JSLHR

Research Article

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

Jarrad H. Van Stan,[a,b,c] Daryush D. Mehta,[a,b,c] Dagmar Sternad,[d]
Robert Petit,[e] and Robert E. Hillman[a,b,c]

Purpose: Ambulatory voice biofeedback has the potential to significantly improve voice therapy effectiveness by targeting carryover of desired behaviors outside the therapy session (i.e., retention). This study applies motor learning concepts (reduced frequency and delayed, summary feedback) that demonstrate increased retention to ambulatory voice monitoring for training nurses to talk softer during work hours.
Method: Forty-eight nurses with normal voices wore the Voice Health Monitor (Mehta, Zañartu, Feng, Cheyne, & Hillman, 2012) for 6 days: 3 baseline days, 1 biofeedback day, 1 short-term retention day, and 1 long-term retention day. Participants were block-randomized into 3 different biofeedback groups: 100%, 25%, and Summary. Performance was measured in terms of compliance time below a participant-specific vocal intensity threshold.
Results: All participants exhibited a significant increase in compliance time (Cohen's $d = 4.5$) during biofeedback days compared with baseline days. The Summary feedback group exhibited statistically smaller performance reduction during both short-term ($d = 1.14$) and long-term ($d = 1.04$) retention days compared with the 100% feedback group.
Conclusions: These findings suggest that modifications in feedback frequency and timing affect retention of a modified vocal behavior in daily life. Future work calls for studying the potential beneficial impact of ambulatory voice biofeedback in participants with behaviorally based voice disorders.

Functional voice use is essential for most activities of daily living, such as communication and occupational demands (Roy, Merrill, Gray, & Smith, 2005). However, 3%–9% of the population is estimated to have impaired vocal function, resulting in limited ability to participate in family, community, and economic activities (Ramig & Verdolini, 1998). Furthermore, the most common voice disorders are related to vocal behavior (Bhattacharyya, 2014). Behaviorally based voice disorders (e.g., vocal fold nodules, polyps, and muscle tension dysphonia) can be difficult to manage solely in the clinical setting because they are thought to result from faulty and/or abusive patterns of vocal behavior exhibited in daily life (Czerwonka, Jiang, & Tao, 2008; Hillman, Holmberg, Perkell, Walsh, & Vaughan, 1989; Van Stan, Mehta, Zeitels, et al., 2015). Voice therapists must typically rely on unreliable patient self-reporting in assessing the contribution of voice use to the patient's voice disorder, as well as to determine whether vocal function is improving outside the therapy session. Ambulatory voice monitoring and biofeedback can both record objective measures of vocal behavior and deliver real-time therapeutic cues during a patient's daily life (Van Stan, Gustafsson, Schalling, & Hillman, 2014). This suggests that ambulatory voice monitoring and biofeedback have significant potential to improve voice therapy effectiveness through targeting the carryover of desired vocal behaviors outside the therapy session and documenting voice therapy compliance in the patient's natural environment.

---

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

Correspondence to Jarrad H. Van Stan: jvanstan@mghihp.edu

Editor: Julie Liss
Associate Editor: Tanya Eadie

Received April 22, 2016
Revision received July 14, 2016
Accepted October 8, 2016
https://doi.org/10.1044/2016_JSLHR-S-16-0164

Disclosure: *The authors have declared that no competing interests existed at the time of publication.*

Only very few studies have attempted to quantify the effect of ambulatory voice biofeedback on vocal motor behavior, and most of these have been case studies with little experimental control (KayPENTAX, 2009; Rubow & Swift, 1985; Stadelman-Cohen, Van Stan, & Hillman, 2014; Van Stan et al., 2017). To date, only two small group-based studies have assessed the effect of ambulatory voice biofeedback on vocal behavior in daily life, and both used simple vocal intensity thresholds (Schalling, Gustafsson, Ternstrom, Bulukin Wilen, & Sodersten, 2013; Van Stan, Mehta, & Hillman, 2015). Both of these studies reported a vocal behavior change only when the biofeedback was present and no indication that the modified loudness behavior remained after biofeedback removal. The absence of a maintained behavior change is significant because successful vocal rehabilitation depends upon a permanent, not temporary, improvement in voice use. Therefore, no indication of maintenance represents a lack of clinical effectiveness.

Potential methods to improve the permanence of a vocal behavior modification may be found in the fields of motor learning and motor control, where empirical studies have demonstrated advantages of various types of feedback schedules for the relative permanence of a motor performance improvement (Schmidt & Lee, 2011). The motor learning literature differentiates between temporary and permanent changes in motor behaviors with the respective labels of "performance improvement" versus "learning." *Performance improvement* refers to the improved execution of a movement during practice; *learning* or *retention* is defined defined as the relatively permanent improvement in a motor skill after a period of time without practice. Different underlying neurophysiological processes are presumed to underlie these temporary or relatively permanent changes (Dayan & Cohen, 2011).

Many motor learning studies have demonstrated an increase in learning (by measuring retention) when decreasing the frequency and manipulating the timing of feedback. For example, providing feedback every fourth trial leads to more effective learning than feedback after every trial, also called 25% or 100% feedback, respectively (Badets & Blandin, 2004; Lee, White, & Carnahan, 1990; Salmoni, Schmidt, & Walter, 1984; Sparrow & Summers, 1992). In addition, delaying the presentation of feedback by providing average summary statistics at predetermined intervals typically has beneficial effects (Anderson, Magill, Sekiya, & Ryan, 2005; Salmoni et al., 1984; Schmidt, Lange, & Young, 1990; Yao, Fischman, & Wang, 1994). However, many of these studies were completed in highly controlled laboratory environments using simple limb-based tasks such as pointing or reaching. The present study investigated the application of these motor learning concepts in a more complex and clinically relevant context: modifying habitual vocal behavior during daily life.

Empirical support for the clinical use of reduced or delayed feedback in voice therapy is currently lacking because most studies have focused on limb movements, not voice and speech-related movements of head/neck structures (Schmidt & Lee, 2011). Furthermore, the few studies focused on how feedback frequency/delay affects speech or vocal learning have not consistently replicated limb-related results (Bislick, Weir, Spencer, Kendall, & Yorkston, 2012; Maas et al., 2008). When the movement to be learned was a speech task, three studies using group-based comparisons demonstrated increased retention when feedback frequency was reduced. The experimental condition of 20% feedback improved retention 2 days postpractice compared with 100% feedback when targeting phrase-length accuracy for participant groups without motor deficits (Adams & Page, 2000) and with Parkinson's disease (Adams, Page, & Jog, 2002), and also improved retention 1 week postpractice when targeting novel Korean phrases in English-speaking individuals (Kim, LaPointe, & Stierwalt, 2012). However, 60% feedback inconsistently improved retention 2–4 weeks posttreatment compared with 100% feedback when targeting phonemes with single-subject designs in individuals with apraxia of speech; more specifically, four of eight patients showed better retention with reduced-frequency feedback (Austermann Hula, Robin, Maas, Ballard, & Schmidt, 2008; Maas, Butalla, & Farinella, 2012). One study resulted in negative findings: No consistent retention advantage was shown between 50% and 100% feedback conditions when targeting phonemes with an individual with apraxia of speech (Katz, McNeil, & Garst, 2010).

In regard to voice-specific learned movements, only one study showed a retention advantage 1 day postpractice for low-frequency feedback (no feedback or 0%, and 50% feedback) compared with 100% feedback when targeting sustained vowel nasalization in participant groups with normal voices (Steinhauer & Grayhack, 2000). One study showed less performance degradation in jitter values 1 week posttraining in a group with normal voices who received no feedback (0%) compared with a 100% feedback group (Ferrand, 1995). One study showed no retention advantage for a summary/delayed feedback condition compared with a real-time feedback condition when targeting phrase-length fundamental frequency contours in participant groups with no voice or speech deficits (Weltens & De Bot, 1984). These conflicting findings point to the need for additional investigations into whether the results of limb-based studies apply to the head and neck structures involved in voice and speech production.

In addition, the cortico-bulbar (head/neck control) and cortico-spinal (core and limb control) sensorimotor systems differ with regard to several anatomical and physiological factors: (a) Many head/neck structures have bilateral cortical input (Simonyan & Horwitz, 2011), whereas limbs predominantly have contralateral cortical input (Kandel, Schwartz, & Jessell, 2000); (b) limbs contain a gamma neuronal system crucial for load bearing and angle sensation (Kandel et al., 2000), whereas gamma neurons have yet to be observed in many head/neck muscles (Brandon et al., 2003); (c) vocal neural circuits are tightly interconnected with respiratory brainstem nuclei (Nishino, 2012); and (d) modification of vocal behavior inherently involves minimal or no visual feedback (i.e., it is not thought that the sensorimotor constituents of phonation routinely engage vision-related neural pathways). Therefore, generalization of motor learning principles to speech- and voice-related movements is a nontrivial endeavor requiring careful empirical study.

Commercial and research ambulatory voice monitors equipped with biofeedback capabilities use routines that provide immediate 100% feedback based on amplitude or fundamental frequency targets (Holbrook, Rolnick, & Bailey, 1974; McGillivray, Proctor-Williams, & McLister, 1994; Van Stan et al., 2014). In order to implement motor learning study designs with ambulatory voice biofeedback schedules, new capabilities were developed for a smartphone-based software application called the Voice Health Monitor (VHM; Llico et al., 2015; Mehta, Zañartu, Feng, Cheyne, & Hillman, 2012). More specifically, the VHM was programmed to directly control modifications to feedback frequency and summary presentation via vibrotactile cues and user interaction with a smartwatch.

The purpose of this study was to systematically assess the effects of modifying the frequency and timing of ambulatory voice biofeedback on retention of a modified vocal behavior (reduced vocal intensity) in participants with normal voices who responded to their respective biofeedback condition. Participants with normal voices have been chosen in order to clearly observe the biofeedback effect without confounding influences (e.g., concomitant voice therapy, abnormal vocal anatomy) that could potentially be introduced by including participants with various types of voice disorders. It was hypothesized that the group who received 100% immediate biofeedback would perform with the highest accuracy (i.e., remain below the vocal intensity threshold for the highest percentage of time) during biofeedback, but that this improved performance would not be retained (i.e., return to baseline vocal intensity levels after feedback removal). In contrast, the other two groups receiving reduced frequency feedback (25% immediate feedback) and summary feedback (summary statistics at defined intervals) were expected to maintain their improved performance after the ambulatory voice biofeedback was removed.

Method

Participants

All participants were registered nurses (RNs) recruited from intensive care or step-down units. This population was targeted for recruitment because of an occupational demand from hospital-initiated quality improvement programs, which ask nursing staff to reduce their vocal intensity while on the units. Both the World Health Organization and the Joint Commission on Accreditation of Healthcare Organizations provide recommendations on environmental noise levels in acute care units because high environmental noise is associated with negative physical and psychological effects for patients (Mazer, 2006, 2012; Nightingale, 1860; World Health Organization, 2009). The most controllable aspect of environmental noise is how loud nurses talk (Kahn et al., 1998), but modifying the volume of nursing conversations has been difficult and often unsuccessful (Kol, Demircan, Erdoğan, Gencer, & Erengin, 2015; Konkani, Oakley, & Penprase, 2014; Qutub & El-Said, 2009; Taylor-Ford, Catlin, LaPlante, & Weinke, 2008). The normal status of all participants was verified via video-stroboscopic laryngeal examination after passing an initial interview demonstrating a history without vocal difficulties, normal auditory-perceptual voice quality as judged by a speech-language pathologist, and no reported history of hearing impairment. The Institutional Review Board of Partners HealthCare System at Massachusetts General Hospital approved all study procedures.

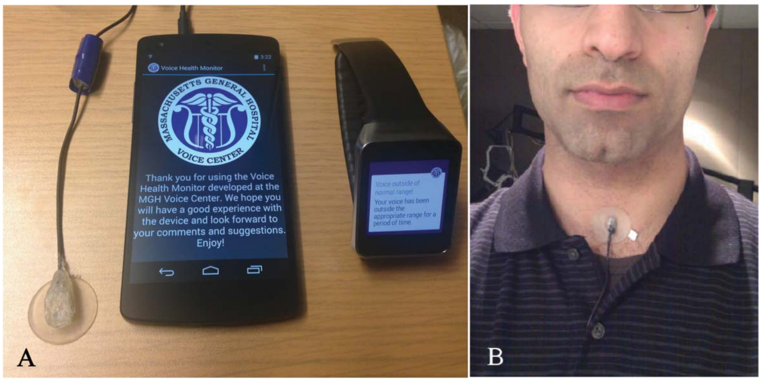
Eighty-five RNs were enrolled into the study; however, four participants failed the endoscopic screening owing to evidence of phonotrauma, 15 participants produced unstable baseline vocal intensity behavior (see Biofeedback Threshold Selection section for more details), and nine participants did not respond to the biofeedback. In addition, participants were terminated for extensive whispering during monitored days ($n = 4$), illness ($n = 1$), taking off equipment before the end of the day ($n = 2$), multiple appointment cancellations ($n = 1$), and being an inherently “soft talker” ($n = 1$). Thus, the total number of participants included in the statistical analysis was 48. The mean (standard deviation) of age was 36.9 years (12.9) for the 100% feedback group, 33.0 years (10.9) for the 25% feedback group, and 30.9 years (9.7) for the Summary feedback group. The number of women and men was, respectively, 12 and three for the 100% feedback group, 14 and three for the 25% feedback group, and 14 and two for the Summary feedback group.

Of note, the baseline vocal intensity distribution of the “soft talker” was quantitatively different (skew of 0.29 and kurtosis of −0.38) from all other participants, who had a group mean (standard deviation) of −0.24 (0.28) for skew and 0.05 (0.40) for kurtosis. Also, at the beginning of the study, five participants were told their percent compliance values after the short-term retention day, potentially confounding vocal behavior on the long-term retention day. This practice was discontinued for the remainder of the study; however, five participants do not have long-term retention data (one in the 100% feedback group, two in the 25% feedback group, and two in the Summary feedback group).

Data Acquisition Platform

The VHM (Mehta et al., 2012) was used to monitor and provide ambulatory voice biofeedback throughout the study. As shown in Figure 1, the VHM attaches a miniature accelerometer (model BU-27135; Knowles Electronics, Itasca, IL) 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 (Gear Live, Samsung, Seoul, South

Figure 1. Voice Health Monitor: (A) Accelerometer, interface cable with circuit encased in epoxy, smartphone, smartwatch; (B) the wired accelerometer mounted on a silicone pad affixed to the neck midway between the thyroid prominence and superior border of the sternum.



Korea; Motorola Moto 360, Lenovo, Morrisville, North Carolina, USA; or the G watch, LG Electronics, Seoul, South Korea), periodic alert capabilities that include system checks (Mehta et al., 2012) and vocal fatigue questions (Nanjundeswaran, Jacobson, Gartner-Schmidt, & Abbott, 2015), voice activity detection settings, and flexible biofeedback schedule settings. The real-time voice activity detection settings used for all days of ambulatory monitoring are listed in Table 1.

The experimenter met each RN prior to every day of monitoring in order to set up the VHM device. A calibration procedure transforming acceleration level to sound pressure level was not performed with each participant; that is, all signal amplitude-based measures were derived directly from the uncalibrated neck-surface acceleration signal (in dB). This approach was chosen because (a) no vocal intensity comparisons were made across individuals, and (b) it was crucial to minimize any signal level error within each participant. The calibration of sound pressure level can introduce on average ± 6–10 dB of measurement error (Švec, Titze, & Popolo, 2005). To further minimize signal level measurement error, on every day of monitoring, the first author placed the accelerometer on the same location of the anterior neck between the thyroid prominence and the superior border of the sternum. The experimenter used participant-specific measurements on the basis of anatomical landmarks (clavicle, wrinkles, thyroid prominence, and the cricoid) to ensure accurate accelerometer placement on the neck surface from day to day.

Study Design

Participants were block-randomized into three different feedback groups: (a) 100% feedback, (b) 25% feedback, and (c) Summary feedback. The overall study design included 6 days of monitoring. The first 3 days established an individual's natural vocal intensity behavior (baseline), the fourth day included ambulatory voice biofeedback for 30 min of phonation time (biofeedback), the fifth day—no biofeedback provided—occurred 1 or 2 days after the biofeedback day (short-term retention), and the sixth day—no biofeedback provided—occurred 5–11 days after the biofeedback day (long-term retention). All short-term retention days were originally scheduled for 1 day postbiofeedback, and all long-term retention days were originally scheduled for between 5 and 9 days postbiofeedback. However, it was not possible to obtain exact daily timing because acute care nursing schedules change frequently; for example, RNs are often asked to cover additional shifts or called off depending upon hospital unit census levels.

Table 1. Parameters for voice activity detection.

Parameter	Setting	Description
Frame duration	50 ms	Duration of a processing frame
Frame interval	0 ms	Interval between successive frames
Level limit	30 dB	Minimal value
Low/high (LH) ratio limit	22 dB	Minimal value
LH ratio low cutoff	70 Hz	Minimal frequency for ratio
LH ratio cutoff	2000 Hz	Low/high frequency separation
LH ratio high cutoff	3930 Hz	Maximal frequency for ratio
f0 lower limit	70 Hz	Minimal value
f0 upper limit	1000 Hz	Maximal value
Autocorrelation peak limit	0.25	Minimal value
Subharmonic peak limit	0.25	Minimal aperiodic voicing value

Biofeedback Threshold Selection

The level (dB) and duration settings for the biofeedback threshold were designed to be as consistent across the three different biofeedback conditions as possible. The level threshold for biofeedback was individually derived for each participant on the basis of his or her 3 days of baseline monitoring. More specifically, biofeedback was triggered/registered when the dB level of the accelerometer signal exceeded the 85th percentile of the pooled distribution of level across each individual's 3 baseline days. The 85th percentile was chosen as the biofeedback threshold level on the basis of our previous experience from studies of ambulatory voice biofeedback (KayPENTAX, 2009; Stadelman-Cohen et al., 2014; Van Stan, Mehta, & Hillman, 2015). Also, piloting with three participants prior to full-scale recruitment helped to establish a percentile-based level threshold that subjectively maximized the potential for a noticeable vocal behavior change (i.e., reduction in loudness) while still allowing functional vocal intensity and minimizing risk related to participant annoyance. The duration of time that the accelerometer signal needed to remain above the threshold level to trigger/register biofeedback was set to a single analysis frame (50-ms duration) in order to detect as many of these events as possible. This differs from previous studies where the threshold level had to be exceeded for multiple consecutive voiced frames to trigger biofeedback, for example, 250 ms (KayPENTAX, 2009; Stadelman-Cohen et al., 2014; Van Stan, Mehta, & Hillman, 2015). The previous approach increased the likelihood that some phonatory segments exceeding the threshold (i.e., voiced segments less than 250 ms in duration) would not trigger/register biofeedback and could potentially introduce an unacceptable amount of variability into the feedback paradigms.

Participants needed to demonstrate reasonable day-to-day stability of their vocal intensity behavior across all baseline days to complete the study. The stability requirement was imposed after we realized that excessive variability across the 3 baseline days would limit our ability to observe the establishment and retention of a biofeedback effect. This became clear after a few early participants produced 2 baseline days of comparable vocal intensity and 1 baseline day of significantly lower or higher vocal intensity. When the overall 85th percentile vocal intensity biofeedback threshold was applied to individual baseline days for these participants, the level threshold would range above the 90th percentile for 1 day. Consequently, it became difficult to demonstrate a statistically significant increase in percent compliance during biofeedback and retention days because percent compliance has a ceiling of 100%. Stability of baseline vocal intensity behavior was determined by finding the overall 85th percentile of the entire baseline vocal intensity period (pooling phonation levels across all 3 baseline days), and individual baseline days were then assigned level values corresponding to the overall 85th percentile. If the level for any baseline day was above the 90th percentile for that day, the participant would be terminated from the study. The 90th percentile cutoff was chosen because it capped participant variability to one third (5 percentage points) of the available percentage improvement (15 percentage points), increasing the likelihood of identifying a significant improvement during biofeedback and/or retention monitoring.

Because a primary purpose of the study was to assess the effect of different biofeedback schedules on retention, it only made sense for participants to continue in the study if they first demonstrated a biofeedback effect. Because the baseline variability limit was set at 90%, participants were only considered to have demonstrated a biofeedback effect if their percent compliance during the biofeedback day was at or above 91%. If a participant failed to achieve ≥ 91% compliance during a biofeedback day, the person was terminated from the study and did not undergo retention monitoring.

Biofeedback Delivery

During biofeedback days for the 100% and 25% feedback groups, the VHM provided a 250-ms vibrotactile cue via a smartwatch whenever the participants exceeded the accelerometer-level threshold after every 50-ms voiced frame (100% feedback group) or every fourth voiced frame they exceeded the threshold (25% feedback group). The VHM would automatically stop recording after registering 36,000 voiced frames (30 min of phonation) so each participant had equal exposure time to biofeedback cuing. Table 2 lists all settings used for biofeedback parameters throughout the study.

Those 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 compliance 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 compliance$\left(\frac{\text{\# of voiced frames inside desired range}}{\text{\# of total voiced frames}}\right)$ at adjustable time frames 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 2, 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 continue only after the participant used the keypad on the smartwatch to accurately enter/replicate the displayed percent compliances. 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 when the user looked at the responses, as well as if the participant accurately recalled his or her compliance percentages.

Statistical Analysis

Consistent with the performance-retention paradigm (Salmoni et al., 1984), statistical analysis was performed

Table 2. Biofeedback settings for ambulatory voice biofeedback.

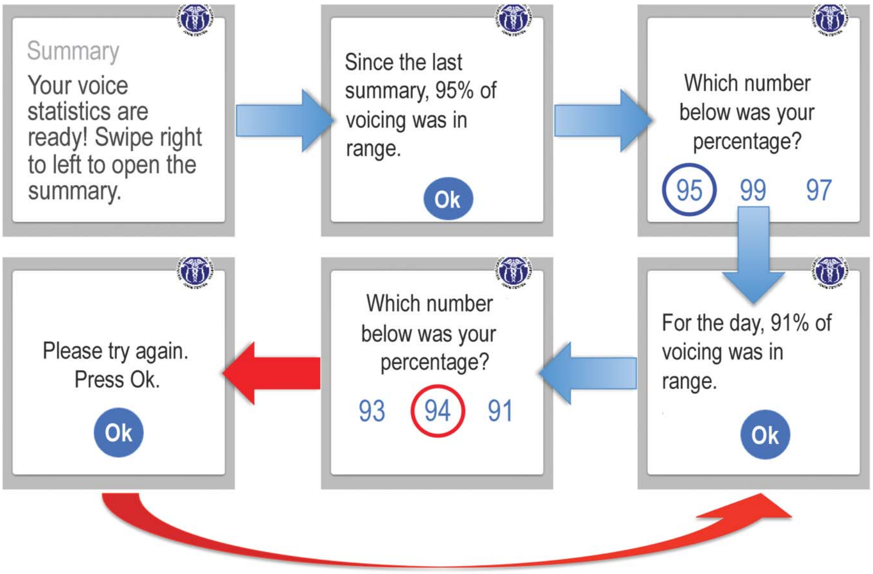
Parameter	Setting	Description
Feature	Level	Selected vocal feature with which to control biofeedback
Lower limit	30 dB	Frame counted toward duration threshold when below lower limit
Upper limit[a]	Variable	Frame counted toward duration threshold when above upper limit
Duration threshold	50 ms	Duration that feature must be outside limit range to trigger biofeedback
Duration hold	0 ms	Duration of voiceless/in-range voiced frames to wait before resetting duration threshold
Relative frequency[b]	1 or 4	Number of times duration threshold must be exceeded to trigger biofeedback
Cue duration	250 ms	Duration of vibrotactile cue administered through smartwatch and phone
Display summary statistics[c]	True	Enable the ability to provide summary statistics regarding feature
Statistics time frame[c]	120 s	Duration specifying analysis window and interval between summary display
Recording limit	30 min	Stop recording at specified duration
Count voicing only	True	Apply only voiced frames to recording limit

[a] The upper limit was personalized according to each participant's baseline behavior. [b] The value 1 was used for 100% frequency feedback and 4 for 25% frequency feedback. [c] Only "true" when providing summary statistics for participants in that feedback group.

separately for (a) acquisition (baseline and biofeedback), (b) short-term retention, and (c) long-term retention phases. This is because performance scores during biofeedback days theoretically represent both temporary changes owing to biofeedback cues and permanent changes from learning, whereas the postbiofeedback days would only represent learning-related changes. Therefore, direct comparisons between the biofeedback and retention days would be confounding (Schmidt & Lee, 2011; Winstein & Schmidt, 1990). For the acquisition phase, a mixed analysis of variance (ANOVA) 3 (Group) × 4 (Day [Baseline Day 1, Baseline Day 2, Baseline Day 3, and biofeedback) was completed with the percent compliance data. When statistical significance was found ($p < .05$), Bonferroni-corrected post hoc tests were conducted on the main effects; univariate follow-up tests were conducted on interaction effects; and, when significant, appropriate post hoc tests were performed.

Two one-way ANOVAs were completed for short- and long-term retention days individually using (a) percent compliance data and (b) performance change postbiofeedback (*change scores* is defined as percent compliance during biofeedback day minus percent compliance on short- and long-term retention days). When statistical significance was found for the main effect of Group, Bonferroni-corrected

Figure 2. 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.



Summary

Your voice statistics are ready! Swipe right to left to open the summary.

Since the last summary, 95% of voicing was in range.

Which number below was your percentage?

For the day, 91% of voicing was in range.

Which number below was your percentage?

Please try again. Press Ok.

post hoc tests were performed. Cohen's *d* was used as an effect size metric for all statistically significant pairwise comparisons such that effect sizes less than 0.2 were interpreted as small, 0.2 to 0.8 as medium, and greater than 0.8 as large (Cohen, 1988). All statistics were calculated using SPSS (version 22.0, IBM, Armonk, NY).

Results

Table 3 lists descriptive statistics of vocal behavior for participants within each biofeedback group. For the Summary group, the mean (*SD*) time between summary presentations was 20:05 (12:23) min:s, the mean (*SD*) time participants took to attend to the smartwatch display of their summaries was 16 (27) s, and the mean (*SD*) time it took for participants to enter answers to the prompts was 7 (8) s.

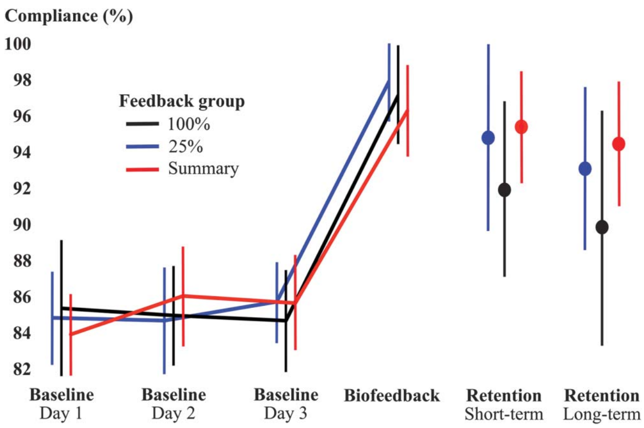
Acquisition

Figure 3 summarizes the group means and standard deviations of the individual feedback groups for each day of monitoring. The results of the 3 × 4 mixed ANOVA indicated a statistically significant main effect of Day; Wilks's lambda, $F(3, 43) = 368.53$, $p < .001$; and post hoc testing demonstrated a statistically significant increase in percent compliance of 12 percentage points (pp) during biofeedback days compared with each baseline day ($p < .001$, $d = 4.5$). Of note, both the Group × Day interaction effect and Group main effect were nonsignificant. Therefore, no differences between groups were identified during baseline or biofeedback days.

Short-Term Retention

The results of the one-way ANOVA using percent compliance data indicated no significant main effect of Group ($p = .094$). However, the one-way ANOVA using change scores indicated a significant main effect of Group; Wilks's lambda, $F(2, 46) = 3.733$, $p = .032$; and the post hoc testing demonstrated a statistically significant difference between the 100% feedback group and the Summary feedback group ($p = .027$, $d = 1.14$). More specifically, the Summary feedback group's performance change postbiofeedback ($M = -0.96$ pp; $SD = 2.50$ pp) was smaller than the 100% feedback group's performance change ($M = -5.15$ pp; $SD = 4.55$ pp). In terms of absolute phonation time, the 100% feedback group phonated approximately two minutes and 21 seconds more above the biofeedback threshold compared with the Summary feedback group.

Figure 3. Ambulatory biofeedback effect showing group-based means of percent compliance pooled within monitoring periods and respective feedback groups. Error bars represent ± 1 *SD* from the mean.



Long-Term Retention

The results of the one-way ANOVA using percent compliance data indicated a near-significant main effect of Group; Wilks's lambda, $F(2, 41) = 3.097$, $p = .056$; and

Table 3. Descriptive statistics of voice use within feedback groups.

Descriptive characteristic	Feedback group, *M* (*SD*) 100%	 25%	 Summary
Long-term retention (days postbiofeedback)	7.1 (1.7)	7.0 (1.3)	6.9 (1.0)
Baseline days			
Monitored duration (hr:min:s)	7:04:38 (3:25:16)	7:37:41 (3:10:33)	8:03:09 (2:57:08)
Phonation time (%)	14.1 (5.7)	13.2 (5.8)	12.3 (3.0)
Biofeedback day			
Monitored duration (hr:min:s)	4:42:51 (2:17:18)	5:01:41 (2:13:41)	4:55:43 (1:28:43)
Phonation time (%)	12.6 (5.1)	12.4 (6.7)	11.0 (3.1)
Number of cues provided	1072.8 (1002.9)	196.6 (212.2)	15 (0)
Short-term retention day			
Monitored duration (hr:min:s)	6:59:21 (3:18:57)	7:47:47 (3:04:18)	7:49:38 (2:24:49)
Phonation time (%)	14.5 (5.4)	12.2 (5.2)	11.1 (2.7)
Long-term retention day			
Monitored duration (hr:min:s)	7:22:54 (3:22:03)	9:05:46 (2:36:03)	7:41:12 (2:37:04)
Phonation time (%)	14.7 (7.0)	10.9 (3.1)	11.8 (2.7)

Note. (hr:min:s) = hours:minutes:seconds.

the post hoc testing demonstrated a near-significant difference between the 100% feedback group and the Summary feedback group ($p = .061$, $d = 0.87$). The Summary feedback group's percent compliance (94.4%) was higher than the 100% feedback group's percent compliance (89.8%). The results of the one-way ANOVA using change scores indicated a significant main effect of Group; Wilks's lambda, $F(2, 41) = 4.262$, $p = .021$; and the post hoc testing demonstrated a significant difference between the 100% feedback group and the Summary feedback group ($p = .017$, $d = 1.04$). The Summary feedback group's performance change post-biofeedback ($M = -2.20$ pp; $SD = 2.90$ pp) was smaller than the 100% feedback group's performance change ($M = -7.21$ pp; $SD = 6.17$ pp). In terms of absolute phonation time, the 100% feedback group phonated approximately 2 min and 52 s more above the biofeedback threshold compared with the Summary feedback group.

Potential Confounding Variables

The study design initially provided monetary incentive to maximize participant compliance during biofeedback and retention days. However, four out of 20 participants informed the experimenters that they whispered throughout their shifts to help ensure that they received the highest payment possible. This vocal behavior was not desirable given the study goals. Thus, these four participants' data were removed from analysis, the financial incentive was discontinued, and the remaining participants ($n = 32$) completed the study protocol with only an occupational motivation to speak softer. The analysis included 16 participants who received the monetary incentive (five in the 100% group; six in the 25% group; five in the Summary group). However, none of the incentivized participants performed statistically better than their non-incentivized peers (z-score of percent compliance values < 2.0), providing empirical support that the financial incentive did not confound the results for these individuals.

Other potential confounding variables include total monitoring time or phonation time. For example, those with shorter monitoring periods during biofeedback and retention days may have had higher compliance time because they did not need to "perform" for as long; that is, they had less chance to experience mental or physical fatigue. Also, participants could have avoided speaking in certain situations or reduced the amount of talking in general to improve their performance. However, these potential confounds do not seem likely as Pearson correlations between daily percent compliance values and total monitoring time or phonation time were statistically nonsignificant.

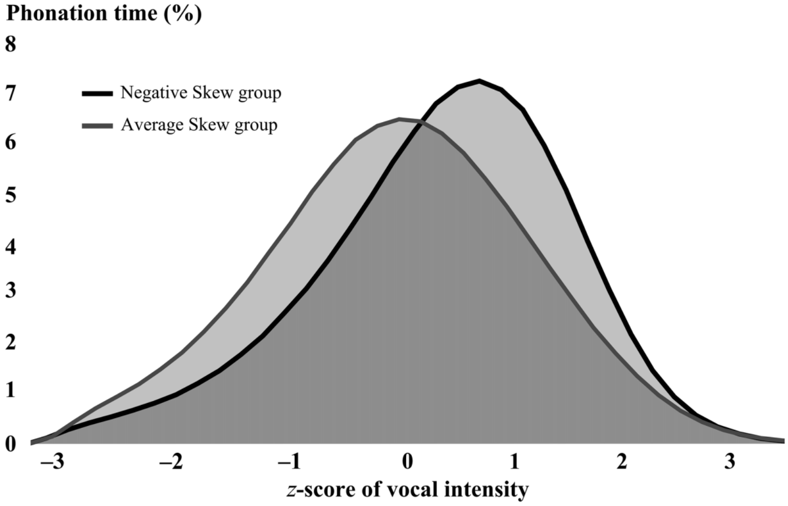
Throughout data collection, some RNs were reported to be "louder than most" by themselves and/or their peers. Therefore, a subset of the participants may have been significantly louder at baseline and consequently had an advantage during biofeedback and retention days; that is, they could have achieved a higher compliance value by speaking at a louder sound pressure level than those who were not louder than average at baseline. Because the study focused on vocal intensity changes within an individual, no daily loudness calibration was completed to convert the units of acceleration level to units of sound pressure level (Švec et al., 2005). Thus, across-individual comparisons of loudness were not possible. However, those who had baseline vocal intensity histograms that were negatively skewed (less than −0.2; Hildebrand, 1986) produced over 97.0% compliance on 22 of the 26 retention days. Because a strong negatively skewed vocal intensity histogram means that the bulk of their behavior was concentrated toward higher vocal intensities—and most of the high-performing retention days came from this group—there could exist a subset of "louder" participants who performed better than the rest of the group.

To further investigate the effects of baseline vocal intensity on subsequent performance, the participants were divided into two groups on the basis of baseline vocal intensity skew values where the Negative Skew group exhibited skews ≤ −0.2 ($N = 25$, overall skew = −0.449) and the Average Skew group exhibited skews between 0.2 and −0.2 ($N = 20$, overall skew = −0.051). Three participants had baseline vocal intensity histograms with a skew > 0.2 and are not included in either group. Figure 4 shows a visualization of the two groups' overall average baseline histograms normalized via z-scores. An independent-samples t test between the two groups using all biofeedback and retention percent compliance values indicated that the Negative Skew group produced significantly higher percent compliance values than those in the Average Skew group ($p = .005$, $d = 0.51$). Therefore, the Negative Skew and Average Skew groups were further subdivided into their respective feedback groups to examine any qualitatively different biofeedback/retention patterns compared with the entire data set. Table 4 shows the mean percent compliance values and associated standard deviations across biofeedback and retention days for each subgroup. Both subgroups displayed similar performance patterns across days compared with the entire data set, so it does not appear as though any subgroup individually influenced the overall results.

Discussion

The main objective of this study was to examine whether modifications in ambulatory voice biofeedback frequency or timing could influence the performance and retention of a reduced vocal intensity behavior in daily life with participants who responded to their respective biofeedback condition. Response to the ambulatory biofeedback was strong, as evidenced by the statistically significant difference and large Cohen's d effect size compared with the baseline days; however, the hypothesis that the 100% feedback group would supply the highest percent compliance during biofeedback was not observed. In fact, the 25% feedback group exhibited the highest group-averaged percent compliance and lowest standard deviation during biofeedback—although this was not a statistically significant difference. Perhaps the amount of cuing provided from 100% feedback was too much (on average 1,072 times

Figure 4. Average histograms of vocal intensity for each skew-determined subgroup. The histograms were obtained by characterizing each individual's pooled vocal intensity histogram from 3 baseline days into 40 *z*-score bins. All individuals' bins were averaged together if the overall baseline vocal intensity skew was < −0.2 (Negative Skew group, *N* = 25) or between −0.2 to 0.2 (Average Skew group, *N* = 20).



per day), eventually becoming distracting or annoying and resulting in a decrement in overall performance. Although no reports of annoyance during biofeedback days were communicated to the study staff, participants were not formally interviewed regarding their biofeedback experience.

The second hypothesis stated that less feedback in general would produce higher retention and a smaller performance decrement postbiofeedback removal. The one-way ANOVAs from the short- and long-term retention data provide support for this hypothesis because the 100% feedback group's performance postbiofeedback deteriorated significantly more than the Summary feedback group with a large associated effect size (d = 1.14 and 1.04 for short- and long-term retention, respectively). Furthermore, the 25% feedback group produced a mean percent compliance value between the 100% and Summary feedback groups during both short- and long-term retention (which would likely become significantly higher than the 100% feedback group with more power/participants). Group variability (standard deviation) was lowest in the Summary feedback group during both retention days, lending further support for the beneficial effect of reduced feedback on short- and long-term retention.

Table 4. Subgroups according to baseline vocal intensity skewness show differences in average percent compliance, but similar retention patterns across groups.

	Feedback group, *M* (*SD*)		
Parameter	**100%**	**25%**	**Summary**
Biofeedback day			
Negative skew	98.1 (1.7)	98.3 (1.7)	97.0 (2.4)
Average skew	96.0 (3.0)	98.0 (1.8)	95.3 (2.6)
Total data set	97.0 (2.7)	97.8 (2.2)	96.2 (2.5)
Short-term retention day			
Negative skew	91.6 (7.7)	96.0 (5.0)	96.6 (2.5)
Average skew	91.8 (2.3)	93.9 (4.8)	94.3 (3.2)
Total data set	91.9 (4.9)	94.7 (5.2)	95.3 (3.1)
Long-term retention day			
Negative skew	90.6 (8.8)	95.3 (3.4)	95.4 (3.6)
Average skew	89.0 (6.0)	91.4 (3.8)	93.1 (2.3)
Total data set	89.8 (6.6)	93.0 (4.5)	94.4 (3.4)

The guidance hypothesis is a popular theory of motor learning that can provide some theoretical explanation regarding the overall average performance of the three groups (Salmoni et al., 1984). This principle states that a learner requires feedback to improve motor performance, but too much feedback creates dependence on the external feedback and minimizes attention to intrinsic aspects of motor performance (Schmidt, Young, Swinnen, & Shapiro, 1989). Motivated by this principle, we surmise that the 100% feedback group may have significantly degraded in average performance postbiofeedback because their internal model of correctness was underdeveloped compared with the other two groups who received less feedback.

Principles from the field of reinforcement learning can be directly applied to the findings as well. For example, operant conditioning concerns the provision of reward or punishment after a volitional behavior to increase or decrease the likelihood of future occurrence (Skinner, 1953). To be more specific, the biofeedback delivered for the 100% and 25% feedback frequency groups could be viewed as negative reward or "punishment" because a consequence (a vibrotactile sensation on the wrist) occurred after an undesirable behavior (talking too loudly) to decrease the likelihood of future loud talking. However, the summary feedback category is not as straightforward to describe because the percent compliance value could represent reward if the number were high, such as 99%, or punishment if the number were low, such as 50%. Furthermore, summary feedback was not delivered immediately after the occurrence of the undesirable behavior. Future investigations could investigate how to improve the short- and long-term behavioral responses to feedback conditions (especially ones associated with punishment) by integrating various reward-related functions used by persuasive systems in this era of the quantified self (Choe, Lee, Lee, Pratt, & Kientz, 2014; Fritz, Huang, Murphy, & Zimmermann, 2014; Manson et al., 1999; Oinas-Kukkonen & Harjumaa, 2008).

When compared to those who responded to the biofeedback, the nine participants who did not respond to biofeedback (and were terminated from further study) performed significantly worse throughout the entire biofeedback day. To provide quantitative evidence of their poor performance, every participant's biofeedback day was divided into 15 consecutive individual subunits of 2,400 voiced frames and 15 cumulative subunits (e.g., 2,400, 4,800, 7,200 voiced frames)—which replicated the feedback that the summary group received. The mean (*SD*) number of subunits that exhibited ≥ 91% compliance time for the nonresponder group was 33.3% (24.8 pp), which was statistically lower than the compliance time of 93.1% (12.0 pp) in the responder group. Compliance time during the cumulative subunits for the nonresponder group was 19.4% (25.7 pp), which was statistically lower than the compliance time of 98.1% (5.7 pp) for the responder group. However, most of the participants who did not respond to the ambulatory biofeedback were from the Summary feedback group (seven of nine), indicating that it may have been more difficult to modify one's vocal behavior in this group because of the minimal amount of feedback provided. It could even be hypothesized that the higher retention exhibited by the Summary feedback group may be a result of filtering out participants with lower levels of motivation, vocal awareness, or general vocal skill.

Clinical Application

Although the results of this study indicate statistically significant differences consistent with motor learning theory and empirical studies, these differences may not be considered clinically meaningful depending on the patient and his or her vocal rehabilitation needs. However, 2.5 to almost 3 min of phonation (the approximate difference between the Summary and 100% feedback groups during retention monitoring) is a clinically meaningful difference when considering the reduction of abusive vocal behaviors in a patient with vocal fold nodules. Assuming a female modal fundamental frequency of 200 Hz, reduction of high (>85th percentile) vocal intensities for 3 min of voicing would directly affect 36,000 vocal fold collisions/oscillations. This effect is even more salient when one recognizes that fundamental frequency is positively correlated with vocal intensity and phonatory forces are thought to become increasingly more damaging as vocal intensity increases (Hillman et al., 1989; Zañartu et al., 2014).

If a patient does not respond to the feedback provided by a clinician or ambulatory voice monitor (i.e., a nonresponder), he or she cannot be simply terminated from therapy (as was done in this study). Therefore, empirically informed use of ambulatory voice biofeedback in a clinical setting will require future studies that provide guidance on how to maximize the chances of successful behavior change (e.g., decisions such as when to introduce ambulatory biofeedback, what objective measure should be targeted by the biofeedback, and where and how to establish a biofeedback threshold).

The variables of ambulatory voice biofeedback were specifically applied in a consistent, controlled manner throughout this study. Retention may be even stronger with different biofeedback settings, such as increased exposure to the biofeedback (dose), a higher or lower vocal intensity biofeedback threshold, or different feedback frequency (e.g., 50% or 33%). As an alternative, the delivery of multiple feedback schedules simultaneously (e.g., 25% feedback combined with summary statistics) or consecutively (e.g., fading from 100% feedback on Day 1 to 25% feedback on Day 2) may improve long-term retention. In addition, within-day feedback frequency could be modulated according to probabilistic instead of deterministic timing (e.g., programmatic fading over the course of the day, where the beginning of the day is closer to 100% feedback and frequency tapers to minimal feedback by the end of the day, resulting in an overall relative frequency of approximately 25%). It is also possible that the impact of different feedback parameters could vary depending on the individual characteristics of patients, including specific diagnoses, concomitant medical conditions, and cognitive/affective characteristics. Therefore, further work is warranted on how to fine-tune feedback settings to maximize the potential of retention on the basis of group and individual attributes.

A final point: Follow-up work must focus on the addition of features or measures for ambulatory voice biofeedback that expand the clinical relevance/application beyond the currently limited options of fundamental frequency and vocal intensity. With respect to the VHM, biofeedback may be based on periodicity measures or estimates of glottal aerodynamic parameters (e.g., maximum flow declination rate and unsteady flow) from the acceleration signal on the basis of a vocal system model (Llico et al., 2015). The flexibility of the smartphone-based platform offers the potential to use additional clinically salient measures such as cepstral peak prominence (Awan, Roy, Jetté, Meltzner, & Hillman, 2010; van Leer, Pfister, & Zhou, 2017), relative fundamental frequency (Stepp, Sawin, Smith, Awan, & Eadie, 2012), multidimensional features (Dejonckere & Lebacq, 2001), and features derived from machine learning algorithms (Ghassemi et al., 2014).

Conclusions

Overall, the results of this study demonstrate that feedback frequency and timing significantly affect the retention of a modified vocal behavior in participants with normal voices. Future work should focus on investigating the application of this approach in participants with voice disorders to assess viability for clinical use and the improved treatment of behaviorally based voice disorders. It is also expected that these clinical studies would incorporate an expanded set of vocal function parameters that could be used as a basis for more versatile and functionally relevant biofeedback than simple thresholds for fundamental frequency and vocal intensity.

Acknowledgments

This study was supported by the Voice Health Institute, Boston, Massachusetts, the National Institute on Deafness and Other Communication Disorders Grants R33 DC011588 (awarded to Dr. Hillman) and F31 DC014412 (awarded to Dr. Van Stan), and the National Institute of Child Health and Human Development R01 HD087089 (awarded to Dr. Sternad). 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

Adams, S., & Page, A. (2000). Effects of selected practice and feedback variables on speech motor learning. *Journal of Medical Speech-Language Pathology, 8,* 215–220.

Adams, S., Page, A., & Jog, M. (2002). Summary feedback schedules and speech motor learning in Parkinson's disease. *Journal of Medical Speech-Language Pathology, 10,* 215–220.

Anderson, D. I., Magill, R. A., Sekiya, H., & Ryan, G. (2005). Support for an explanation of the guidance effect in motor skill learning. *Journal of Motor Behavior, 37,* 231–238.

Austermann Hula, S. N., Robin, D. A., Maas, E., Ballard, K. J., & Schmidt, R. A. (2008). Effects of feedback frequency and timing on acquisition, retention, and transfer of speech skills in acquired apraxia of speech. *Journal of Speech, Language, and Hearing Research, 51,* 1088–1113.

Awan, S. N., Roy, N., Jetté, M. E., Meltzner, G. S., & Hillman, R. E. (2010). Quantifying dysphonia severity using a spectral/cepstral-based acoustic index: Comparisons with auditory-perceptual judgements from the CAPE-V. *Clinical Linguistics & Phonetics, 24,* 742–758.

Badets, A., & Blandin, Y. (2004). The role of knowledge of results frequency in learning through observation. *Journal of Motor Behavior, 36,* 62–70. https://doi.org/10.3200/JMBR.36.1.62-70

Bhattacharyya, N. (2014). The prevalence of voice problems among adults in the United States. *The Laryngoscope, 124,* 2359–2362.

Bislick, L. P., Weir, P. C., Spencer, K., Kendall, D., & Yorkston, K. M. (2012). Do principles of motor learning enhance retention and transfer of speech skills? A systematic review. *Aphasiology, 26,* 709–728.

Brandon, C. A., Rosen, C., Georgelis, G., Horton, M. J., Mooney, M. P., & Sciote, J. J. (2003). Staining of human thyroarytenoid muscle with myosin antibodies reveals some unique extrafusal fibers, but no muscle spindles. *Journal of Voice, 17,* 245–254.

Choe, E. K., Lee, N. B., Lee, B., Pratt, W., & Kientz, J. A. (2014, April). *Understanding quantified-selfers' practices in collecting and exploring personal data*. Paper presented at the Proceedings of the 32nd annual ACM conference on Human Factors in Computing Systems, Toronto, Canada.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Czerwonka, L., Jiang, J. J., & Tao, C. (2008). Vocal nodules and edema may be due to vibration-induced rises in capillary pressure. *The Laryngoscope, 118,* 748–752.

Dayan, E., & Cohen, L. G. (2011). Neuroplasticity subserving motor skill learning. *Neuron, 72,* 443–454.

Dejonckere, P., & Lebacq, J. (2001). Plasticity of voice quality: A prognostic factor for outcome of voice therapy? *Journal of Voice, 15,* 251–256.

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.

Fritz, T., Huang, E. M., Murphy, G. C., & Zimmermann, T. (2014, April). *Persuasive technology in the real world: A study of long-term use of activity sensing devices for fitness*. Paper presented at the Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Toronto, Canada.

Ghassemi, M., Van Stan, J. H., Mehta, D. D., Zañartu, M., Cheyne, H. A., II, Hillman, R. E., & Guttag, J. V. (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. https://doi.org/10.1109/tbme.2013.2297372

Hildebrand, D. K. (1986). *Statistical thinking for behavioral scientists*. Boston, MA: Duxbury Press.

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.

Kahn, D. M., Cook, T. E., Carlisle, C. C., Nelson, D. L., Kramer, N. R., & Millman, R. P. (1998). Identification and modification of environmental noise in an ICU setting. *CHEST Journal, 114,* 535–540.

Kandel, E. R., Schwartz, J. H., & Jessell, T. M. (2000). *Principles of neural science* (4th ed.). New York, NY: McGraw-Hill.

Katz, W. F., McNeil, M. R., & Garst, D. M. (2010). Treating apraxia of speech (AOS) with EMA-supplied visual augmented feedback. *Aphasiology, 24,* 826–837.

KayPENTAX. (2009). *Ambulatory phonation monitor: Applications for speech and voice*. Lincoln Park, NJ: Author.

Kim, I.-s., LaPointe, L. L., & Stierwalt, J. A. (2012). The effect of feedback and practice on the acquisition of novel speech behaviors. *American Journal of Speech-Language Pathology, 21,* 89–100.

Kol, E., Demircan, A., Erdoğan, A., Gencer, Z., & Erengin, H. (2015). The effectiveness of measures aimed at noise reduction in an intensive care unit. *Workplace Health & Safety, 63,* 539–545.

Konkani, A., Oakley, B., & Penprase, B. (2014). Reducing hospital ICU noise: A behavior-based approach. *Journal of Healthcare Engineering, 5,* 229–246.

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.

Llico, A. F., Zañartu, M., González, A. J., Wodicka, G. R., Mehta, D. D., Van Stan, J. H., & Hillman, R. E. (2015). Real-time estimation of aerodynamic features for ambulatory voice biofeedback. *The Journal of the Acoustical Society of America, 138,* EL14–EL19.

Maas, E., Butalla, C. E., & Farinella, K. A. (2012). Feedback frequency in treatment for childhood apraxia of speech. *American Journal of Speech-Language Pathology, 21,* 239–257.

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.

Manson, J. E., Hu, F. B., Rich-Edwards, J. W., Colditz, G. A., Stampfer, M. J., Willett, W. C., . . . Hennekens, C. H. (1999).

A prospective study of walking as compared with vigorous exercise in the prevention of coronary heart disease in women. *The New England Journal of Medicine, 341,* 650–658.

Mazer, S. E. (2006). Increase patient safety by creating a quieter hospital environment. *Biomedical Instrumentation & Technology, 40,* 145–146.

Mazer, S. E. (2012). Creating a culture of safety: Reducing hospital noise. *Biomedical Instrumentation & Technology, 46,* 350–355.

McGillivray, R., Proctor-Williams, K., & McLister, B. (1994). Simple biofeedback device to reduce excessive vocal intensity. *Medical and 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.

Nanjundeswaran, C., Jacobson, B. H., Gartner-Schmidt, J., & Abbott, K. V. (2015). Vocal Fatigue Index (VFI): Development and validation. *Journal of Voice, 29,* 433–440.

Nightingale, F. (1860). *Notes on nursing: What it is, and what it is not*. New York, NY: D Appleton and Company.

Nishino, T. (2012). The swallowing reflex and its significance as an airway defensive reflex. *Frontiers in Physiology, 3,* 489. https://doi.org/10.3389/fphys.2012.00489

Oinas-Kukkonen, H., & Harjumaa, M. (2008, February). *Towards deeper understanding of persuasion in software and information systems*. Paper presented at the Advances in Computer-Human Interaction, 2008 First International Conference, Sainte-Luce, Martinique.

Qutub, H. O., & El-Said, K. F. (2009). Assessment of ambient noise levels in the intensive care unit of a university hospital. *Journal of Family & Community Medicine, 16,* 53–57.

Ramig, L. O., & Verdolini, K. (1998). Treatment efficacy: Voice disorders. *Journal of Speech, Language, and Hearing Research, 41,* S101–S116.

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,* 1988–1995.

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., Ternstrom, S., Bulukin Wilen, F., & Sodersten, 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., Lange, C., & Young, D. E. (1990). Optimizing summary knowledge of results for skill learning. *Human Movement Science, 9,* 325–348.

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

Schmidt, R. A., Young, D. E., Swinnen, S., & Shapiro, D. C. (1989). Summary knowledge of results for skill acquisition: Support for the guidance hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15,* 352–359.

Simonyan, K., & Horwitz, B. (2011). Laryngeal motor cortex and control of speech in humans. *The Neuroscientist, 17,* 197–208.

Skinner, B. F. (1953). *Science and human behavior*. New York, NY: Simon and Schuster.

Sparrow, W. A., & Summers, J. J. (1992). Performance on trials without knowledge results (KR) in reduced relative frequency presentations of KR. *Journal of Motor Behavior, 24,* 197–209.

Stadelman-Cohen, T., Van Stan, J. H., & Hillman, R. E. (2014). Use of ambulatory biofeedback to supplement traditional voice therapy for treating primary muscle tension dysphonia in an adult female. In J. Stemple & E. Hapner (Eds.), *Voice therapy: Clinical case studies* (4th ed., pp. 157–163). San Diego, CA: Plural Publishing.

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.

Stepp, C. E., Sawin, D. E., Smith, A., Awan, S., & Eadie, T. L. (2012). The relationship between perception of vocal effort and relative fundamental frequency during voicing offset and onset. *Journal of Speech, Language, and Hearing Research, 55,* 1887–1896.

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.

Taylor-Ford, R., Catlin, A., LaPlante, M., & Weinke, C. (2008). Effect of a noise reduction program on a medical–surgical unit. *Clinical Nursing Research, 17,* 74–88.

van Leer, E., Pfister, R. C., & Zhou, X. (2017). An iOS-based cepstral peak prominence application: Feasibility for patient practice of resonant voice. *Journal of Voice, 31,* 131.e9–131.e16.

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,* 80–86.

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,* 713–721.

Van Stan, J. H., Mehta, D. D., Petit, R., 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,* 1–10.

Van Stan, J. H., Mehta, D. D., Zeitels, S. M., Burns, J. A., Barbu, A. M., & Hillman, R. E. (2015). Average ambulatory measures of sound pressure level, fundamental frequency, and vocal dose do not differ between adult females with phonotraumatic lesions and matched control subjects. *Annals of Otology, Rhinology, and Laryngology, 124,* 864–874.

Weltens, B., & De Bot, K. (1984). Visual feedback of intonation II: Feedback delay and quality of feedback. *Language and Speech, 27,* 79–88.

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

World Health Organization. (2009). *Global health risks: Mortality and burden of disease attributable to selected major risks.* Geneva, Switzerland.

Yao, W. X., Fischman, M. G., & Wang, Y. T. (1994). Motor skill acquisition and retention as a function of average feedback, summary feedback, and performance variability. *Journal of Motor Behavior, 26,* 273–282.

Zañartu, M., Galindo, G. E., Erath, B. D., Peterson, S. D., Wodicka, G. R., & Hillman, R. E. (2014). Modeling the effects of a posterior glottal opening on vocal fold dynamics with implications for vocal hyperfunction. *The Journal of the Acoustical Society of America, 136,* 3262–3271.