JSLHR

Research Note

The Effect of Voice Ambulatory Biofeedback on the Daily Performance and Retention of a Modified Vocal Motor Behavior in Participants With Normal Voices

Jarrad H. Van Stan,[a,b] Daryush D. Mehta,[a,b,c] and Robert E. Hillman[a,b,c]

Purpose: Ambulatory biofeedback has potential to improve carryover of newly established vocal motor behaviors into daily life outside of the clinic and warrants systematic research that is lacking in the literature. This proof-of-concept study was designed to establish an empirical basis for future work in this area by formally assessing whether ambulatory biofeedback reduces daily vocal intensity (performance) and the extent to which this change remains after biofeedback removal (retention).
Method: Six participants with normal voices wore the KayPENTAX Ambulatory Phonation Monitor for 3 baseline days followed by 4 days with biofeedback provided on odd days.
Results: Compared to baseline days, participants exhibited a statistically significant decrease in mean vocal intensity (4.4 dB) and an increase in compliance (16.8 percentage points) when biofeedback was provided above a participant-specific intensity threshold. After biofeedback removal, mean vocal intensity and compliance reverted back to baseline levels.
Conclusions: These findings suggest that although current ambulatory biofeedback approaches have potential to modify a vocal motor behavior, the modified behavior may not be retained after biofeedback removal. Future work calls for the testing of more innovative ambulatory biofeedback approaches on the basis of motor control and learning theories to improve retention of a desired vocal motor behavior.

Successful voice therapy relies upon the patient learning new—or relearning previous—vocal motor behaviors (e.g., loudness, pitch, voice quality, respiration, and efficiency) in order to decrease communication-related disabilities in daily life. However, little is known about how various components of voice therapy, such as verbal feedback, biofeedback, practice, and dosing, affect the robustness with which patients learn modifications to their vocal motor behaviors (c.f. Ferrand, 1995; Steinhauer & Grayhack, 2000; Wong, Ma, & Yiu, 2011; Yiu, Verdolini, & Chow, 2005). One reason is that most motor control/learning studies have focused on limb or postural motor behaviors, which differ significantly in terms of biomechanics and sensorimotor components compared to the vocal or bulbar neuromuscular system (c.f. Maas et al., 2008). A second reason is that designing a vocal learning task in which the target can be accurately measured quantitatively is challenging; currently most vocal motor behaviors targeted in therapy rely on the clinician's perceptual assessment of "correctness." In this research note, "learning" is considered specific to the procedural, or motor, aspects of learning (c.f. Cohen & Squire, 1980).

Studies regarding the effects of biofeedback on motor learning typically are conducted in clinical or research laboratory settings (c.f. Maryn, De Bodt, & Van Cauwenberge, 2006; Schmidt & Lee, 2011). Wearable devices have been recently developed that can provide biofeedback related to vocal sound pressure level (SPL) and fundamental frequency (f0) as individuals go about their daily routine. Biofeedback throughout a patient's activities of daily living has the potential to significantly improve one of the most challenging aspects of voice therapy—carryover of newly established vocal motor behaviors outside of the therapy session (Van Stan, Gustafsson, Schalling, & Hillman, 2014). Advances in ambulatory monitoring technology can potentially extend voice therapy concepts into multiple contexts and

[a]MGH Institute of Health Professions, Boston, MA
[b]Massachusetts General Hospital, Boston
[c]Harvard Medical School, Boston, MA
Correspondence to Jarrad Van Stan: jvanstan@mghihp.edu

Editor: Jody Kreiman
Associate Editor: Scott Thomson
Received June 10, 2014
Revision received October 20, 2014
Accepted January 10, 2015
DOI: 10.1044/2015_JSLHR-S-14-0159

Disclosure: *Robert E. Hillman has a financial interest in the APM based on a contractual agreement between Sensimetrics, Inc. (R&D for the initial version of the APM) and KayPENTAX, Inc. (manufacturer of the APM).*

types of voice use (e.g., singing, conversing, and practicing therapy exercises) and provide the capability to store information regarding vocal motor behavior modifications due to biofeedback.

Most studies have not directly quantified the effect of ambulatory biofeedback on vocal motor behavior but instead have described the effect qualitatively because the devices did not store ambulatory vocal SPL or f0 data for later analysis (Holbrook, Rolnick, & Bailey, 1974; Lancioni & Markus, 1999; Lancioni, Markus, & Behrendt, 1998; Lancioni, Van Houten, & Ten Hoopen, 1997; McGillivray, Proctor-Williams, & McLister, 1994). Quantifying long-term effects of biofeedback in real-life settings is important because patients need to learn (or relearn) sustainable healthy vocal motor behaviors instead of only temporary behavior changes observable during therapy sessions or elicited solely in the presence of biofeedback. If patients' vocal motor behavior does not change permanently, then they are still at significant risk for recurrence of their voice/communication disorder (Hillman, Holmberg, Perkell, Walsh, & Vaughan, 1989). Because learning is a latent variable, learning can be indirectly estimated through measures of retention, which represent a relatively permanent change in a motor skill, that is, the accurate execution of a task after feedback is removed or after a period of time without practice (c.f. Schmidt & Lee, 2011). Measures of performance represent short-term effects that reflect how well a skill is executed during practice or in the presence of feedback.

A few studies have attempted to quantify the effect of voice ambulatory biofeedback on vocal motor behavior (KayPENTAX, 2009; Rubow & Swift; 1985; Schalling, Gustafsson, Ternström, Wilen, & Södersten, 2013; Stadelman-Cohen, Van Stan, & Hillman, 2014). However, only one published study (Schalling et al., 2013) provided insight into the specific effects of ambulatory biofeedback on vocal performance and retention without the confounding influences of voice therapy. That pilot study monitored six participants with Parkinson's disease using the VoxLog ambulatory voice monitor (Sonvox AB, Umea, Sweden) to investigate if ambulatory biofeedback delivered via a vibrotactile cue would increase vocal intensity and if the increased vocal intensity would be retained after biofeedback was removed. Results showed a statistically significant increase in mean vocal intensity across the entire group when biofeedback was applied compared to vocal intensity during nonbiofeedback days; however, the increased intensity was not retained after removing the biofeedback. Despite statistically significant differences between biofeedback and nonbiofeedback days, the vocal intensity of participants only differed by a group average of 1.5 dB. This difference may not have been a clinically significant behavioral change (i.e., minimal clinically important change) or even represent a noticeable behavior change (i.e., minimal detectable change; Haley & Fragala-Pinkham, 2006) for two reasons: (a) Typical Parkinson's-focused intervention studies measuring vocal intensity have demonstrated in-clinic dB increases ranging from 5 to > 10 dB post-therapy (c.f. Ramig, Shapir, Fox, & Countryman, 2001), and (b) pilot data from ambulatory monitoring have demonstrated that patients with Parkinson's were approximately 10 dB softer in vocal intensity than age-matched controls (Boudreaux, 2011).

As the existence of an ambulatory biofeedback effect on vocal motor behavior has yet to be clearly demonstrated, the purpose of this initial prospective study is simply to establish if ambulatory biofeedback has a consistent effect on vocal motor behavior in participants with normal voices and determine if any behavioral change is retained after the removal of biofeedback. Such formal proof-of-concept testing is viewed as necessary to establish an empirical basis for future work in this area.

Methods

Participants

Nine adult participants (five women and four men) with normal voices were recruited at the Center for Laryngeal Surgery and Voice Rehabilitation, Massachusetts General Hospital (MGH Voice Center). Upon contacting the MGH Voice Center, participants were screened over the phone by a licensed speech-language pathologist specializing in voice disorders with more than 5 years of experience to ensure they met the following study criteria: (a) a history without vocal difficulties, (b) normal auditory–perceptual voice quality as judged by the speech-language pathologist, (c) consistent voice use for at least 4 consecutive days, (d) an occupation not requiring loud voice use, and (e) no reported history of hearing impairment. One female participant voluntarily withdrew from the study after 1 day of monitoring; another female participant did not phonate enough to obtain reliable measures for biofeedback implementation, and one male participant experienced hardware failure. Thus, six participants (three men and three women, average age 25.2 years, *SD* 4.4 years) successfully completed the study and were assigned identification labels representing their gender and study number (i.e., F1, F2, F3, M1, M2, and M3). Occupations of those who completed the study included a Ph.D.-level research engineer, three master's-level speech pathology students, and two administrative assistants. Participants were financially compensated for every day that they participated in the study, and the study protocol was approved by the institutional review board at Massachusetts General Hospital.

Data Acquisition Platform

The Ambulatory Phonation Monitor (APM Model 3200; KayPENTAX, Montvale, NJ) was used to acquire all voicing data throughout this study. The APM attaches a miniature accelerometer (ACC) via double-sided tape at the base of the neck above the sternal notch to sense phonation (Cheyne, Hanson, Genereux, Stevens, & Hillman, 2003). The ACC is connected to the APM system box that is worn in a satchel around the waist. The thin wire connecting the ACC and APM system box was run underneath clothing to prevent it from being accidentally caught and broken.

Effort was made to place the ACC on the same spot of the neck every day to facilitate consistent monitoring of vocal intensity across days.

A calibration procedure using a microphone to sense oral sound pressure as a reference was completed each morning prior to each day of monitoring. The calibration procedure required each participant to produce a loudness glide on an /a/ vowel to establish the relationship between the amplitude of the ACC signal and the oral sound pressure level so that acoustic SPL could be estimated from skin acceleration level (Švec, Titze, & Popolo, 2005). Participants were required to meet one of the investigators every morning of monitoring to complete this system calibration for SPL and ensure the quality of the previous day's data.

The APM real-time processing settings are hard-coded. To be specific, the device records values for vocal intensity (dB SPL) and f0 (Hz) over nonoverlapping frames of 50 ms in duration. Each 50-ms frame is divided into two 25-ms subframes, and the entire frame is considered voiced if the levels of both subframes exceed 5 dB (uncalibrated). SPL is then recomputed over the entire frame using the SPL calibration factors. f0 for each voiced frame is equal to the reciprocal of the first peak location in the normalized autocorrelation function when the peak amplitude exceeds a threshold of 0.25. Otherwise, frames are considered unvoiced (e.g., silence, unvoiced speech, or other nonspeech energy), and SPL and f0 values are set to zero. In addition, the APM is hard-coded with a lower f0 cutoff of 60 Hz, meaning that the frame was set to zero SPL and f0 if the reciprocal of the autocorrelation peak was 60 Hz or less. The data were filtered further according to individualized lower limits of SPL and f0. The individualized SPL lower limit was determined as the point at which the left tail of the SPL histogram became level with the noise floor. The individualized f0 lower limit was determined as the lowest f0 histogram bin registering 1% of the total phonation. There was no upper SPL limit, and the upper f0 limit was fixed at 500 Hz across all participants.

Study Design

The objective of this study was to analyze the effect of biofeedback on SPL changes compared to each participant's baseline vocal behavior. Participants wore the APM for 7 days. During the first 3 days, participants wore the APM to yield baseline vocal intensity levels without any biofeedback applied. During the subsequent 4 days, ambulatory biofeedback was provided on the first and third days. When biofeedback was triggered, it consisted of a vibrotactile cue that was provided by a pager vibrator worn on the belt and connected via a wire to the APM system box. Ambulatory biofeedback was designed to reduce the vocal intensity of participants by providing cues when their SPL estimate was above a participant-specific threshold.

Figure 1 shows how the biofeedback threshold was personalized and applied. Figure 1a illustrates how the threshold (intensity level above which cueing occurred) was determined by adding 5 dB to the mean baseline intensity level (derived from data pooled across all three baseline days). This threshold definition was empirically determined and allowed the participant adequate intensity for functional oral communication and reduced the likelihood of participant frustration and annoyance. On average, a threshold 5 dB above the participant's mean SPL provided access to the lower 66% of their baseline vocal intensity range (SPL at or below the threshold) and restricted/modified access to the upper 33% of their baseline vocal intensity range (SPL above the threshold).

Figure 1b depicts the interaction between participant vocal intensity and the biofeedback threshold. Ambulatory biofeedback consisted of a 250-ms vibrotactile cue when the estimated SPL of the participant was greater than the biofeedback threshold for 500 ms of continuous phonation. This duration was determined in previous case studies in which biofeedback appeared to trigger too often below 500 ms of continuous phonation and not frequently enough above 500 ms of continuous phonation (KayPENTAX, 2009; Stadelman-Cohen et al., 2014). Also, the duration of 500 ms allowed participants access to vocal intensities greater than 5 dB above their mean for short bursts of time; that is, they were able to phonate at an SPL above the biofeedback threshold without receiving a cue if the duration was less than 500 ms. The duration of the vibrotactile cue was set to 250 ms, which was shown in trial-and-error testing and previous case studies to minimize overlapping of cues (KayPENTAX, 2009; Stadelman-Cohen et al., 2014).

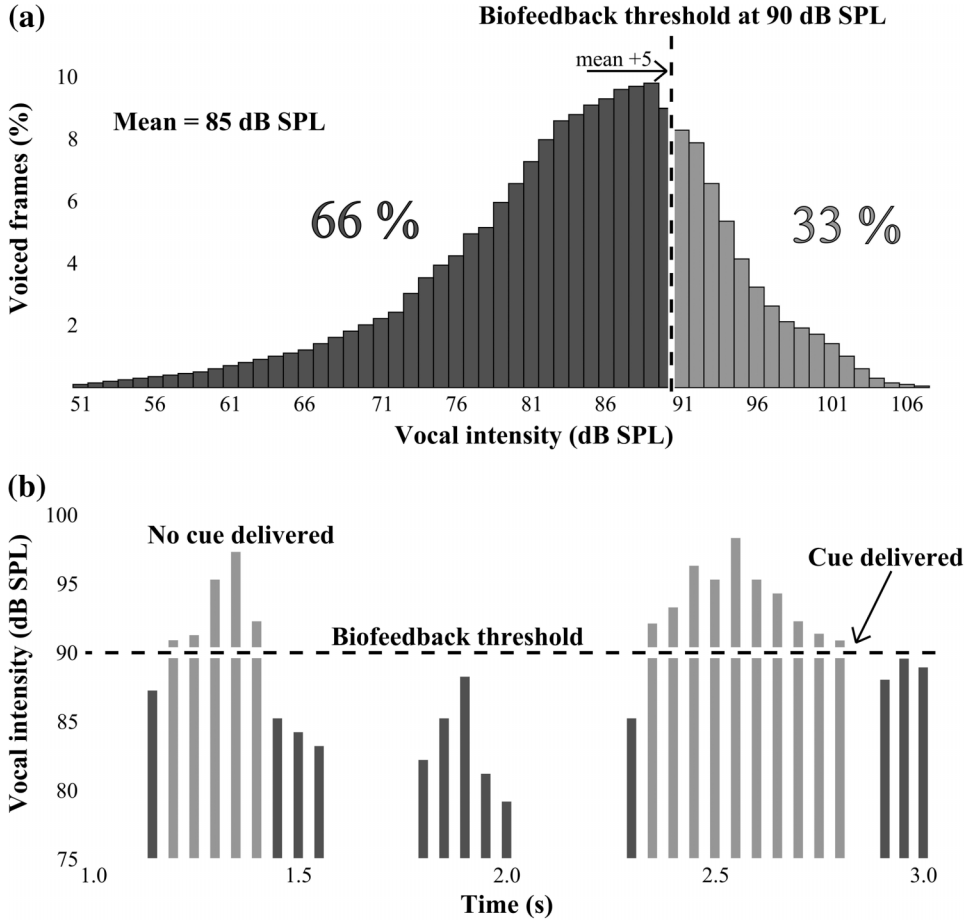
Data Postprocessing

Each morning following a day of monitoring, participants met with one of the investigators to upload data, discuss their activities from the previous day, and document those activities in a daily activity log. The daily activity log was used to screen data for confounding situations in which the participant would have been unable to remain below the SPL threshold for a prolonged period due to extenuating circumstances (e.g., attending a rock concert, socializing in a noisy restaurant, and talking on the telephone on public transportation).

Because the SPL calibration process has a known variability of ± 6–10 dB (Švec et al., 2005), the calibration factors (multiplier and offset) were averaged across all days for each participant to obtain a more stable estimate of vocal intensity. Upon applying average calibration factors, one of M2's baseline days significantly decreased in volume (67.5 dB SPL to 54.5 dB SPL). According to field notes, the ACC was placed slightly inferior on his neck during baseline Day 2 due to skin irritation from the previous day. Because placing the ACC further inferior from the larynx would have caused a decrease in overall skin acceleration level, this baseline day for M2 was removed from analysis.

The resulting daily data files were used to obtain traditional measures of mean intensity (dB SPL) and compliance (percentage of voiced frames exhibiting vocal intensity less than or equal to the SPL threshold). SPL data were

Figure 1. Illustration of the biofeedback threshold definition and the timing of cues delivered. (a) Example baseline histogram of vocal intensity derived from the Ambulatory Phonation Monitor (APM) with a biofeedback threshold defined at 5 dB above the mean intensity of 85 dB SPL. With this definition, approximately 66% of voiced frames (dark gray) tended to lie below the threshold, and approximately 33% of voiced frames (light gray) tended to lie above the threshold. (b) An example sequence of APM frames during a biofeedback day demonstrates two scenarios when vocal intensity crosses the biofeedback threshold: No cue is delivered when vocal intensity exceeds the threshold for only four consecutive frames (200 ms) whereas a vibrotactile cue is delivered when vocal intensity exceeds the threshold for 10 consecutive frames (500 ms).



both pooled across days within each monitoring period (baseline, biofeedback, and postbiofeedback) and within biofeedback and postbiofeedback days individually. Because the SPL biofeedback threshold was individualized for each participant according to his or her vocal behavior on baseline days, daily mean SPL data for the four post-baseline days were normalized by subtracting SPL values from average baseline SPL levels. The normalized mean vocal intensity values did not produce different statistical results from non-normalized vocal intensity values; however, the normalized values are used in the statistical analyses for conceptual consistency.

Statistical Analysis

Statistics included two within-group repeated-measure analysis of variance (RM-ANOVA) tests comparing the normalized mean vocal intensity and compliance among three levels (pooled baseline, biofeedback, and postbiofeedback days) and two within-group RM-ANOVA tests comparing normalized mean vocal intensity and compliance among five levels (pooled baseline, individual biofeedback, and individual postbiofeedback days). Due to multiple hypotheses tested, Bonferroni corrections modified the α to .025 in the three-level RM-ANOVAs and to .015 in the five-level RM-ANOVAs. When statistical significance was found in an RM-ANOVA, post hoc testing was completed with paired *t* tests. Consistent with the performance-retention paradigm (Salmoni, Schmidt, & Walter, 1984), all post hoc comparisons were in reference to the baseline monitoring period. This is because performance scores during biofeedback days theoretically represent both temporary changes from the biofeedback and permanent changes from learning, and the postbiofeedback days would only represent learning-related changes. Therefore direct comparisons between the two conditions would be confounding (Schmidt & Lee, 2011;

Winstein & Schmidt, 1990). All statistics were calculated using SPSS (version 22.0, IBM, Armonk, NY).

Some participants registered more voiced frames than others during biofeedback days, resulting in uneven exposure to biofeedback across the participants. To ensure that the percentage compliance values for postbiofeedback days were not affected, a correlation was completed between the number of voiced frames recorded during a biofeedback day and the subsequent percentage compliance from the postbiofeedback day.

Results

Figure 2a summarizes the three-level RM-ANOVA results. The results of the first three-level RM-ANOVA indicated a statistically significant change in normalized mean intensity across monitoring periods, $F(2, 10) = 17.61$, $\eta^2 = .78$, $p = .001$, and the post hoc testing demonstrated a statistically significant decrease of 4.4 dB during biofeedback days compared to baseline days ($p = .005$). The results of the second three-level RM-ANOVA indicated a significant change in compliance across monitoring periods, $F(2, 10) = 22.61$, $\eta^2 = .82$, $p < .001$, and the post hoc testing demonstrated a statistically significant increase of 16.8 percentage points for compliance during biofeedback days compared to baseline days ($p = .003$). Normalized mean vocal intensity and compliance were statistically identical during baseline and postbiofeedback days.

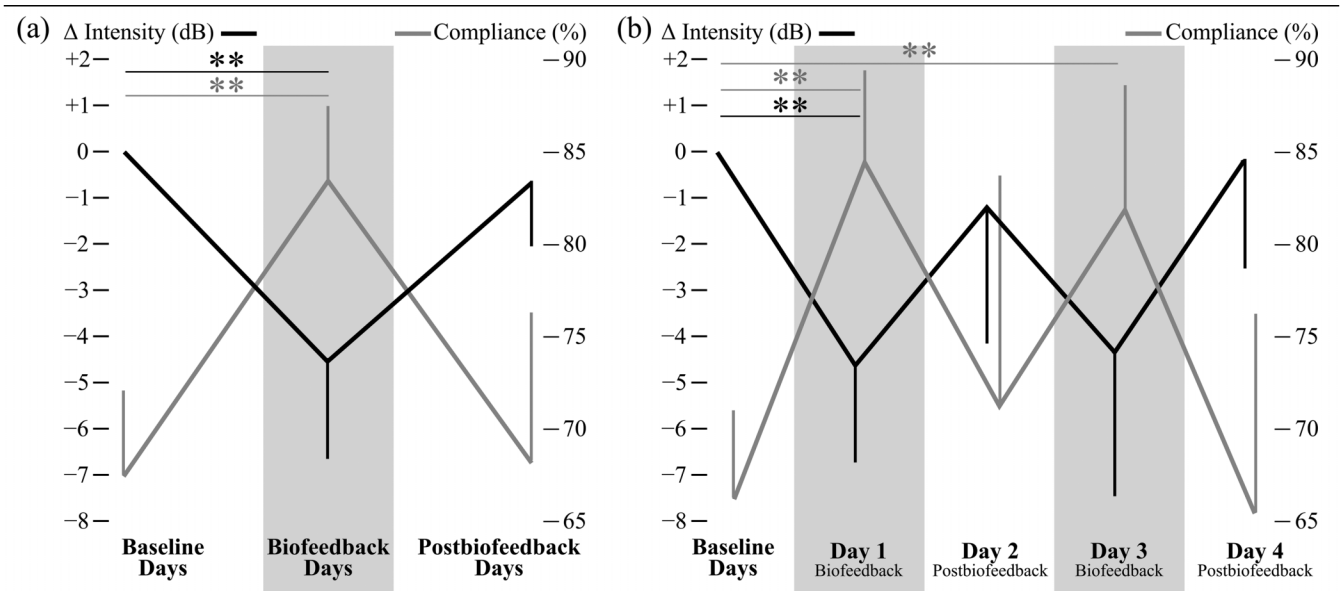
Figure 2b summarizes the five-level RM-ANOVA results. The results of the first five-level RM-ANOVA indicated a statistically significant change in normalized mean intensity across monitoring periods, $F(4, 20) = 5.39$, $\eta^2 = .52$, $p = .004$, and the post hoc testing demonstrated a statistically significant decrease of 4.5 dB during Day 1 (biofeedback) compared to baseline days ($p = .004$). There was a large decrease of 4.3 dB during Day 3 (biofeedback) compared to baseline days; however, the significance did not meet the modified α of .015 ($p = .02$). The results of the second five-level RM-ANOVA indicated a statistically significant change in compliance across monitoring, $F(4, 20) = 9.09$, $\eta^2 = .65$, $p < .001$, and the post hoc testing demonstrated a statistically significant increase compared to baseline days for compliance of 17.8 percentage points during Day 1 biofeedback ($p = .003$) and 16.3 percentage points during Day 3 biofeedback ($p = .007$). Normalized mean vocal intensity and compliance were statistically identical during baseline and postbiofeedback days.

There was no statistically significant correlation between the number of voiced frames during biofeedback days and the percentage compliance of the following postbiofeedback day ($r = .076$, $p = .836$). Two of the 12 points were outliers ($|z\text{-score}| > 2$ on one or both axes). Last, note that each subject's individual SPL mean, SPL mode, SPL standard deviation, and percentage compliance across monitoring periods can be found in Appendix A and Appendix B.

Discussion

One objective of this study was to establish if ambulatory biofeedback could have a consistent effect on a target vocal motor behavior (in this case, decreased average vocal intensity) during the daily life of participants with normal

Figure 2. Ambulatory biofeedback effect using group-based averages of mean vocal intensity difference (Δ) relative to baseline and compliance (a) pooled within monitoring periods and (b) pooled within individual biofeedback and postbiofeedback days. Error bars represent *SD* (paired *t* test: ** $p < .01$).



voices. Large effect sizes for statistically significant changes in vocal motor behavior were demonstrated during biofeedback days compared to baseline days for the reduction of mean intensity and increase in compliance. The group statistics demonstrated a significantly different vocal motor behavior (softer phonation) during biofeedback days compared to nonbiofeedback days. This finding is even more encouraging when one considers that the SPL calibration process is inherently variable (Švec et al., 2005), which in turn increases the variability of the resulting data and biases against finding significant differences between experimental conditions.

The second objective of this study was to determine if the vocal motor behavior change was retained after ambulatory biofeedback was removed. Group statistics demonstrated no significant difference in normalized mean intensity and compliance when comparing baseline days to postbiofeedback days. This finding appears to indicate that the change in behavior exhibited on biofeedback days was not retained on days when biofeedback was removed. This interpretation is also supported by the statistical trends in Figure 2b that demonstrated washout patterns across alternating biofeedback and postbiofeedback days for normalized mean intensity and compliance, respectively. It is interesting to note that on Day 3 (biofeedback), although normalized mean vocal intensity was close to being statistically different from baseline vocal intensity ($p = .02$), percentage compliance increased to a statistically significant degree. This was due to some subjects using a "burst-like" voicing pattern at higher vocal intensities (voicing occurred above the intensity threshold for less than 500 consecutive milliseconds) to avoid the biofeedback vibrotactile cue. This can most clearly be seen with F3's and M3's Day 3 in Appendix B; percentage compliance minimally increased, and mean SPL did not change when compared to Day 2. Therefore, compared to baseline, the group's normalized mean SPL did not significantly decrease, and percentage compliance significantly increased on Day 3.

Many studies in motor learning have demonstrated that high-frequency (e.g., providing feedback 100% of the time—or nearly every time—a person exceeds the threshold) and immediate feedback schedules may significantly increase performance of a motor skill when compared to low-frequency feedback and delayed-feedback schedules. However, lower-frequency and delayed-feedback schedules have been shown to produce significantly higher long-term retention after feedback removal than immediate and frequent-feedback schedules (Lavery & Suddon, 1962; Lee, White, & Carnahan, 1990; Salmoni et al., 1984; Schmidt & Lee, 2011). The feedback used in this study was meant to closely approximate a high-frequency and immediate type of schedule, which appears to have been accomplished given the demonstration of the type of performance and retention results that would be expected for this type of feedback schedule. Strictly speaking, however, the APM feedback schedule was not at 100% nor was it immediate because of the requirement that phonatory segments had to exceed the biofeedback threshold for 500 ms to trigger feedback; that is, phonation that exceeded the threshold for less than 500 ms did not trigger feedback.

Although the number of days within each monitoring period was maintained across participants, the amount of phonation per day was not controlled. Therefore, some participants had more exposure to biofeedback than others, which could have confounded the results of retention. However, a correlation between the total number of voiced frames during biofeedback and compliance during postbiofeedback was not found to be statistically significant, supporting that the length of time exposed to ambulatory biofeedback did not significantly affect the retention-related results in this study. Nonetheless, controlling for total voicing duration is a desirable study design for future ambulatory biofeedback studies.

Because the APM does not record the raw neck skin acceleration signal, there is always the possibility of non-voice-related sounds being included in the analysis, which would have most significantly affected low intensity levels. As the focus of this study was to decrease high-intensity voiced frames, this potential noise source had minimal or no effects. Future studies using voice monitors that record the raw signal would allow for more sophisticated voice activity detection and improved data quality. The small sample size is a limitation as usual, but encouraging results were obtained.

This initial proof-of-concept study provides support for empirically testing more sophisticated types of ambulatory biofeedback characteristics and their effect on the performance and retention of a vocal motor behavior. Using a recently developed ambulatory monitoring application on a smartphone-based platform (Ghassemi et al., 2014; Mehta, Zañartu, Feng, Cheyne, & Hillman, 2012), sophisticated ambulatory biofeedback algorithms and schedules may be developed to evaluate how to provide ambulatory biofeedback in the most effective manner (i.e., resulting in the highest degree of retention after short-term exposure to biofeedback). Some examples of more sophisticated ambulatory biofeedback include (a) delayed feedback through summary statistics upon participant request or after a predetermined amount of time; (b) lower-frequency feedback by providing biofeedback in a staggered manner (e.g., every second, third, or fourth time—50%, 33%, and 25%, respectively) instead of every time the threshold is passed; (c) "fading" of the biofeedback frequency and timing over consecutive days (e.g., Day 1 = 100% biofeedback, Day 2 = 25% biofeedback, and Day 3 = one-time summary feedback at end of day); (d) the incorporation of probability-based algorithms that can predict when a feedback threshold is inappropriate for specific situations or how strongly the participant is reacting to the biofeedback cues; (e) the development of more sensitive measures of error during ambulatory biofeedback, such as constant error, variable error, and total variability (Chapanis, 1951; Henry, 1975); and (f) the development of measures specifically sensitive to biofeedback cues that have multiple parameters (e.g., time and SPL; Hancock, Butler, & Fischman, 1995). Item (f)

may be especially salient because the traditional measures of mean intensity and compliance used in the present study rely solely upon SPL estimates and do not take into account how many continuous phonatory frames occurred above the SPL threshold to trigger a biofeedback cue. For example, there were times when the participants phonated above the intensity threshold for 50–450 continuous milliseconds without being cued, which increased mean SPL and decreased compliance levels during biofeedback even though participants were being successfully compliant (i.e., they did not trigger the vibrotactile cue).

Overall, this study indicates that ambulatory biofeedback may be a useful clinical tool for enacting a temporary vocal motor behavior change during the context of an individual's daily life. Future work warrants the testing of innovative ambulatory biofeedback approaches on the basis of motor control and learning theories to improve retention of a desired vocal motor behavior.

Acknowledgments

This work was supported by the Voice Health Institute and the National Institutes of Health (NIH) National Institute on Deafness and Other Communication Disorders under Grants R33 DC011588 and F31 DC014412. The article's contents are solely the responsibility of the authors and do not necessarily represent the official views of the NIH.

References

Boudreaux, D. M. (2011). *Using the Ambulatory Phonation Monitor to measure the vocal parameters of older people with and without Parkinson's disease* (Unpublished master's thesis). Louisiana State University, Baton Rouge, LA.

Chapanis, A. (1951). Theory and methods for analyzing errors in man-machine systems. *Annals of the New York Academy of Sciences, 51,* 1179–1203.

Cheyne, H., Hanson, H., Genereux, R., Stevens, K., & Hillman, R. E. (2003). Development and testing of a portable vocal accumulator. *Journal of Speech, Language, and Hearing Research, 46,* 1457–1467.

Cohen, N. J., & Squire, L. R. (1980, October 10). Preserved learning and retention of pattern-analyzing skill in amnesia: Dissociation of knowing how and knowing that. *Science, 210,* 207–210.

Ferrand, C. T. (1995). Effects of practice with and without knowledge of results on jitter and shimmer levels in normally speaking women. *Journal of Voice, 9,* 419–423.

Ghassemi, M., Van Stan, J. H., Mehta, D. D., Zañartu, M., Cheyne, H. A., Hillman, R. E., & Guttag, J. (2014). Learning to detect vocal hyperfunction from ambulatory neck-surface acceleration features: Initial results for vocal fold nodules. *IEEE Transactions on Biomedical Engineering, 61,* 1668–1675.

Haley, S. M., & Fragala-Pinkham, M. A. (2006). Interpreting change scores of tests and measures used in physical therapy. *Physical Therapy, 86,* 735–743.

Hancock, G. R., Butler, M. S., & Fischman, M. G. (1995). On the problem of two-dimensional error scores: Measures and analyses of accuracy, bias, and consistency. *Journal of Motor Behavior, 27,* 241–250.

Henry, F. M. (1975). Absolute error vs "E" in target accuracy. *Journal of Motor Behavior, 7,* 227–228.

Hillman, R. E., Holmberg, E. B., Perkell, J. S., Walsh, M., & Vaughan, C. (1989). Objective assessment of vocal hyperfunction: An experimental framework and initial results. *Journal of Speech and Hearing Research, 32,* 373–392.

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,* 298–303.

KayPENTAX. (2009). *Ambulatory Phonation Monitor: Applications for Speech and Voice*. Lincoln Park, NJ: KayPENTAX.

Lancioni, G. E., & Markus, S. (1999). A deaf woman learning to control her excessive vocal loudness through a portable feedback system. *Perceptual and Motor Skills, 88,* 1347–1349.

Lancioni, G. E., Markus, S., & Behrendt, M. (1998). A portable vibratory-feedback device for reducing excessive vocal loudness. *Behavioural and Cognitive Pyschotherapy, 26,* 371–376.

Lancioni, G. E., Van Houten, K., & Ten Hoopen, G. T. (1997). Reducing excessive vocal loudness in persons with mental retardation through the use of a portable auditory-feedback device. *Journal of Behavioral Therapy & Experimental Psychiatry, 28,* 123–128.

Lavery, J. J., & Suddon, F. H. (1962). Retention of simple motor skills as a function of the number of trails by which KR is delayed. *Perceptual and Motor Skills, 15,* 231–237.

Lee, T. D., White, M. A., & Carnahan, H. (1990). On the role of knowledge of results in motor learning: Exploring the guidance hypothesis. *Journal of Motor Behavior, 22,* 191–208.

Maas, E., Robin, D. A., Austermann Hula, S. N., Freedman, S. E., Wulf, G., Ballard, K. J., & Schmidt, R. A. (2008). Principles of motor learning in treatment of motor speech disorders. *American Journal of Speech-Language Pathology, 17,* 277–298.

Maryn, Y., De Bodt, M., & Van Cauwenberge, P. (2006). Effects of biofeedback in phonatory disorders and phonatory performance: A systematic literature review. *Applied Psychophysiology and Biofeedback, 31,* 65–83.

McGillivray, R., Proctor-Williams, K., & McLister, B. (1994). Simple biofeedback device to reduce excessive vocal intensity. *Medical & Biological Engineering and Computing, 32,* 348–350.

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,* 3090–3096.

Ramig, L. O., Shapir, S., Fox, C., & Countryman, S. (2001). Changes in vocal loudness following intensive voice treatment (LSVT) in individuals with Parkinson's disease: A comparison with untreated patients and normal age-matched controls. *Movement Disorders, 16,* 79–83.

Rubow, R., & Swift, E. (1985). A microcomputer-based wearable biofeedback device to improve transfer of treatment in Parkinsonian dysarthria. *Journal of Speech and Hearing Disorders, 50,* 178–185.

Salmoni, A. W., Schmidt, R. A., & Walter, C. B. (1984). Knowledge of results and motor learning: A review and critical reappraisal. *Psychological Bulletin, 95,* 355–386.

Schalling, E., Gustafsson, J., Ternström, S., Wilen, F. B., & Södersten, M. (2013). Effects of tactile biofeedback by a portable voice accumulator on voice sound level in speakers with Parkinson's disease. *Journal of Voice, 27,* 729–737.

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

Stadelman-Cohen, T., Van Stan, J. H., & Hillman, R. E. (2014). Use of ambulatory biofeedback to supplement traditional voice therapy for treating primary MTD in an adult female.

In J. Stemple & E. Hapner (Eds.), *Voice therapy: Clinical case studies* (4th ed.) (pp. 157–163). San Diego, CA: Plural.

Steinhauer, K., & Grayhack, J. P. (2000). The role of knowledge of results in performance and learning of a voice motor task. *Journal of Voice, 14,* 137–145.

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,* 1386–1394.

Van Stan, J. H., Gustafsson, J., Schalling, E., & Hillman, R. E. (2014). Direct comparison of three commercially available devices for voice ambulatory monitoring and biofeedback. *Perspectives on Voice and Voice Disorders, 24*(2), 80–86.

Winstein, C. J., & Schmidt, R. A. (1990). Reduced frequency of knowledge of results enhances motor skill learning. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 16,* 677–691.

Wong, A. Y.-H., Ma, E. P.-M., & Yiu, E. M.-L. (2011). Effects of practice variability on learning of relaxed phonation in vocally hyperfunctional speakers. *Journal of Voice, 25,* e103–e113.

Yiu, E. M.-L., Verdolini, K., & Chow, L. P.-Y. (2005). Electromyographic study of motor learning for a voice production task. *Journal of Speech, Language, and Hearing Research, 48,* 1254–1268.

Appendix A

Summary statistics for each participant's vocal intensity data (dB SPL), including mean (*M*), mode, and standard deviation (*SD*)

Subject ID	Statistic	Baseline	Day 1	Day 2	Day 3	Day 4
F1	*M*	68	65	66	67	68
	Mode	69	65	63	65	67
	SD	10	9	10	10	10
F2	*M*	69	62	73	63	67
	Mode	66*	63	57*	68	57*
	SD	16	11	19	11	14
F3	*M*	68	61	65	65	67
	Mode	71*	63	68	65	66
	SD	9	7	9	8	9
M1	*M*	75	72	76	70	73
	Mode	78	76	78	66	74
	SD	13	10	12	9	11
M2	*M*	68	63	62	58	67
	Mode	68	60*	64	58	68
	SD	13	10	10	9	12
M3	*M*	68	66	66	66	71
	Mode	68	66	66	68	73
	SD	11	9	10	10	12

Note. Baseline statistics are reported over the three baseline days combined. Subject identification (ID) indicates gender of participant as female (F) or male (M).

*Bimodal distribution observed.

Appendix B

Ambulatory biofeedback effect demonstrated within each participant using averages of mean vocal intensity and compliance for (a) all three female participants (F1, F2, and F3) and (b) all three male participants (M1, M2, and M3). Baseline statistics are reported over the three baseline days combined.