Reducing Vocal Fatigue With Bone Conduction Devices

Nudelman, Charles; Udd, Daniela; Åhlander, Viveka Lyberg; Bottalico, Pasquale

Published in:
Journal of Speech, Language, and Hearing Research

DOI:
10.1044/2023_JSLHR-23-00409

Published: 01/11/2023

Document Version
Accepted author manuscript

Document License
Publisher rights policy

Link to publication

Please cite the original version:
Nudelman, C., Udd, D., Åhlander, V. L., & Bottalico, P. (2023). Reducing Vocal Fatigue With Bone Conduction Devices: Comparing Forbrain and Sidetone Amplification. *Journal of Speech, Language, and Hearing Research*, *66*(11), 4380-4397. https://doi.org/10.1044/2023_JSLHR-23-00409

Reducing Vocal Fatigue with Bone Conduction Devices: Comparing Forbrain® and Sidetone Amplification

Charles Nudelman[1*], Daniela Udd[2], Viveka Lyberg Åhlander[2], and Pasquale Bottalico[1]

[1]Department of Speech and Hearing Science
University of Illinois Urbana-Champaign
Champaign, Illinois, United States of America

[2]Department of Speech and Language Pathology Faculty of Arts, Psychology and Theology
Åbo Akademi University, Turku, Finland

[*]Author to whom correspondence should be addressed. Electronic mail: nudelma2@illinois.edu

ABSTRACT

Purpose: Altered auditory feedback research aims to identify methods to improve speakers' awareness of their own voicing behaviors, diminish their perception of vocal fatigue, and improve their voice production. The present study aims to compare the effects of two bone conduction devices that provide altered auditory feedback.

Methods: Twenty participants (19 − 33 years; mean (SD) 25.5 (3.85) years) participated in a vocal loading task using a standard Forbrain® device which provides filtered auditory feedback via bone conduction, a modified Forbrain® device which provides only sidetone amplification, and a control condition with no device was also included. They rated their vocal fatigue on a visual analog scale every two-minutes during the vocal loading task. Additionally, pre- and post-loading voice samples were analyzed for acoustic voice parameters.

Results: Across all participants, the use devices that increase the sidetone resulted in a lower vocal fatigue when compared to the condition with no feedback. During the pre- and post- voice samples, the sound pressure level decreased significantly during feedback conditions, and spectral measures significantly changed during feedback conditions.

Conclusions: The results promote bone conduction as a possible preventative tool that may reduce self-reported vocal fatigue and compensatory voice production for healthy individuals without voice disorders.

Keywords: vocal fatigue, bone conduction, vocal loading task, altered auditory feedback

INTRODUCTION

Across the lifespan, approximately 30% of the population will experience impairments in voice production, resulting in a voice disorder (Roy et al., 2004, 2005), and currently, approximately 7% - 17% of the population is experiencing a voice disorder (Behlau et al., 2012; Lyberg-Åhlander et al., 2019; OECD, 2014). One of the most commonly treated voice disorder is vocal hyperfunction (Bhattacharyya, 2014; Coyle et al., 2001; Herrington-Hall et al., 1988; Zhukhovitskaya et al., 2015), which can be defined as excessive perilaryngeal musculoskeletal activity during phonation (Oates & Winkworth, 2008). In certain occupational groups who use their voice extensively (e.g., teachers, physicians, salespeople, etc.), well over 50% of the workforce has obvious anatomical signs of vocal hyperfunction upon examination of their vocal folds (Tavares & Martins, 2007). Within this group of highly prevalent voice disorders, a possible underlying factor is sensorimotor integration. Sensorimotor integration can be defined as the integration of auditory, visual, and somatosensory information with motor actions – in this case, the motor actions of voice production (Machado et al., 2010).

Recent evidence indicates that individuals with specific types of voice disorders display distinct vocal responses in relation to their auditory perception (Castro et al., 2022; Hillman et al., 2020; Stepp et al., 2017; Weerathunge et al., 2022; Ziethe et al., 2019). This grouping of abnormal vocal responses (coined "auditory-motor phenotype"; (Abur et al., 2021; Weerathunge et al., 2022) is hypothesized to contribute to the pathophysiology of hyperfunctional voice disorders (Castro et al., 2022; Hillman et al., 2020; Stepp et al., 2017; Ziethe et al., 2019). This auditory-motor phenotype originiated from behavioral studies in which patients with voice disorders were presented with intensity-, formant-, or pitch-altered feedback of their voice in real-time during phonation tasks (e.g., Abur et al., 2018; X. Chen et al., 2013; Houde et al., 2019;

Naunheim et al., 2019). Through this methodology, patients' vocal responses were analyzed in an effort to understand the relationship between sensorimotor integration and their voice disorder. Among the symptoms of hyperfunctional voice disorders, vocal fatigue is commonly the first to appear (Mahalingam et al., 2021; Nanjundeswaran et al., 2019). Given that hyperfunctional voice disorders are most common among occupational voice users, and that vocal fatigue symptoms may be tied to differences in voice production (e.g., Fujiki et al., 2021, 2022; Kang et al., 2020; Milbrath & Solomon, 2003; Nanjundeswaran et al., 2017; Shembel & Nanjundeswaran, 2022), it would be informative to explore possible solutions to the auditory-motor phenotype that is influencing the development of voice disorders in this population.

Real-time altered auditory feedback (AAF) is one tool that has been studied in order to better understand voice disorders and contribute to biofeedback paradigms that could address the auditory-motor phenotype present in those with hyperfunctional voice disorders. Typically, AAF is presented in the daily lives of users through headphones, including traditional air conduction headphones (Pelegrín-García & Brunskog, 2012; Sierra-Polanco et al., 2021), and more recently, bone conduction headphones (Escera et al., 2018; Nudelman et al., 2022).

Bone conduction involves mechanical vibration to the bones of the skull, whereas air conduction involves transformation of the auditory sound wave into a mechanical signal within the middle ear (Henry & Letowski, 2007). Both the mechanical vibrations from bone conduction and air conduction are eventually converted into neural impulses within the cochlea (Lowy, 1942; v. Békésy, 1932). In terms of real-time AAF to be delivered in daily life, bone conduction may be a better option, especially for occupational voice users, as there is no occlusion of their ears. This would allow for speakers to hear their communication partner and the ambient sound environment, simultaneously, while receiving real time altered auditory feedback for their voices. Alternatively, air conducted AAF would fully cover the ears, making it difficult for occupational voice users to hear their communication partners.

There is one commercially available AAF device that uses bone conduction to improve speech. Forbrain®, developed by Sound For Life Limited (Soundev) in Luxemburg (model UN38.3; Europe; http://www.forbrain.com) uses a pair of bone conductors and a microphone to provide a speaker with real-time AAF. According to its website and patent registration, the Forbrain® device is ergonomically designed to fit all head sizes comfortably. The device is displayed in Figure 1.

[INSERT FIGURE 1 ABOUT HERE]

Research measuring the effects of the Forbrain® device reports significant improvements in Cepstral Peak Prominence Smoothed (CPPS) and spectral tilt for 32 adult healthy speakers using the device (Escera et al., 2018). According to its patent registration (Guajarengues & Lohmann, 2015), Forbrain® implements a two-band filter, which (from the perspective of the current authors), creates a slightly noticeable perception of altered voice quality (i.e., increased brightness of tone) for the user. The two-band filter applies one of two settings to the voice input, and these two settings are activated by the input sound energy at 1 kHz over a time window of integration ranging 10–200 ms. The resulting output is altered in its frequency spectrum by the two-band filter and is then delivered through bone conduction headphones to the temporal bones (Escera et al., 2018). Moreover, a simplified version of bone conduction AAF (without the two-band filtering) was demonstrated to significantly improve the voice quality of speakers with voice disorders (Nudelman et al., 2022). This simplified device uses sidetone amplification (amplified playback of one's own voice; Garnier et al., 2010; Laukkanen et al., 2004; Tomassi et al., 2021). Sidetone amplification elicits the Fletcher effect, which is a reflexive phenomenon of decreasing vocal loudness (SPL) by 1 dB (SPL) for each 2–3 dB (SPL) increase in auditory feedback (Fletcher, 1918; Lane & Tranel, 1971). To the authors' knowledge, the slim body of work on the Forbrain® device and the other study regarding bone conduction represent the current evidence that has examined AAF devices that use bone conduction. Thus, comparing the existing bone conduction devices may be useful, as these devices ostensibly offer more ecological validity (given that the user's ears remain uncovered/un-occluded compared to air conduction AAF), and may be useful in future clinical biofeedback paradigms, especially in occupational voice users. Moreover, examining AAF provides insights into the higher order mechanisms of planning and producing speech (Weerathunge et al., 2022). Research in this domain has provided critical information regarding how speakers detect and correct errors in their speech. Specifically, this area of research provides insights about the sensorimotor integration process. For example, experimental AAF studies examining the pitch-shift-reflex (a reflexive vocal response to changes in the pitch of voice auditory feedback; Hain et al., 2001) have successfully determined the efficiency and sensitivity of participants' auditory feedback control system, which is responsible for correcting errors in voice production (e.g., Behroozmand et al., 2012; Burnett et al., 1998; Kim & Larson, 2019; Larson & Robin, 2016; Liu & Larson, 2007; Scheerer & Jones, 2018).

Outside of the bone conduction studies, AAF research related to voice production has primarily been examined in daily life using self-reported symptoms as proxies for impaired voice production, especially in occupational voice users (Cutiva et al., 2013; Martins et al., 2014; Nair et al., 2021). However, alterations in voice production secondary to AAF can be more objectively quantified in laboratory settings through vocal loading tasks (VLTs), which aid in understanding how speakers without voice disorders respond vocally to challenging scenarios

(Kelchner et al., 2006; Solomon & DiMattia, 2000; Stemple et al., 1995). A systematic review examined VLTs within the literature and concluded that loud and prolonged reading is the most common task (Fujiki & Sivasankar, 2017). Recent VLTs have measured objective and subjective indicators of vocal fatigue using VLTs that involved between 10 - 60 minutes of loud reading (Echternach et al., 2020; Free et al., 2021; Lei et al., 2020; Xue et al., 2019) and, recently, 60 minutes of loud singing was employed in a VLT study (Devadas et al., 2023). Prior VLTs have induced vocal fatigue by incorporating external acoustic input (i.e., increased background noise; Bottalico et al., 2016; Cipriano et al., 2017; Herndon et al., 2019; Whitling et al., 2015), increasing duration of voice use, intensity of voice production, and adjusting speech patterns so that they are unnatural to the speakers (e.g., Buekers, 1998; Enflo et al., 2013; Fujiki et al., 2017; Kelchner et al., 2006; Nudelman et al., 2021; Remacle et al., 2012; Stemple et al., 1995; Vilkman et al., 1999; Yiu & Chan, 2003; Yiu et al., 2013).

AAF research aims to identify and validate methods to improve speakers' awareness of their own voicing behaviors, diminish their perception of vocal fatigue, and improve their acoustic voice parameters. The present study aims to compare the effects of AAF provided by two bone conduction devices to a control condition. The two bone conduction devices are: 1) sidetone amplification *via* a modified Forbrain® device and 2) Forbrain®'s filtered auditory feedback. Specifically, these devices will be compared through a VLT in terms of acoustic voice parameters derived from the long-term average spectrum produced by healthy participants and their subjective self-ratings of vocal fatigue. We hypothesize that the AAF devices will result in improvements in compensatory voice production (e.g., reduced sound pressure level (SPL), and decreased mean of the long-term average spectrum (LTAS)) and a lower slope of increase in vocal fatigue during the VLT, as compared to the control conditions.

MATERIALS AND METHODS

Twenty participants (19 − 33 years; mean (SD) 25.5 (3.85) years) were enrolled in the study and were recruited through sequential convenience sampling. Ten of the participants were male and ten were female. All the participants were self-described as conversationally proficient American English-speaking adults. Their ethnicities per self-report were "Caucasian" (n = 6), "Asian-Pacific Islander Native" (n = 2), "Hispanic-Latino" (n = 10), and "Black-African-American" (n = 2).

Inclusion criteria for the present study was being over the age of 18 years old, passing a hearing screen and reporting no history of voice, speech, language, or hearing disorders. The hearing screenings were pure tone audiometry tests performed by a certified speech-language pathologist. Hearing loss was considered present when bilateral thresholds were greater than 20 dB HL at octave frequencies from 500-4000 Hz. A voice disorder was considered present when a participant met at least as one of the following criteria: 1) VHI-10 score above 11 (Sund et al., 2021), 2) V-RQOL score above 91.25 (Behlau et al., 2016), or 3) VFI score greater or equal to 24.47 for factor 1, 6.90 for factor 2, or less than or equal to 7.71 for factor 3 (Nanjundeswaran et al., 2015). No participants were excluded from the study.

With protocol approval from the University of Illinois at Urbana-Champaign Institutional Review Board (IRB # 18179), speech samples of each participant were recorded during a VLT in three different AAF conditions. The recordings were performed in a soundproof double-walled Whisper Room (interior dimensions: 226 × 287 cm and h = 203 cm). Reverberation time (T30) was measured for mid-frequencies (500 – 2,000 Hz) to be 0.07 seconds in the Whisper Room and background noise was equal to 25 dB(A). The effects of the type and level of external auditory feedback on 1) the amount of self-reported vocal fatigue on a visual analog scale

(VAS), 2) sound pressure level values (SPL), 3) spectral mean of the long-term average spectrum (LTAS_mean), 4) standard deviation of the long-term average spectrum (LTAS_SD), and 5) skewness of the long-term average spectrum (LTAS_skew). were evaluated.

Vocal Loading Task (VLT) and Conditions

Prior to each AAF condition of the VLT, the participants completed two pre-loading speech tasks, (1) reading aloud the first six sentence of *The Rainbow Passage*, a standardized text in English (Fairbanks, 1960) and (2) sustaining an /a/ vowel for at least 5 seconds. These tasks were completed with the same AAF device (or lack thereof) that was used in the AAF condition and was repeated after each AAF condition as well to gather post-loading data. For example, prior to- and after- the AAF condition which used the standard Forbrain® headset, the participants read aloud *The Rainbow Passage* and sustained an /a/ vowel while wearing the same standard Forbrain® headset used during the main VLT, and the same occurred for the other two AAF conditions. These pre- and post-loading measures were used to provide baseline data for each AAF condition, as well as post-loading data which was intended to reveal the effects of vocal loading associated with each AAF condition. One component of the rationale for implementing these pre- and post-loading tasks was to gather voicing data that was unobscured by the fixed voicing level required during the VLT (see below). For this reason, it was necessary to retain the forthcoming (pre-loading) or preceding (post-loading) AAF condition (i.e., the participant kept wearing the AAF device during these pre- and post-loading tasks, if applicable), in order to assess the effects of the given AAF device (or lack thereof) on the voice signal.

For the VLT, the participants were instructed to read aloud five short stories by L. Frank Baum: *The Glass Dog* (Baum & Ilie, 2011), *The Queen of Quok* (Baum et al., 2005b), *The Magic Bon Bons* (Baum et al., 2005a), *The Capture of Father Time* (Baum, n.d.-a), and *The Wonderful*

Pump (Baum, n.d.-b). A reading task (as opposed to a non-speech VLT) was selected to elicit running speech in order to assess the influence of the AAF devices on a realistic voice signal. These short stories were presented in a random order, to stratify their linguistic content randomly across the AAF conditions. During the VLT, participants' voice level (intensity) was fixed at 73 dB(A) (i.e., a loud vocal effort level (ISO 9921, 2003)), which was achieved through real-time visual feedback from a sound level meter application on an iPad (described below). This visual feedback was incorporated to ensure that the participants were objectively achieving a loud vocal effort level, according to the international standard (ISO 9921, 2003). To this end, the participants were instructed to maintain a level of 73 dB(A) on the sound level meter for the duration of the reading task, consistent with previous studies (e.g., Kelchner et al., 2006). During the VLT, the participants were prompted to rate their vocal fatigue on a visual analog scale (VAS) every two minutes of reading, in a similar manner as a previously validated VLT paradigm (Nudelman et al., 2021). The VAS was 100 mm in length and included the instructions, "Please rate level of vocal fatigue from 0-100." The VAS limits were labelled, "0 – Not at all" (left) and "100 – Extremely" (right). Prior to the start of the VLT, vocal fatigue was defined for the participants as "your perception of a decline in your voice during the voice production task" (Hunter et al., 2020).

The duration of each AAF condition during the VLT was 20 minutes. The three different AAF conditions were (1) a control condition, (2) AAF sidetone amplification *via* a modified Forbrain® device, and (3) filtered AAF *via* a standard Forbrain® headset. The control condition involved performing the VLT without any AAF headset, that is, the participants were speaking with unaltered sidetone during the control condition. All conditions were presented without external noise added. Following each AAF conditions (i.e., every 20 minutes), the participants were offered an optional five-minute water break prior to starting the next condition. This process was repeated three times until the participant had completed the entire VLT (all three AAF conditions). The entire VLT was completed within a single session lasting approximately one hour. The order of administration of the AAF conditions were randomized to control for any unknown confounding variables relating to the task order.

Equipment

All speech material was recorded by an M2211 microphone (NTi Audio, Tigard, OR, United States) which was placed at a 0 degrees azimuth from the speaker at a fixed distance of 30 cm from the mouth (Švec & Granqvist, 2010). The direct digital recording sampled at 44100 Hz was recorded using an external soundboard (UH-7000 TASCAM, Teac Corporation, Montebello, CA, United States) connected to a personal computer (PC) running Audacity 174 3.1.3 (SourceForge, La Jolla, CA). While the participants read the short stories for the VLT on a Dell monitor, an iPad running Too Noisy software (ios), a sound level meter application, was used to display the visual feedback to maintain a loud voice level (73 dB(A)). The iPad was 1 meter from the participants and was calibrated on running speech in a process that involved an author (CN) reading the Rainbow Passage at a loud vocal effort level continuously, while the loud vocal effort level of 73 dB(A) was confirmed *via* a Lingwaves II SPL-meter (WEVOSYS234 hardware (IEC 651 Type 2, ANSI S1.4 Type 2)) by another author (DU). During this calibration process, the specifications of the Too Noisy app were manipulated until the sound level meter accurately responded to a 73 dB(A) voice signal. The specifications within the app for the 'sensitivity' setting was 95% and the 'dampening' setting within the app was 47%. The iPad was always visible during the VLT, thus providing uninterrupted real-time visual feedback which maintained a loud vocal effort level throughout the experiment.

External Auditory Feedback Equipment

Two AAF devices were compared, as well as a control condition. The first AAF device was a modified Forbrain® device, provided at no cost by the manufacturer. In this case, the manufacturer removed their patented filter from the device, and thus the Forbrain® provided only sidetone amplification. That is, the direct microphone output was played back to the participant at a level of 2.7 dB, as determined through the calibration procedure (see below).

The second device was a standard Forbrain® headset, developed by Sound For Life Limited (Soundev) in Luxemburg (model UN38.3; Europe; http://www.forbrain.com). This device uses a pair of bone conductors and a microphone to provide a speaker with external auditory feedback and was provided at no cost by the manufacturer. According to its patent registration and information outlined in previous research (Escera et al., 2018; Guajarengues & Lohmann, 2015), Forbrain® implements a two-band dynamic filter similar to a Baxandall equalizer (Baxandall, 1952). The two bands of the filter are triggered based on the voice energy at 1 kHz (mic input). One of the settings (Setting 1) raises low frequencies (100–800 Hz, +12 dB) while dampening high frequencies (800–15000 Hz, −12 dB) when the input signal energy at 1 kHz exceeds−56 dBV for a trigger time t1=10–50 ms. The other setting (Setting 2) performs the opposite (i.e., dampening low frequencies ranging 100–800 Hz and raises high frequencies ranging 800–15000 Hz) when the input signal at 1 kHz drops below−66 to−70 dBV for a holding time t2=20–200 ms. For the Forbrain® device, the direct microphone output was played back to the participant at a level of approximately 7 dB, as determined through the calibration procedure (see below).

Both AAF devices (sidetone amplification and filtered) were calibrated (post-hoc) through a procedure in which their vibration was captured while a Head and Torso Simulator

(HATS, GRAS 45BB KEMAR) was producing a white noise at a level of 70.6 dB (SPL) at 1 meter in a sound booth (corresponding to the experimental protocol). This white noise was recorded using an identical experimental setup, with each AAF device placed on the HATS, using the same settings, as well as a control condition with no AAF devices used. The white noise produced by the mouth of the HATS and recorded by the ears of the HATS was amplified through both the modified (sidetone amplification) and standard Forbrain® devices by a magnitude 2.7 dB and 7 dB, respectively. Of note, the Forbrain® implements an adaptive filter, which changes based on the energetic content within specified frequencies. In selecting white noise in the present calibration, the Forbrain® was characterized as if it were linear and time invariant, which is not reflective of its overall response to speech. Figure 2 displays the amplitude responses of the Forbrain®, sidetone amplification, and no AAF conditions as recorded during the calibration process.

[INSERT FIGURE 2 ABOUT HERE]

Analysis

All participant recordings (pre- and post-loading tasks and reading tasks associated with the VLT) were processed to calculate the 1) the amount of self-reported vocal fatigue on a visual analog scale (VAS), 2) sound pressure level values (SPL), 3) spectral mean of the long-term average spectrum (LTAS_mean), 4) standard deviation of the long-term average spectrum (LTAS_SD), and 5) skewness of the long-term average spectrum (LTAS_skew).

The recordings were processed with MATLAB R2022b (Mathworks, Natick, 284 MA, USA) and Praat 5.4/5.4.17 (Netherlands). Specifically, a custom MATLAB script was applied to calculate the SPL values on the voiced speech signal. This script included the MATLAB function, detectVoiced (Giannakopoulos, 2009) which extracts 1) the signal's energy and 2) its spectral centroid every 50 ms for the duration of the recorded signal. From these two features, dynamic thresholds are applied in order to detect voiced segments and remove unvoiced segments. To analyze all LTAS, the voiced segments were inputted into Praat 5.4/5.4.17 and a script was applied with the following settings: 512-point fast Fourier transform, 0.0025 seconds of spectrogram window length, Gaussian window shape, and 50 dB of dynamic range. From the LTAS, the spectral moments were calculated with the standard queries of Praat.

Statistical analyses were conducted using R version 4.2.0 (R Development Core Team, 2022) Linear Mixed-Effects (LME) models were fitted by restricted maximum likelihood (REML). Random effects terms were chosen based on variance explained. Tukey's post-hoc pairwise comparisons (Multiple Comparisons of Means: Tukey Contrasts) were performed to examine the differences between all levels of the fixed factors of interest. These are pairwise z tests, where the z statistic represents the difference between an observed statistic and its hypothesized population parameter in units of the standard deviation. The LME output includes the estimates of the fixed effects coefficients, the standard error associated with the estimate, the degrees of freedom (df), the test statistic (t), and the p-value. The Satterthwaite method was used to approximate degrees of freedom and calculate p-values. The models for each dependent variable are included in the Results section.

Variables: Definitions

Vocal fatigue VASs have been associated with alterations in acoustic voice parameters and psychosocial voice impairment in healthy occupational voice users (Castillo-Allendes et al., 2022).

SPL is a measure of the voice signal's physical magnitude (Baken & Orlikoff, 2000), and is related to the amplitude of the sound escaping from the upper airway. SPL is broadly referred to as the intensity of the voice signal.

LTAS_mean, SD, and skew are considered spectral moments, and are measures of the long-term average spectrum (LTAS) of the voice signal. The LTAS displays the power of frequencies within a voice signal and is calculated from a fast Fourier transform (mean of all spectra during a voice sample). These spectral moments have been demonstrated to capture compensatory voice production to some degree (e.g., Hammarberg et al., 1980; Harwardt, 2011; Mendoza et al., 1996).

LTAS_mean is the average value of the LTAS distribution, given that the presence of spectral energy above 5,000 Hz is a strong predictor of dysphonic voice quality (Hartmann & von Cramon, 1984), a lower average value reflects better voice quality. LTAS_SD describes the variance of the spectral distribution, with lower variance indicating better voice quality (Tanner et al., 2005). LTAS_skew describes positive or negative tilt of the LTAS. Positive skewness results in a "tail" of values extending to the right of the bell curve, which moves the overall shape of the spectrum towards lower frequencies. Negatively skewness results in the opposite (i.e., tail extending to left of the bell curve). There remains a lack of consensus surround LTAS_skew as an indicator of voice quality, with some studies indicating no significant influences on voice (e.g., Tanner et al., 2005) and others indicating significant associations with dysphonic voice quality (e.g., Hillenbrand & Houde, 1996).

RESULTS

Results: Vocal Fatigue

To assess the effects of the AAF devices on vocal fatigue, a linear mixed effect model was fitted with 'vocal fatigue' as response variable. The predictors used in the model are (1) time (a numerical variable including the time in minutes from 2 to 20 within each AAF condition) and (2) AAF condition (a factor with 3 levels: ForBrain, SideTone amplification, Control condition). The interaction between (1) and (2) was not statistically singificant. We included the participant id and the randomized order of AAF condition as random factor to remove their variance from the model.

Across all participants, the use of the Forbrain® and sidetone amplification AAF devices had a statistically significant effect on self-reported vocal fatigue ratings (elicited through a VAS). Specifically, across participants in the Forbrain® conditions, the vocal fatigue VAS ratings were approximately 9 points (i.e., 9 mm) lower ($p = .002$) when compared to the condition with no AAF. Across participants in the sidetone amplification condition, the vocal fatigue VAS ratings were approximately 15 points lower ($p < .001$) when compared to the condition with no AAF. Post-hoc comparisons confirmed that the increases in the vocal fatigue ratings during the Forbrain® conditions compared to non-AAF conditions (Estimate = -8.63, SE = 2.83, z = -3.05, $p = .006$) and comparing the sidetone amplification conditions to the non-AAF conditions (Estimate = -14.83, SE = 2.83, z = -5.23, $p < .001$) are statistically significant.

Additionally, over the course of the VLT, self-reported vocal fatigue increased by approximately 4 points on the vocal fatigue VAS each time a rating was made (i.e., every two minutes). Figure 3 and Table 1 display the results.

[INSERT FIGURE 3 ABOUT HERE]

[INSERT TABLE 1 ABOUT HERE]

Results: Pre- and Post-Loading

The following results represent the data analyzed from the pre- and post-loading tasks which occurred immediately prior to and immediately following each of the three AAF conditions. Table 2 summarizes multiple LME models fit by REML for each of the response variables (SPL, LTAS_mean, LTAS_SD, and LTAS_skew). The predictors used in all the four models are the 2 factors: (1) pre/post-loading (2 levels factor) and (2) AAF condition (3 levels: ForBrain, SideTone amplification, Control condition). The interaction between (1) and (2) was not significant. We included the participant id, gender, and task (reading and sustained vowel) as random factors to remove their variance from the model.

Figures 4-7 represent mean and the standard error for each response variable measured pre- and post-loading for vowel and connected speech tasks.

[INSERT TABLE 2 ABOUT HERE]

Sound Pressure Level. The use of the Forbrain® and sidetone amplification AAF devices during the pre- and post- loading tasks had a statistically significant effect on SPL, with SPL decreasing by approximately 1.2 dB (*p* = .048) during Forbrain® conditions and 1.5 dB (*p* = .015) during sidetone amplification conditions compared to non-AAF conditions. Comparing the pre- and post- loading tasks themselves, there was a significant increase in SPL by approximately 2.5 dB (*p* < .001) when speaking in the post-tasks. Post-hoc comparisons confirmed that the changes in SPL during the sidetone amplification conditions compared to non-AAF conditions (Estimate = -1.53, SE = 0.63, z = -2.44, *p* = .039) and comparing the pre- to the post-loading conditions (Estimate = 2.46, SE = 0.51, z = 4.83, *p* < .001) are statistically significant.

Spectral Mean of the Long-term Average Spectrum. The use of an AAF device had a statistically significant effect on LTAS_mean during pre- and post- loading tasks, with

LTAS_mean decreasing by approximately 92.52 Hz (*p* < .001) during Forbrain® conditions compared to non-AAF conditions and by approximately 86.58 Hz (*p* < .001) during sidetone amplification conditions compared to non-AAF conditions. Comparing the pre- and post- loading tasks, there was no detectable relationship between LTAS_mean and order. Post-hoc comparisons confirmed that the decreases in LTAS_mean comparing the Forbrain® conditions to the non-AAF conditions (Estimate = -92.52, SE = 23.25, z = -3.98, *p* < .001) and the sidetone amplification conditions to the non-AAF conditions (Estimate = -86.58. SE = 23.25, z = -3.72, *p* < .001), are statistically significant.

Standard Deviation of the Long-term Average Spectrum. The use of an AAF device had a statistically significant effect on LTAS_SD during pre- and post- loading tasks, with LTAS_ SD decreasing by approximately 66.66 Hz (*p* = .030) during Forbrain® conditions compared to non-AAF conditions and by approximately 70.98 Hz (*p* = .021) during sidetone amplification conditions compared to non-AAF conditions. Comparing the pre- and post- loading tasks, there was no detectable relationship with LTAS_ SD. Post-hoc comparisons did not confirm any LTAS_SD results with significance.

Skewness of the Long-term Average Spectrum. The use of an AAF device had a statistically significant effect on LTAS_skew during pre- and post- loading tasks, with LTAS_ skew increasing by approximately 0.83 Hz (*p* < .013) during Forbrain® conditions compared to non-AAF conditions and by approximately 0.66 Hz (*p* < .049) during sidetone amplification conditions compared to non-AAF conditions. Comparing the pre- and post- loading tasks, there was no detectable relationship between LTAS_ skew and order. Post-hoc comparisons confirmed that the increases in LTAS_ skew comparing the Forbrain® conditions to the non-AAF conditions (Estimate = 0.83, SE = 0.33, z = 2.50, *p* = .034) are statistically significant.

[INSERT FIGURE 4 ABOUT HERE]

[INSERT FIGURE 5 ABOUT HERE]

[INSERT FIGURE 6 ABOUT HERE]

[INSERT FIGURE 7 ABOUT HERE]

Results: During VLT

The following results represent the data analyzed from the reading tasks associated with the VLT. Within these analyses, the variable, "time" corresponds to the passage of time associated with the successive vocal fatigue ratings that the participants made every two-minutes during the VLT. Table 3 summarizes multiple LME models fit by REML for each of the response variables (LTAS_mean, LTAS_SD, and LTAS_skew) recorded during each condition of the VLT. The predictors used in all the four models are: (1) time (a numerical variable including the time in minutes from 2 to 20 within each AAF condition), and (2) AAF condition (3 levels: ForBrain, SideTone amplification, Control condition). The interaction between (1) and (2) was not significant. We included the participant id, gender, and the randomization order of the 3 AAF condition as random factors to remove their variance from the model.

[INSERT TABLE 3 ABOUT HERE]

Spectral Mean of the Long-term Average Spectrum. During each condition of the VLT, as time progressed, LTAS_mean decreased significantly by approximately 1.85 Hz (p = .001). During the VLT in Forbrain® conditions, LTAS_mean was significantly lower by approximately 53 Hz ($p < .001$) compared to non-AAF conditions and by approximately 15 Hz during VLT sidetone amplification conditions compared to non-AAF conditions (p = 0.031). Post-hoc comparisons confirmed that the decreases in LTAS_mean comparing the Forbrain® conditions to the non-AAF conditions (Estimate = -53.00, SE = 6.73, z = -7.74, $p < .001$) and the

Forbrain® conditions to the sidetone amplification conditions (Estimate = 37.57, SE = 6.64, z = 5.66, *p* < .001) are statistically significant. Post-hoc comparisons did not confirm that the decreases in LTAS_mean comapring sidetone amplification conditions to the non-AAF conditions with significance (Estimate = -14.53, SE = 6.73, z = -2.16, p = 0.079).

Standard Deviation of the Long-term Average Spectrum. During the VLT, as time progressed, LTAS_SD decreased significantly by approximately 1.5 Hz (p = 0.025). During the VLT in Forbrain® conditions, LTAS_sd was significantly lower by approximately -32 Hz (*p* < .001) compared to non-AAF conditions. Post-hoc comparisons confirmed that the decreases in LTAS_SD comparing the Forbrain® conditions to the non-AAF conditions (Estimate = -32.43, SE = 9.66, z = -3.36, *p* = .002) are statistically significant. There were no detectable relationships between the other LTAS_sd and the other conditions during the VLT.

Skewness of the Long-term Average Spectrum. During the VLT, there was no detectable relationship with the progression of time and LTAS_skew. During the VLT in Forbrain® conditions, LTAS_skew was significantly higher by approximately 0.43 Hz (*p* < .001) compared to non-AAF conditions and by approximately 0.17 Hz during VLT sidetone amplification conditions compared to non-AAF conditions (*p* = .015). Post-hoc comparisons confirmed that the decreases in LTAS_skew comparing the Forbrain® conditions to the non-AAF conditions (Estimate = 0.43, SE = 0.07, z = 6.21, *p* < .001), the sidetone amplification conditions to the non-AAF conditions (Estimate = 0.17, SE = 0.06, z = 2.43, *p* = .040), and the Forbrain® conditions to the sidetone amplification conditions (Estimate = 0.25, SE = 0.07, z = -3.82, *p* = .001) are statistically significant.

DISCUSSION

The primary aim of the present study was to compare the effects of AAF on voice production and self-reported vocal fatigue when the AAF was provided by two bone conduction devices: 1) sidetone amplification via a modified Forbrain® device and 2) Forbrain®'s filtered auditory feedback through a VLT. The results demonstrated that both the sidetone amplification and Forbrain®'s filtered AAF resulted in significantly decreased self- reported vocal fatigue during the VLT.

For the for the pre- and post-loading tasks, use of both devices resulted in significantly decreased spectral mean compared to the no AAF condition, the Forbrain® device resulted in significantly increased skewness of the LTAS, and the sidetone amplification device resulted in significantly decreased SPL. Of note, the increases in SPL with the Forbrain® and decreases in standard deviation of the LTAS with both types of AAF during the pre- and post- loading tasks were not confirmed as statistically significant.

For the voice recordings captured during the VLT tasks, both the sidetone amplification and Forbrain®'s filtered AAF resulted in significantly decreased spectral mean, and significantly increased skewness of the LTAS. Additionally, during the VLT, the standard deviation of the LTAS was significantly reduced in the Forbrain® conditions compared to the non-AAF conditions.

A secondary finding of the study was that the VLT paradigm that was employed resulted in significant increases in vocal fatigue over time and significant increases in SPL in the post-tasks compared to pre-loading tasks, which has been verified as an objective marker of vocal fatigue in prior VLT paradigms (e.g., Anand et al., 2021; Bottalico, 2017). Thus, the VLT can be considered valid in eliciting subjective and objective measures of vocal fatigue.

Vocal Fatigue

Compared to the no AAF condition, the use of AAF devices significantly reduced the amount of self-reported vocal fatigue during the VLT. Vocal fatigue ratings were 9 points lower on the VAS comparing the Forbrain® to the no AAF condition and were 15 points lower on the VAS during the sidetone amplification condition compared to the no AAF condition. These results have clinical significance and support the possibility that AAF devices that target sensorimotor integration could reduce voice symptoms and possibly serve as a preventative tool. In contrast, the current preventative recommendation mainly provided by voice clinicians is to use personal voice amplification systems (Bovo et al., 2013; Jónsdottir et al., 2001, 2002, 2003; McCormick & Roy, 2002; Roy et al., 2003). However, it has been empirically demonstrated that personal voice amplification devices actually worsen voice-related outcomes, ostensibly due to the relationship between amplification devices and increased room noise (a risk factor for voice disorders) (Banks et al., 2022; Nudelman et al., 2023). In addition to voice amplification, other devices exist to help speakers, such as sound field amplification (SFA), which aims to help speakers' voices reach listeners. In classrooms, SFA has been empirically demonstrated to aid in students' speech perception of their teacher (Trinite & Astolfi, 2021). One advantage that AAF has over SFA is the ability for teachers to move around their classroom without the restraint of electrical cables, which are commonly included within SFA devices.

Sound Pressure Level

Previous literature has demonstrated that AAF targeting increased bone conduction reduces vocal SPL in healthy speakers and those with voice disorders (Bauer et al., 2006; Heinks-Maldonado et al., 2006; Nudelman et al., 2022; Tomassi et al., 2021). Regarding SPL, the present results demonstrate that the sidetone amplification AAF resulted in significant decreases in voice intensity compared to the no AAF condition. These results reinforce bone conduction AAF as a tool that can successfully reduce the vocal SPL in running speech for users, ostensibly reducing their risk for sustaining a voice disorder (e.g., Chen et al., 2010). The magnitude of change in vocal SPL during the sidetone amplification condition was 1.53 dB (SPL). While this magnitude may seem trivial, previous research has correlated changes in vocal SPL ranging from 0.8 dB (SPL) – 2.0 dB (SPL) to the accumulation of vocal loading during a workday in occupational voice users (Jónsdottir et al., 2002; Laukkanen et al., 2008).

Spectral Measures: Mean, Standard Deviation, and Skewness

In the present study, measures of the LTAS had significant relationships with AAF devices, to varying degrees. That is, LTAS_mean was significantly lower during Forbrain® conditions, LTAS_SD was significantly lower during Forbrain® condition compared to the sidetone amplification conditions during the VLT, and LTAS_skew was significantly higher during the use of AAF, with the Forbrain® conditions having significantly higher LTAS_skew compared to the sidetone amplification conditions. Overall, LTAS_mean has been demonstrated to be the primary spectral voice measure that consistently accounts for the majority of variance in changes in compensatory voice production (Tanner et al., 2005). In the present study, the Forbrain® conditions resulted in significant decreases in LTAS_mean, indicating more energy near the fundamental frequency, as opposed to more energy in the higher frequency ranges. According to prior research, this shift in the spectral energy during the Forbrain® conditions implies that the participants were using less vocal effort compared to the non-AAF control condition (Harwardt, 2011). In the context of this evidence, vocal effort was strongly linked with increasing one’s vocal intensity in response to a communication demand. This is interesting, as the present experiment employed a task that required a loud vocal effort, as well. In light of these results from Harwardt, (2011), which associates decreased vocal effort with decreased

LTAS_mean, we can gain insight into the effects of bone conduction AAF. The present results affirm that in vocally demanding scenarios, the use of bone conduction AAF contributes significantly to decreases in the acoustic correlates of vocal effort, even at a loud vocal intensity. Additionally, previous studies have found significant decreases in spectral mean in patients' voices after they had completed successful voice therapy (Tanner et al., 2005). Of note, spectral measures also tend to have a positive relationship with SPL and fundamental frequency. That is, increases in SPL have been demonstrated to lead to increases in fundamental frequency (Gramming et al., 1988), and ostensibly contributes increased energy in the higher frequencies in the voice signal (i.e., increased LTAS_mean). In the present study, the opposite occurred during AAF conditions, as participants demonstrated reduced SPL and reduced LTAS_mean, particularly during the pre- and post-loading tasks. These results imply that AAF provided via bone conduction may also benefit a listener, as the improved voice quality is associated with increased signal-to-noise ratio (i.e., there is less high frequency noise in the voice signal based on the reduced LTAS_mean; Evitts et al., 2016; Lallh & Rochet, 2000; Lyberg-Åhlander et al., 2015).

Applying Bone Conduction AAF in Practice

The AAF devices used in the present study are publicly available and comparable in price to typical air conduction headphones used in voice science. Based on the present results, and previous studies which examined the effects of bone conduction AAF on voice production (e.g., Lee et al., 2019; Nudelman, Codino, et al., 2022), these devices seem to be promising tools in the prevention and treatment of voice disorders, as they appear to augment auditory-motor integration during speech in a positive way for both healthy speakers and those with voice disorders. An important consideration for future research and clinical applications with these devices has to do with the microphone. At present, the direct voice signal cannot be recorded from the Forbrain® device's attached microphone. With this in mind, the use of a contact microphone (Bottalico & Nudelman, 2023) or an accelerometer (Mehta et al., 2012) placed on the neck may be necessary to capture voice recordings in daily life. With such data, researchers and clinicians could ostensibly assess the effectiveness of the use of bone conduction AAF in daily communication scenarios.

Limitations

There are a few limitations of the study that should be acknowledged. The first limitation involves generalizability. Only vocally healthy participants were included, and thus it is unclear how the results may differ in individuals with voice disorders, including categories of voice disorders that have been linked to impaired sensorimotor integration (e.g., hyperfunctional voice disorders; Castro et al., 2022; Hillman et al., 2020; Stepp et al., 2017; Ziethe et al., 2019). Previous research has found that bone conduction sidetone amplification results in consistent adaptation in the SPL values and mean pitch strength in patients with the vocal hyperfunction, glottal insufficiency and organic/neurological laryngeal pathologies compared to conditions with no feedback (Nudelman et al., 2022). However, these results in a clinical population remain to be verified during a VLT and during daily communication scenarios.

Another limitation has to do with the ecological validity of the VLT used in the present study. Given that reading at a fixed intensity level likely does not simulate everyday vocal loading, the present results should be interpreted with caution and should not be generalized to imply that AAF will result in reduced vocal fatigue or reduced compensatory voice production during daily communication scenarios. Recent studies have examined vocal loading with the goal of utilizing more ecologically valid approaches (Nusseck et al., 2022; Sandage et al., 2022;

Trinite et al., 2022). These studies each implemented ecologically valid vocal loading in different ways, such as reading tasks (Nusseck et al., 2022), providing fixed background noise levels in real-life rooms (Trinite et al., 2022), and pre-/post-measures after unstandardized daily vocal loading tasks in daily life (Sandage et al., 2022). It would be useful to examine the efficacy of AAF devices in ecologically valid VLTs such as these, and potentially incorporate multivariate objective measures (e.g., the Daily Phonotrauma Index; (C. J. Nudelman, Ortiz, et al., 2022; Van Stan et al., 2021, 2023) in association with self-reported vocal status ratings.

Along similar lines, a final limitation is associated with the possibility that the VLT employed in the present study did not effectively fatigue the voice mechanism. Previous research has verified that individuals' self-rated vocal status is prone to inaccuracy and error (e.g., Mehta et al., 2016). Additionally, there is inconsistency in outcome data from VLTs employing loud reading tasks for an hour or less in duration. Prior VLTs with short, loud reading tasks failed to elicit objective measures of vocal fatigue (e.g., Buekers, 1998; Whitling et al., 2015), while more recent studies (which elicited loud reading ranging from 10 – 60 minutes) successfully achieved objective measures of vocal fatigue (Echternach et al., 2020; Free et al., 2021; Lei et al., 2020; Xue et al., 2019). These inconsistent results have prompted recommendations that a short (less than one hour) VLT using a loud reading task should involve at least one additional factor to produce measurable change in voice (Fujiki & Sivasankar, 2017). Such additional factors could involve altering vocal quality, eliciting nonhabitual speech, or implementing environmental perturbations. Additionally, it has been recommended that VLTs are assessed in a multi-system manner. That is, the suggested outcome measures from VLTs should capture the physiologic fatigue-response of the respiratory, laryngeal, and supralaryngeal subsystems (e.g., Sundarrajan et al., 2017). The present VLT failed to follow these two guidelines. Thus, to better extend the present results, future VLT experiments utilizing AAF would benefit from employing a similar reading task that is longer in duration and also involves additional factors in the elicitation of multi-system vocal fatigue.

CONCLUSION

This study provides evidence that both AAF sidetone amplification *via* a modified Forbrain® device and filtered AAF *via* a standard Forbrain® headset devices contribute to significantly reduced self-reported vocal fatigue, significantly decreased LTAS mean and standard deviation, increased skewness of the LTAS, and decreased SPL during a vocal loading task. During the vocal loading task, the Forbrain® device resulted in improved voice quality in regard to the standard deviation and skewness of the LTAS compared to the sidetone amplification condition. These results promote bone conduction AAF as a possible preventative tool that may reduce self-reported vocal fatigue and compensatory voice production for healthy individuals without voice disorders.

ACKNOWLEDGEMENTS

The authors would like to thank the participants for their valuable cooperation and interest.

DECLARATION OF INTEREST STATEMENT

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 STATEMENT

All data obtained and/or analyzed are available from the authors upon reasonable request.

REFERENCES

Abur, D., Lester-Smith, R. A., Daliri, A., Lupiani, A. A., Guenther, F. H., & Stepp, C. E. (2018). Sensorimotor adaptation of voice fundamental frequency in Parkinson's disease. *PLOS ONE*, *13*(1), e0191839. https://doi.org/10.1371/journal.pone.0191839

Abur, D., Subaciute, A., Kapsner-Smith, M., Segina, R. K., Tracy, L. F., Noordzij, J. P., & Stepp, C. E. (2021). Impaired auditory discrimination and auditory-motor integration in hyperfunctional voice disorders. *Scientific Reports*, *11*(1), 1–11. https://doi.org/10.1038/s41598-021-92250-8

Anand, S., Bottalico, P., & Gray, C. (2021). Vocal Fatigue in Prospective Vocal Professionals. *Journal of Voice*, *35*(2), 247–258. https://doi.org/10.1016/j.jvoice.2019.08.015

Baken, R. J., & Orlikoff, R. F. (2000). *Clinical measurement of speech and voice*. Cengage Learning.

Banks, R. E., Cantor-Cutiva, L. C., & Hunter, E. (2022). Factors Influencing Teachers' Experience of Vocal Fatigue and Classroom Voice Amplification. *Journal of Voice*, S0892199722001928. https://doi.org/10.1016/j.jvoice.2022.06.026

Bauer, J. J., Mittal, J., Larson, C. R., & Hain, T. C. (2006). Vocal responses to unanticipated perturbations in voice loudness feedback: An automatic mechanism for stabilizing voice amplitude. *The Journal of the Acoustical Society of America*, *119*(4), 2363–2371.

Baum, L. F. (n.d.-a). *The Capture Of Father Time*.

Baum, L. F. (n.d.-b). *The Wonderful Pump*.

Baum, L. F., Fernandez, F., & Batucan, L. (2005a). *The Magic Bonbons*. Wualou Limited.

Baum, L. F., Fernandez, F., & Batucan, L. (2005b). *The Queen of Quok*. Wualou Limited.

Baum, L. F., & Ilie, E. (2011). *The glass dog*. Paralela 45.

Baxandall, P. J. (1952). Negative-feedback tone control. *Wireless World*, *58*(10), 402–405.

Behlau, M., Madazio, G., Moreti, F., Oliveira, G., Dos Santos, L. de M. A., Paulinelli, B. R., & Junior, E. de B. C. (2016). Efficiency and cutoff values of self-assessment instruments on the impact of a voice problem. *Journal of Voice*, *30*(4), 506-e9.

Behlau, M., Zambon, F., Guerrieri, A. C., & Roy, N. (2012). Epidemiology of Voice Disorders in Teachers and Nonteachers in Brazil: Prevalence and Adverse Effects. *Journal of Voice*, *26*(5), 665.e9-665.e18. https://doi.org/10.1016/j.jvoice.2011.09.010

Behroozmand, R., Korzyukov, O., Sattler, L., & Larson, C. R. (2012). Opposing and following vocal responses to pitch-shifted auditory feedback: Evidence for different mechanisms of voice pitch control. *The Journal of the Acoustical Society of America*, *132*(4), 2468–2477.

Bhattacharyya, N. (2014). The prevalence of voice problems among adults in the United States. *Laryngoscope*, *124*(10), 2359–2362. https://doi.org/10.1002/lary.24740

Bottalico, P. (2017). Speech Adjustments for Room Acoustics and Their Effects on Vocal Effort. *Journal of Voice: Official Journal of the Voice Foundation*, *31*(3), 392.e1-392.e12. https://doi.org/10.1016/j.jvoice.2016.10.001

Bottalico, P., Graetzer, S., & Hunter, E. J. (2016). Effects of speech style, room acoustics, and vocal fatigue on vocal effort. *Journal of the Acoustical Society of America*, *139*(5), 2870–2879. Scopus. https://doi.org/10.1121/1.4950812

Bottalico, P., & Nudelman, C. J. (2023). Do-It-Yourself Voice Dosimeter Device: A Tutorial and Performance Results. *Journal of Speech, Language, and Hearing Research*, 1–15.

Bovo, R., Trevisi, P., Emanuelli, E., & Martini, A. (2013). Voice amplification for primary school teachers with voice disorders: A randomized clinical trial. *International Journal of*

Occupational Medicine and Environmental Health, *26*(3), 363–372. https://doi.org/10.2478/s13382-013-0115-1

Buekers, R. (1998). Are voice endurance tests able to assess vocal fatigue? *Clinical Otolaryngology & Allied Sciences*, *23*(6), 533–538.

Burnett, T. A., Freedland, M. B., Larson, C. R., & Hain, T. C. (1998). Voice F0 responses to manipulations in pitch feedback. *The Journal of the Acoustical Society of America*, *103*(6), 3153–3161.

Castillo-Allendes, A., Guzmán-Ferrada, D., Hunter, E. J., & Fuentes-López, E. (2022). Tracking Occupational Voice State with a Visual Analog Scale: Voice Quality, Vocal Fatigue, and Effort. *The Laryngoscope*.

Castro, C., Prado, P., Espinoza, V. M., Testart, A., Marfull, D., Manriquez, R., Stepp, C. E., Mehta, D. D., Hillman, R. E., & Zañartu, M. (2022). Lombard Effect in Individuals With Nonphonotraumatic Vocal Hyperfunction: Impact on Acoustic, Aerodynamic, and Vocal Fold Vibratory Parameters. *Journal of Speech, Language, and Hearing Research*, *65*(8), 2881–2895. https://doi.org/10.1044/2022_JSLHR-21-00508

Chen, S. H., Chiang, S.-C., Chung, Y.-M., Hsiao, L.-C., & Hsiao, T.-Y. (2010). Risk factors and effects of voice problems for teachers. *Journal of Voice*, *24*(2), 183–192. https://doi.org/10.1016/j.jvoice.2008.07.008

Chen, X., Zhu, X., Wang, E. Q., Chen, L., Li, W., Chen, Z., & Liu, H. (2013). Sensorimotor control of vocal pitch production in Parkinson’s disease. *Brain Research*, *1527*, 99–107. https://doi.org/10.1016/j.brainres.2013.06.030

Cipriano, M., Astolfi, A., & Pelegrín-García, D. (2017). Combined effect of noise and room acoustics on vocal effort in simulated classrooms. *The Journal of the Acoustical Society of America*, *141*(1), EL51. https://doi.org/10.1121/1.4973849

Coyle, S. M., Weinrich, B. D., & Stemple, J. C. (2001). Shifts in relative prevalence of laryngeal pathology in a treatment-seeking population. *Journal of Voice*, *15*(3), 424–440. https://doi.org/10.1016/S0892-1997(01)00043-1

Cutiva, L. C. C., Vogel, I., & Burdorf, A. (2013). Voice disorders in teachers and their associations with work-related factors: A systematic review. *Journal of Communication Disorders*, *46*(2), 143–155.

Devadas, U., Vinod, D., & Maruthy, S. (2023). Immediate effects of straw phonation in water exercises on parameters of vocal loading in carnatic classical singers. *Journal of Voice*, *37*(1), 142-e13.

Echternach, M., Huseynov, J., Döllinger, M., Nusseck, M., & Richter, B. (2020). The impact of a standardized vocal loading test on vocal fold oscillations. *European Archives of Oto-Rhino-Laryngology*, *277*, 1699–1705.

Enflo, L., Sundberg, J., Romedahl, C., & McAllister, A. (2013). Effects on vocal fold collision and phonation threshold pressure of resonance tube phonation with tube end in water. *Journal of Speech, Language, and Hearing Research*, *56*(5), 1530–1538. https://doi.org/10.1044/1092-4388(2013/12-0040)

Escera, C., López-Caballero, F., & Gorina-Careta, N. (2018). The potential effect of forbrain as an altered auditory feedback device. *Journal of Speech, Language, and Hearing Research*, *61*(4), 801–810. https://doi.org/10.1044/2017_JSLHR-S-17-0072

Evitts, P. M., Starmer, H., Teets, K., Montgomery, C., Calhoun, L., Schulze, A., MacKenzie, J., & Adams, L. (2016). The impact of dysphonic voices on healthy listeners: Listener reaction times, speech intelligibility, and listener comprehension. *American Journal of Speech-Language Pathology*, *25*(4), 561–575.

Fairbanks, G. (1960). The rainbow passage. *Voice and Articulation Drillbook*, *2*, 127–127.

Fletcher, H. (1918). Study of the effects of different sidetones in the telephone set. *(No Title)*.

Free, N., Stemple, J. C., Smith, J. A., & Phyland, D. J. (2021). The impact of a vocal loading task on voice characteristics of female speakers with benign vocal fold lesions. *Journal of Voice*.

Fujiki, R. B., Chapleau, A., Sundarrajan, A., McKenna, V., & Sivasankar, M. P. (2017). The Interaction of Surface Hydration and Vocal Loading on Voice Measures. *Journal of Voice: Official Journal of the Voice Foundation*, *31*(2), 211–217. https://doi.org/10.1016/j.jvoice.2016.07.005

Fujiki, R. B., Huber, J. E., & Sivasankar, M. P. (2021). Mitigating the effects of acute vocal exertion in individuals with vocal fatigue. *The Laryngoscope*, *131*(12), 2732–2739.

Fujiki, R. B., Huber, J. E., & Sivasankar, M. P. (2022). The effects of vocal exertion on lung volume measurements and acoustics in speakers reporting high and low vocal fatigue. *Plos One*, *17*(5), e0268324.

Fujiki, R. B., & Sivasankar, M. P. (2017). A Review of Vocal Loading Tasks in the Voice Literature. *Journal of Voice : Official Journal of the Voice Foundation*, *31*(3), 388.e33-388.e39. https://doi.org/10.1016/j.jvoice.2016.09.019

Garnier, M., Henrich, N., & Dubois, D. (2010). Influence of sound immersion and communicative interaction on the lombard effect. *Journal of Speech, Language, and Hearing Research*, *53*(3), 588–608. https://doi.org/10.1044/1092-4388(2009/08-0138)

Giannakopoulos, T. (2009). A method for silence removal and segmentation of speech signals, implemented in Matlab. Department of Informatics and Telecommunications, University of Athens, Greece. *Computational Intelligence Laboratory (CIL), Insititute of Informatics and Telecommunications (IIT), NCSR DEMOKRITOS, Greece*.

Gramming, P., Sundberg, J., Ternström, S., Leanderson, R., & Perkins, W. H. (1988). Relationship between changes in voice pitch and loudness. *Journal of Voice*, *2*(2), 118–126.

Guajarengues, T., & Lohmann, K. (2015). *Apparatus and method for active voice training* (Patent Cooperation Treaty Patent WO2015067741A1). http://www.google.ch/patents/WO2015067741A1?cl=en

Hain, T. C., Burnett, T. A., Larson, C. R., & Kiran, S. (2001). Effects of delayed auditory feedback (DAF) on the pitch-shift reflex. *The Journal of the Acoustical Society of America*, *109*(5), 2146–2152.

Hammarberg, B., Fritzell, B., Gaufin, J., Sundberg, J., & Wedin, L. (1980). Perceptual and acoustic correlates of abnormal voice qualities. *Acta Oto-Laryngologica*, *90*(1–6), 441–451.

Hartmann, E., & von Cramon, D. (1984). Acoustic measurement of voice quality in central dysphonia. *Journal of Communication Disorders*, *17*, 425–440.

Harwardt, C. (2011). *Comparing the impact of raised vocal effort on various spectral parameters*. Twelfth Annual Conference of the International Speech Communication Association.

Heinks-Maldonado, T. H., Nagarajan, S. S., & Houde, J. F. (2006). Magnetoencephalographic evidence for a precise forward model in speech production. *Neuroreport*, *17*(13), 1375–1379. https://doi.org/10.1097/01.wnr.0000233102.43526.e9

Henry, P., & Letowski, T. (2007). Bone conduction: Anatomy , physiology , and communication. *Army Research Lab*, *ARL-TR-4138*.

Herndon, N. E., Sundarrajan, A., Sivasankar, M. P., & Huber, J. E. (2019). Respiratory and laryngeal function in teachers: Pre-and postvocal loading challenge. *Journal of Voice*, *33*(3), 302–309.

Herrington-Hall, B. L., Lee, L., Stemple, J. C., Niemi, K. R., & McHone, M. M. (1988). Description of laryngeal pathologies by age, sex, and occupation in a treatment-seeking sample. *Journal of Speech & Hearing Disorders*, *53*(1), 57–64.

Hillenbrand, J., & Houde, R. A. (1996). Acoustic correlates of breathy vocal quality: Dysphonic voices and continuous speech. *Journal of Speech and Hearing Research*, *39*(2), 311–321.

Hillman, R. E., Stepp, C., Van Stan, J. H., Zanartu, M., & Mehta, D. D. (2020). An Updated Theoretical Framework for Vocal Hyperfunction. *American Journal of Speech Language Pathology*, *29*(4), 2254–2260.

Houde, J. F., Gill, J. S., Agnew, Z., Kothare, H., Hickok, G., Parrell, B., Ivry, R. B., & Nagarajan, S. S. (2019). Abnormally increased vocal responses to pitch feedback perturbations in patients with cerebellar degeneration. *The Journal of the Acoustical Society of America*, *145*(5), EL372–EL378. https://doi.org/10.1121/1.5100910

Hunter, E. J., Cantor-Cutiva, L. C., Van Leer, E., Van Mersbergen, M., Nanjundeswaran, C. D., Bottalico, P., Sandage, M. J., & Whitling, S. (2020). Toward a Consensus Description of Vocal Effort, Vocal Load, Vocal Loading, and Vocal Fatigue. *Journal of Speech, Language, and Hearing Research*, *63*(2), 509–532.

ISO 9921. (2003). *ISO 9921-1 Ergonomic assessment of speech communication Part 1*.

Jónsdottir, V., Laukkanen, A. M., Ilomäki, I., Roininen, H., Alastalo-Borenius, M., & Vilkman, E. (2001). Effects of amplified and damped auditory feedback on vocal characteristics. *Logopedics, Phoniatrics, Vocology*, *26*(2), 76–81. https://doi.org/10.1080/140154301753207449

Jónsdottir, V., Laukkanen, A.-M., & Siikki, I. (2003). Changes in teachers' voice quality during a working day with and without electric sound amplification. *Folia Phoniatrica et Logopaedica: Official Organ of the International Association of Logopedics and Phoniatrics (IALP)*, *55*(5), 267–280. https://doi.org/10.1159/000072157

Jónsdottir, V., Laukkanen, A.-M., & Vilkman, E. (2002). Changes in teachers' speech during a working day with and without electric sound amplification. *Folia Phoniatrica et Logopaedica*, *54*(6), 282–287.

Kang, J., Xue, C., Lou, Z., Scholp, A., Zhang, Y., & Jiang, J. J. (2020). The therapeutic effects of straw phonation on vocal fatigue. *The Laryngoscope*, *130*(11), E674–E679.

Kelchner, L. N., Toner, M. M., & Lee, L. (2006). Effects of prolonged loud reading on normal adolescent male voices. *Language, Speech, and Hearing Services in Schools*, *37*(2), 96–103. https://doi.org/10.1044/0161-1461(2006/012)

Kim, J. H., & Larson, C. R. (2019). Modulation of auditory-vocal feedback control due to planned changes in voice fo. *The Journal of the Acoustical Society of America*, *145*(3), 1482–1492.

Lallh, A. K., & Rochet, A. P. (2000). The effect of information on listeners' attitudes toward speakers with voice or resonance disorders. *Journal of Speech, Language, and Hearing Research*, *43*(3), 782–795.

Lane, H., & Tranel, B. (1971). The Lombard sign and the role of hearing in speech. *Journal of Speech and Hearing Research*, *14*(4), 677–709.

Larson, C. R., & Robin, D. A. (2016). Sensory processing: Advances in understanding structure and function of pitch-shifted auditory feedback in voice control. *AIMS Neuroscience*, *3*(1), 22–39.

Laukkanen, A.-M., Ilomäki, I., Leppänen, K., & Vilkman, E. (2008). Acoustic measures and self-reports of vocal fatigue by female teachers. *Journal of Voice*, *22*(3), 283–289.

Laukkanen, A.-M., Mickelson, N. P., Laitala, M., Syrjä, T., Salo, A., & Sihvo, M. (2004). Effects of HearFones on speaking and singing voice quality. *Journal of Voice*, *18*(4), 475–487.

Lee, S. H., Yu, J. F., Fang, T. J., & Lee, G. S. (2019). Vocal fold nodules: A disorder of phonation organs or auditory feedback? *Clinical Otolaryngology*, *44*(6), 975–982. https://doi.org/10.1111/coa.13417

Lei, Z., Fasanella, L., Martignetti, L., Li-Jessen, N. Y.-K., & Mongeau, L. (2020). Investigation of vocal fatigue using a dose-based vocal loading task. *Applied Sciences*, *10*(3), 1192.

Liu, H., & Larson, C. R. (2007). Effects of perturbation magnitude and voice F0 level on the pitch-shift reflex. *Journal of the Acoustical Society of America*, *122*(6), 3671–3677.

Lowy, K. (1942). Cancellation of the Electrical Cochlear Response with Air-and Bone-Conducted Sound. *The Journal of the Acoustical Society of America*, *14*(2), 156–158.

Lyberg-Åhlander, V., Brännström, K. J., & Sahlén, B. S. (2015). On the interaction of speakers' voice quality, ambient noise and task complexity with children's listening comprehension and cognition. *Frontiers in Psychology*, *6*, 871.

Lyberg-Åhlander, V., Rydell, R., Fredlund, P., Magnusson, C., & Wilén, S. (2019). Prevalence of Voice Disorders in the General Population, Based on the Stockholm Public Health Cohort. *Journal of Voice*, *33*(6), 900–905. https://doi.org/10.1016/j.jvoice.2018.07.007

Machado, S., Cunha, M., Velasques, B., Minc, D., Teixeira, S., Domingues, C. A., Silva, J. G., Bastos, V. H., Budde, H., & Cagy, M. (2010). Sensorimotor integration: Basic concepts, abnormalities related to movement disorders and sensorimotor training-induced cortical reorganization. *Rev Neurol*, *51*(7), 427–436.

Mahalingam, S., Boominathan, P., Arunachalam, R., Venkatesh, L., & Srinivas, S. (2021). Cepstral measures to analyze vocal fatigue in individuals with hyperfunctional voice disorder. *Journal of Voice*, *35*(6), 815–821.

Martins, R. H. G., Pereira, E. R. B. N., Hidalgo, C. B., & Tavares, E. L. M. (2014). Voice disorders in teachers. A review. *Journal of Voice*, *28*(6), 716–724.

McCormick, C. A., & Roy, N. (2002). The ChatterVox portable voice amplifier: A means to vibration dose reduction? *Journal of Voice: Official Journal of the Voice Foundation*, *16*(4), 502–508. https://doi.org/10.1016/s0892-1997(02)00126-1

Mehta, D. D., Cheyne, H. A. 2nd, Wehner, A., Heaton, J. T., & Hillman, R. E. (2016). Accuracy of Self-Reported Estimates of Daily Voice Use in Adults With Normal and Disordered

Voices. *American Journal of Speech-Language Pathology*, *25*(4), 634–641. https://doi.org/10.1044/2016_AJSLP-15-0105

Mehta, D. D., Zañartu, M., Feng, S. W., Cheyne, H. A., & Hillman, R. E. (2012). Mobile voice health monitoring using a wearable accelerometer sensor and a smartphone platform. *IEEE Transactions on Bio-Medical Engineering*, *59*(11), 3090–3096. https://doi.org/10.1109/TBME.2012.2207896

Mendoza, E., Muñoz, J., & Naranjo, N. V. (1996). The long-term average spectrum as a measure of voice stability. *Folia Phoniatrica et Logopaedica*, *48*(2), 57–64.

Milbrath, R. L., & Solomon, N. P. (2003). *Do vocal warm-up exercises alleviate vocal fatigue?*

Nair, C. B., Nayak, S., Maruthy, S., Krishnan, J. B., & Devadas, U. (2021). Prevalence of Voice Problems, Self-Reported Vocal Symptoms and Associated Risk Factors in Call Center Operators (CCOs): A Systematic Review. *Journal of Voice*. Scopus. https://doi.org/10.1016/j.jvoice.2021.07.022

Nanjundeswaran, C., Jacobson, B. H., Gartner-Schmidt, J., & Verdolini Abbott, K. (2015). Vocal Fatigue Index (VFI): Development and Validation. *Journal of Voice*, *29*(4), 433–440. https://doi.org/10.1016/j.jvoice.2014.09.012

Nanjundeswaran, C., van Mersbergen, M., & Morgan, K. (2019). Restructuring the Vocal Fatigue Index Using Mokken Scaling: Insights Into the Complex Nature of Vocal Fatigue. *Journal of Voice*, *33*(1), 110–114. https://doi.org/10.1016/j.jvoice.2017.09.008

Nanjundeswaran, C., VanSwearingen, J., & Abbott, K. V. (2017). Metabolic mechanisms of vocal fatigue. *Journal of Voice*, *31*(3), 378-e1.

Naunheim, M. L., Yung, K. C., Schneider, S. L., Henderson-Sabes, J., Kothare, H., Mizuiri, D., Klein, D. J., Houde, J. F., Nagarajan, S. S., & Cheung, S. W. (2019). Vocal motor control and central auditory impairments in unilateral vocal fold paralysis. *The Laryngoscope*, *129*(9), 2112–2117. https://doi.org/10.1002/lary.27680

Nudelman, C. J., Bottalico, P., & Cantor-Cutiva, L. C. (2023). The Effects of Room Acoustics on Self-reported Vocal Fatigue: A Systematic Review. *Journal of Voice*.

Nudelman, C. J., Codino, J., Fry, A. C., Bottalico, P., & Rubin, A. D. (2022). Voice Biofeedback via Bone Conduction Headphones: Effects on Acoustic Voice Parameters and Self-Reported Vocal Effort in Individuals With Voice Disorders. *Journal of Voice*.

Nudelman, C. J., Ortiz, A. J., Fox, A. B., Mehta, D. D., Hillman, R. E., & Van Stan, J. H. (2022). Daily Phonotrauma Index: An objective indicator of large differences in self-reported vocal status in the daily life of females with phonotraumatic vocal hyperfunction. *American Journal of Speech-Language Pathology*, *31*(3), 1412–1423.

Nudelman, C., Webster, J., & Bottalico, P. (2021). The Effects of Reading Speed on Acoustic Voice Parameters and Self-reported Vocal Fatigue in Students. *Journal of Voice*. https://doi.org/10.1016/j.jvoice.2021.05.021

Nusseck, M., Immerz, A., Richter, B., & Traser, L. (2022). Vocal Behavior of Teachers Reading with Raised Voice in a Noisy Environment. *International Journal of Environmental Research and Public Health*, *19*(15), 8929. https://doi.org/10.3390/ijerph19158929

Oates, J., & Winkworth, A. (2008). Current knowledge, controversies and future directions in hyperfunctional voice disorders. *International Journal of Speech-Language Pathology*, *10*(4), 267–277.

Patel, R. R., Awan, S. N., Barkmeier-Kraemer, J., Courey, M., Deliyski, D., Eadie, T., Paul, D., Švec, J. G., & Hillman, R. (2018). Recommended protocols for instrumental assessment of voice: American Speech-Language-Hearing Association expert panel to develop a protocol for instrumental assessment of vocal function. *American Journal of Speech-Language Pathology*, *27*(3), 887–905.

Pelegrín-García, D., & Brunskog, J. (2012). Speakers' comfort and voice level variation in classrooms: Laboratory research. *The Journal of the Acoustical Society of America*, *132*(1), 249–260. https://doi.org/10.1121/1.4728212

Publishing, OECD. (2014). *OECD Employment Outlook 2014. September*, 294–294. https://doi.org/10.1787/empl

Remacle, A., Finck, C., Roche, A., & Morsomme, D. (2012). Vocal impact of a prolonged reading task at two intensity levels: Objective measurements and subjective self-ratings. *Journal of Voice*, *26*(4), e177–e186.

Roy, N., Merrill, R. M., Gray, S. D., & Smith, E. M. (2005). Voice disorders in the general population: Prevalence, risk factors, and occupational impact. *The Laryngoscope*, *115*(11), 1988–1995.

Roy, N., Merrill, R. M., Thibeault, S., Parsa, R. A., Gray, S. D., & Elaine, S. (2004). Prevalence of voice disorders in teachers and the general population. *Journal of Speech, Language, and Hearing Research*, *47*(2), 281–293.

Roy, N., Weinrich, B., Gray, S. D., Tanner, K., Stemple, J. C., & Sapienza, C. M. (2003). Three treatments for teachers with voice disorders: A randomized clinical trial. *Journal of Speech, Language, and Hearing Research*, *46*(3), 670–688.

Sandage, M. J., Hamby, H. A., Barnett, L. A., Harris, M. L., Parker, C. R., & Allison, L. H. (2022). Vocal function differences before and after sorority recruitment. *Journal of Voice*, *36*(2), 212–218.

Scheerer, N. E., & Jones, J. A. (2018). The role of auditory feedback at vocalization onset and mid-utterance. *Frontiers in Psychology*, *9*, 2019.

Shembel, A. C., & Nanjundeswaran, C. (2022). Potential Biophysiological Mechanisms Underlying Vocal Demands and Vocal Fatigue. *Journal of Voice*, S089219972200220X. https://doi.org/10.1016/j.jvoice.2022.07.017

Sierra-Polanco, T., Cantor-Cutiva, L. C., Hunter, E. J., & Bottalico, P. (2021). Changes of Voice Production in Artificial Acoustic Environments. *Frontiers in Built Environment*, *7*, 666152. https://doi.org/10.3389/fbuil.2021.666152

Solomon, N. P., & DiMattia, M. S. (2000). Effects of a vocally fatiguing task and systemic hydration on phonation threshold pressure. *Journal of Voice*, *14*(3), 341–362.

Stemple, J. C., Stanley, J., & Lee, L. (1995). Objective measures of voice production in normal subjects following prolonged voice use. *Journal of Voice: Official Journal of the Voice Foundation*, *9*(2), 127–133. https://doi.org/10.1016/s0892-1997(05)80245-0

Stepp, C. E., Lester-Smith, R. A., Abur, D., Daliri, A., Pieter Noordzij, J., & Lupiani, A. A. (2017). Evidence for auditory-motor impairment in individuals with hyperfunctional voice disorders. *Journal of Speech, Language, and Hearing Research*, *60*(6), 1545–1550. https://doi.org/10.1044/2017_JSLHR-S-16-0282

Sund, L. T., Collum, J. A., Bhatt, N. K., & Hapner, E. R. (2021). VHI-10 scores in a treatment-seeking population with dysphonia. *Journal of Voice*.

Sundarrajan, A., Huber, J. E., & Sivasankar, M. P. (2017). Respiratory and laryngeal changes with vocal loading in younger and older individuals. *Journal of Speech, Language, and Hearing Research*, *60*(9), 2551–2556.

Svec, J. G., & Granqvist, S. (2010). Guidelines for selecting microphones for human voice production research. *American Journal of Speech-Language Pathology*, *19*(4), 356–368.

Tanner, K., Roy, N., Ash, A., & Buder, E. H. (2005). Spectral moments of the long-term average spectrum: Sensitive indices of voice change after therapy? *Journal of Voice*, *19*(2), 211–222.

Tomassi, N. E., Castro, M. E., Timmons Sund, L., Díaz-Cádiz, M. E., Buckley†, D. P., & Stepp, C. E. (2021). Effects of Sidetone Amplification on Vocal Function During Telecommunication. *Journal of Voice*. https://doi.org/10.1016/j.jvoice.2021.03.027

Trinite, B., & Astolfi, A. (2021). The impact of sound field amplification systems on speech perception of pupils with and without language disorders in natural conditions. *Applied Acoustics*, *175*, 107824.

Trinite, B., Barute, D., Blauzde, O., Ivane, M., Paipare, M., Sleze, D., & Valce, I. (2022). Choral Conductors Vocal Loading in Rehearsal Simulation Conditions. *Journal of Voice*.

v. Békésy, G. (1932). Zur theorie des hörens bei der schallaufnahme durch knochenleitung. *Annalen Der Physik*, *405*(1), 111–136.

Van Stan, J. H., Burns, J., Hron, T., Zeitels, S., Panuganti, B. A., Purnell, P. R., Mehta, D. D., Hillman, R. E., & Ghasemzadeh, H. (2023). Detecting Mild Phonotrauma in Daily Life. *The Laryngoscope*, lary.30750. https://doi.org/10.1002/lary.30750

Van Stan, J. H., Ortiz, A. J., Marks, K. L., Toles, L. E., Mehta, D. D., Burns, J. A., Hron, T., Stadelman-Cohen, T., Krusemark, C., Muise, J., & others. (2021). Changes in the Daily Phonotrauma Index following the use of voice therapy as the sole treatment for phonotraumatic vocal hyperfunction in females. *Journal of Speech, Language, and Hearing Research*, *64*(9), 3446–3455.

Vilkman, E., Lauri, E.-R., Alku, P., Sala, E., & Sihvo, M. (1999). Effects of prolonged oral reading on F0, SPL, subglottal pressure and amplitude characteristics of glottal flow waveforms. *Journal of Voice*, *13*(2), 303–312. https://doi.org/10.1016/S0892-1997(99)80036-8

Weerathunge, H. R., Tomassi, N. E., & Stepp, C. E. (2022). What Can Altered Auditory Feedback Paradigms Tell Us About Vocal Motor Control in Individuals With Voice Disorders? *Perspectives of the ASHA Special Interest Groups*, 1–18. https://doi.org/10.1044/2022_PERSP-21-00195

Whitling, S., Rydell, R., & Åhlander, V. L. (2015). Design of a clinical vocal loading test with long-time measurement of voice. *Journal of Voice*, *29*(2), 261. e13-261. e27.

Xue, C., Kang, J., Hedberg, C., Zhang, Y., & Jiang, J. J. (2019). Dynamically Monitoring Vocal Fatigue and Recovery Using Aerodynamic, Acoustic, and Subjective Self-Rating Measurements. *Journal of Voice*, *33*(5), 809.e11-809.e18. Scopus. https://doi.org/10.1016/j.jvoice.2018.03.014

Yiu, E. M. L., & Chan, R. M. M. (2003). Effect of hydration and vocal rest on the vocal fatigue in amateur karaoke singers. *Journal of Voice : Official Journal of the Voice Foundation*, *17*(2), 216–227. https://doi.org/10.1016/s0892-1997(03)00038-9

Yiu, E. M.-L., Wang, G., Lo, A. C. Y., Chan, K. M.-K., Ma, E. P.-M., Kong, J., & Barrett, E. A. (2013). Quantitative high-speed laryngoscopic analysis of vocal fold vibration in fatigued voice of young karaoke singers. *Journal of Voice: Official Journal of the Voice Foundation*, *27*(6), 753–761. https://doi.org/10.1016/j.jvoice.2013.06.010

Zhukhovitskaya, A., Battaglia, D., Khosla, S. M., Murry, T., & Sulica, L. (2015). Gender and age in benign vocal fold lesions. *The Laryngoscope*, *125*(1), 191–196. https://doi.org/10.1002/lary.24911

Ziethe, A., Petermann, S., Hoppe, U., Greiner, N., Brüning, M., Bohr, C., & Döllinger, M. (2019). Control of fundamental frequency in dysphonic patients during phonation and speech. *Journal of Voice*, *33*(6), 851–859.

TABLE TITLES

Table 1: LME models output run with vocal fatigue as the response variable and AAF condition and time as fixed factors.

Table 2: Summary of the pre- and post-loading data including LME models with all response variables, the AAF conditions, and order (post-tasks) as fixed factors. Participant ID and participant sex were the random effects terms, and the reference levels were no AAF for the condition and pre- loading for the task.

Table 3: Summary of the during VLT voice data, including LME models with all response variables and the AAF conditions and time as the fixed factors. Participant ID and participant sex were the random effects terms, and the reference level was no AAF for the condition.

FIGURE CAPTIONS

Figure 1: Overhead view of the Forbrain® device (A), with placement of the headset (B), and one of the two bone conductors (C).

Figure 2: Differences in spectra recorded by a Head and Torso Simulator (HATS) during the calibration procedure for the Forbrain® device, Sidetone Amplification device, and no AAF condition.

Figure 3: Mean and standard error (SE) of vocal status ratings for the AAF conditions and their change over time during the VLT.

Figure 4: Mean and standard error (SE) of sound pressure level for the AAF conditions and pre- versus post-loading values.

Figure 5: Mean and standard error (SE) of the spectral mean of the long-term average spectrum for the AAF conditions and pre- versus post-loading values.

Figure 6: Mean and standard error (SE) of the standard deviation of the long-term average spectrum for the AAF conditions and pre- versus post-loading values.

Figure 7: Mean and standard error (SE) of the skewness of the long-term average spectrum for the AAF conditions and pre- versus post-loading values.