ASHA American Speech-Language-Hearing Association

AJSLP

Research Note

Ambulatory Voice Biofeedback: Acquisition and Retention of Modified Daily Voice Use in Patients With Phonotraumatic Vocal Hyperfunction

Jarrad H. Van Stan,[a,b,c] Andrew J. Ortiz,[a] Dagmar Sternad,[d] Daryush D. Mehta,[a,b,c] Chuanbing Huo,[a,c] and Robert E. Hillman[a,b,c]

[a] Massachusetts General Hospital, Boston [b] Harvard Medical School, Boston, MA [c] MGH Institute of Health Professions, Boston, MA [d] Northeastern University, Boston, MA

ARTICLE INFO

Article History:
Received May 13, 2021
Revision received August 2, 2021
Accepted September 1, 2021

Editor-in-Chief: Katherine C. Hustad
Editor: Nancy Pearl Solomon

https://doi.org/10.1044/2021_AJSLP-21-00141

ABSTRACT

Purpose: Voice ambulatory biofeedback (VAB) has potential to improve carryover of therapeutic voice use into daily life. Previous work in vocally healthy participants demonstrated that motor learning inspired variations to VAB produced expected differences in acquisition and retention of modified daily voice use. This proof-of-concept study was designed to evaluate whether these VAB variations have the same desired effects on acquisition and retention in patients with phonotraumatic vocal hyperfunction (PVH).
Method: Seventeen female patients with PVH wore an ambulatory voice monitor for 6 days: three baseline days, one biofeedback day, one short-term retention day, and one long-term retention day. Short- and long-term retention were 1- and 7-days postbiofeedback, respectively. Patients were block-randomized to receive one of three types of VAB: 100%, 25%, and Summary. Performance was measured in terms of adherence time below a subject-specific vocal intensity threshold.
Results: All three types of VAB produced a biofeedback effect with 13 out of 17 patients displaying an increase in adherence time compared to baseline days. Additionally, multiple patients from each VAB group increased their adherence time during short- and/or long-term retention monitoring compared to baseline.
Conclusions: These findings show that VAB can be associated with acquisition and retention of desired voice use in patients with PVH. Specifically, all three feedback types improved multiple patients' performance and retention for up to 1 week after biofeedback removal. Future work can investigate the impact of incorporating VAB into voice therapy.

Phonotraumatic vocal hyperfunction (PVH) is a group of voice disorders characterized by obvious trauma-related tissue changes on the contact surfaces of the true vocal folds; for example, nodules and polyps (Van Stan, Mehta, Ortiz, Burns, Marks, et al., 2020). Because these vocal fold lesions are ostensibly caused by and/or associated with chronic elevated levels of phonotrauma during daily voice use, treatment for PVH often attempts to change how the patient uses his/her voice in daily life (Thomas & Stemple, 2007). However, carryover of therapeutically desired behaviors from the clinic into the patient's daily life has been noted by clinicians and patients as the most difficult barrier to successful voice therapy (Ziegler et al., 2014). This finding is reinforced by research showing a high prevalence of nonadherence and premature dropouts in voice therapy (Hapner et al., 2009; van Leer & Connor, 2015). Although voice therapy effectiveness for these patients is inextricably linked to modified

Correspondence to Jarrad H. Van Stan: jvanstan@mgh.harvard.edu.
Disclosure: *Robert Hillman, Daryush Mehta, and Steven Zeitels have financial interests in InnoVoyce LLC, a company focused on developing and commercializing technologies for the prevention, diagnosis, and treatment of voice-related disorders. Their interests were reviewed and are managed by Massachusetts General Hospital and Mass General Brigham in accordance with their conflict-of-interest policies. The other authors have declared that no other competing financial or nonfinancial interests existed at the time of publication.*

ambulatory voice use, treatment remains nearly entirely dependent upon episodic, in-clinic delivery. Adding ambulatory voice monitoring with real-time biofeedback to standard care has great potential to facilitate a paradigm shift in voice therapy, as it can extend treatment principles into the patient's daily life (KayPENTAX, 2009). Ideally, this could make the therapeutic process more efficient in many ways such as facilitating (and verifying) the faster adoption of voice therapy goals outside of the clinical environment and possibly providing a means for patients to "recalibrate" themselves and prevent relapse.

How to best use voice ambulatory biofeedback (VAB) in general is unclear. The first studies that systematically evaluated the effects of VAB provided it with 100% frequency, such that a vibrotactile or auditory cue is delivered every time a threshold is exceeded (Holbrook et al., 1974; Morsomme & Remacle, 2021; Van Stan, Mehta, & Hillman, 2015; Zicker et al., 1980). The results from these studies suggest that 100% frequency feedback often produces temporary changes in daily vocal performance that quickly fade after removing the VAB. This is not a preferable outcome because voice therapy effectiveness relies upon producing relatively permanent changes in the patient's vocal behavior, that is, retention of the learned behavior (Schmidt & Lee, 2011). If a therapeutic vocal behavior is not retained after feedback removal, the patient would be at significant risk of disorder recurrence. In our previous work, we used motor learning principles to guide the development and testing of more flexible VAB paradigms with increased potential to elicit retention (Van Stan, Mehta, Petit, et al., 2017). Specifically, relative frequency —defined as the number of trials that feedback was provided divided by the number of trials that feedback could have been provided—proved important for acquisition and retention. In addition, delayed or summary presentation —defined as withholding feedback for a block of trials and then presenting a summary of those trials—has a strong relationship with acquisition and retention of newly learned motor skills (Salmoni et al., 1984). In general, higher feedback frequencies result in better acquisition than lower feedback frequencies. However, lower feedback frequencies result in better retention than higher feedback frequencies, for example, 25% frequency (feedback after every fourth trial) versus 100% frequency (Weeks & Kordus, 1998). The literature has tended to demonstrate, compared to high-frequency feedback, decreased acquisition and increased retention as delayed feedback summaries encompass more trials (Schmidt et al., 1990).

A recent study investigated the effect of varying VAB according to these motor learning principles on the retention of a modified vocal behavior (reduced vocal intensity) in 48 participants with healthy voices (verified via history, perceptual assessment, and laryngeal videostroboscopy; Van Stan, Mehta, Sternad, et al., 2017). These participants were block-randomized into three different biofeedback groups: 100% (a cue every time voicing was too loud), 25% (a cue every fourth time voicing was too loud), and Summary (summary statistics on percent adherence every 2 min of voiced time). All participants were occupationally motivated to avoid excessively loud vocal intensities. Specifically, they were registered nurses in intensive care units or step-down units with quality improvement initiatives based on the World Health Organization's (WHO's) and the Joint Commission on Accreditation of Healthcare Organization's recommendation to reduce environmental noise levels (WHO, 2009). The results of this study supported predictions based on motor learning principles. All three biofeedback approaches resulted in significantly improved performance (reduced voicing in the participant's highest vocal intensities). When the nurses were monitored 1 week after removing the biofeedback, as hypothesized, the 100% feedback group demonstrated the most deterioration (worst retention) and the summary feedback group demonstrated the least deterioration (best retention).

It is probably more difficult for patients with PVH, compared to vocally health participants, to modify their daily voice use for multiple reasons, for example, the presence of phonotraumatic vocal fold lesions. Therefore, it cannot be assumed that these previously tested variations in biofeedback will result in similar acquisition and retention patterns in patients with PVH as the vocally healthy participants. The purpose of this prospective study was to simply test if individual patients with PVH differentially respond to various VAB schedules in accordance with motor learning principles. It is hoped that these observations begin to provide a basis for developing recommendations about how to incorporate VAB into clinical practice and help to identify additional factors that could be explored to further enhance the clinical effectiveness of VAB. All data were collected as part of a larger ongoing project aimed at a better understanding of the prevention, assessment, and treatment of hyperfunctional voice disorders. The governing institutional review board approved all experimental aspects related to the use of human subjects for this study.

Method

Participants

Only female participants were selected for this study to ensure a homogenous sample of a group that has a significantly higher incidence of phonotraumatic vocal fold lesions (Coyle et al., 2001). To simply investigate whether motor learning–inspired variations in VAB produced consistent or variable acquisition/retention responses across individual patients, the study aimed to recruit 15 total patients

(five per group). Seventeen female patients with a diagnosis of vocal fold nodules were recruited through sequential convenience sampling. Diagnoses were based on a comprehensive team evaluation (laryngologist and speech-language pathologist) at the Center for Laryngeal Surgery and Voice Rehabilitation at Massachusetts General Hospital (MGH Voice Center) that included (a) the collection of a complete case history, (b) endoscopic imaging of the larynx, (c) completion of the Voice-Related Quality of Life (V-RQOL) questionnaire (Hogikyan & Sethuraman, 1999), (d) an auditory-perceptual evaluation using the Consensus Auditory-Perceptual Evaluation of Voice (CAPE-V; Kempster et al., 2009), and (e) aerodynamic and acoustic assessments of vocal function (Patel et al., 2018). The mean age of the patients was approximately 26 years (range: 18–62 years old). Table 1 reports subscale scores for the self-reported V-RQOL and clinician-judged CAPE-V ratings for the patients. These subjective scales are reported only for the purpose of generally describing the severity level of the patients, not for statistical analysis or results reporting. Therefore, reliability was not addressed. V-RQOL scores are normalized ordinal ratings that lie between 0 and 100, with higher scores indicating a higher voice-related quality of life. CAPE-V scores are visual analog scale ratings that range from 0 to 100, with 0 indicating normality and 100 indicating the most extreme example of deviance for a particular voice quality characteristic. The CAPE-V measurement for each patient came from one rater—a voice-specialized speech-language pathologist's single rating using the CAPE-V standard reading and sustained vowel samples. Both perceptual scales qualitatively indicate that these patients were generally dysphonic and complained of voice-related impairments in their daily life.

Data Collection

A smartphone-enabled voice health monitor (VHM) was used to monitor and provide VAB throughout the study (Mehta et al., 2012). The VHM attaches a miniature accelerometer (model BU-27135, Knowles Electronics) via double-sided tape at the base of the neck (subglottal) above the sternal notch to sense phonation. The sensor is connected to a custom smartphone application as the data acquisition platform, and the system records the unprocessed acceleration signal at 11025-Hz sampling rate, 16-bit quantization, and 80-dB dynamic range to obtain frequency content of neck surface vibrations up to 5000 Hz. The VHM application provides a user-friendly interface for starting/stopping recording, daily sensor calibration, smartwatch coupling (Samsung Gear Live, Motorola Moto 360, or the LG G watch), voice activity detection settings, and flexible biofeedback schedule settings. The real-time voice activity detection settings used for all days of ambulatory monitoring were the same as in previous studies (Van Stan, Mehta, Petit, et al., 2017).

Table 1. Self-reported quality-of-life impact due to their voice disorder using the Voice-Related Quality of Life (V-RQOL) subscales and voice quality as judged by a speech-language pathologist using the Consensus Auditory-Perceptual Evaluation of Voice (CAPE-V) form for 17 patients with phonotraumatic vocal hyperfunction.

Scale	*M* (range)
V-RQOL	
Social–Emotional	74 (12–100)
Physical Functioning	69 (8–96)
Total Score	71 (10–95)
CAPE-V	
Overall Severity	27 (1–68)
Roughness	20 (3–65)
Breathiness	19 (0–65)
Strain	22 (0–50)

The experimenter virtually interacted with each patient prior to every day of monitoring to set up the VHM device including neck placement of the accelerometer. Since the neck-surface acceleration magnitude (NSAM) is highly correlated with the oral sound pressure level (SPL), NSAM is often converted to oral SPL based on a calibration procedure (Švec et al., 2005). However, for the present biofeedback application, a calibration procedure transforming acceleration level to SPL was not performed. All signal amplitude-based measures (in dB) are derived directly from the uncalibrated NSAM. This approach was chosen because (a) no comparisons of vocal intensity were made across individuals and (b) the calibration of the SPL can introduce a measurement error of on average ± 6–10 dB (Švec et al., 2005); it was crucial to minimize any signal level error within each subject. To further minimize measurement error in the signal level, every day of monitoring the patient sent the investigators a picture of their anterior neck showing that the accelerometer was placed on the same location between the thyroid prominence and the superior border of the sternum. The experimenter used subject-specific anatomical landmarks (clavicle, wrinkles, thyroid prominence, and the cricoid) to ensure accurate accelerometer placement on the neck surface from day to day. If the photo showed an incorrect accelerometer placement (i.e., too high, low, and/or too lateral), the experimenter asked the patient to reposition the sensor, provided written guidance by marking the photo with the correct placement, and asked the patient to send another picture with the new accelerometer placement. This process continued until the accelerometer was correctly placed. Incorrect accelerometer placement occurred in < 5% of days and was quickly resolved every time.

Biofeedback Threshold Selection

The NSAM and duration settings for the biofeedback threshold were designed to be as consistent across

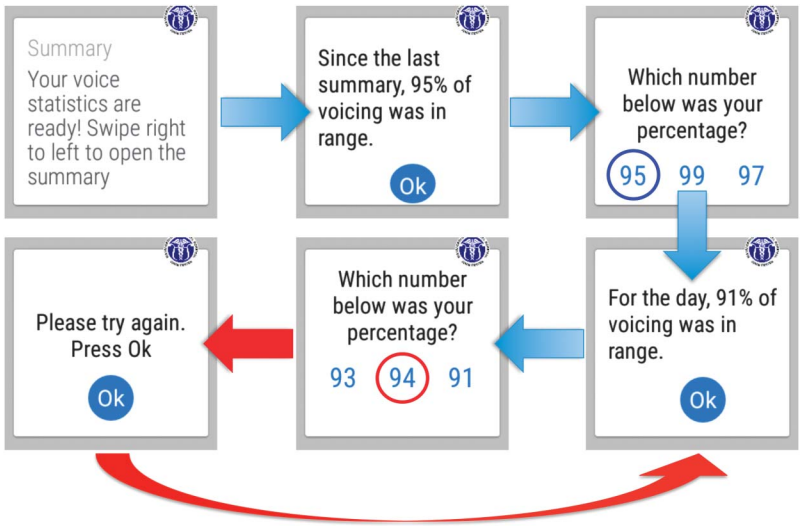
the three biofeedback conditions as possible. The NSAM threshold for biofeedback was individually derived for each participant based upon her 3 days of baseline monitoring. Biofeedback was triggered/registered when the dB level of the NSAM signal exceeded the 85th percentile of the pooled distribution of level across each individual's three baseline days. The duration of time that the NSAM signal needed to remain above the threshold level to trigger/register biofeedback was set to a single analysis frame (50-ms duration) to detect as many of these events as possible. The 85th percentile was chosen as the biofeedback threshold level based on our previous experience from studies of VAB (Van Stan, Mehta, Sternad, et al., 2017). Specifically, previous work established this percentile-based level threshold to subjectively maximize the potential for a noticeable vocal behavior change (i.e., reduction in loudness) while still allowing functional vocal intensity and minimizing risk from annoying the participant. Additionally, it was clinically desirable to provide patients with PVH this biofeedback (avoid their upper 15th percentile of loudness) because these patients (compared to matched controls) tend to spend more time at higher vocal intensities (Van Stan, Mehta, Ortiz, Burns, Toles, et al., 2020). Upper percentile vocal intensities are physiologically associated with higher potential for phonotrauma (Jiang et al., 2001; Titze, 1994).

Biofeedback Delivery

During biofeedback days for the 100% and 25% feedback groups, the VHM provided a 250-ms vibrotactile cue via a smartwatch and phone whenever they exceeded the NSAM level threshold after every 50-ms voiced frame (100% feedback group) or every fourth voiced frame when they exceeded the threshold (25% feedback group). The VHM would automatically stop recording after registering 50,400 voiced frames (42 min of phonation) so each participant had equal exposure time to biofeedback cueing. All other biofeedback settings were the same as previous studies (Van Stan, Mehta, Petit, et al., 2017).

Patients in the summary feedback group received a continuous vibrotactile cue on their smartwatch and phone after 2,400 voiced frames (i.e., 2 min of phonation) to alert them to look at their percentage adherence values for (a) the entire day and (b) the most recent period of voicing. More specifically, the VHM software displayed a simple summary statistic on the smartwatch screen called percent adherence $\left(\frac{\#\text{ of voiced frames inside desired range}}{\#\text{ of total voiced frames}}\right)$ at adjustable timeframes to replicate the concept of summary feedback. The participants were provided summary statistics after every 2 min of voiced time, which corresponded to approximately every 20 min of monitoring (10% phonation time was typical). As shown in Figure 1, to ensure that the participant adequately comprehended their summary statistics, multiple screens for user interaction were provided on the phone and smartwatch whenever the statistics were displayed. Monitoring could only continue after the participant used the smartwatch to accurately enter/replicate the displayed percent adherences. Participant responses were recorded in a text document on the smartphone that allowed documentation of how much time passed between the summary statistic presentation and

Figure 1. Screenshots taken from an LG G smartwatch when displaying summary statistics for the Summary feedback group. The blue arrows/circle denote the path taken when the user inputs a correct answer, and the red arrows/circle denote the path taken when the user inputs an incorrect answer. The summary statistics screens will not finish until the participant correctly enters all statistics.



when the user looked at them, as well as if the participant accurately recalled their adherence percentages.

Study Design

Patients were block-randomized to receive one of three different types of feedback: (a) 100% feedback, (b) 25% feedback, and (c) Summary feedback. Each patient was monitored for a total of 6 days of monitoring. The first 3 days established an individual's natural vocal intensity behavior (baseline), the fourth day included VAB for 42 min of phonation time (biofeedback), the fifth day included no biofeedback and occurred 1 day after the biofeedback day (short-term retention), and the sixth day included no biofeedback and occurred 7 days after the biofeedback day (long-term retention). Forty-two minutes of phonation per day was chosen because, based on previous work (Van Stan, Mehta, Ortiz, Burns, Toles, et al., 2020), it is the least amount of voicing that can be realistically expected across all patients. Historically, patients have worn the device for 10 hr/day (on average) and exhibited phonation percentages of 7% or higher (i.e., 7% phonation = 42 min of voicing over 10 hr). Since a primary purpose of the study was to assess the effect of different biofeedback schedules on retention, it only made sense for patients to continue the study if they first demonstrated a biofeedback effect. As in previous work, subjects were only considered to have demonstrated a biofeedback effect if their percent adherence during the biofeedback day was at or above 90% (Van Stan, Mehta, Sternad, et al., 2017). If a subject failed to achieve ≥ 90% adherence during a biofeedback day, they were terminated from the study and did not undergo retention monitoring.

Since the study aimed to evaluate the effect of only VAB variations (not in combination with treatment) on the acquisition and retention of modified daily voice use, all patients participated in the study before undergoing any voice therapy and/or laryngeal surgery. Before starting a biofeedback day, a voice-specialized speech-language pathologist provided approximately 15 min of education regarding the biofeedback. First, the patients were asked to avoid voicing at their "loudest loud" (i.e., 85th percentile or louder) for the entire biofeedback and two retention days. They were told that this was not a treatment study, so avoiding their loudest voicing may not have any effect on their voice disorder or vocal symptoms. However, there are no indications that avoiding loud voicing will worsen their disorder or vocal symptoms. Patients were asked to let study staff know if any worsening of vocal symptoms occurred during biofeedback and retention. Then, the patients were acquainted with their "loudest loud" by producing 30 s of voicing in spontaneous speech at three loudness levels (softer-than-usual, usual, and louder-than-usual) with 100% frequency feedback. Subsequently, patients receiving 100% feedback were told that vibrations will occur on their smartwatch every time they voiced at their loudest loud (i.e., 100%), even for a very short time (e.g., throat clearing). Patients receiving 25% feedback were told that vibrations will occur every fourth time they voiced at their loudest loud (i.e., 25%). It was stressed that every cue represented four instances of being too loud, not just one. Patients receiving Summary feedback were told that vibrations will only occur after they voiced for 2 min, informing them that their summary statistics were ready for viewing. It was emphasized that higher percentages were desired and represented successful avoidance of their loudest loud. Furthermore, these patients were told that percentages of ≥ 90% must be attained to demonstrate some change in voicing (i.e., a biofeedback effect) and they would be terminated from the study if their total adherence percentage was < 90% (i.e., they would not undergo retention monitoring). Finally, looking at and answering the summary feedback was completed 3 times (30 s of voicing at softer-than-usual, usual, and louder-than-usual vocal intensities). Biofeedback monitoring did not start until the patient verbally reported and vocally demonstrated that they understood the desired behavior (avoid their loudest loud voicing) and their specific biofeedback type.

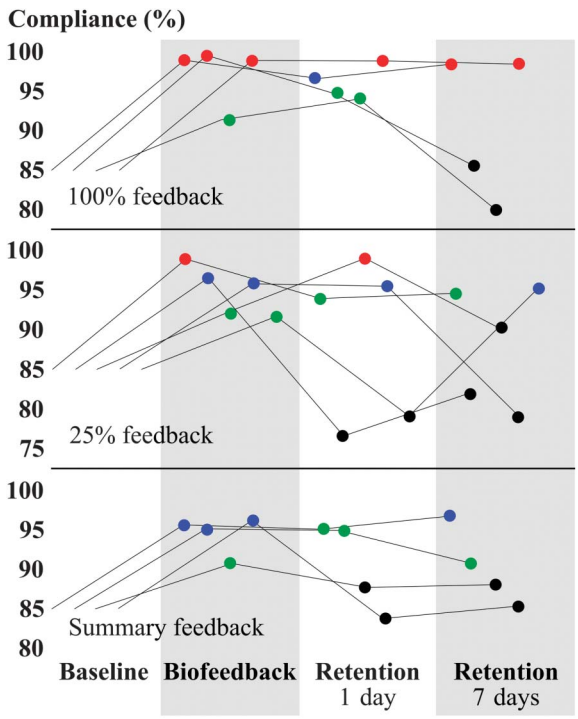
Statistical Analysis

Within individual patients, differences between baseline monitoring (all 3 days pooled together) versus the biofeedback day, short-term retention day, or long-term retention day were assessed via percentage adherence and odds ratios (*OR*s). Chi-square tests were all significant ($p < .001$) because each time period contained thousands of 50-ms voice frames, rendering high statistical power. Therefore, *OR*s were used as an effect size metric to demonstrate which comparisons resulted in a meaningful difference; small, medium, and large differences were $OR > 1.68$ (i.e., 90.5% adherence), 3.47 (i.e., 95.2% adherence), or 6.71 (i.e., 97% adherence), respectively (Chen et al., 2010). If any statistically significant difference produced an $OR < 1.68$, it was considered not meaningfully different than baseline, that is, no biofeedback effect during a biofeedback day or no retention during a retention day.

Results

Out of the 17 patients, four patients did not respond to the biofeedback and were not monitored for retention. Two of the nonresponders received 100% feedback, one received 25% feedback, and one received Summary feedback. Figure 2 illustrates the biofeedback and retention effects of the remaining 13 patients who responded to the

Figure 2. Ambulatory biofeedback effect showing individual patient percent adherence values across monitoring periods for the 100% (top panel), 25% (middle panel), and Summary (bottom panel) feedback groups. Colored circles represent individual patient large (red), medium (blue), and small (green) effect sizes, or no effect (black).



biofeedback. During the biofeedback day, those receiving 100% feedback performed with the largest effect sizes (*OR* = 52.14, 18.39, 17.87, 1.98, with respective adherence = 99.66, 99.05, 99.02, 91.81%), those receiving 25% feedback performed with the next largest effect sizes (*OR* = 16.64, 5.08, 4.06, 2.15, 1.97, with respective adherence = 98.95, 96.64, 95.80, 92.40, 91.78%), and those receiving Summary feedback performed with the smallest effect sizes (*OR* = 4.21, 3.95, 3.57, 1.89, with respective adherence = 95.97, 95.72, 95.29, 90.78%).

During short-term retention monitoring, 1 day after removing the biofeedback, all four of the patients who received 100% feedback retained their new behavior with one large effect size (*OR* = 16.67; % adherence = 98.95), one medium effect size (*OR* = 5.28; % adherence = 96.77), and two small effect sizes (*OR* = 3.11, 2.74; % adherence = 94.62, 93.96). Three out of the five patients who received 25% feedback retained their new behavior, with one patient showing a large effect size (*OR* = 16.08; % adherence = 98.91), one patient showing a medium effect size (*OR* = 3.73; % adherence = 95.48), and one patient showing a small effect size (*OR* = 2.75; % adherence = 93.97). Two out of the four patients who received Summary feedback retained their new behavior, both with small effect sizes (*OR* = 3.43, 3.36; % adherence = 95.11, 95.01).

During long-term retention monitoring, 7 days after removing the biofeedback, two of the four patients who received 100% feedback retained their new behavior, both with large effect sizes (*OR* = 12.35, 12.25; % adherence = 98.59, 98.57). Both of these patients retained their new behavior across short- and long-term retention monitoring. Two of the five patients who received 25% feedback retained their new behavior, with one patient showing a medium effect size (*OR* = 3.91; % adherence = 95.68) and one patient showing a small effect size (*OR* = 3.09; % adherence = 94.60). Of note, only one of these patients retained across both short- and long-term retention monitoring. Two of the four patients who received Summary feedback retained their new behavior, with one patient showing a medium effect size (*OR* = 5.28; % adherence = 96.76) and one patient showing a small effect size (*OR* = 1.75; % adherence = 90.85). Both of these patients retained their new behavior across short- and long-term retention monitoring.

Discussion

Biofeedback Acquisition Effects

The main objective of this work was to demonstrate how motor learning–inspired variations in VAB can affect the acquisition and retention of a modified vocal behavior—for example, avoiding the patient's highest vocal intensities—in individual patients with PVH. The patients' responses to the biofeedback were, in some ways, expected according to previous VAB studies (Van Stan, Mehta, Petit, et al., 2017; Van Stan, Mehta, Sternad, et al., 2017) and results from motor learning (Salmoni et al., 1984; Schmidt & Lee, 2011). As shown in previous studies, daily voice use was significantly changed when patients received VAB. Specifically, most patients, 13 out of 17, avoided their loudest vocal intensities during biofeedback. The biofeedback responses for the patients were ranked according to the expectations of motor learning principles: Highest to lowest adherence values were seen in the patients receiving 100%, 25%, and Summary feedback; large mean *OR* = 22.60, % adherence = 97.38; medium mean *OR* = 5.98, % adherence = 95.11; and small mean *OR* = 3.40, % adherence = 94.44, respectively. However, this is markedly different than previous findings in healthy voices where all three types of feedback produced similar results with medium-to-large effect sizes: % adherence = 97.04, 97.84, and 96.2, respectively (Van Stan, Mehta, Sternad, et al., 2017). The patients generally benefitted more from higher doses of feedback than lower doses, and the vocally healthy participants generally modified their voice use by a similar amount regardless of feedback dosing. Perhaps the desired behavior of avoiding the upper 15th percentile of vocal intensity was more difficult for patients than for vocally healthy participants. The literature points to multiple potential explanations for why modified daily voice use

would be more difficult for patients than vocally healthy individuals: self-regulation (Vinney & Turkstra, 2013), personality (Roy et al., 2000), altered vocal fold biomechanics due to the lesions (Bastian et al., 1990; Verdolini et al., 2003), and so forth.

Retention Effects

The results for short- and long-term retention indicate that all three types of feedback can produce significant changes in vocal behavior, lasting at least 1 week after biofeedback removal. This seems to contradict previous findings in vocally healthy participants, where 100% feedback demonstrated the worst short- and long-term retention (Van Stan, Mehta, Sternad, et al., 2017). However, many vocally healthy participants also retained their modified behavior at 1-week postbiofeedback across all three types of feedback: six out of 15 for 100% feedback, 10 out of 17 for 25% feedback, and 10 out of 16 individuals for Summary feedback. Also, variability across the patients receiving different types of feedback was similar to our previous study. Specifically, those patients receiving Summary feedback displayed less variable short- and long-term retention (minimum and maximum % adherence = 84–97) than those patients receiving 25% and 100% feedback (% adherence = 77–99 and 80–99, respectively). For the vocally healthy subjects, the difference in percentage adherence range for those receiving the Summary, 25%, and 100% feedbacks was similar: % adherence = 89.7–99.6, 83.8–99.9, and 78.2–99.6, respectively (Van Stan, Mehta, Sternad, et al., 2017). For both vocally healthy and PVH participants, these differences in variability were driven more by the lower % adherence values—that is, participants receiving 100% and 25% feedback performed worse than baseline—than by the higher % adherence values; that is, differences in patients retained a strong biofeedback effect. Therefore, summary feedback may produce the most stable modifications in voice use with the least risk of an adverse reaction upon biofeedback removal, such as performing worse than baseline. Furthermore, patients receiving Summary feedback got the lowest dose of feedback, which could ostensibly be the primary cause of the smaller effect sizes. If the weaker Summary feedback effects were due to decreased dosing, future work could attempt to increase the strength of these relatively stable effects by providing Summary feedback over longer time scales than a single day.

Biofeedback Nonresponders

There does appear to be a difference in the proportion of nonresponders to biofeedback between the patients with PVH and the previously tested vocally healthy group. Specifically, noticeably more patients with PVH did not respond to the biofeedback (four out of 17, ~25%) compared to the vocally healthy subjects (nine out of 57, ~15%). Three of the four patients who did not respond to the biofeedback were briefly interviewed about their experience to gain additional insights into factors that may have interfered with the intervention. There appeared to be two main barriers to the biofeedback effect: reduced self-awareness of vocal intensity in daily life and lack of trust in the VAB. Table 2 provides quotes from the patients regarding these two barriers. Regarding the first barrier, these three patients were surprised that they had not significantly modified their vocal intensity during the biofeedback. Furthermore, the patients thought that they had changed their voice use in some manner, even if it was not vocal intensity. To evaluate this, the patients' ambulatory voice data were analyzed to extract percent phonation, fundamental frequency, cepstral peak prominence, and the difference in amplitude between the first and second harmonics (H1-H2). No significant differences were found between baseline and biofeedback monitoring for any of these voice-use measures. This decreased self-awareness of voice use may ostensibly result from known group-based differences in personality between patients with PVH and vocally healthy controls. Specifically, patients with PVH more often exhibit personality profiles with high extroversion, high neuroticism, and low impulse control (Roy et al., 2000). These three patients also did not appear to trust the biofeedback, as they suggested it was either not set correctly or providing incorrect vocal intensity values. A biofeedback malfunction in the field cannot be absolutely excluded, as the experimenters were not present throughout the patient's daily life. However, this is unlikely because the equipment passed in-clinic quality checks before and after the patient participated in the study as well as in-field quality checks before and after each day of monitoring. To address this barrier, before providing VAB in the field, the clinician could provide information about the validity and quality of the VAB to gain the

Table 2. Patient statements underlying the two barriers toward achieving a biofeedback effect.

Barriers	Quotes
Reduced vocal awareness	"Surprising poor performance, [I] felt like I was not talking very loud. . ."
	"[I] was surprised that I didn't hit the 90% goal."
Lack of trust in the VAB	"I know my voice very well, [then she demonstrated loud voicing and said] but apparently that is loud?"
	". . . I would try to talk a little quieter and did not think I was talking that loud in the first place."
	"I would get excited and then the biofeedback would buzz, but I didn't feel like I was being louder."
	"When the phone was buzzing, [it was] frustrating because . . . I would be constantly asking myself 'Am I really too loud?' If the buzzing was true, I'd be talking too loud all the time."

Note. VAB = voice ambulatory biofeedback.

patient's buy-in or trust. The information could be provided in many ways, ranging from a simple lecture (e.g., "The device is rather accurate, so trust the device if it tells you that you're being loud") to trialing the device with the patient, for example, having the patient speak at various loudness levels while receiving the VAB and saying, "See how the device only cues you when you're louder than necessary?"

Relation to Pathophysiology

The objective measure (vocal intensity) and threshold (stay below the 85th percentile of vocal intensity) for biofeedback were chosen because of their indirect relationship to putative pathophysiological mechanisms such as phonotrauma. Since vocal intensity feedback is indirectly targeting the patient's PVH, the obvious concern is that a patient could use hyperfunctional behaviors to maintain adequate adherence with the device and defeat the purpose of an ambulatory intervention. The patients in this study were asked about their vocal status (i.e., fatigue, effort, discomfort) after biofeedback and retention monitoring. No one reported any vocal deterioration, so it is unlikely that the patients in this study engaged in grossly maladaptive behaviors during their biofeedback and retention days. Future work could focus on incorporating biofeedback measures that are more directly related to the hypothesized etiology and/or pathophysiology of PVH, for example, inverse filtered measures such as AC Flow or maximum flow declination rate (Sapienza & Stathopoulos, 1995), indirect estimates of vocal fold collision (Titze & Hunter, 2015), or subglottal pressure (Wokurek & Pützer, 2009). The Daily Phonotrauma Index (DPI) is a promising measure related to PVH pathophysiology because it has been demonstrated to change toward "normal" in expected ways following surgery and/or voice therapy (Van Stan, Mehta, Ortiz, Burns, Marks, et al., 2020; Van Stan, Mehta, Ortiz, Burns, Toles, et al., 2020; Van Stan et al., 2021). However, the DPI is based on two features—NSAM skew and H1-H2 standard deviation—that require distributions of multiple voiced frames, meaning that it could only be incorporated into Summary biofeedback and not immediate feedback paradigms like 100% or 25% frequency.

Limitations

This work has a few limitations that should be considered. First, because the number of patients receiving each type of feedback was relatively small, no attempt was made to statistically test for group-based differences —that is, 100% versus 25% versus Summary—and between studies—that is, patients with PVH in this study versus vocally healthy individuals from a previous study—because such tests would be underpowered. Thus, what appear to be group-based differences between different types of biofeedback (e.g., nonresponders, biofeedback effects, and retention effects) might not remain, or change in effect size, once more patients with PVH are monitored and provided biofeedback. However, the 17 case studies with PVH here can provide preliminary indications regarding how to incorporate ambulatory voice biofeedback into a voice therapy regimen for an individual patient. For example, if a patient is having difficulty generalizing therapeutic voicing into their daily life, perhaps the clinician should try 100% feedback first. This is because 100% feedback has been the most likely paradigm to elicit a large biofeedback effect with some possibility of long-term retention. If the new vocal behavior is not retained after turning off the biofeedback, the clinician can fade the ambulatory biofeedback by applying Summary feedback. This is because Summary feedback has provided the most evidence of retention across past studies and the most stable vocal performance (Van Stan, Mehta, Sternad, et al., 2017).

Another limitation of this study is that the patients who did not respond to the biofeedback were simply stopped from continuing the design. However, when patients do not respond to feedback in a clinical setting, their treatment program cannot be simply discontinued. Since the current study design only provided a single type of biofeedback to each individual patient, it cannot provide guidance regarding how to modify the ambulatory biofeedback paradigms to improve the patient's responsiveness. Future work will be required to evaluate how variables related to VAB may be modified according to patient performance. For example, when should VAB be introduced (e.g., when the patient achieves a certain level of mastery in voice therapy), what objective measure or combination of measures should be targeted by the biofeedback, what is the impact on retention of varying biofeedback dosing, how should the clinician find/establish a biofeedback threshold (e.g., negative practice), and so forth.

Conclusions

Overall, the results of this study suggest that VAB has potential to improve the carryover of therapeutically desired voice use in the treatment of patients with PVH. Specifically, all three modifications in feedback frequency and timing improved one or more individual patients' performance and retention of avoiding their highest vocal intensities. Future clinical studies should investigate the impact of incorporating VAB into voice therapy treatment. Such studies could integrate additional voice use measures that are more closely related to the underlying putative pathophysiology of the disorder being treated than simple thresholds based on vocal frequency and intensity, for example, the DPI for PVH, glottal inverse filtering features, and multidimensional thresholds based on improvements noted during the therapy session.

Acknowledgments

This work was supported by the Voice Health Institute and the National Institutes of Health, National Institute on Deafness and Other Communication Disorders under Grants R33 DC011588 (PI: Robert Hillman) and P50 DC015446 (PI: Robert Hillman). The article's contents are solely the responsibility of the authors and do not necessarily represent the official views of the National Institutes of Health.

References

Bastian, R. W., Keidar, A., & Verdolini-Marston, K. (1990). Simple vocal tasks for detecting vocal fold swelling. *Journal of Voice, 4*(2), 172–183. https://doi.org/10.1016/S0892-1997(05)80144-4

Chen, H., Cohen, P., & Chen, S. (2010). How big is a big odds ratio? Interpreting the magnitudes of odds ratios in epidemiological studies. *Communications in Statistics—Simulation and Computation, 39*(4), 860–864. https://doi.org/10.1080/03610911003650383

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

Hapner, E., Portone-Maira, C., & Johns, M. M., III (2009). A study of voice therapy dropout. *Journal of Voice, 23*(3), 337–340. https://doi.org/10.1016/j.jvoice.2007.10.009

Hogikyan, N. D., & Sethuraman, G. (1999). Validation of an instrument to measure voice-related quality of life (V-RQOL). *Journal of Voice, 13*(4), 557–569. https://doi.org/10.1016/S0892-1997(99)80010-1

Holbrook, A., Rolnick, M. I., & Bailey, C. W. (1974). Treatment of vocal abuse disorders using a vocal intensity controller. *Journal of Speech and Hearing Disorders, 39*(3), 298–303. https://doi.org/10.1044/jshd.3903.298

Jiang, J. J., Shah, A. G., Hess, M. M., Verdolini, K., Banzali, F. M., Jr., & Hanson, D. G. (2001). Vocal fold impact stress analysis. *Journal of Voice, 15*(1), 4–14. https://doi.org/10.1016/S0892-1997(01)00002-9

KayPENTAX. (2009). *Ambulatory phonation monitor: Applications for speech and voice*.

Kempster, G. B., Gerratt, B. R., Verdolini Abbott, K., Barkmeier-Kraemer, J., & Hillman, R. E. (2009). Consensus Auditory-Perceptual Evaluation of Voice: Development of a standardized clinical protocol. *American Journal of Speech-Language Pathology, 18*(2), 124–132. https://doi.org/10.1044/1058-0360(2008/08-0017)

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

Morsomme, D., & Remacle, A. (2021). Can ambulatory biofeedback help a transgender woman speak at a higher pitch. *Logopedics Phoniatrics Vocology*. Advance online publication. https://doi.org/10.1080/14015439.2021.1881158

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. https://doi.org/10.1044/2018_AJSLP-17-0009

Roy, N., Bless, D. M., & Heisey, D. (2000). Personality and voice disorders: A multitrait-multidisorder analysis. *Journal of Voice, 14*(4), 521–548. http://doi.org/10.1016/S0892-1997(00)80009-0

Salmoni, A. W., Schmidt, R. A., & Walter, C. B. (1984). Knowledge of results and motor learning: A review and critical reappraisal. *Psychological Bulletin, 95*(3), 355–386. https://doi.org/10.1037/0033-2909.95.3.355

Sapienza, C. M., & Stathopoulos, E. T. (1995). Speech task effects on acoustic and aerodynamic measures of women with vocal nodules. *Journal of Voice, 9*(4), 413–418. https://doi.org/10.1016/S0892-1997(05)80203-6

Schmidt, R. A., Lange, C., & Young, D. E. (1990). Optimizing summary knowledge of results for skill learning. *Human Movement Science, 9*(3–5), 325–348. https://doi.org/10.1016/0167-9457(90)90007-Z

Schmidt, R. A., & Lee, T. D. (2011). *Motor control and learning: A behavioral emphasis* (5th ed.). Human Kinetics.

Švec, J. G., Titze, I. R., & Popolo, P. S. (2005). Estimation of sound pressure levels of voiced speech from skin vibration of the neck. *The Journal of the Acoustical Society of America, 117*(3), 1386–1394. https://doi.org/10.1121/1.1850074

Thomas, L. B., & Stemple, J. C. (2007). Voice therapy: Does science support the art. *Communicative Disorders Review, 1*(1), 49–77.

Titze, I. R. (1994). Mechanical stress in phonation. *Journal of Voice, 8*(2), 99–105. https://doi.org/10.1016/S0892-1997(05)80302-9

Titze, I. R., & Hunter, E. J. (2015). Comparison of vocal vibration-dose measures for potential-damage risk criteria. *Journal of Speech, Language, and Hearing Research, 58*(5), 1425–1439. https://doi.org/10.1044/2015_JSLHR-S-13-0128

van Leer, E., & Connor, N. P. (2015). Predicting and influencing voice therapy adherence using social–cognitive factors and mobile video. *American Journal of Speech-Language Pathology, 24*(2), 164–176. https://doi.org/10.1044/2015_AJSLP-12-0123

Van Stan, J. H., Mehta, D. D., & Hillman, R. E. (2015). The effect of voice ambulatory biofeedback on the daily performance and retention of a modified vocal motor behavior in participants with normal voices. *Journal of Speech, Language, and Hearing Research, 58*(3), 713–721. https://doi.org/10.1044/2015_JSLHR-S-14-0159

Van Stan, J. H., Mehta, D. D., Ortiz, A. J., Burns, J. A., Marks, K. L., Toles, L. E., Stadelman-Cohen, T., Krusemark, C., Muise, J., Hron, T., Zeitels, S. M., Fox, A. B., & Hillman, R. E. (2020). Changes in a Daily Phonotrauma Index after laryngeal surgery and voice therapy: Implications for the role of daily voice use in the etiology and pathophysiology of phonotraumatic vocal hyperfunction. *Journal of Speech, Language, and Hearing Research, 63*(12), 3934–3944. https://doi.org/10.1044/2020_JSLHR-20-00168

Van Stan, J. H., Mehta, D. D., Ortiz, A. J., Burns, J. A., Toles, L. E., Marks, K. L., Vangel, M., Hron, T., Zeitels, S., & Hillman, R. E. (2020). Differences in weeklong ambulatory vocal behavior between female patients with phonotraumatic lesions and matched controls. *Journal of Speech, Language, and Hearing Research, 63*(2), 372–384. https://doi.org/10.1044/2019_JSLHR-19-00065

Van Stan, J. H., Mehta, D. D., Petit, R. J., Sternad, D., Muise, J., Burns, J. A., & Hillman, R. E. (2017). Integration of motor learning principles into real-time ambulatory voice biofeedback and example implementation via a clinical case study with vocal fold nodules. *American Journal of Speech-Language Pathology, 26*(1), 1–10. https://doi.org/10.1044/2016_AJSLP-15-0187

Van Stan, J. H., Mehta, D. D., Sternad, D., Petit, R., & Hillman, R. E. (2017). Ambulatory voice biofeedback: Relative frequency and summary feedback effects on performance and retention of reduced vocal intensity in the daily lives of participants

with normal voices. *Journal of Speech, Language, and Hearing Research, 60*(4), 853–864. https://doi.org/10.1044/2016_JSLHR-S-16-0164

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., Fox, A. B., Nudelman, C., Zeitels, S., & Hillman, R. E. (2021). Changes in the Daily Phonotrauma Index (DPI) 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. https://doi.org/10.1044/2021_JSLHR-21-00082

Verdolini, K., Branski, R. C., Rosen, C. A., & Hebda, P. A. (2003). Shifts in biochemical markers associated with wound healing in laryngeal secretions following phonotrauma: A preliminary study. *Annals of Otology, Rhinology & Laryngology, 112*(12), 1021–1025. https://doi.org/10.1177/000348940311201205

Vinney, L. A., & Turkstra, L. S. (2013). The role of self-regulation in voice therapy. *Journal of Voice, 27*(3), 390.e1–390.e11. https://doi.org/10.1016/j.jvoice.2013.01.003

Weeks, D. L., & Kordus, R. N. (1998). Relative frequency of knowledge of performance and motor skill learning. *Research Quarterly for Exercise and Sport, 69*(3), 224–230. https://doi.org/10.1080/02701367.1998.10607689

Wokurek, W., & Pützer, M. (2009). Accelleration sensor measurements of subglottal sound pressure for modal and breathy phonation quality. In C. Manfredi (Ed.), *6th International Workshop on Models and Analysis of Vocal Emissions for Biomedical Applications (MAVEBA 2009)* (pp. 153–156). Firenze University Press.

World Health Organization. (2009). *Global health risks: Mortality and burden of disease attributable to selected major risks*. https://apps.who.int/iris/handle/10665/44203

Zicker, J. E., Tompkins, W. J., Rubow, R. T., & Abbs, J. H. (1980). A portable microprocessor-based biofeedback training device. *IEEE Transactions on Biomedical Engineering, 27*(9), 509–515. https://doi.org/10.1109/TBME.1980.326665

Ziegler, A., Dastolfo, C., Hersan, R., Rosen, C. A., & Gartner-Schmidt, J. (2014). Perceptions of voice therapy from patients diagnosed with primary muscle tension dysphonia and benign mid-membranous vocal fold lesions. *Journal of Voice, 28*(6), 742–752. https://doi.org/10.1016/j.jvoice.2014.02.007