

The Journal of the Acoustical Society of America
How Understanding Voice Naturalness Includes Authenticity and Further Aspects of Self-Voice Processing: A Comment on Pinheiro (2025)
--Manuscript Draft--

Manuscript Number:	JASA-13582
Full Title:	How Understanding Voice Naturalness Includes Authenticity and Further Aspects of Self-Voice Processing: A Comment on Pinheiro (2025)
Article Type:	Letters to the editor
Corresponding Author:	Christine Nussbaum Friedrich Schiller University Jena: Friedrich-Schiller-Universität Jena GERMANY
First Author:	Christine Nussbaum
Order of Authors:	Christine Nussbaum Sascha Fröhholz 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 comment, 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

DO NOT
DISTRIBUTE



FRIEDRICH-SCHILLER-
UNIVERSITÄT
JENA

Institut für Psychologie

Lehrstuhl für Allgemeine Psychologie und
kognitive Neurowissenschaften

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

Univ.-Prof. Dr. Stefan R. Schweinberger

Christine Nussbaum, PhD

To the editors of

Journal of the Acoustical Society of America (JASA)

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

Jena, 7. Februar 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 Comment on Pinheiro (2025)”, by Christine Nussbaum, Sascha Frühholz and Stefan R. Schweinberger.

This letter is a comment on 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 three 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 comment 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



1 **How Understanding Voice Naturalness Includes Authenticity and Further**

2 **Aspects of Self-Voice Processing: A Comment on Pinheiro (2025)**

3

4 Christine Nussbaum^{1,2}, Sascha Fröhholz^{3,4}, and Stefan R. Schweinberger^{1,2,5}

5

6 ¹Department for General Psychology and Cognitive Neuroscience, Friedrich Schiller University Jena,
7 07743 Jena, Germany

8 ²Voice Research Unit, Friedrich Schiller University, 07743 Jena, Germany

9 ³Department of Psychology, University of Oslo, 0371 Oslo, Norway

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

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

12

13 Correspondence should be addressed to Christine Nussbaum (<https://www.fsv.uni-jena.de/en/63723/christine-nussbaum@uni-jena-de>), Department for General Psychology and
14 Cognitive Neuroscience, Friedrich Schiller University Jena, Am Steiger 3/1, 07743 Jena, Germany. Tel:
15 +49 (0) 3641 945934, E-Mail: christine.nussbaum@uni-jena.de.

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

25

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

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

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

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

59 **Voice naturalness and its links to self-voice processing**

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

78 framework." (page 4046). But in fact, this is the essence of our model (cf. Figure 2, Nussbaum et al.
79 2025, page 473).

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

88 On a related note, while we welcome Pinheiro's call for more scientific attention towards
89 multifaceted aspects that contribute to the impression of a voice as one's own, closer contact with
90 Nussbaum et al. (2025) reveals how this call can be easily integrated into the existing framework.
91 Specifically, our distinction between *naturalness* and *authenticity* represents an obvious entry point
92 (cf. Figure 3, Nussbaum et al. 2025, page 475). Whereas naturalness refers to the holistic impression
93 of a voice (i.e. "does a voice sound natural or unnatural?"), authenticity has been used to refer to
94 specific social signals, such as emotion, gender or identity (i.e. "does this voice sound like an
95 authentic or posed expression of anger?"). Identity authenticity, specifically, is assessed with regard
96 to a specific speaker: "does this voice sound authentically like speaker X?" (e.g., Roswadowitz et al.,
97 2024). Self-voice assessments can be seen as a special case of identity authenticity: "Does this voice
98 authentically sound *like me?*" Note that this does not imply that self-voice assessment is equivalent
99 to assessment of authenticity for other speakers' identities. Pinheiro (2025) listed several reasons
100 why the self-voice could be a special case, which we will even expand upon below (see also Frühholz
101 & Schweinberger, 2021). We therefore do not object, but we quite simply find that a conceptual
102 clarification linking it to the concepts of voice naturalness and authenticity eradicates the need to
103 declare self-relevance as a complementary dimension.

104 **Current Challenges for Understanding Self-Voice Processing**

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

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

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

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

139 **Acknowledgments**

140 We thank Manuel Pöhlmann for helpful suggestions about a previous version of this work. The
141 authors gratefully acknowledge the award of funding through an EU-MSCA doctoral network “Voice
142 Communication Sciences” (action 101168998).

143 **Author Declarations**

144 The authors have no conflicts to disclose.

145 **Data Availability**

146 There is no data linked to this manuscript.

147 **References**

- 148 Belin, P., Bestelmeyer, P. E. G., Latinus, M., & Watson, R. (2011). Understanding voice perception.
149 *British Journal of Psychology*, 102(4), 711–725. <https://doi.org/10.1111/j.2044-8295.2011.02041.x>
- 150
151 DeBruine, L. M., Smith, F. G., Jones, B. C., Roberts, S. C., Petrie, M., & Spector, T. D. (2009). Kin
152 recognition signals in adult faces. *Vision Res*, 49(1), 38–43.
153 <https://doi.org/10.1016/j.visres.2008.09.025>

- 154 Frühholz, S., & Belin, P. (2018). *The Oxford Handbook of Voice Perception*. Oxford University Press.
- 155 <https://doi.org/10.1093/oxfordhb/9780198743187.001.0001>
- 156 Frühholz, S., & Schweinberger, S. R. (2021). Nonverbal auditory communication - Evidence for
- 157 integrated neural systems for voice signal production and perception. *Progress in*
- 158 *Neurobiology*, 199, 101948. <https://doi.org/10.1016/j.pneurobio.2020.101948>
- 159 Hyppa-Martin, J., Lilley, J., Chen, M., Friese, J., Schmidt, C., & Bunnell, H. T. (2024). A large-scale
- 160 comparison of two voice synthesis techniques on intelligibility, naturalness, preferences, and
- 161 attitudes toward voices banked by individuals with amyotrophic lateral sclerosis.
- 162 *Augmentative and Alternative Communication*, 40(1), 31–45.
- 163 <https://doi.org/10.1080/07434618.2023.2262032>
- 164 Kapolowicz, M. R., Guest, D. R., Montazeri, V., Baese-Berk, M. M., & Assmann, P. F. (2022). Effects of
- 165 Spectral Envelope and Fundamental Frequency Shifts on the Perception of Foreign-Accented
- 166 Speech. *Language and Speech*, 65(2), 418–443.
- 167 <https://doi.org/10.1177/00238309211029679>
- 168 Kovács, G., Göschel, L., Pawlik, S. M., & Tramacere, A. (2025). The neural dynamics of current and
- 169 past self-face perception: Challenging the privilege access hypothesis. *Cortex; a Journal*
- 170 *Devoted to the Study of the Nervous System and Behavior*, 194, 77–90.
- 171 <https://doi.org/10.1016/j.cortex.2025.11.006>
- 172 Lavan, N., Irvine, M., Rosi, V., & McGettigan, C. (2025). Voice clones sound realistic but not (yet)
- 173 hyperrealistic. *PLoS One*, 20(9), e0332692. <https://doi.org/10.1371/journal.pone.0332692>
- 174 Lee, E.-J. (2010). The more humanlike, the better? How speech type and users' cognitive style affect
- 175 social responses to computers. *Computers in Human Behavior*, 26(4), 665–672.
- 176 <https://doi.org/10.1016/j.chb.2010.01.003>
- 177 Maloney, L. T., & Dal Martello, M. F. (2006). Kin recognition and the perceived facial similarity of
- 178 children. *Journal of Vision*, 6(10), 1047–1056. <https://doi.org/10.1167/6.10.4>
- 179 Mateo, J. M. (2015). Perspectives: Hamilton's Legacy: Mechanisms of Kin Recognition in Humans.
- 180 *Ethology*, 121(5), 419–427. <https://doi.org/10.1111/eth.12358>

- 181 Maurer, D., & Landis, T. (1990). Role of bone conduction in the self-perception of speech. *Folia*
182 *Phoniatrica*, 42(5), 226–229. <https://doi.org/10.1159/000266070>
- 183 Mitchell, W. J., Szerszen, K. A., Lu, A. S., Schermerhorn, P. W., Scheutz, M., & Macdorman, K. F.
184 (2011). A mismatch in the human realism of face and voice produces an uncanny valley. *I-*
185 *Perception*, 2(1), 10–12. <https://doi.org/10.1068/i0415>
- 186 Nussbaum, C., Fröhholz, S., & Schweinberger, S. R. (2025). Understanding voice naturalness. *Trends*
187 *in Cognitive Sciences*, 29(5), 467–480. <https://doi.org/10.1016/j.tics.2025.01.010>
- 188 Orepic, P., Kannape, O. A., Faivre, N., & Blanke, O. (2023). Bone conduction facilitates self-other voice
189 discrimination. *Royal Society Open Science*, 10(2), 221561.
190 <https://doi.org/10.1098/rsos.221561>
- 191 Orepic, P., & Pinheiro, A. (2025). *From Voice to Self: An Integrative Framework on Self-Voice*
192 *Processing*. PsyArXiv (preprint) https://doi.org/10.31234/osf.io/kg4ns_v2
- 193 Pandey, A., Le Maguer, S., & Harte, N. (2025). What is Naturalness? In *13th edition of the Speech*
194 *Synthesis Workshop* (pp. 215–221). ISCA. <https://doi.org/10.21437/SSW.2025-33>
- 195 Pinheiro, A. P. (2025). Beyond acoustics: Self-relevance as a key to voice naturalness (L). *The Journal*
196 *of the Acoustical Society of America*, 158(5), 4045–4047.
197 <https://doi.org/10.1121/10.0039927>
- 198 Pinheiro, A. P., Sarzedas, J., Roberto, M. S., & Kotz, S. A. (2023). Attention and emotion shape self-
199 voice prioritization in speech processing. *Cortex; a Journal Devoted to the Study of the*
200 *Nervous System and Behavior*, 158, 83–95. <https://doi.org/10.1016/j.cortex.2022.10.006>
- 201 Platek, S. M., & Kemp, S. M. (2009). Is family special to the brain? An event-related fMRI study of
202 familiar, familial, and self-face recognition. *Neuropsychologia*, 47(3), 849–858.
203 <https://doi.org/10.1016/j.neuropsychologia.2008.12.027>
- 204 Roswandowicz, C., Kathiresan, T., Pellegrino, E., Dellwo, V., & Fröhholz, S. (2024). Cortical-striatal
205 brain network distinguishes deepfake from real speaker identity. *Communications Biology*,
206 7(1), 711. <https://doi.org/10.1038/s42003-024-06372-6>

- 207 Schweinberger, S. R., Casper, C., Hauthal, N., Kaufmann, J. M., Kawahara, H., Kloth, N.,
- 208 Robertson, D. M., Simpson, A. P., & Zäske, R. (2008). Auditory adaptation in voice perception.
- 209 *Curr Biol*, 18(9), 684–688. <https://doi.org/10.1016/j.cub.2008.04.015>
- 210 Sui, J., & Humphreys, G. W. (2017). The ubiquitous self: What the properties of self-bias tell us about
- 211 the self. *Ann N Y Acad Sci*, 1396(1), 222–235. <https://doi.org/10.1111/nyas.13197>
- 212 Tsuchiya, A., & Schweinberger, S. R. (2022). Erkennen von Verwandtschaft zu sich Selbst und bei
- 213 Anderen [Recognition of Kinship to Self and in Others]: Part 2, Issue 1/2022. *In-Mind*
- 214 *Magazine*, 1(2). [https://de.in-mind.org/article/erkennen-von-verwandtschaft-zu-sich-selbst-](https://de.in-mind.org/article/erkennen-von-verwandtschaft-zu-sich-selbst-und-bei-anderen)
- 215 [und-bei-anderen](#)
- 216 Yamagishi, J., Veaux, C., King, S., & Renals, S. (2012). Speech synthesis technologies for individuals
- 217 with vocal disabilities: Voice banking and reconstruction. *Acoustical Science and Technology*,
- 218 33(1), 1–5. <https://doi.org/10.1250/ast.33.1>
- 219 Young, A. W., Frühholz, S., & Schweinberger, S. R. (2020). Face and voice perception: Understanding
- 220 commonalities and differences. *Trends Cogn Sci*, 24(5), 398–410.
- 221 <https://doi.org/10.1016/j.tics.2020.02.001>
- 222