forum acusticum 2026
8-12 September · Graz, Austria

EFFECT OF HEADPHONES AND HOARSENESS-INDUCED AUDITORY FEEDBACK ON VOICE QUALITY: A MULTIPLE CASE STUDY

Isabel S. Schiller [1], Karolin Krüger [2], Gerhard Schmidt [2],
Antonia Chacon[3], Catherine Madill [3]

[1] *Work and Engineering Psychology, RWTH Aachen University, Germany*
[2] *Digital Signal Processing and System Theory, Kiel University, Germany*
[3] *The University of Sydney Voice Lab, The University of Sydney, Australia*

This paper is a revised version of the authors' submitted manuscript that was peer-reviewed by at least two anonymous reviewers.

Abstract

Real-time auditory feedback modulation (AFM) alters how speakers perceive their own voice during speech, enabling the study of vocal motor control. This multiple-case study specifically examines how auditory feedback with artificially increased hoarseness (i.e., hoarseness-induced feedback) affects acoustic voice quality in healthy speakers, relative to two control conditions: unaltered headphone feedback and no headphones. Eleven participants completed a total of three sustained-vowel production sessions – one per condition (AFM and controls). Acoustic voice quality was assessed using smoothed cepstral peak prominence (CPPS). Results show that hoarseness-induced AFM elicited a small but significant improvement in voice quality. In the control conditions, voice quality slightly declined with unaltered headphone feedback over the course of the task, while it improved without headphones. These patterns were closely linked to changes in voice intensity, with higher RMS values associated with higher CPPS. Overall, our findings indicate that hoarseness-induced AFM may enhance acoustic voice quality and suggest that including no-headphone controls may provide a more thorough picture of factors influencing vocal motor control.

Keywords: *auditory feedback, motor control, voice quality*

1. Introduction

When hearing our own voice altered in real time, we tend to adjust it automatically. For instance, speakers typically lower their pitch when it is artificially shifted upward via headphones [1]. In line with the Directions Into Velocities of Articulators (DIVA) model by Guenther et al. [2, 3], such compensatory responses occur when the brain detects a mismatch between expected and perceived auditory voice feedback and adjusts motor commands to the phonatory muscles. This mechanism suggests that auditory feedback modulation (AFM) could be used to shape voice production in a targeted manner, highlighting its potential relevance for voice therapy and vocal training. However, while it is well established that pitch shifts [4, 5, 6] and amplitude shifts [6, 7] elicit compensatory vocal responses in the opposite direction, little is known about how degradations in voice quality in auditory feedback affect voice production. Given that the primary symptom of many voice disorders is dysphonia (hoarseness), we argue that it is crucial to investigate whether hoarseness-related alterations in auditory feedback can also elicit vocal compensation.

To address this gap, we recently introduced *VQ-Synth*, a voice resynthesis system designed to degrade speakers' voice quality in their auditory feedback to examine the resulting phonatory responses [8, 9]. Results from an initial AFM experiment with healthy speakers, who repeatedly sustained /a:/ while receiving hoarseness-induced auditory feedback, showed that *VQ-Synth* elicited a significant improvement in acoustic voice quality [9]. This was reflected in a 0.5 dB increase in smoothed cepstral peak prominence (CPPS) – an acoustic measure of voice quality in which higher values indicate clearer, more harmonic phonation [10]. By comparison, a control group that received unaltered auditory feedback showed a gradual decline in CPPS, consistent with emerging vocal fatigue. This study was the first to indicate that hoarseness-induced auditory feedback may be applied to enhance acoustic voice quality. What

Corresponding author: isabel.schiller@psych.rwth-aachen.de.

remains unclear, however, is to what extent hearing one's own voice via headphones – even without perturbation – affects voice production compared to natural phonation. Additionally, little is known about individual response patterns to AFM and the factors that may predict them.

The present multiple-case study therefore aims to advance our understanding of AFM by (1) replicating the previously observed effect of hoarseness-induced auditory feedback on acoustic voice quality, (2) examining the effect of hearing one's own voice through headphones without perturbation, and (3) exploring individual variability and potential explanatory factors in responses to AFM.

2. Method

A real-time auditory feedback experiment was conducted in a soundproof booth at The University of Sydney Voice Lab, Australia. Data were collected between October and November 2025. In a within-subject design, participants completed the same phonation task across three sessions under different conditions: hoarseness-induced AFM, unaltered headphone feedback, and phonation without headphones. Each session lasted approximately 40 minutes. All participants provided their written informed consent.

2.1 Participants

Eleven participants (6 female), aged 26-55 years ($M = 40$) took part in this study. Table 1 presents their demographic characteristics, including gender, age, Acoustic Voice Quality Index (AVQI; [11]), hearing ability, instrumental and singing training, and performance in a pitch discrimination (PD) task [12]. Considering commonly reported AVQI thresholds of 2.72–3.33 for discriminating between healthy and dysphonic voices [13], most participants fell within the normal range, with IDs 9 and 11 near the threshold. Two participants demonstrated a mild hearing impairment (IDs 2 and 6). Two (IDs 1 and 7) had not received any instrumental or singing training; three (IDs 3, 5, and 6) were highly trained (instrument and singing), and the rest were moderately trained. PD accuracy ranged from 35% to 95%, with the lowest scores in IDs 7 and 9, and the highest in IDs 1 and 3.

2.2 Task

In each of the three sessions, participants performed the same task: sustaining the vowel /a:/ for 4 seconds across 140 trials (preceded by five practice trials) at a comfortable loudness and pitch. The sequence of auditory feedback conditions was balanced across sessions. For the AFM condition, we used *VQ-Synth* and adopted the same phase design as Schiller et al. [9]: 20 baseline trials without modulation, 50 ramp trials with stepwise increases in hoarseness, 50 hold trials at maximum modulation, and 20 after trials without modulation. In both control conditions (headphones and no headphones), auditory feedback remained unaltered. Participants' headphone output was calibrated to approximately match the perceived intensity of their naturally produced voice (see [9] for details). Fig. 1 shows a participant performing the task in one of the headphone conditions.

Figure 1. Participant performing the auditory feedback task. The headset microphone (DPA 4066) is positioned on the left side of the participant's face and is therefore not visible in this image.

Table 1. Participant demographics. *PD = pitch discrimination

ID	Gender	Age	AVQI	Hearing	Instrumental Training	Singing Training	PD* Accuracy
1	female	30	2.66	normal	none	none	95%
2	male	42	1.88	mildly impaired	moderate	moderate	70%
3	male	45	2.06	normal	high	high	90%
4	female	37	2.23	normal	moderate	high	80%
5	female	26	1.68	normal	high	high	65%
6	female	55	1.67	mildly impaired	high	high	85%
7	female	48	2.37	normal	none	none	40%
8	female	49	1.65	normal	moderate	none	50%
9	male	37	2.77	normal	high	none	35%
10	male	38	2.71	normal	high	moderate	45%
11	male	33	2.75	normal	moderate	moderate	70%

2.3 Procedure

Before the main task, participants completed an audiometric screening, a pitch discrimination task [12], and a questionnaire about prior experience with instrumental or singing training. They then received instructions for the phonation task and were fitted with a headset microphone (DPA 4066; 2.5 cm mouth–microphone distance) and, when applicable for the respective condition, closed headphones (Sennheiser HD 280 Pro). The task was performed while sitting in front of a laptop that guided the procedure. Voice input was recorded at 44.1 kHz with 24-bit resolution for later acoustic analysis. The overall system latency (software and hardware) for the auditory feedback was 8.7 ms.

2.4 CPPS calculation

CPPS values were extracted using Parselmouth (v0.4.7; [14]) interfacing with Praat (v6.1.38; [15]). A PowerCepstrogram was computed using a 60 Hz pitch floor, 2 ms time step, and 5000 Hz maximum frequency, and CPPS was calculated using Praat's standard settings. For trials with unaltered auditory feedback (baseline and after trials in the AFM condition, and all trials in the control conditions), CPPS was calculated from 2-s steady-state mid-vowel segments. In the AFM group's ramp and hold phases, CPPS was calculated from a 300-ms window spanning 100–400 ms after modulation onset.

2.5 Statistical analysis

Statistical analyses were conducted in R [16] using linear mixed-effects models (LMM; lme4, [17]), with CPPS as the dependent variable and random intercepts for participant (ID). In the AFM condition, CPPS was modeled as a function of Phase (Baseline, Ramp, Hold, After), controlling for RMS amplitude and fundamental frequency (f_0). In the controls, CPPS was modeled as a function of Condition, Trial, and their interaction, again including RMS amplitude and f_0 as covariates. Fixed effects were evaluated using Type II ANOVA with Satterthwaite's approximation (lmerTest, [18]), with partial η^2 reported. Post-hoc tests were Tukey-adjusted (emmeans, [19]). Individual variability was examined descriptively.

3. Results

The sections below present results for three aspects of AFM examined in this study: the effect of hoarseness-induced auditory feedback on acoustic voice quality (measured as CPPS), the impact of unaltered headphone feedback compared with natural phonation without headphones, and individual variability in responses to AFM.

3.1 Impact of AFM on CPPS

Focussing on the AFM condition, Fig. 2 shows the changes of CPPS as a function of Baseline, Ramp, Hold, and After phase. Results from the LMM showed that Phase had a small but significant effect on CPPS, $F(3,\ 1502.55) = 8.20$, $p < .001$, $\eta_p^2 = .02$. Tukey-adjusted pairwise comparisons indicated that CPPS was significantly higher during Ramp ($\Delta = 0.28$, $p = .001$, $d = 0.30$) and Hold ($\Delta = 0.31$, $p < .001$, $d = 0.34$) compared to Baseline. The After phase did not differ from Baseline ($p = .78$). However, CPPS was significantly lower during After than during Hold ($\Delta = -0.23$, $p = .014$, $d = 0.25$), whereas Ramp and Hold did not differ ($p = .95$). RMS amplitude positively predicted CPPS, $F(1,\ 1501.56) = 138.02$, $p < .001$, $\eta_p^2 = .08$ ($\beta = 0.19$), while f_0 showed a negative association, $F(1,\ 451.89) = 37.66$, $p < .001$, $\eta_p^2 = .08$ ($\beta = -0.03$). Random intercept variance indicated substantial between-participant variability ($SD = 1.93$).

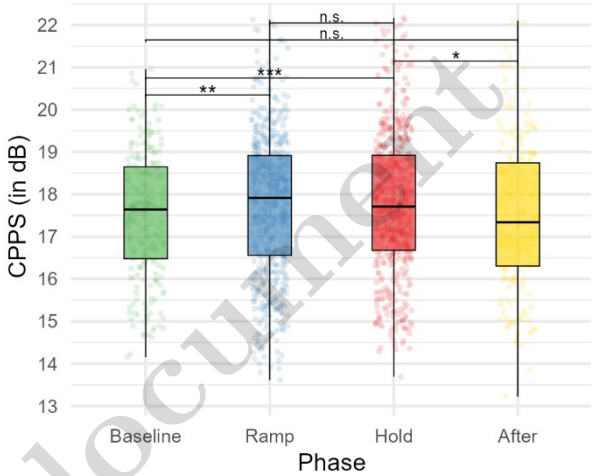

Figure 2. CPPS distributions across experimental phases in the AFM condition. Boxplots depict medians and interquartile ranges with individual trial points overlaid. Tukey-adjusted pairwise comparisons from the LMM are indicated (*$p < .05$, **$p < .01$, ***$p < .001$; n.s. = not significant).

3.2 Headphone-related voice changes

Results from the LMM showed that the Condition × Trial interaction had a small but significant effect on CPPS, $F(1,\ 3036.8) = 49.40$, $p < .001$, $\eta_p^2 = .02$, indicating different CPPS trajectories across conditions. In the no-headphone condition, CPPS significantly increased across trials ($\beta = 0.0048$, 95% CI [0.00368, 0.00597]), whereas a non-significant decrease was observed in the headphone condition ($\beta = -0.001$, 95% CI [−0.00213, 0.00017]). Voice RMS amplitude strongly and positively predicted CPPS, $F(1,\ 2992.1) = 447.29$, $p < .001$, $\eta_p^2 = .13$ ($\beta = 0.19$), while f_0 showed a small negative association, $F(1,\ 2015.3) = 31.94$, $p < .001$, $\eta_p^2 = .02$ ($\beta = -0.01$). Substantial between-participant variability was observed ($SD = 1.21$). Fig. 3 illustrates the close association between CPPS and RMS amplitude, with both increasing across trials without headphones and slightly decreasing with headphones.

3.3 Individual variability

Across participants, responses to hoarseness-induced AFM showed substantial individual variability. While the group-level analysis indicated increased CPPS during Ramp and Hold (see section 3.1), individual

deltas relative to Baseline strongly varied in both magnitude and direction. The strongest compensatory vocal responses were observed for IDs 8 (ΔRamp = +1.98 dB; ΔHold = +2.70 dB) and 2 (ΔRamp = +0.77 dB; ΔHold = +0.84 dB), with additional small-to-moderate increases for IDs 1 (+0.34 dB in Hold), 3 (+0.57 dB in Hold), and 11 (+1.02 dB in Hold). In contrast, IDs 6 and 10 showed clear decreases during both Ramp and Hold (ID6: −0.82 / −0.62 dB; ID10: −1.06 / −1.70 dB), indicating following rather than compensatory responses (i.e., decreased harmonicity). Other participants displayed mixed patterns. A descriptive inspection of participant characteristics (gender, age, AVQI value, hearing ability, instrumental and singing training, and pitch discrimination accuracy; Tab. 1) did not reveal a clear pattern explaining individual differences in CPPS responses to AFM.

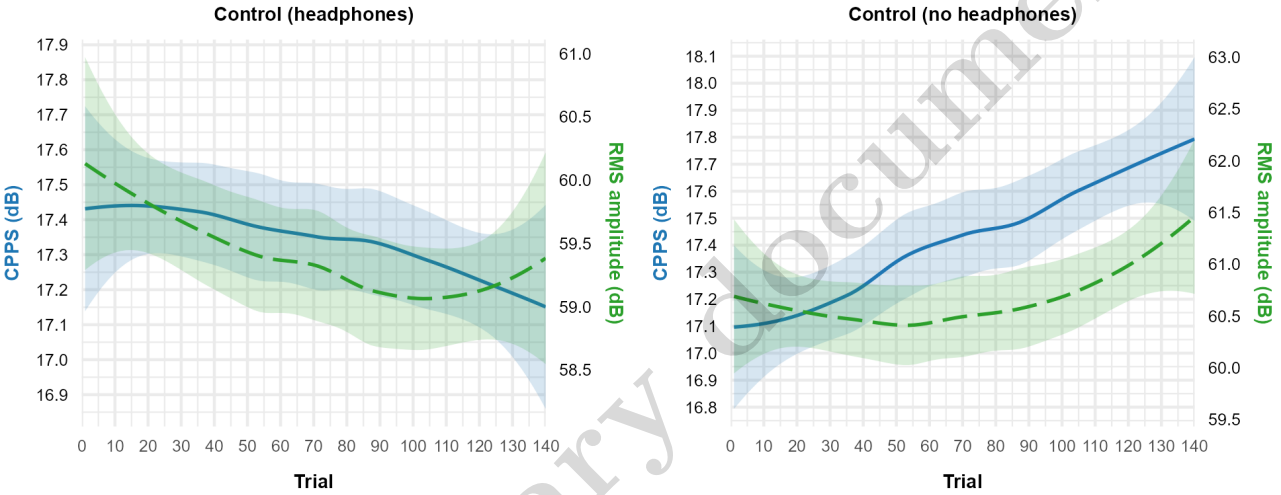

Figure 3. Smoothed trajectories of CPPS (blue, solid line) and RMS amplitude (green, dashed line) across trials in the two control conditions: unaltered auditory feedback with headphones (left panel) versus no headphones (right panel). Shaded areas represent loess-based 95% confidence intervals.

4. Discussion and conclusion

This study set out to replicate the previously reported effects of hoarseness-induced auditory feedback on acoustic voice quality, examine how hearing one's own unaltered voice via headphones influences CPPS trajectories over repeated phonation, and descriptively explore individual variability in responses to AFM.

Our findings on the impact of hoarseness-induced AFM replicate the general pattern reported by Schiller et al. [9], showing that hearing one's own degraded voice quality during phonation elicited compensatory increases in CPPS across participants, indicating improved acoustic voice quality. Within the DIVA model framework [2,3], these effects can be interpreted as auditory error-driven motor adjustments, whereby a mismatch between expected and perceived voice output results in corrective phonatory control. However, the 0.3 dB CPPS increase from Baseline to Hold observed here was lower than the 0.5 dB increase reported in [9]. This may be attributable to the more heterogeneous sample in the present study, including participants with slight hearing impairments, elevated AVQI values, and clinical and research experience in voice disorders, which may have influenced both sensitivity to feedback modulation and the magnitude of phonatory responses. Future work on the *VQ-Synth* system should examine whether modifications to the AFM method can elicit stronger CPPS improvements. Another difference to [9] was that CPPS did not remain elevated in the After phase but returned to baseline. This may reflect shorter-lived adaptation effects or faster recalibration of internal voice targets once feedback returned to normal, potentially related to musically or vocally trained participants in the sample. However, given the small sample size and the lack of information on training in [9], this interpretation remains tentative.

Regarding headphone-related effects, the results from the comparison of the two control groups importantly indicate that headphone use itself influences CPPS trajectories over time, associated with changes in voice amplitude. While CPPS as well as RMS amplitude showed a slight decline when participants heard their unaltered voice via headphones, both parameters increased across trials in the no-headphone condition. Thus, even without explicit manipulation, headphone-mediated feedback can alter participants' vocal motor output. A likely explanation for the different CPPS trajectories across the two control conditions concerns changes in the balance between air- and bone-conducted feedback.

In the no-headphone condition, participants monitored their voice through natural air- and bone-conducted feedback, which may have supported gradual optimisation of phonation across repetitions. Subtle increases in voice intensity to maintain clarity and stability could account for the rise in CPPS over time, consistent with the association between voice intensity and cepstral measures observed here and previously reported in the literature [20]. In contrast, the use of closed headphones may have altered the air–bone conduction balance, potentially shifting sensory

weighting during phonation. During natural phonation without headphones, we perceive our own voice through both air and bone conduction. Because bone conduction emphasises lower-frequency components of the voice signal, we typically perceive our voice as fuller or lower during natural phonation than when listening to a recording of it [21]. Closed headphones such as those used in this study can also produce an occlusion effect [22, 23], i.e., a perceived increase in low-frequency sound arising from bone-conducted vibrations within the occluded ear canal. This change in the relative contribution of air- and bone-conducted cues may have prompted participants to reduce vocal intensity when phonating with headphones, thereby affecting RMS amplitude and CPPS.

The substantial individual variability in CPPS responses to AFM indicates that compensatory responses to auditory feedback modulation are not uniform across speakers. While CPPS increased at the group level, individual changes ranged from relatively strong vocal compensation to clear decreases in acoustic voice quality, suggesting differences in sensorimotor control mechanisms. This variability may reflect inter-individual differences in reliance on auditory versus somatosensory feedback or baseline differences in voice quality. The absence of clear associations with demographic factors, musical training, or pitch discrimination ability suggests that compensatory behavior cannot be readily explained by these variables – at least in our small sample – underscoring the need to identify other factors underlying individual response patterns.

In conclusion, the findings from this multiple-case study confirm that hoarseness-induced auditory feedback can elicit compensatory increases in CPPS toward improved acoustic voice quality, and that headphone use alone influences voice quality and intensity over time. The notable inter-individual variability suggests that vocal responses to AFM differ markedly across speakers and should be examined in larger samples to better predict individual response patterns.

5. Funding information

This study was funded under the Excellence Strategy of the Federal Government and the Länder (StUpPD_445-23 and GSO091) and by philanthropic funding from The Dr Liang Voice Program at the University of Sydney.

6. Acknowledgments

The authors sincerely thank Sarah Breusch for her assistance in preparing the study materials and calculating the AVQI values. We also thank Duy Duong Nguyen for his valuable guidance analysing the pitch discrimination data. Finally, we are grateful to all participants for their time and effort in contributing to this study.

References

[1] M. D. C. Alves, P. C. Mancini, and L. C. Teixeira, "Modifications of auditory feedback and its effects on the voice of adult subjects: a scoping review," *CoDAS*, vol. 36, no. 1, p. e20220202, 2024, doi: 10.1590/2317-1782/20232022202en.

[2] F. H. Guenther, "Cortical interactions underlying the production of speech sounds," *J. Commun. Disord.*, vol. 39, no. 5, pp. 350–365, 2006, doi: 10.1016/j.jcomdis.2006.06.013.

[3] F. H. Guenther and G. Hickok, "Role of the auditory system in speech production," in *Handbook of Clinical Neurology*, vol. 129, G. G. Celesia and G. Hickok, Eds. Amsterdam, The Netherlands: Elsevier, 2015, pp. 161–175, doi: 10.1016/B978-0-444-62630-1.00009-3.

[4] L.-H. Ning, "Sensorimotor adaptation and aftereffect to frequency-altered feedback in Mandarin-speaking vocalists and non-vocalists," *Concentric Stud. Linguist.*, vol. 46, no. 2, pp. 125–147, 2020, doi: 10.1075/consl.00015.nin.

[5] N. E. Scheerer and J. A. Jones, "Detecting our own vocal errors: An event-related study of the thresholds for perceiving and compensating for vocal pitch errors," *Neuropsychologia*, vol. 114, pp. 158–167, 2018, doi: 10.1016/j.neuropsychologia.2017.12.007.

[6] C. R. Larson, J. Sun, and T. C. Hain, "Effects of simultaneous perturbations of voice pitch and loudness feedback on voice F and amplitude control," *J. Acoust. Soc. Am.*, vol. 121, no. 5, pp. 2862–2872, 2007, doi: 10.1121/1.2715657.

[7] J. J. Bauer, J. Mittal, C. R. Larson, and T. C. Hain, "Vocal responses to unanticipated perturbations in voice loudness feedback: An automatic mechanism for stabilizing voice amplitude," *J. Acoust. Soc. Am.*, vol. 119, no. 4, Art. no. 4, 2006, doi: 10.1121/1.2173513.

[8] I. S. Schiller, A. Schnapka, C. Eggert, P. Birkholz, and S. Stone, "VQ-Synth: Development and perceptual evaluation of a system for voice quality modification," in *Proc. 10th Convention Eur. Acoust. Assoc. Forum Acusticum 2023*, Turin, Italy, Eur. Acoust. Assoc., 2023, pp. 3533–3540, doi: 10.61782/fa.2023.0250.

[9] I. S. Schiller, K. Krüger, P. Weede, M. T. Sopha, and G. Schmidt, "Enhancing Acoustic Voice Quality Through Real-Time Auditory Feedback Modulation," *J. Voice*, p. S0892199726000226, 2026, doi: 10.1016/j.jvoice.2026.01.020.

[10] J. R. Cortés Ponce, L. Á. Garza Montelongo, J. E. Juárez Silva, and J. L. Trevino González, "Smoothed Cepstral Peak Prominence: A Comparison Between Dysphonic and Non-

dysphonic Mexican Adults Employing the Praat Software," *Cureus*, 2024, doi: 10.7759/cureus.72292.

[11] Y. Maryn, P. Corthals, P. Van Cauwenberge, N. Roy, and M. De Bodt, "Toward improved ecological validity in the acoustic measurement of overall voice quality: Combining continuous speech and sustained vowels," *J. Voice*, vol. 24, no. 5, 2010, doi: 10.1016/j.jvoice.2008.12.014.

[12] D. D. Nguyen, A. M. Chacon, D. Novakovic, N. J. Hodges, P. N. Carding, and C. Madill, "Pitch Discrimination Testing in Patients with a Voice Disorder," *J. Clin. Med.*, vol. 11, no. 3, p. 584, 2022, doi: 10.3390/jcm11030584.

[13] T. Jayakumar and J. J. Benoy, "Acoustic Voice Quality Index (AVQI) in the Measurement of Voice Quality: A Systematic Review and Meta-Analysis," *J. Voice*, p. S0892199722000844, 2022, doi: 10.1016/j.jvoice.2022.03.018.

[14] Y. Jadoul, B. Thompson, and B. De Boer, "Introducing Parselmouth: A Python interface to Praat," *J. Phon.*, vol. 71, pp. 1–15, 2018, doi: 10.1016/j.wocn.2018.07.001.

[15] Boersma, Paul and Weenink, David, *Praat: Doing phonetics by computer*. (2024). [Online]. Available: https://praat.org

[16] R Core Team, *R: A Language and Environment for Statistical Computing*. (2024). R Foundation for Statistical Computing, Vienna, Austria. [Online]. Available: https://www.R-project.org/

[17] D. Bates, M. Mächler, B. Bolker, and S. Walker, "Fitting Linear Mixed-Effects Models Using lme4," *J. Stat. Softw.*, vol. 67, no. 1, 2015, doi: 10.18637/jss.v067.i01.

[18] A. Kuznetsova, P. B. Brockhoff, and R. H. B. Christensen, "lmerTest Package: Tests in Linear Mixed Effects Models," *J. Stat. Softw.*, vol. 82, no. 13, 2017, doi: 10.18637/jss.v082.i13.

[19] R. V. Lenth, "Least-Squares Means: The *R* Package lsmeans," *J. Stat. Softw.*, vol. 69, no. 1, 2016, doi: 10.18637/jss.v069.i01.

[20] M. Brockmann-Bauser, J. H. Van Stan, M. Carvalho Sampaio, J. E. Bohlender, R. E. Hillman, and D. D. Mehta, "Effects of Vocal Intensity and Fundamental Frequency on Cepstral Peak Prominence in Patients with Voice Disorders and Vocally Healthy Controls," *J. Voice*, vol. 35, no. 3, pp. 411–417, 2021, doi: 10.1016/j.jvoice.2019.11.015.

[21] P. Orepic, O. A. Kannape, N. Faivre, and O. Blanke, "Bone conduction facilitates self-other voice discrimination," *R. Soc. Open Sci.*, vol. 10, no. 2, p. 221561, 2023, doi: 10.1098/rsos.221561.

[22] S. Stenfelt and S. Reinfeldt, "A model of the occlusion effect with bone-conducted stimulation," *Int. J. Audiol.*, vol. 46, no. 10, pp. 595–608, 2007, doi: 10.1080/14992020701545880.

[23] S. Reinfeldt, P. Östli, B. Håkansson, and S. Stenfelt, "Hearing one's own voice during phoneme vocalization—Transmission by air and bone conduction," *J. Acoust. Soc. Am.*, vol. 128, no. 2, pp. 751–762, 2010, doi: 10.1121/1.3458855.