Integration of Motor Learning Principles Into Real-Time Ambulatory Voice Biofeedback and Example Implementation Via a Clinical Case Study With Vocal Fold Nodules

Jarrad H. Van Stan,a,b Daryush D. Mehta,a,b,c Robert J. Petit,d Dagmar Sternad,e Jason Muise,a James A. Burns,a,c and Robert E. Hillmana,b,c

Purpose: Ambulatory voice biofeedback (AVB) has the potential to significantly improve voice therapy effectiveness by targeting one of the most challenging aspects of rehabilitation: carryover of desired behaviors outside of the therapy session. Although initial evidence indicates that AVB can alter vocal behavior in daily life, retention of the new behavior after biofeedback has not been demonstrated. Motor learning studies repeatedly have shown retention-related benefits when reducing feedback frequency or providing summary statistics. Therefore, novel AVB settings that are based on these concepts are developed and implemented.

Method: The underlying theoretical framework and resultant implementation of innovative AVB settings on a smartphone-based voice monitor are described. A clinical case study demonstrates the functionality of the new relative frequency feedback capabilities.

Results: With new technical capabilities, 2 aspects of feedback are directly modifiable for AVB: relative frequency and summary feedback. Although reduced-frequency AVB was associated with improved carryover of a therapeutic vocal behavior (i.e., reduced vocal intensity) in a patient post-excision of vocal fold nodules, causation cannot be assumed.

Conclusions: Timing and frequency of AVB schedules can be manipulated to empirically assess generalization of motor learning principles to vocal behavior modification and test the clinical effectiveness of AVB with various feedback schedules.

Systematic study of ambulatory voice biofeedback has potential to dramatically and directly affect current clinical practice paradigms because voice therapy remains wholly dependent on episodic and visit-based treatment delivery despite the obvious need for a more integrated approach into the daily lives of patients. Ambulatory voice biofeedback can help transform therapeutic practice toward a fundamental shift of extending therapy principles and techniques throughout the course of a patient’s daily life. For example, a ubiquitous goal of voice therapy is the permanent retraining of a patient’s vocal behavior (e.g., improved voice quality, improved vocal efficiency, reduced vocal fatigue), yet one of the most challenging aspects of voice therapy is achieving carryover—including generalization and long-term retention—of newly established vocal behaviors outside the clinic (Ziegler, Dastolfo, Hersan, Rosen, & Gartner-Schmidt, 2014). It is likely that ambulatory voice biofeedback has significant potential to address this carryover challenge by providing patients with timely information about their vocal behavior throughout daily life (Stadelman-Cohen, Van Stan, & Hillman, 2014; Van Stan, Mehta, & Hillman, 2015).

However, how to best provide ambulatory voice biofeedback is an unanswered question because it has been nearly unexplored and commercial ambulatory voice monitors provide biofeedback in a similar manner: a vibrotactile or auditory cue every time a threshold is exceeded (Van Stan, Gustafsson, Schalling, & Hillman, 2014). Furthermore, only two formal pilot studies using ambulatory voice biofeedback exist in the literature (Schalling, Gustafsson, Ternstrom, Bulukin Wilen, & Sodersten, 2013; Van Stan, Mehta, & Hillman, 2015). Their findings imply that the current

Disclosure: The authors have declared that no competing interests existed at the time of publication.
method of providing ambulatory voice biofeedback induces only a temporary change in daily vocal performance and that retention is poor when feedback is removed. This lack of retention is clinically significant because successful rehabilitation is achieved only by a permanent modification of the patient’s vocal function in daily life. Therefore, a lack of retention represents a lack of clinical effectiveness and an increased risk of disease and/or symptom recurrence.

The field of motor learning may provide theoretical and practical insights for structuring ambulatory voice biofeedback in a way that maximizes the retention (or learning) of new vocal behaviors (Schmidt & Lee, 2011). This literature is rich with empirical studies advocating advantages of various types of feedback schedules through modification of frequency and/or type of feedback. For example, many laboratory-based studies on limb movements have demonstrated an increase in learning when decreasing the frequency of feedback (termed relative frequency; Lee, White, & Carnahan, 1990) and/or delaying the presentation of feedback after multiple trials (termed summary feedback; Yao, Fischman, & Wang, 1994). Relative frequency is commonly defined as the frequency that feedback was provided divided by the total number of trials for which feedback could have been provided (Salmoni, Schmidt, & Walter, 1984). The effect of various feedback frequencies is commonly studied by comparing retention metrics of subjects in two groups: a baseline group that receives feedback every trial (100% feedback) and a comparison group that receives feedback at a reduced frequency, such as after every other trial (50% feedback) or after every fourth trial (25% feedback; Adams & Page, 2000; Badets & Blandin, 2004; Badets, Blandin, Wright, & Shea, 2006; Lee et al., 1990; Salmoni et al., 1984; Sidaway et al., 2008; Sparrow & Summers, 1992; Sullivan, Kantak, & Burtner, 2008; Vander Linden, Cauraugh, & Greene, 1993; Weeks & Kordus, 1998; Weeks, Zelaznik, & Beyak, 1993; Weinstein & Schmidt, 1990).

Summary feedback has most commonly been tested using a method in which feedback is withheld for a block of trials, and after the last trial the subject is presented with either (a) a graph of trials over time or (b) an overall average of all trials (Anderson, Magill, & Sekiya, 1994, 2001; Anderson, Magill, Sekiya, & Ryan, 2005; Gable, Shea, & Wright, 1991; Guadagnoli, Dornier, & Tandy, 1996; Lavery & Suddon, 1962; Salmoni et al., 1984; Schmidt, Lange, & Young, 1990; Schmidt, Young, Swinnen, & Shapiro, 1989; Weeks & Sherwood, 1994; Yao et al., 1994; Young & Schmidt, 1992). Both options appear equally valid because both have demonstrated statistically identical improvements in retention compared to no trial delay. The literature has tended to demonstrate increased retention in simple motor tasks as feedback becomes less frequent and summaries encompass more trials (Guadagnoli et al., 1996).

Although improved retention or learning through decreased feedback appears to be counterintuitive, there are multiple accounts to explain this commonly reported phenomenon. The guidance hypothesis proposes that feedback is required to improve or modify performance, but too much feedback causes the learner to become dependent on the feedback and minimizes attention to intrinsic aspects of motor performance (Salmoni et al., 1984). Therefore, 100% feedback may result in the highest performance during biofeedback, but when feedback is removed, performance degrades due to a weak internal model of correctness. Another popular theory is the stability hypothesis, which states that stable trial-to-trial performance allows easier identification and storage of consistent movement patterns (Lai & Shea, 1998). Higher frequency feedback schedules result in more variability, or trial-to-trial changes, than low-frequency or summary feedback because subjects modify their performance after each bout of feedback. This increases the difficulty of extracting stable movement patterns from 100% feedback and subsequently degrades retention.

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 and neck structures (Schmidt & Lee, 2011). Furthermore, the few studies that have focused on how feedback frequency or delay affect 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 supported generalization of limb-based findings (Adams & Page, 2000; Adams, Page, & Jog, 2002; Kim, LaPointe, & Sterewalt, 2012), two provided inconsistent support (Austermann Hula, Robin, Maas, Ballard, & Schmidt, 2008; Friedman, Hancock, Schulz, & Bamdad, 2010; Maas, Butalla, & Farinella, 2012), and one resulted in negative findings (Kat, McNeil, & Garst, 2010). In regards to voice-specific learned movements, the results of studies have been mostly negative (Weltens & De Bot, 1984; Yiu, Verdolini, & Chow, 2005); one study provided inconsistent support (Ferrand, 1995), and another replicated limb study findings (Steinhauer & Grayhack, 2000). Therefore, replicating the results of limb-based studies in the head and neck structures involved in voice and speech production is a much-needed effort.

This may be a nontrivial step because the corticobulbar (which controls the head and neck) and corticospinal (which controls the core and limbs) sensorimotor systems differ in many ways. In regards to anatomical and physiological differences, (a) many head and neck structures have bilateral cortical input (Simonyan & Horwitz, 2011), whereas limbs typically have predominantly contralateral cortical input (Kandel, Schwartz, & Jessell, 2000); (b) gamma neurons have yet to be noted in many head and neck muscles (Brandon et al., 2003), whereas limbs contain a gamma neuronal system that is crucial for load bearing and angle sensation (Kandel et al., 2000); and (3) unlike limb-related corticospinal circuits, vocal neural circuits are tightly interconnected with respiratory brainstem nuclei (Nishino, 2012). Also, unlike the perceptual aspects of limb movements, vocal behaviors (e.g., modification of voice quality or vocal efficiency) involve minimal or no visual feedback. Last, studies in motor learning are primarily carried out in highly controlled experimental conditions, with low-skill
movements such as pointing or reaching targeted as outcomes in rehabilitation rather than complex, functional skills (e.g., vocal intensity production throughout a person’s daily life).

To our knowledge, all current ambulatory voice biofeedback capabilities are limited to 100% relative frequency and immediate timing (i.e., a vibrotactile or auditory cue is given within milliseconds every time a level threshold is exceeded). Therefore, it has not been possible to empirically test modifiable feedback frequency and delays in ambulatory voice biofeedback based on motor learning principles (Van Stan et al., 2014). The purpose of this clinical focus article is to describe new motor learning–inspired biofeedback capabilities that are incorporated into an ambulatory monitoring software application on a smartphone platform so that reduced frequency and summary feedback can be provided. To demonstrate the functionality and potential clinical use of these novel biofeedback settings, a clinical case study is presented in which a patient is provided with ambulatory biofeedback in addition to therapy after surgical excision of vocal fold nodules with the specific goal of decreasing vocal intensity in daily life.

Method

Voice Health Monitor

Figure 1 shows the Voice Health Monitor (VHM; Mehta, Zañartu, Feng, Cheyne, & Hillman, 2012), which was the device used to implement all novel biofeedback settings and to provide ambulatory voice biofeedback throughout the case study. The VHM attaches a miniature accelerometer (ACC; model BU-27135, Knowles Electronics, Itasca, IL) via double-sided tape to the base of the neck above the sternal notch to sense phonation. The sensor is connected to a custom smartphone application (VHM) as the data acquisition platform, and the system records the unprocessed ACC signal at an 11,025-Hz sampling rate, 16-bit quantization, and 80-dB dynamic range to obtain frequency content of neck surface vibrations up to 5,000 Hz. The VHM application provides a user-friendly interface for starting and stopping recording, daily sensor calibration, smartwatch coupling, and periodic alert capabilities that include system checks (Mehta et al., 2012) and vocal fatigue questions (Nanjundeswaran, Jacobson, Gartner-Schmidt, & Abbott, 2015). An Android-based software platform permitted the modification of previous versions of the VHM app to incorporate novel voice activity detection, recording settings, and biofeedback settings.

Ambulatory Voice Biofeedback Based on Motor Learning Principles

Table 1 outlines all modifiable ambulatory voice biofeedback settings, which result in the provision of visual statistics or vibrotactile cues provided by an Android-based smartphone (Nexus 5 or Samsung Galaxy S3) and/or a smartwatch (e.g., Samsung Gear Live, Motorola Moto 360, or the LG G Watch). In order to provide biofeedback, real-time voice activity detection was implemented on the smartphone; the voice activity detection details are outlined in Appendix A, and the related temporal decision tree is outlined in Appendix B. When the ambulatory biofeedback section of the settings menu is enabled, the feature option provides the ability to pick which voice feature is used to control biofeedback. Level and fundamental frequency currently are the only features available for biofeedback, but the software is implemented in a way that other features of interest can easily be incorporated in the future (Llico et al., 2015).

The relative frequency setting in Table 1 requires a whole-number input that represents the number of times a feature must exceed the duration threshold before triggering a biofeedback event. Therefore, the default relative frequency of 1 will provide a vibrotactile trigger every time the duration threshold is passed (i.e., 100% relative frequency). However, a relative frequency of 4 will trigger a vibrotactile cue every fourth time the duration threshold is passed (i.e., 25% relative frequency). Motor learning studies have repeatedly demonstrated that, in general, retention is improved by decreasing relative frequency of feedback (Salmoni et al., 1984).
Table 1. Modifiable parameters and their default values for ambulatory voice biofeedback.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Default</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Feature&lt;sup&gt;a&lt;/sup&gt;</td>
<td>Level</td>
<td>Select feature to control biofeedback</td>
</tr>
<tr>
<td>Lower limit&lt;sup&gt;b&lt;/sup&gt;</td>
<td>45 dB</td>
<td>Frame counted toward duration threshold when below limit</td>
</tr>
<tr>
<td>Upper limit&lt;sup&gt;b&lt;/sup&gt;</td>
<td>90 dB</td>
<td>Frame counted toward duration threshold when above limit</td>
</tr>
<tr>
<td>Duration threshold&lt;sup&gt;d&lt;/sup&gt;</td>
<td>50 ms</td>
<td>Duration that feature must be outside limit range</td>
</tr>
<tr>
<td>Duration hold&lt;sup&gt;b&lt;/sup&gt;</td>
<td>0 ms</td>
<td>Duration of voiceless or in-range frames to wait before reset duration threshold count</td>
</tr>
<tr>
<td>Relative frequency</td>
<td>1</td>
<td>Number of times duration threshold must be exceeded to trigger biofeedback</td>
</tr>
<tr>
<td>Cue duration&lt;sup&gt;a&lt;/sup&gt;</td>
<td>200 ms</td>
<td>Duration of time to vibrate when cueing the user</td>
</tr>
<tr>
<td>Summary statistics</td>
<td>True</td>
<td>Enable the ability to provide summary statistics regarding targeted feature</td>
</tr>
<tr>
<td>Statistics time frame</td>
<td>300 s</td>
<td>Time used to calculate and show summary statistics</td>
</tr>
<tr>
<td>Feature file</td>
<td>True</td>
<td>Comma separated value feature file produced</td>
</tr>
<tr>
<td>Recording limit</td>
<td>45 min</td>
<td>Stop recording at specified time</td>
</tr>
<tr>
<td>Voiced frames only</td>
<td>True</td>
<td>Apply only voiced frames to recording limit</td>
</tr>
</tbody>
</table>

<sup>a</sup>Derived from the Ambulatory Phonation Monitor software (APM Model 3200; KayPENTAX, Montvale, NJ).

Figure 2 displays an example of relative frequency (as well as other novel biofeedback settings) applied to hypothetical data.

The VHM software provides a simple summary statistic called percent compliance (number of voiced frames inside the desired range divided by the total voiced frames, and multiplied by 100) at adjustable time frames to replicate the concept of summary feedback. Motor learning studies provide feedback regarding the targeted motor behavior, not both the motor behavior and the time between trials; therefore, the statistics time frame (i.e., how many trials pass before providing summary statistics) counts only voiced frames toward the block of time and excludes nonvoiced frames. For example, if the time frame were set for 3 min, the summary statistics will appear after every 3 min of voiced frames. This would correspond to every 30 min if the person spoke at 10% phonation time. To ensure that the subject attends to and correctly processes their summary statistics, multiple screens for user interaction are provided on the phone and smartwatch whenever the statistics are displayed. Figure 3 outlines the flow of screens on the smartwatch. Monitoring can continue only with accurate manual input of the statistics by the user. Patient responses are recorded in a text document on the VHM app that allows documentation of (a) how much time passed between the summary statistic presentation and when the user looked at it and (b) whether the subject accurately recalled their percent compliance.

Three other novel biofeedback-related features (explained in Table 1) have been developed for practicality. The inclusion of a lower limit and upper limit threshold for every feature will permit biofeedback provision that is based on a desired range that avoids extremes; all current voice monitors can provide feedback only above a threshold or below a threshold, not both. The potential benefit of two limits can be illustrated in a hypothetical example using cepstral peak prominence (CPP), a popular clinically used measure (Awan et al., 2010; Shue, Chen, & Alwan, 2010; Murphy, 2006; Murphy & Akande, 2007). Avoiding extremes of CPP may be clinically beneficial because low values of CPP are correlated to breathy or rough voicing and high values of CPP have been associated with pressed voicing (Awan et al., 2010; Shue, Chen, & Alwan, 2010). Duration hold allows for duration threshold times that are longer than average speech-related voiced durations; for example, the most frequently occurring voiced duration in speech is approximately 150 ms to 200 ms (Titze, Hunter, & Švec, 2007;
levels of technical expertise. Waveforms, which can be time consuming and require high processing algorithms on hours of ambulatory acceleration regardless of the relative frequency or duration hold settings. The feature file option avoids the need for using post-monitoring analysis of counting the number of vibrotactile cues provided regardless of the relative frequency or duration hold settings. Also, the feature file option avoids the need for using post-processing algorithms on hours of ambulatory acceleration waveforms, which can be time consuming and require high levels of technical expertise.

Clinical Case Study

A 34-year-old woman presented with recent increased vocal strain and fatigue corresponding to the addition of new teaching responsibilities that required multiple hours-long lectures per week. Eight years previously she underwent surgical excision of bilateral fibrovascular vocal fold lesions as well as a successful course of voice therapy. As seen in Figure 4A, the patient’s most recent transoral rigid stroboscopy revealed a recurrence of bilateral fibrovascular vocal fold lesions (left > right), and the lesions were excised in a follow-up suspension microlaryngoscopy. She began voice therapy approximately 3 weeks postsurgery and, despite progressing in her use of therapeutic voicing techniques during the therapy session, reported continued occasions of strain and vocal deterioration associated with talking too loudly. The patient also reported that her friends, family, and coworkers complained frequently that she spoke too loudly. In addition, her treating speech-language pathologist (SLP) and the patient herself noted a decreased ability to monitor her loudness and a habitually louder-than-normal speaking voice. Therefore, the treating SLP and patient decided to use ambulatory voice biofeedback, targeting reduced vocal intensity measured via decibels of skin acceleration level (Svec, Titze, & Popolo, 2005), in hopes of improving the patient’s self-monitoring skills and her application of therapeutic voicing in daily life. The objective measure used for biofeedback (vocal intensity) was not the only therapy target, but it was a measure that demonstrated a strong association with therapeutic improvement. Because vocal intensity feedback indirectly targets the patient’s vocal hyperfunction, the obvious concern is that a patient could use hyperfunctional behaviors to maintain adequate compliance with the device and defeat the purpose of an ambulatory intervention. Therefore, biofeedback was provided only after the patient had reliably learned strategies during therapy that decreased both vocal hyperfunction and vocal intensity.

Study Design

The patient wore the VHM for 2 days to obtain a representation of her typical vocal intensity behavior, which was characterized by the pooled histogram of all of her vocal intensity data points from both days of monitoring (i.e., all data were derived from speech in daily life). Through discussion with the treating SLP and the patient as well as trialing during voice therapy, it was decided to place the threshold at the 90th percentile of her baseline vocal intensity histogram (i.e., the patient would receive a vibrotactile cue whenever she phonated in the upper 10% of her vocal intensity). The following choices were made when setting the vocal intensity threshold. First, the duration threshold was 50 ms because the treating SLP and patient wanted to know every time she phonated in her “louder” voice. This also allowed more fidelity to both 100% and 25% feedback frequency. Second, during trialing, the shortest vibrotactile cue the patient felt comfortable detecting was 250 ms. Last, through therapeutic activities, the 90th percentile was chosen because the patient and SLP felt it allowed functional vocal intensity while also providing frequent cueing during phonation that was perceptually judged to be loud.
Results

Figure 5 shows the histograms for each of the six time periods for which the patient wore the VHM. The patient’s average (standard deviation) vocal intensity was as follows: baseline period, 61.17 dB (5.17 dB); short-term monitoring after 25% feedback, 65.88 dB (4.34 dB); mid-term monitoring after 25% feedback, 65.24 dB (4.34 dB); mid-term monitoring after 100% feedback, 56.07 dB (5.37 dB); long-term monitoring after 25% feedback, 56.47 dB (5.69 dB); long-term monitoring after 100% feedback, 57.09 dB (5.52 dB); 100% retention period, 59.90 dB (5.37 dB); 25% biofeedback period, 57.56 dB (5.37 dB); and 25% retention period, 56.51 dB (5.06 dB).

Per effect size calculations, both baseline days were not significantly different from each other (Cohen’s $d = 0.30$). The patient maintained a higher percent compliance compared with baseline when the 100% frequency feedback was turned on (Cohen’s $d = 0.86$ for short-term monitoring periods and 0.76 for long-term monitoring periods), but this behavior was not retained when the feedback was turned off (Cohen’s $d = 0.23$). When 25% frequency feedback was provided, the patient maintained a high percent compliance during the biofeedback days and the retention periods (Cohen’s $d = 0.68$ and 0.91, respectively), indicating that she did retain the vocal behavior after biofeedback was removed. Also, before being informed of her percent compliance results, the patient described her retention monitoring after 100% feedback as vocally fatiguing and reported difficulty performing therapeutic vocal behaviors. In contrast, she described minimal difficulties throughout her retention monitoring after 25% feedback.

Discussion

Novel ambulatory voice biofeedback settings inspired by motor learning principles have been implemented on the VHM in order to provide adjustable amounts of relative frequency and summary biofeedback. This advancement in technological capabilities can provide a platform for empirically testing the generalization of motor learning studies that are based on limb movements to those of voice- and speech-related movements and may help assess the clinical effectiveness of various feedback schedules—that is, whether feedback structure directly affects the degree of carryover or learning of new vocal behaviors.

The motor learning–inspired biofeedback capabilities were derived from what is known about learning new motor behaviors in general; therefore, the potential benefits are not isolated to those patients with the same diagnosis as in the case study (i.e., vocal fold nodules). Other vocal rehabilitative endeavors that require behavioral training can apply the same principles, such as treating muscle tension dysphonia.
Parkinson’s disease, or maladaptive voicing postmedialization for vocal fold paralysis. Even habilitative endeavors, such as pitch modification for transgender voice patients, are under the purview of motor learning principles. The implementation of reduced frequency and summary feedback provides ambulatory voice biofeedback flexibility for a clinician. This means that commonly applied clinical techniques such as fading can be used—for example, the patient can be provided a week of 100% frequency feedback, then fading feedback in the following week to 25% frequency feedback, and finally fading feedback even more in the next week to summary feedback every 5 min of voicing. Individualization for each patient’s needs can be considered when using the modifiable feedback settings. This is significantly different from current capabilities, which require the same feedback schedule (100% immediate feedback) for every patient undergoing ambulatory voice biofeedback (Van Stan et al., 2014). However, attempts have been made to indirectly represent modified frequency of feedback by varying temporal and feature-specific thresholds with 100% immediate-feedback schedules (Gustafsson, Ternstrom, Sodersten, & Schalling, 2015). This approach unfortunately does not exert direct control on the variable of feedback frequency, meaning that the resulting frequency a patient received could vary widely depending on uncontrollable factors such as the individual patient’s day-to-day vocal variability, environmental noise exposure, acoustics of the room or surroundings, and so on. Implementation of programmable settings for the direct control of feedback frequency and summary statistics will ensure that a desired feedback modification will be delivered per the clinician’s—and patient’s—preferences.

The clinical case study provides a demonstration of the new biofeedback functionality—specifically, the use of reduced relative frequency. Although the patient clearly showed increased retention after 25% feedback compared with after 100% feedback, it is not possible to exclusively assign the positive outcome to a reduction in relative frequency or fading of feedback. This is because the patient was theoretically improving over time, meaning that the improved retention could be attributable to the fact that the patient had more time to practice her modified vocal behavior before her “after 25% feedback retention” period than before her “after 100% feedback retention” period. As an alternative, the improved retention could be due to an overall accumulation of cueing received over time and not specifically the reduced frequency of feedback.

Future work should replicate limb-based results of improved retention with reduced relative frequency and summary feedback in groups of subjects with typical voices. This is an essential first step to clearly observing biofeedback and retention effects without confounding influences (e.g., concomitant voice therapy, abnormal vocal anatomy) that potentially could be introduced by including subjects with various types of voice disorders. Then, efficacy or effectiveness studies in patients undergoing voice therapy could test ambulatory voice biofeedback methods shown to optimize retention of therapeutic vocal behaviors. However, it is expected that any clinical studies would need to incorporate an expanded set of vocal function parameters for more versatile

Figure 5. The results of the patient’s weeks of monitoring. All histograms are pooled across multiple days and oriented vertically so that low intensity levels (dB) are at the bottom and high intensity levels are at the top. The same baseline histogram (black) is on the left and right for comparison of biofeedback (gray scale) and retention (red). All 100% frequency histograms and the subsequent retention histogram are on the left; the 25% frequency histogram and the subsequent retention histogram are on the right. *Cohen’s $d > 0.30$ compared with baseline.
and functionally relevant biofeedback than simple thresholds for fundamental frequency and vocal intensity. Such novel biofeedback targets could be based on additional measures that can be extracted from the accelerometer, such as those generated from machine learning (Ghassemi et al., 2014), cepstral- and spectral-based measures (Awan et al., 2010), parameters using impedance-based inverse filtering (Lico et al., 2015; Zañartu, Ho, Mehta, Hillman, & Wodicka, 2013), and combinations of different subsets of these measures. Last, because the VHM records the raw acceleration signal, which is then stored in a database, it can be reprocessed and used in simulations to gain new insights as novel measures are developed or as new salient aspects of the acceleration waveform are determined.

Conclusions

Novel ambulatory voice biofeedback settings that are based on principles of motor learning have been implemented on a smartphone-based ambulatory voice monitor and have been beneficially applied to a subject experiencing difficulties with carryover of therapeutic behaviors outside the therapy session. It is now possible to systematically study the effect of feedback modifications on the performance and retention of vocal motor behaviors in groups of subjects with typical voices or those with voice disorders, which is a crucial step toward building an evidence base for the eventual clinical adoption of sophisticated ambulatory voice biofeedback.

Acknowledgments

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

References


# Appendix A

## Modifiable Parameters and Their Default Values for Voice Activity Detection

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Default</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frame length(^a)</td>
<td>50 ms</td>
<td>Length of a processing frame</td>
</tr>
<tr>
<td>Frame interval</td>
<td>0 ms</td>
<td>Interval or overlap between frames</td>
</tr>
<tr>
<td>Level limit</td>
<td>45 dB</td>
<td>Minimal value for voiced frames</td>
</tr>
<tr>
<td>Calibration multiplier</td>
<td>1</td>
<td>Multiplier for calibrated level</td>
</tr>
<tr>
<td>Calibration offset</td>
<td>0</td>
<td>Offset for calibrated level</td>
</tr>
<tr>
<td>LH ratio limit</td>
<td>22 dB</td>
<td>Minimal value for voiced frames</td>
</tr>
<tr>
<td>LH ratio low cutoff</td>
<td>70 Hz</td>
<td>Minimal frequency for ratio</td>
</tr>
<tr>
<td>LH ratio cutoff</td>
<td>2000 Hz</td>
<td>Low- and high-frequency separation</td>
</tr>
<tr>
<td>LH ratio high cutoff</td>
<td>3930 Hz</td>
<td>Maximal frequency for ratio</td>
</tr>
<tr>
<td>f0 lower limit</td>
<td>70 Hz</td>
<td>Minimal value for voiced frames</td>
</tr>
<tr>
<td>f0 upper limit</td>
<td>1000 Hz</td>
<td>Maximal value for voiced frames</td>
</tr>
<tr>
<td>Autocorrelation peak limit</td>
<td>0.25</td>
<td>Minimal value for voiced frames</td>
</tr>
<tr>
<td>Subharmonic peak limit</td>
<td>0.25</td>
<td>Minimal aperiodic voicing value</td>
</tr>
<tr>
<td>Feature file</td>
<td>True</td>
<td>Comma separated value feature file produced</td>
</tr>
<tr>
<td>Recording limit</td>
<td>45 min</td>
<td>Stop recording at specified time.</td>
</tr>
<tr>
<td>Voiced frames only</td>
<td>True</td>
<td>Apply only voiced frames to recording limit.</td>
</tr>
</tbody>
</table>

\(^a\)Dropped samples may result from frame lengths less than 50 ms due to smartphone processing limitations.

**Note.** LH = low to high; f0 = fundamental frequency.
Appendix B
Decision Tree for Real-Time Voice Activity Detection

Real-time voice activity detection decision tree. The accelerometer signal is analyzed in nonoverlapping frames of 50 ms in duration. All thresholds are modifiable in the application settings menu. If any thresholds are not satisfied, the frame is considered nonvoiced. LH ratio = low-to-high frequency spectral power ratio; \( f_0 \) = fundamental frequency; \( T_A \) = time lag (ms) of highest non-zero correlation coefficient; \( T_B \) = time lag (ms) of highest correlation coefficient at approximately half the duration of \( T_A \).