sources of information about real persons, and that voice perception theory continues to benefit from face perception models (e.g. Belin et al., 2011; Young et al., 2020). One important aspect of self-voice, not mentioned by Pinheiro, is kinship recognition, which is deeply rooted in our biology (Mateo, 2015) and well-studied in the face domain (Maloney & Dal Martello, 2006), but virtually unstudied in human voice perception. There are good reasons to believe that perceived self-similarity is one of the key mechanisms of kinship recognition (DeBruine et al., 2009; Tsuchiya & Schweinberger, 2022), and that brain mechanisms for discriminating self and kin from unfamiliar persons may be partially overlapping (Platek & Kemp, 2009). But even after decades of self-face perception research, the question of whether the self is special does remain controversial in psychology, neuroscience and philosophy (Kovács et al., 2025; Sui & Humphreys, 2017). Against the background of far fewer empirical studies, proposals that self-*voice* perception is special are

intriguing, but require more systematic evidence, as well as an integrative framework. We explicitly acknowledge ongoing work in that direction (Orepic & Pinheiro, 2025).

In conclusion, we appreciate that the response to this letter gave us the chance to clarify how the self-voice can be seamlessly linked to our existing framework of voice naturalness. We look forward to empirical contributions that link phenomena of self-voice processing to this framework and expect that these will be helpful to further develop and refine the model. We also reflected on the challenges of self-voice research more broadly, both methodologically and conceptually. With increasingly sophisticated studies, we are confident that future research programs will address these challenges and promote further progress that will mutually benefit a better understanding of both self-voice processing and voice naturalness impressions.

Conflict of interest and Funding

References

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

DeBruine, L. M., Smith, F. G., Jones, B. C., Roberts, S. C., Petrie, M., & Spector, T. D. (2009). Kin recognition signals in adult faces. *Vision Res*, *49*(1), 38–43. https://doi.org/10.1016/j.visres.2008.09.025

Frühholz, S., & Belin, P. (2018). *The Oxford Handbook of Voice Perception*. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780198743187.001.0001

Frühholz, S., & Schweinberger, S. R. (2021). Nonverbal auditory communication - Evidence for

   integrated neural systems for voice signal production and perception. *Progress in*

   Neurobiology, *199*, 101948. https://doi.org/10.1016/j.pneurobio.2020.101948

Hyppa-Martin, J., Lilley, J., Chen, M., Friese, J., Schmidt, C., & Bunnell, H. T. (2024). A large-scale

   comparison of two voice synthesis techniques on intelligibility, naturalness, preferences, and

   attitudes toward voices banked by individuals with amyotrophic lateral sclerosis.

   Augmentative and Alternative Communication, *40*(1), 31–45.

   https://doi.org/10.1080/07434618.2023.2262032

Kapolowicz, M. R., Guest, D. R., Montazeri, V., Baese-Berk, M. M., & Assmann, P. F. (2022). Effects of

   Spectral Envelope and Fundamental Frequency Shifts on the Perception of Foreign-Accented

   Speech. *Language and Speech*, *65*(2), 418–443.

   https://doi.org/10.1177/00238309211029679

Kovács, G., Göschel, L., Pawlik, S. M., & Tramacere, A. (2025). The neural dynamics of current and

   past self-face perception: Challenging the privilege access hypothesis. *Cortex; a Journal*

   Devoted to the Study of the Nervous System and Behavior, *194*, 77–90.

   https://doi.org/10.1016/j.cortex.2025.11.006

Lavan, N., Irvine, M., Rosi, V., & McGettigan, C. (2025). Voice clones sound realistic but not (yet)

   hyperrealistic. *PLoS One*, *20*(9), e0332692. https://doi.org/10.1371/journal.pone.0332692

Lee, E.-J. (2010). The more humanlike, the better? How speech type and users' cognitive style affect

   social responses to computers. *Computers in Human Behavior*, *26*(4), 665–672.

   https://doi.org/10.1016/j.chb.2010.01.003

Maloney, L. T., & Dal Martello, M. F. (2006). Kin recognition and the perceived facial similarity of

   children. *Journal of Vision*, *6*(10), 1047–1056. https://doi.org/10.1167/6.10.4

Mateo, J. M. (2015). Perspectives: Hamilton's Legacy: Mechanisms of Kin Recognition in Humans.

   Ethology, *121*(5), 419–427. https://doi.org/10.1111/eth.12358

Maurer, D., & Landis, T. (1990). Role of bone conduction in the self-perception of speech. *Folia*

   Phoniatrica, *42*(5), 226–229. https://doi.org/10.1159/000266070

Mitchell, W. J., Szerszen, K. A., Lu, A. S., Schermerhorn, P. W., Scheutz, M., & Macdorman, K. F.

    (2011). A mismatch in the human realism of face and voice produces an uncanny valley. *I-*

    Perception, *2*(1), 10–12. https://doi.org/10.1068/i0415

Nussbaum, C., Frühholz, S., & Schweinberger, S. R. (2025). Understanding voice naturalness. *Trends*

    in Cognitive Sciences, *29*(5), 467–480. https://doi.org/10.1016/j.tics.2025.01.010

Orepic, P., Kannape, O. A., Faivre, N., & Blanke, O. (2023). Bone conduction facilitates self-other voice

    discrimination. *Royal Society Open Science*, *10*(2), 221561.

    https://doi.org/10.1098/rsos.221561

Orepic, P., & Pinheiro, A. (2025). *From Voice to Self: An Integrative Framework on Self-Voice*

    Processing. PsyArXiv (preprint) https://doi.org/10.31234/osf.io/kg4ns_v2

Pandey, A., Le Maguer, S., & Harte, N. (2025). What is Naturalness? In *13th edition of the Speech*

    Synthesis Workshop (pp. 215–221). ISCA. https://doi.org/10.21437/SSW.2025-33

Pinheiro, A. P. (2025). Beyond acoustics: Self-relevance as a key to voice naturalness (L). *The Journal*

    of the Acoustical Society of America, *158*(5), 4045–4047.

    https://doi.org/10.1121/10.0039927

Pinheiro, A. P., 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

Platek, S. M., & Kemp, S. M. (2009). Is family special to the brain? An event-related fMRI study of

    familiar, familial, and self-face recognition. *Neuropsychologia*, *47*(3), 849–858.

    https://doi.org/10.1016/j.neuropsychologia.2008.12.027

Roswandowitz, C., Kathiresan, T., Pellegrino, E., Dellwo, V., & Frühholz, S. (2024). Cortical-striatal

    brain network distinguishes deepfake from real speaker identity. *Communications Biology*,

    7(1), 711. https://doi.org/10.1038/s42003-024-06372-6

Schweinberger, S. R., Casper, C., Hauthal, N., Kaufmann, J. M., Kawahara, H., Kloth, N.,

    Robertson, D. M., Simpson, A. P., & Zäske, R. (2008). Auditory adaptation in voice perception.

    Curr Biol, *18*(9), 684–688. https://doi.org/10.1016/j.cub.2008.04.015

Sui, J., & Humphreys, G. W. (2017). The ubiquitous self: What the properties of self-bias tell us about the self. *Ann N Y Acad Sci*, *1396*(1), 222–235. https://doi.org/10.1111/nyas.13197

Tsuchiya, A., & Schweinberger, S. R. (2022). Erkennen von Verwandtschaft zu sich Selbst und bei Anderen [Recognition of Kinship to Self and in Others]: Part 2, Issue 1/2022. *In-Mind Magazine*, *1*(2). https://de.in-mind.org/article/erkennen-von-verwandtschaft-zu-sich-selbst-und-bei-anderen

Yamagishi, J., Veaux, C., King, S., & Renals, S. (2012). Speech synthesis technologies for individuals with vocal disabilities: Voice banking and reconstruction. *Acoustical Science and Technology*, *33*(1), 1–5. https://doi.org/10.1250/ast.33.1

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