JSLHR

Clinical Focus

Quantitative Assessment of Learning and Retention in Virtual Vocal Function Exercises

Jarrad H. Van Stan,[a,b,c] Se-Woong Park,[d] Matthew Jarvis,[e] Joseph Stemple,[f] Robert E. Hillman,[a,b,c] and Dagmar Sternad[g]

Purpose: Successful voice therapy requires the patient to learn new vocal behaviors, but little is currently known regarding how vocal motor skills are improved and retained. To quantitatively characterize the motor learning process in a clinically meaningful context, a virtual task was developed based on the Vocal Function Exercises. In the virtual task, subjects control a computational model of a ball floating on a column of airflow via modifications to mean airflow (L/s) and intensity (dB-C) to keep the ball within a target range representing a normative ratio (dB × s/L).
Method: One vocally healthy female and one female with nonphonotraumatic vocal hyperfunction practiced the task for 11 days and completed retention testing 1 and 6 months later. The mapping between the two execution variables (airflow and intensity) and one error measure (proximity to the normative ratio) was evaluated by quantifying distributional variability (tolerance cost and noise cost) and temporal variability (scaling index of detrended fluctuation analysis).
Results: Both subjects reduced their error over practice and retained their performance 6 months later. Tolerance cost and noise cost were positively correlated with decreases in error during early practice and late practice, respectively. After extended practice, temporal variability was modulated to align with the task's solution manifold.
Conclusions: These case studies illustrated, in a healthy control and a patient with nonphonotraumatic vocal hyperfunction, that the virtual floating ball task produces quantitative measures characterizing the learning process. Future work will further investigate the task's potential to enhance clinical assessment and treatments involving voice control.
Supplemental Material: https://doi.org/10.23641/asha.13322891

Successful voice therapy requires the patient to modify or learn new vocal behaviors, but little is currently known regarding how humans with or without voice disorders learn new vocal motor skills. One reason is that studies investigating vocal motor learning mainly focused on perturbation or adaptation of well-learned (i.e., habituated) vocal behaviors, for examples, sustained vowels, glissandos (Larson et al., 2000; Stepp et al., 2017; Zarate et al., 2010), syllables, or speech (Chen et al., 2007; Guenther et al., 2006). Also, the design of clinical voice treatment studies has largely focused on average differences between isolated time points, for example, before versus after surgery/voice therapy (Ramig & Verdolini, 1998). While the results of treatment studies provide empirical support for the effectiveness of voice treatments, they rarely offer insights into how patients learn and improve behaviors. Of additional interest is how long the measured improvements will last after discontinuing therapy, that is, carryover or retention.

The field of motor control and learning is rich with theories attempting to quantify how the central nervous system (CNS) controls, adapts, and learns new movements. In particular, studying how humans learn motor tasks with redundancy—tasks with infinitely many ways to achieve success—has the potential to offer insights into how people establish new vocal motor behaviors (Cusumano & Cesari, 2006; Müller & Sternad, 2009; Scholz et al., 2000). Redundant motor tasks can be described by how execution variables relate to the desired result of the task, for example, error

[a]Massachusetts General Hospital, Boston
[b]Harvard Medical School, Boston, MA
[c]MGH Institute of Health Professions, Boston, MA
[d]University of Texas, San Antonio
[e]Newark, DE
[f]University of Kentucky, Lexington
[g]Northeastern University, Boston, MA
Correspondence to Jarrad H. Van Stan: jvanstan@mgh.harvard.edu

Editor-in-Chief: Bharath Chandrasekaran
Editor: Jack J. Jiang
Received June 23, 2020
Revision received September 3, 2020
Accepted September 17, 2020
https://doi.org/10.1044/2020_JSLHR-20-00357

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

from a target. When multiple execution variables map into one result variable, this creates an infinite number of combinations of execution variables that achieve a given result or error value. For example, "vocal efficiency" can be considered a redundant motor behavior as subglottal pressure, mean airflow, and acoustic sound pressure level (SPL; multiple execution variables) during voicing can all differ to create the same "target" level of vocal efficiency (zero error; Holmberg et al., 1988). Mathematically speaking, the relation between execution variables and the result creates a null space or a manifold of solutions—comprising those executions that all lead to zero error. The vocal rehabilitation example may illustrate this task analysis. The solution manifold represents all the different ways—that is, all the different combinations of SPL, mean airflow, and subglottal pressure—with which a patient can achieve the desired vocal efficiency. This normative relation between execution and result variables predicts vocal efficiency from the combination of the three execution variables subglottal pressure, mean airflow, and SPL.

In order to systematically study such motor tasks in a controlled quantitative manner, several lines of research in the area of human motor control have developed virtual environments where the physics of the task is mathematically modeled and fully defined. Then, the solution manifold is exactly defined and can be derived either analytically or numerically. Data can be analyzed against the derived solution manifold. For example, Müller and Sternad developed an experimental paradigm that modeled a throwing task so that a small set of execution variables fully determine the error (Müller & Sternad, 2004). Using this approach, investigators could quantify how subjects learned the motor skill by relating practice-based performance improvements (reduction in error) to changes in the execution variables. This study will take this methodological approach to quantify learning of a vocal motor skill.

An extensive area of research in motor learning is also dedicated to investigating the structure of variability as an avenue to gain insight into learning processes. A decrease in variability with practice is a characteristic feature of any improvement. However, variability can not only decrease, but it can also change in structure. "Structure" refers to the different types of distributions it can have (e.g., Gaussian, anisotropy) or the different types of temporal changes it can take (e.g., Brownian motion, pink noise; Ajemian et al., 2013; Faisal et al., 2008; Sternad, 2018; Sternad et al., 2014). When considering performance in the context of execution and result variables, analyzing the variability of those execution and result variables can shed light on control strategies of the CNS. Sternad and colleagues have developed multiple metrics that assess how distributional and temporal variability among execution variables directly affects resulting performance (Abe & Sternad, 2013; R. Cohen & Sternad, 2009; Van Stan et al., 2017). Changes in distributional characteristics over the course of practice and learning are quantified by tolerance cost (T-Cost) and noise cost (N-Cost). T-Cost quantifies how subjects find the most error-tolerant solutions for their performance. N-Cost quantifies how subjects modify the dispersion of their distribution to minimize error. As shown in previous literature, subjects first improve performance exponentially through finding an error-tolerant space on the solution manifold for the multidimensional distribution of their execution variables (T-Cost). Further improvement proceeds at a slower time scale by modifying the dispersion of their execution variables to align with the solution manifold (N-Cost).

Temporal variability, or changes from trial to trial in the execution variables, is quantified by the detrended fluctuation analysis (DFA) scaling index (SCI) that estimates how the patient is error-correcting (or not) across all directions of the execution space. Specifically, "control" (or error correction) is indicated when the motor output is stabilized around a set value (i.e., if a measure increases, the next moment it will decrease). Decreases in SCI values represent increases in control. When applying DFA SCI to learning novel movements, variability among execution variables tends to have minimal directional preference in execution space during the early stages of practice (Abe & Sternad, 2013; Van Stan et al., 2017). However, later in practice, variability is selectively channeled into error-irrelevant directions—parallel to the solution manifold, where variability does not affect error. Variability in error-relevant directions—orthogonal to the solution manifold—is reduced as it can obviously affect error. In other words, the sensorimotor system may not be primarily concerned with simply decreasing variability, but selectively channeling variability according to the task demands.

Recently, a virtual throwing paradigm with redundancy was adapted from the motor skill literature to vocal motor learning and replicated results from the limb-based literature (Van Stan et al., 2017). Specifically, 10 vocally healthy subjects practiced throwing projectiles at a target with a sling, controlled via modifications in fundamental frequency and vocal intensity. The two-to-one mapping between the execution variables (frequency and intensity) and error (distance between the projectile and the target) was evaluated using analyses that quantified distributional and temporal properties of the subjects' variability. The result of this study indicated that vocal motor learning is very similar to limb motor learning—especially how variability is modified over practice to reduce the overall error. This result is significant as findings from motor control/learning studies based on limb movements often do not transfer to bulbar movements/skills (e.g., speech, swallowing, voice); exceptions are perturbation or adaptation paradigms (for reviews, see Bislick et al., 2012; Maas et al., 2008).

The purpose of this study was to extend the same methodology into a clinical context. We chose two treatment components from the Vocal Function Exercise (VFE) program: the Sustained Vowel Exercises 1 and 4 (Stemple et al., 1994). Inspired by the clinically used "flowball" device, this study developed a computational model of a floating ball and rendered it into a virtual interactive exercise (Lã et al., 2017). Specifically, this study will (a) describe the newly developed voice-controlled virtual task and (2) demonstrate through two case studies—one with a normal voice

and one with nonphonotraumatic vocal hyperfunction (NPVH; Hillman et al., 2020)—that the task can be learned. Through demonstrating that the task can be learned, this focus article aims to evaluate whether (a) subjects can significantly reduce their error in the task; (b) if so, how long it takes to minimize error; and (3) whether the distributional and temporal variability metrics from previous work generalize to the floating ball task. The governing institutional review board approved all experimental aspects related to the use of human subjects for this study.

Method

VFE Program

The VFE treatment protocol will be partially described according to the Rehabilitation Treatment Specification System (RTSS; Hart et al., 2019; Van Stan et al., 2019). Specifically, the RTSS labels the smallest unit of treatment as a "treatment component" consisting of three parts: (a) the ingredient(s)—clinician actions intended to change a specific patient function, (b) the target—the singular patient function directly modified by the clinician action(s), and (c) the mechanism(s) of action—the observed or hypothesized way that the ingredients affect the target. The VFE program is a standardized voice treatment approach that has demonstrated effectiveness in the habilitation of subjects with normal voices and rehabilitation of patients with a variety of voice disorders (Angadi et al., 2019). The virtual floating ball task was designed to represent two treatment components from the VFE program (Exercises 1 and 4) that both have the same target. Exercise 1 includes two practice trials of sustained voicing on the musical note F4, and Exercise 4 includes 10 practice trials of sustained voicing at five musical notes in increasing order: C4, C4, D4, D4, E4, E4, F4, F4, G4, and G4. Correct production of a practice trial includes (a) engaged voicing (chest > head registration), (b) at a soft intensity (low decibels/dB), (c) using an inverted megaphone posture (closed lips and open pharynx), (d) increased forward resonance (i.e., increased vibrotactile sensations in the facial mask), (e) maximal abdominal-based inhalation before beginning each practice trial, and (f) voicing until all air is exhaled, despite any vocal instabilities or breaks. The target of both exercises is to sustain each practice trial for a desired duration range (seconds), where the upper and lower limit for the target's duration range is equal to the individual subject's vital capacity in liters (L) divided by 0.08 L/s and 0.1 L/s, respectively. The VFE normative flow range was based on normative data (Hirano & McCormick, 1986b). Describing the entire VFE protocol is outside the scope of this clinical focus article (Stemple, 2005).

Experimental Setup

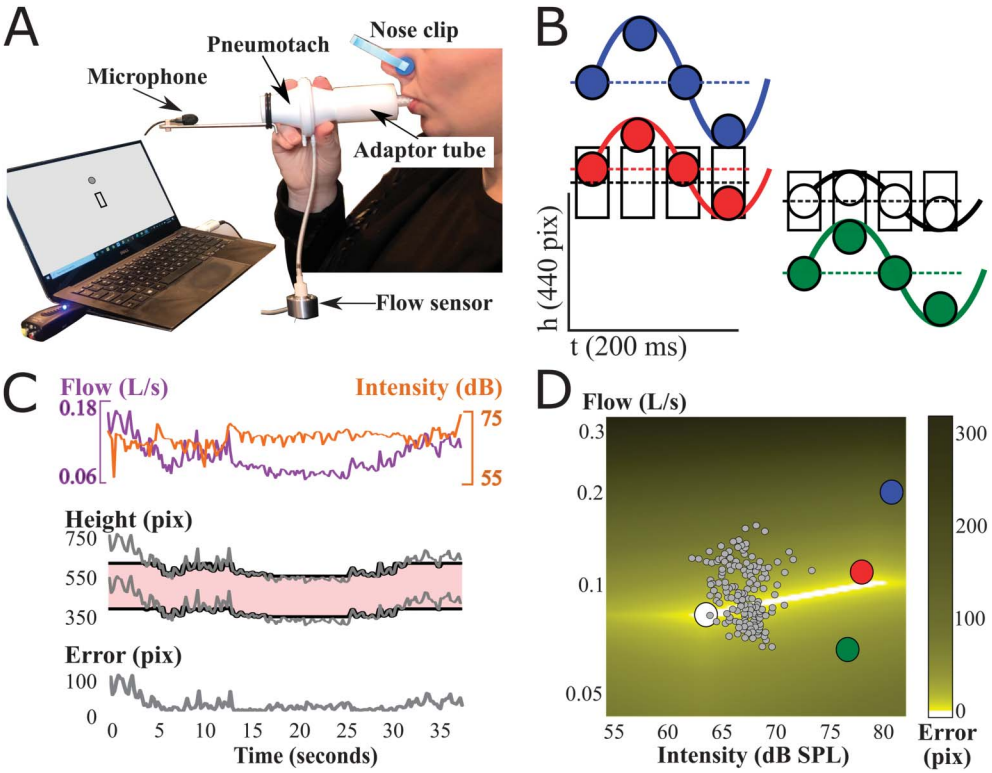
Figure 1A shows the experimental setup for playing the virtual floating ball task. Each subject's airflow was recorded using a customized pneumotachograph (Phonatory Aerodynamic System, Model 6600, PENTAX) and a Glottal Enterprises flow sensor (pressure transducer: PT-2E, and Model MS-110, Glottal Enterprises) attached to the pressure tap most proximal to the subject's mouth. Vocal intensity was recorded using a unidirectional condenser microphone (MKE104, Sennheiser Electronic GmbH) placed 10 cm from the distal end of the pneumotachograph. The mouth-to-microphone distance is kept constant during game play as the participants wrap their lips around a tube; the microphone is at the end of the tube, yielding a fixed distance between mouth and microphone of 10 cm. Both signals were input into a laptop (XPS 13, Dell) running Windows 10 (Microsoft). The native laptop sound card was bypassed by using an external soundcard adaptor (ICUSBAUDIO, StarTech). The dimensions of the custom-made adaptor tube that connected to the pneumotachograph were chosen to ensure fidelity with theoretically important aspects of the VFE program; specifically, the /o/ vowel and its associated semi-occluded vocal tract posture. The outer diameter of the tube was 12 mm, consistent with MRI studies on the diameter of lip opening for /o/ vowels (e.g., see Story et al., 1998). The inner diameter of the tube was 9 mm, based on the commonly used LaxVox tube for semi-occluded vocal tract treatments (Andrade et al., 2016).

Before starting a session, the signal from the pressure transducer (flow sensor) was calibrated for estimating airflow in units of liters per second (L/s) using reference airflow levels (MCU-4 Pneumotach Calibration Unit, Glottal Enterprises). The acoustic signal was calibrated using two complex tones at increasing intensity levels measured by a Class 2 sound-level meter (NL-20, RION) to map the uncalibrated voltage signal to units of pascal and C-weighted decibels (dB C) at 10 cm. The software processed both airflow and microphone signals (recorded with a 10-kHz low-pass filter, 22050-Hz sampling rate, and 16-bit quantization) every 50 ms to produce quasi-real-time estimates of mean airflow (L/s) and vocal intensity (dB C). The MS-110's amplitude modulation (8-kHz carrier frequency) was used to preserve the offset in the flow signal, and the signal was demodulated in real time by the custom virtual task software on the laptop. Subjects wore a noseclip to prevent airflow through the nose during voicing.

Floating Ball Task

In the virtual floating ball task, we developed a task model such that the ball dynamics were fully determined by the subject's vocal intensity α (ball oscillation amplitude in pixels) and mean airflow β (mean ball height/offset in pixels). The task model was initially based on the aerodynamics of a floating ball activity in which vertical airflow underneath a ping-pong ball makes the ball float and oscillate vertically (Lã et al., 2017). Equation 1 reflects the aerodynamic properties between the airflow and mean ball height; the linear relationship is also experimentally shown in another publication (Lã et al., 2017). Because vocal intensity has a complex relationship with airflow (Tanaka & Gould, 1983), we simplified the model such that vocal intensity exclusively affects the ball oscillation amplitude. Specifically, the ball position along the *y*-axis (y)—oscillating

Figure 1. (A) Subjects control a computational model of a floating ball by modifying their mean airflow and intensity during sustained voicing. (B) Four examples illustrate how mean flow (ball height; colored dotted horizontal lines) and intensity (ball amplitude; amplitude of the colored sine waves) control the floating ball (200-ms samples, each box = 50-ms time frame). The target boxes show the desired mean flow (height, black dotted horizontal lines) and intensity (distance between the bottom and top of the box) based on the normative ratio. (C) One Vocal Function Exercise trial lasting approximately 35 s in the virtual environment. Top panel: purple line (mean flow); orange line (vocal intensity). Middle panel: gray lines = top/bottom of the ball oscillation produced by the subject's flow and intensity. Black lines and shading = the top/bottom of the target box. Bottom panel: error (in pixels) every 200 ms. (D) The execution space indicates the amount of error for all combinations of airflow and vocal intensity. The four examples in B are represented by large colored circles, and the example time series in C is represented by small gray circles (one circle = 200 ms).



at a frequency (f) of 2 Hz (La et al., 2017)—at any point in time (t) during a sustained voicing trial is determined by α and β according to Equation 1:

$$y(t) = \alpha_i \cos(2\pi f t) + \beta_i. \quad (1)$$

The target box for the floating ball task was designed to reflect the treatment target for the VFEs 1 and 4. Specifically, the height of the target box (distance between the top and bottom of the box) is determined by a normative ratio of vocal intensity divided by mean flow (box vertical position) within the mean flow range of 0.08–0.1 L/s. The ratio was chosen to represent the treatment target because it theoretically corresponded to a "normal" relationship between mean airflow and SPL, which is related to vocal efficiency (i.e., Colton et al., 2006; Hirano & McCormick, 1986a; Holmberg et al., 1988) and correct vocal production of the exercise (e.g., chest > head registration, forward resonance). To acquire the normative vocal intensity per mean flow value, a previously acquired database was analyzed consisting of 23 vocally healthy females (endoscopically verified) who produced soft, sustained /o/ vowels. Ratios of vocal intensity divided by mean flow were only calculated for the 50-ms frames that were between 0.08 and 0.1 L/s. This resulted in a normally distributed histogram, where the mean ratio was approximately 800 dB*s/L. Of note, this ratio approximates findings from previous studies in normative female populations where the mean ratio of minimum vocal intensity (since the task is to be done as softly as possible) divided by the minimum mean flow was 780 dB*s/L. The values used to obtain the overall mean were: 744 dB*s/L at normal fundamental frequency and 800 dB*s/L at high fundamental frequency (both at normal vocal intensity; Holmberg et al., 1989), 651 dB*s/L at soft vocal intensity, 743 dB*s/L at normal intensity, and 963 dB*s/L at loud intensity (both at normal fundamental frequency; Holmberg et al., 1988).

During game play, when the subject produces voicing within the desired flow range, the target box dynamically moves up and down on the screen to represent the subject's mean airflow. The box height dynamically increases and decreases representing the desired (not subject produced)

vocal intensity, that is, the mean flow multiplied by 800. When the subject produces a mean flow above or below the 0.08–0.1-L/s zone, the ball travels above or below the target box because the target box remains static at a position offset from the bottom of the screen. This position is equal to 0.1 L/s or 0.08 L/s, and the target box height remains equal to 80 dB or 64 dB, respectively. Therefore, the target box's vertical position offset from the bottom of the screen (representing the desired mean flow) only moves up and down if the subject voices within the desired flow range of 0.08–0.1 L/s, that is, zero error is only possible within this range of mean flow. Figure 1B illustrates hypothetical examples of the ball oscillating above the box (red and blue examples) and below the box (green example). Figure 1C shows an example trial where the subject exhibited a behavior causing the ball to be above the box at the beginning of the trial (0–2 s) and below the box in the middle of the trial (17–25 s).

Error (E) was evaluated in pixels at all points in time (t) by calculating the Euclidean distance between the subject-produced mean ball height (β_S) and mean oscillation amplitude (α_S) versus the target box's height and distance between the upper/lower bounds (β_T and α_T, respectively).

$$E(t) = \sqrt{(\beta_S - \beta_T)^2 + (\alpha_S - \alpha_T)^2}\,. \quad (2)$$

Error was calculated as the mean of four consecutive analysis frames (200 ms) with 0% analysis frame overlap. The analysis frame length of 200 ms was chosen since it is approximately the shortest time duration associated with volitional control/error correction (Kandel et al., 2000). The 0% overlap between analysis frames was chosen to reduce correlation across windows. Since the virtual environment allowed a direct mathematical mapping between the two execution variables (mean airflow and vocal intensity) and the resulting error, this error could be portrayed with a color code on a two-dimensional execution space (see Figure 1D).

For each trial, a simple voice activity detection algorithm was implemented to determine when voicing began and ended. Specifically, a trial automatically began and ended when six consecutive frames or four consecutive frames of mean airflow were above or below the threshold of 0.02 L/s, respectively. Also, the experimenter monitored the real-time values during game play to constantly evaluate if the flow signal did not return to < 0.01 ml/s between trials. When the flow signal did not return to < 0.01 ml/s, it was recalibrated before a subject continued with the next trial. During piloting, this occurred only twice during game play. To remove nonstationarities associated with vocal onsets and offsets, all trials were analyzed for error and variability metrics after the first and last second of voicing were removed.

Participants

Two female participants were consented for participation in this study. One participant had no history of a voice disorder, and the second participant had a diagnosis of NPVH. Both were female because NPVH is known to be sex-imbalanced—females account for over 70% of the patients seen for NPVH (Coyle et al., 2001; Kridgen et al., 2020)—and the virtual environment requires sex-specific normative values for mean flow during voicing (Holmberg et al., 1988).

A patient with NPVH was chosen for this study because (a) NPVH is one of the most commonly treated voice disorders (Bhattacharyya, 2014), (b) NPVH is believed to be predominantly behaviorally based (Hillman et al., 1989), (c) multiple studies have demonstrated that VFEs are effective with this patient population (Angadi et al., 2019), and (d) patients with NPVH have anatomically normal vocal folds (Hillman et al., 2020); that is, physiologically, they should be capable of matching the game's normative mean flow–vocal intensity ratio. The NPVH diagnosis was based on a complete team evaluation by laryngologists and speech-language pathologists (SLPs) that included endoscopic visualization of the larynx, objective measures of voice production (aerodynamic and acoustic), auditory-perceptual judgments of voice quality using the Consensus Auditory Perceptual Evaluation–Voice (CAPE-V; Kempster et al., 2009), and patient-reported ratings of voice difficulty using the Voice-Related Quality of Life (V-RQOL) scale (Hogikyan & Sethuraman, 1999). The patient's voice quality (CAPE-V) and self-reported quality of life (V-RQOL) were assessed before playing the game for the first time. These subjective scales are reported only for the purpose of generally describing the severity level of the patient, not for statistical analysis or results reporting. Therefore, reliability was not addressed. V-RQOL scores are normalized ordinal ratings that lie between 0 and 100, with higher scores indicating a higher voice-related quality of life. CAPE-V scores are visual analog scale ratings that range from 0 to 100, with 0 indicating normality and 100 indicating the most extreme example of deviance for a particular voice quality characteristic. The CAPE-V measurement for the patient with NPVH came from one rater—a voice-specialized SLP's single rating based on the CAPE-V standard reading and sustained vowel samples. According to the clinical evaluation, the patient was of mild severity. Specifically, her voice quality was 13 (overall dysphonia), 8 (roughness), 5 (breathiness), and 10 (strain), and her V-RQOL was 85 out of 100 possible points. Laryngeal videostrobosopy identified mild anterior–posterior constriction only during voicing. Aerodynamic measures of estimated subglottal pressure were increased for both comfortable and loud phonation (8.68 cm-H_2O and 13.57 cm-H_2O, respectively) compared to normative data. All acoustic measures were within normal limits (i.e., fundamental frequency, vocal intensity, and cepstral peak prominence).

To qualify for study participation, the vocally healthy subject passed an in-person screening with a voice-specialized SLP. The screening contained the following questions, and all questions had to be answered negatively: (a) Have you ever seen a medical professional for your voice? (b) Have you had any recent hoarseness or voice loss? (c) Does the SLP perceptually note any dysphonia?

Study Design

Both subjects completed a total of 13 sessions: 11 practice sessions that could be maximally separated by 3 days and two long-term retention sessions at 1 month and 6 months after Practice Session 11. Each session lasted approximately 20–30 min, included 18 repetitions of sustained voicing in the virtual environment (two practice repetitions of Exercise 1, 10 practice repetitions of Exercise 2, three additional practice repetitions of the highest note G4, and three additional practice repetitions of the lowest note C4). Before the first practice session, a licensed SLP with expertise in voice instructed the subjects to, for every trial, (a) take a maximally deep breath and voice, (b) as softly as possible, (c) without stopping (even if their voice dropped out), until all air was exhaled. The SLP provided the subjects with examples of different ways of voicing with associated labels of high error (pressed, breathy, rough, fry, cul-de-sac resonance) and low error (forward resonance). Subjects were instructed to try to voice in a way that made the ball oscillate inside the box and hit the top/bottom edges of the box. This would make the ball turn white, representing zero error. The SLP was present during all sessions in case the trial needed to be stopped and repeated due incorrect execution of an exercise. An individual exercise was considered incorrect if the subject did a shallow breath before starting the trial, produced loud voicing (> 80 dB-C), strayed from the required pitch, stopped the prolonged voicing before running out of air, or if the subject produced multiple consecutive seconds of moderate (or worse) roughness, strain, breathiness, or nonmodal phonation (e.g., vocal fry, diplophonia). When an exercise was considered incorrect by the SLP, the subject was stopped, informed of what was incorrect, and asked to start the trial again. The vocally healthy subject never produced a grossly erroneous trial and was never stopped and asked to restart a trial. The patient with NPVH was stopped 4 times during Practice Session 8, and all occurrences were for voicing too loudly.

Analysis of Distributional Variability

Figure 2 illustrates the distributional variability metrics of T-Cost and N-Cost. Of note, there is one other cost that was not used: covariation cost (R. Cohen & Sternad, 2009). Due to the simple, linear solution manifold for the floating ball task, N-Cost captured most and, in some cases, all of the covariation between variables. This resulted in zero, or practically zero, covariation cost values.

T-Cost captures a cost due to the data not being at the best "place" in the execution space. T-Cost is estimated by generating an optimized data set in which the mean vocal intensity and flow were shifted in execution space to the location yielding the best overall result (Sternad et al., 2014). More specifically, the execution space was parsed into a grid of 1500 × 1500 points (the boundaries of this grid were determined by the limits of the task). The data set is then shifted across this grid through every possible center point and evaluated its mean result at each location. Note that the dispersion in execution space is preserved during this process. When data points extended beyond the grid limits, the values were calculated on the extrapolated execution space. The location that produced the best (lowest) overall mean error was compared to the actual data set. The algebraic difference between the actual mean error and the optimal mean error defined T-Cost and expresses how much the data could have improved its performance if it had been at a different location in execution space.

N-Cost is the cost to overall performance due to nonoptimal stochastic variability in the execution space. N-Cost is estimated by generating an optimized data set in which variability is reduced in a step-wise manner to achieve the least possible mean error, while leaving overall mean vocal intensity and flow unchanged. Though one would expect that all data sets should be best when reduced to a single point (the mean vocal intensity and flow), this expectation does not hold as a data set with a small distribution may produce the lowest mean error depending on the geometry of the solution space. In the numerical procedure, the radial distance for every data point to its mean was divided into 100 steps. Then, all data points were shrunk toward their mean at 1% intervals, and the mean error was evaluated at each interval. The algebraic difference between the mean of the interval that produced the lowest mean error (optimized data set) and the original data set defined N-Cost. This value expresses how much the data could have improved if only their dispersion had been reduced.

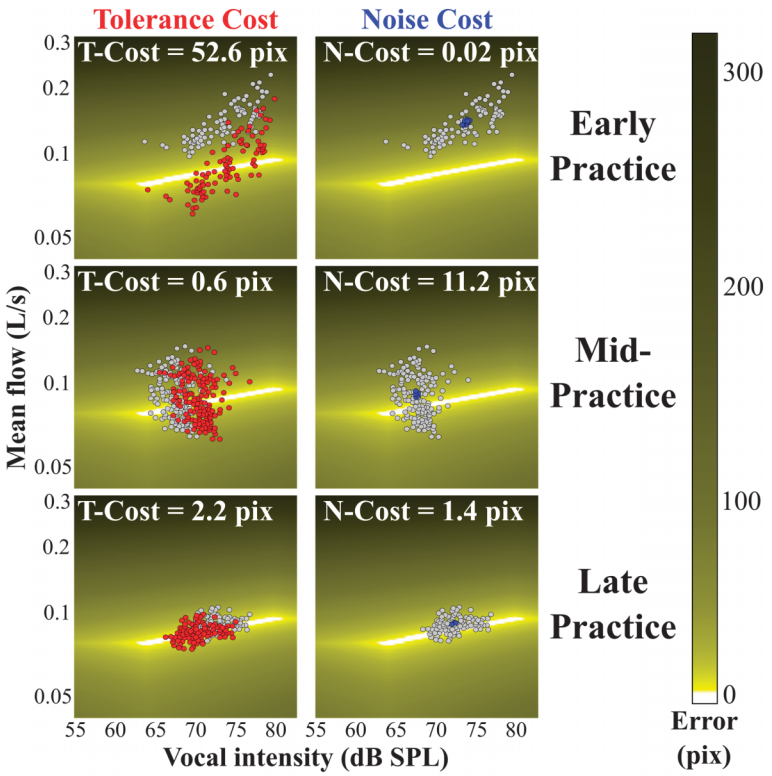
Analysis of Directional Variability in Execution Space

A second type of analysis focuses on temporal variability in the execution space as illustrated in Figure 3. This analysis was developed in previous studies (Abe & Sternad, 2013; Van Stan et al., 2017). To investigate the temporal structure in vocal intensity and mean flow, the two axes of the execution space had to be normalized since they have different units. To this end, the data for each 50-ms analysis frame (per individual sustained vowel exercise) was transformed into z scores according to the mean and standard deviation of mean flow and vocal intensity for that exercise. To evaluate whether trial-to-trial variability was channeled into preferential directions on the solution manifold, the two-dimensional data of each block were projected onto a single line through the center of the data set using the following equation:

$$x_\theta(i) = x_1(i)\cos\theta + x_2(i)\sin\theta. \quad (3)$$

The trial index is i, and x_θ (i) denote the new time series after projection onto the line. The variables x_1 and x_2 denote the z score of vocal intensity and mean flow, respectively. The angle θ of this line was zero when parallel to the horizontal x_1-axis (variability of vocal intensity, denoted as a black line in Figure 3A) and 90° when parallel to the vertical x_2-axis (variability of mean flow). The center

Figure 2. Example and optimized sets of three practice trials from one normal subject. The left column shows data optimized in terms of tolerance cost (T-Cost), the right column shows data optimized in terms of noise cost (N-Cost). Gray circles represent 200-ms segments of the subject's actual sustained voicing, and the red/blue circles represent surrogate data with one component optimized (T- or N-Cost, respectively). The top, middle, and bottom panels show data from Practice Days 1, 4, and 10, respectively.



of the data was defined by the mean of vocal intensity and mean flow for each sustained vowel trial for each individual. This line was then rotated through 180° in 180 steps. At each rotation angle θ, the data were projected onto the line and the time series of the projected data was evaluated using the DFA. The angle of the direction parallel (error-irrelevant) to the solution manifold was defined as θ-PAR, and the direction orthogonal (error-relevant) to the solution manifold was defined as θ-ORT (denoted as red lines in Figure 3A).

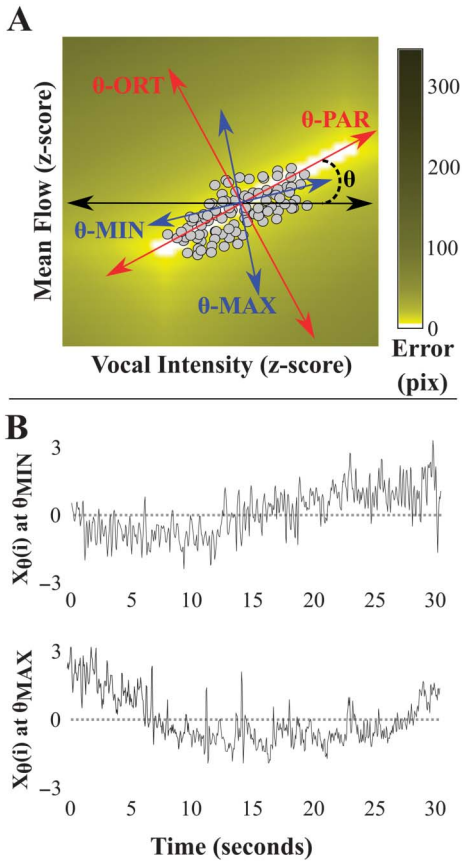
Analysis of Temporal Variability

The temporal structure of $x_\theta(i)$ obtained for all angles θ was evaluated by DFA. This analysis method has been chosen because it provides statistical quantifications of temporal persistence (when future fluctuations are likely to be in the same direction as current fluctuations) and antipersistence (when currently observed fluctuations are in the opposite direction of future fluctuations) on longer time scales (Peng et al., 1995). The DFA is a modification of the root-mean-square analysis of a random walk that is relatively insensitive to nonstationarities and noise in the data (Peng et al., 1995). Specifically, the time series was cumulatively summed to obtain an integrated signal and was then detrended with linear regression within windows of a number of trials n. The root-mean-square of the detrended time series $F(n)$ was then calculated for windows of n trials. Plotting $F(n)$ versus n in log-log coordinates, the DFA SCI was obtained from the slope of a linear regression (Peng et al., 1995). Each sustained vowel trial (between 500 and 800 data points; corresponding to 25–40 s) was used in the analysis of directionality in execution space. Temporal variability was classified as either uncorrelated white noise (SCI = 0.5), antipersistence denoting stable dynamic behavior and error correction (SCI < 0.5), or persistence denoting potentially unstable dynamic behavior and lack of error correction (SCI > 0.5).

Statistical Analysis

Performance improvement across practice was evaluated by fitting exponential functions to the error: $y = a\, e^{-bx} + c$, where y denotes the error and x denotes time; a, b, and c are fitting constants. Fits were performed for each participant, calculated from absolute mean values of 18 trials per practice session. Retention was assessed at 1 and 6 months with a "savings" score (Schmidt & Lee, 2011). Specifically, a subject would be considered to have demonstrated retention if the number of practice trials to reach

Figure 3. (A) Execution space and the rotation axes used to analyze temporal fluctuations in different directions. The black line shows 0° (*x*-axis). The red lines show directions parallel (PAR) and orthogonal (ORT) to the solution manifold (i.e., the white space). The gray data points represent 200-ms nonoverlapping analysis frames from one sustained voicing trial. The one-dimensional time series $x_\theta(i)$ is obtained by projecting the two variables onto the direction as expressed in Equation 3. The angles θ associated with the highest (MAX) and lowest (MIN) scaling index are represented by blue lines. (B) The time series at the top shows $x_\theta(i)$ for angle θ with the minimum scaling index. The bottom time series shows $x_\theta(i)$ for angle θ with the maximum scaling index.



asymptotic performance during retention was less than the number of practice trials to reach asymptotic performance during early practice. The original design was to have subjects complete retention testing for as many days as needed to attain an average error as on Practice Day 11 (± 1 *SD*). However, only 1 day of retention testing was needed for both subjects at the 1- and 6-month time points.

The individual contribution of each cost toward error reduction was evaluated through a Pearson correlation coefficient (T-Cost or N-Cost vs. mean error) for the first 5 and last 5 days of practice representing early and late practice, respectively. Previous work in virtual tasks has shown that T-Cost contributes to the initial decreases in error and N-Cost contributes to longer-term reductions of error later in practice. Therefore, it was hypothesized that T-Costs and N-Cost's correlation with error would change from early to late practice: The contribution of T-Cost would decrease with practice days, while N-Cost would increase with practice days.

Changes in the subjects' sensitivity to directions in the execution space were assessed as the mean values from the first and 10th days of practice. The daily mean SCI was calculated for all directions, derived from 18 individual trials per day; approximately 500–800 frames (50 ms) per practice trial. One-tailed paired *t* tests were used to assess the difference between the mean directions and mean amplitudes of SCI at θ-MIN (the lowest SCI among all directions) and θ-MAX (the highest SCI among all directions) for Practice Session 1 versus 10. If subjects developed sensitivity to how trial-to-trial fluctuations related to error, the directions associated with the largest reductions in SCI (Practice Day 10 minus Practice Day 1) will be near to θ-ORT. Additionally, the directions with minimal changes (or even increases) in SCI (Practice Day 10 minus Practice Day 1) will be near to θ-PAR. This will be interpreted as increased CNS control, over the course of practice, in the θ-ORT compared to θ-PAR.

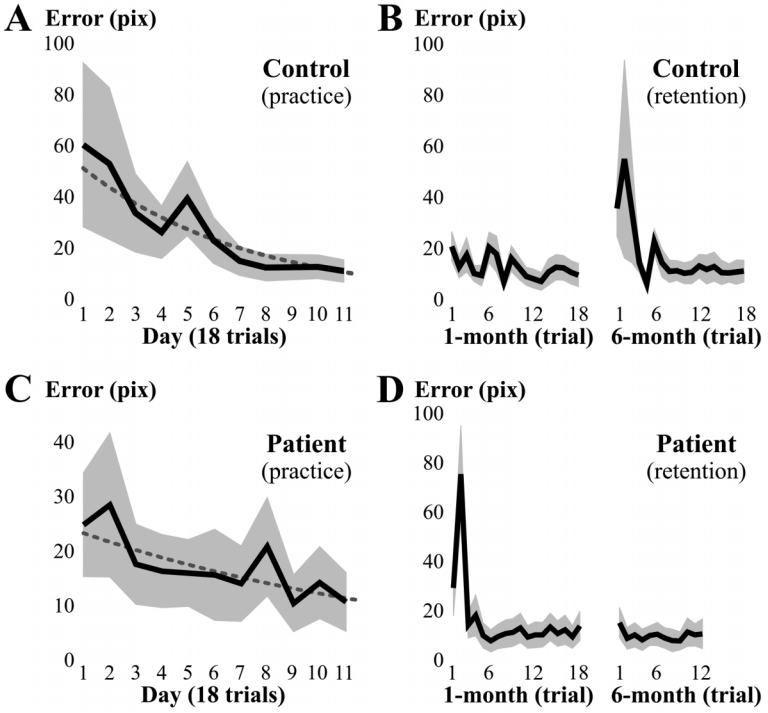
Cohen's *d* was used as an effect size metric for all statistically significant pairwise comparisons such that effect sizes less than 0.20 were interpreted as small, between 0.20 and 0.80 as medium, and greater than 0.80 as large (J. Cohen, 1988). All statistics were calculated using SPSS software (Version 22.0, IBM).

Results

Figure 4 shows the progression of error for the two female subjects across practice. The figure includes performance at the long-term retention tests. Both subjects significantly reduced their error over practice. The exponential fits have r^2 values of .9 (control) and .6 (NPVH). The control subject exhibited a mean (standard deviation) error of 60.4 (36.6) pixels and 11.1 (1.7) pixels on Practice Days 1 and 11, respectively. The patient with NPVH exhibited 24.9 (10.4) pixels and 10.7 (2.8) pixels on Practice Days 1 and 11, respectively. The control subject and patient both retained the new vocal behavior during the 1-month retention test (three trials vs. four trials to perform within 1 *SD* of Practice Day 11, respectively). Even further, they both retained performance after 6 months without practice (four trials vs. one trial to perform within 1 *SD* of Practice Day 11, respectively).

These two case studies give first evidence that the virtual task has an adequate level of difficulty (not too easy or too hard). Both subjects took approximately 10 days to significantly reduce their error; neither of the subjects appeared to have "plateaued" by Practice Session 11 (indicating they were likely to improve with more practice), and both reported that the target vocal behavior was still challenging to perform accurately after 11 practice sessions. Although the average error metrics in Figure 4 shows smooth, exponential improvements over time, the first low-error vocalizations (the oscillating ball was white for

Figure 4. Left panels illustrate mean error (solid lines), exponential fit (dotted lines), and interquartile range (gray shading) per practice day (18 trials per day) for a vocally healthy control, (A) a patient with nonphonotraumatic vocal hyperfunction (C). Right panels illustrate mean error (solid lines) and interquartile range (gray shading) per trial during retention days at 1 and 6 months after the Practice Session 11 for a vocally healthy control (B) and a patient with nonphonotraumatic vocal hyperfunction (D). The patient only completed 12 out of 18 practice trials at the 6-month retention testing due to time limitations. Pix = pixels.

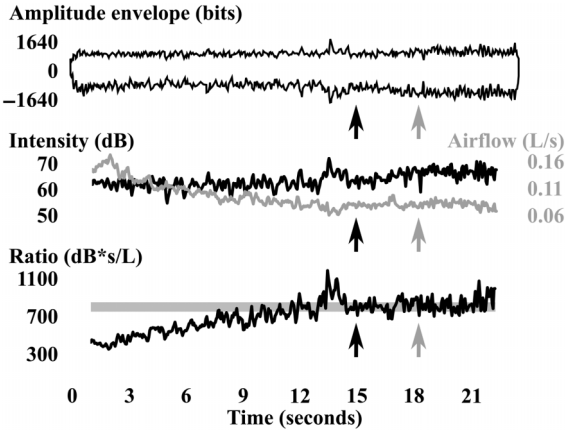


an extended period of time) often occurred unexpectedly/abruptly in the middle of trials. An example trial can be seen in Figure 5 and heard in a recording included in Supplemental Material S1. Auditory-perceptually, these initial low-error voicing segments/trials were obvious and sudden enough that the subjects would be visibly surprised by their mid-trial change in voicing. Specifically, from the SLP's perception, the sudden change was characterized by increased vocal intensity and modified resonance (often referred to as "forward resonance" in the VFE protocol; Stemple, 2005). After completing the study, the two participants were asked to describe how it felt when the target vocal behavior changed, that is, when the ball turned white. Both subjects qualitatively reported decreased physical effort to voice (i.e., vocal effort) as well as increased vibrotactile sensations in the nose (i.e., forward resonance) and soft palate area (i.e., the inverse megaphone, as the vocally healthy subject described the location as "where you yawn").

Figure 2 showed exemplary data from the control subject. Comparing Practice Sessions 1, 4, and 10 in Figure 2, it can be seen that the distribution of execution variables moved toward a more error-tolerant location on the solution manifold (T-Cost). In addition, the dispersion of the data cloud decreased and the anisotropy of the data changed to align with the solution manifold (N-Cost). As shown in Table 1, the correlation between error and T-Cost decreased in late practice (Sessions 7–11) versus early practice (Sessions 1–5) for both the control and patient subjects: early practice $r = .78$ and .75, late practice $r = .35$ and .51, respectively. Additionally, the correlation between error and N-Cost increased in late practice compared to early practice for both the control and patient subjects: early practice $r = .00$ and .27, late practice $r = .52$ and .63, respectively.

Figure 6 summarizes the results of the temporal analyses with respect to direction θ for the control and patient subjects (see Figures 6A and 6B). For both subjects in early practice, the SCI values for θ-MAX were closer to θ-ORT and for θ-MIN were closer to θ-PAR. Therefore, we tested if early practice SCI values at θ-MAX would decrease and SCI values at θ-MIN would either stay the same or increase in late practice. According to the paired t tests, the SCI values at θ-MAX were significantly lower during Practice Session 10 than Practice Session 1; control: $t(17) = 6.88$, $p < .001$, $d = 1.62$; patient: $t(17) = 1.98$, $p = .03$, $d = 0.47$. Also, neither of the subjects showed a change in the SCI values at θ-MIN. Therefore, the temporal correlation measure (SCI) demonstrated significantly different trial-by-trial dynamics between Practice Session 1 and Session 10, that is, movement toward more stable dynamics at θ-MAX near θ-ORT and no change in dynamics at θ-MIN near θ-PAR, indicative of changes in

Figure 5. Example of a single sustained vowel trial from the subject with a normal voice during Practice Session 3. This practice trial represents the first time the subject produced the desired ratio of vocal intensity and mean airflow for an extended period of time, which appeared suddenly and mildly at the black arrow (~15 s) and abruptly increased in accuracy at the gray arrow (~18 s). The upper panel shows the amplitude envelope of the acoustic waveform. The middle panel shows the mean acoustic vocal intensity (decibels, dB) in black and mean airflow (L/s) in gray every 50 ms. The bottom panel shows the subject-produced ratio of vocal intensity divided by mean airflow (black line) in relation to the target ratio (gray horizontal box – representing ratios from 750 to 850 dB*s/L). The audio file (WAV) for this trial can be heard in Multimedia 1.



selective control over practice depending on the direction in solution space.

Discussion

This study illustrated, in a healthy control and a patient with NPVH, that a newly developed and therapeutically based video game produced quantitative measures (T-Cost, N-Cost, SCI) capable of characterizing and quantifying the motor learning process. The correlation between error and the cost metrics evolved over practice as demonstrated in previous motor control and learning experiments. Specifically, performance improvements during the first half of practice were strongly associated with finding an error-tolerant area in the execution space, that is, T-Cost. Decreases in error during the last half of practice were significantly related to fine-tuning variability so that it aligned with the solution manifold, that is, N-Cost. Temporal variability (SCI) in the two-dimensional execution space illustrated how subjects' trial-to-trial behavior became more sensitive to error-relevant directions in the solution manifold. Finally, both subjects accurately performed the desired vocal skill after 6 months of no practice, lending credence to the task's potential utility in fostering long-term retention of a therapeutically desirable manner of voicing.

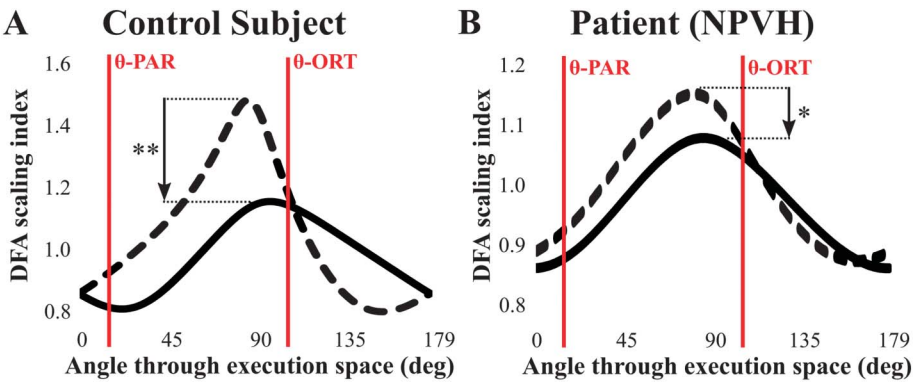
Table 1. Pearson correlation coefficient (*r*) between error and tolerance cost (T-Cost) or noise cost (N-Cost) per practice trial for the control subject and patient with nonphonotraumatic vocal hyperfunction (NPVH) during the beginning half of practice (Days 1–5) and the last half of practice (Days 7–11).

	Control		Patient with NPVH	
Practice time period	**T-Cost**	**N-Cost**	**T-Cost**	**N-Cost**
Correlation with error (*r*)				
Days 1–5	0.78	< 0.001	0.75	0.27
Days 7–11	0.35	0.52	0.51	0.63

The resulting subject data will now be discussed according to the RTSS's three-part treatment component (ingredients ➔ mechanisms of action ➔ target; Hart et al., 2019). The subject data provide initial support that virtual practice repetitions (Ingredient 1) combined with real-time feedback from the game (Ingredient 2) can be associated with acquisition and months-long retention of a therapeutically desirable vocal skill (the "target" of a ratio between vocal intensity and mean airflow). This vocal learning occurred despite the absence of many aspects of the VFE protocol. For example, an SLP traditionally delivering the VFE protocol would attempt to shape the patient's vocal practice by repeatedly providing "how to" instructions (verbal or physical models of anterior resonance, inverted megaphone position, chest registration, etc.) and verbal feedback on the accuracy of their vocal performance. However, the subjects in this study were only provided very general "how to" instructions once before the first practice session (e.g., demonstration of different ways to voice and the goal of the game) and no feedback on performance from the clinician. In fact, no specific instructions were given on the desired VFE behavior, for example, amount of desired airflow, chest versus head voice registration, an inverted megaphone posture, or verbal feedback on patient performance.

To investigate the mechanisms of action connecting the ingredients (practice and feedback) with changes in the target (ratio of vocal intensity and mean flow), distributional (costs) and temporal (SCI) variability metrics provided quantitative insights into how subjects reduced their error and how well the subjects implicitly dealt with the task's redundancy. The strong relation between T-Cost and error in early practice captured how the patient, when producing prolonged vowel trials, was closer/further away from the solution manifold. This could imply undesirable loudness (too soft or loud), undesirable mean flow (too much or too little), or a combination of the two. In auditory-perceptual terms, T-Cost and error correlations were highest when voicing was excessively soft (e.g., low vocal intensity) and/or degraded voice quality (breathiness, strain, fry, etc.). An increased correlation between N-Cost and error (as well as a decreased correlation between T-Cost and error) in late practice corresponded to the patient producing prolonged vowel trials that were centered on, and more or less shaped like, the solution manifold. This meant that her performance was in the desired mean airflow and loudness ranges, but she more or less frequently voiced at the desired ratio. In auditory-perceptual terms, the strongest relation between N-Cost and error occurred when practice fluctuated between voicing trials without forward resonance (i.e., on average, voicing was within the desired

Figure 6. Scaling index of the detrended fluctuation analysis (DFA) as a function of rotation angle in execution space. Practice Day 1 is shown by dotted lines, and Practice Day 10 is shown by solid lines (all days represent the mean of 18 trials). Downward arrows represent significant reductions in the maximum scaling index (i.e., increased stability/error correction). (A) Control data. (B) Patient data. *y*-axis scaling is different per panel. **Large effect sizes (Cohen's $d > 0.8$). *Medium-to-large effect sizes ($d > 0.5$). deg = degrees; NPVH = nonphonotraumatic vocal hyperfunction.



limits, but rarely at the desired ratio) and with some degree of forward resonance (i.e., on average, voicing was at the desired ratio). Finally, the practice sessions with no strong correlation between N-Cost or T-Cost and error occurred when the subject consistently produced vocal practice trials that were, on average, at the desired ratio (i.e., the best, or lowest error, practice sessions).

As seen in Figure 6, the directions in the execution space with the strongest modulations across practice were approximately orthogonal to the solution manifold, suggesting that the subjects specifically increased their sensitivity to the most error-salient directions. The directions that changed the most over practice are close to and sometimes even overlapping with the 90° direction, that is, the direction that represents only the temporal dynamics of airflow. In contrast, the directions that changed the least over practice are close to and sometimes overlapping with the 0° direction, that is, the direction that represents only the temporal dynamics of vocal intensity. This potentially indicates that mean airflow during voicing became more overtly/volitionally controlled throughout practice while volitional control of vocal intensity was little changed. Both subjects were likely highly skilled in controlling vocal intensity before practicing this game, as variations from very soft to very loud are easily produced, perceived, and useful to convey information during conversation. In contrast, both subjects probably had very little skill in controlling mean airflow during voicing before practicing the game, as variations from low to high airflows are often not perceptually pertinent for routine vocal demands.

The purpose of this study was to evaluate if subjects with and without a voice disorder could learn the virtual floating ball task, not to compare performance differences between the control subject and patient with NPVH. However, there are differences between the two subjects, the most paradoxical being that the patient began practice with noticeably lower mean error than the control. One reason for this difference could be that the patient had four voice therapy sessions 6 months previously, where she practiced voicing with increased forward resonance as well as gargling water during voicing to improve her control of mean airflow. Of note, the patient reported a similar level of mild impairment during her pretherapy evaluation 6 months previous: CAPE-V ratings were 10 (overall dysphonia), 10 (roughness), 0 (breathiness), and 9 (strain); Voice Handicap Index-10 (Jacobson et al., 1997; Rosen et al., 2004) was 17 out of 40 possible points, where higher scores represent more severe impacts on quality of life. Regardless of previous voice therapy, she did make significant improvements (decreases in error) in the virtual floating ball task.

There are also some limitations to this study. Most notably, the results here are based on only two case studies and may not accurately represent the average or typical time course of how patients learn this game. To thoroughly characterize the learning in this task, as well as identify aberrant learning patterns, data will need to be collected on groups of subjects with healthy voices and with NPVH. Also, all virtual practice of the floating ball task relies on sustained phonation with a single vowel, and it is unclear if the improved voicing will generalize to connected speech in daily life. However, the exercises are part of the VFE protocol, a protocol that has demonstrated broadly improved patient-reported outcomes in spontaneous speech and daily life (e.g., quality of life, handicap, auditory-perception of voice quality) in more than 10 studies (e.g., Berg et al., 2008; Gillivan-Murphy et al., 2006; Kaneko et al., 2015; Kapsner-Smith et al., 2015; Nguyen & Kenny, 2009; Pasa et al., 2007; Patel et al., 2012; Pedrosa et al., 2016; Roy et al., 2001; Sauder et al., 2010; Tanner et al., 2010; Tay et al., 2012; Teixeira & Behlau, 2015; Ziegler et al., 2014). Therefore, based on this copious evidence that the VFE protocol can generalize to spontaneous speech (to the authors' knowledge, more real-life evidence than any other standardized vocal rehabilitation protocol), generalization from the

virtual task into spontaneous speech is a probable outcome (at the group level). A final caveat and open question is whether a more severely dysphonic patient could improve in this virtual task under the same conditions, for example, minimal cues and feedback on performance from the SLP.

Considerations for Clinical Applications

This study illustrated, in a healthy control and a patient with NPVH, that the virtual floating ball task produces quantitative measures capable of following the motor learning process. Future work could investigate whether the distributional (costs) and temporal variability (SCI) metrics have potential to inform two major clinical difficulties. First, it would be clinically useful to objectively estimate how much treatment (practice) an individual patient may need (i.e., stimulability; Gillespie & Gartner-Schmidt, 2016). Future investigations could assess how variability metrics from early practice predict how quickly a patient may minimize their error. Recent studies have shown that, paradoxically, subjects with larger amounts of variability in early practice tend to reduce their error faster—although "variability" was defined differently in each study (Barbado Murillo et al., 2017; Cardis et al., 2017; He et al., 2016; Mehler et al., 2017; Singh et al., 2016; Sternad, 2018; Wu et al., 2014). It is conjectured that some aspect of baseline variability reflects active exploration, which in turn helps learning. Should early-practice variability metrics correlate with a rate of learning, patients with NPVH could practice the virtual task before starting therapy to indicate how much therapy might be necessary. Second, voice therapy for patients with NPVH relies on the assumption that the newly established therapeutic behavior will remain for a long time after discharge, but objective estimates of long-term retention do not currently exist (to the authors' knowledge). Future work could assess how late-practice variability metrics correlate to retention months or years later. For example, many limb-motor studies using redundant tasks have shown that expert performance is associated with variability channeled in ways that do not affect error; subjects exploit the solution manifold instead of reducing overall variability (Abe & Sternad, 2013; R. Cohen & Sternad, 2009; Van Stan et al., 2017). This can improve the behavior's robustness (i.e., it makes noise "matter less") and increase the probability of long-term retention. Therefore, if variability metrics from late practice correlate with long-term retention, patients could practice the virtual task throughout the course of therapy to assist in discharge planning.

While the floating ball task was developed in hopes of providing therapeutically meaningful insights into the process of vocal motor learning, its general addition into the VFE protocol has potential to improve the standardized intervention's efficacy/effectiveness. For example, the game provides objective feedback that can be as implicit as needed (i.e., minimal didactic content). In practice, using the virtual task in voice therapy could be anywhere on a continuum of mostly implicit (minimal cues and explanation) to mostly explicit (providing cues and explanations while playing the game) depending upon the patient. In contrast, the current VFE approach relies on the clinician's subjective perceptual judgments associated with explicit descriptions. In multiple studies, implicit feedback has been associated with better skill acquisition and retention compared to explicit feedback (Dienes & Berry, 1997; Jie et al., 2018; Kal et al., 2018; Masters, 1992). Also, adding the floating ball task into the VFE may improve overall treatment adherence, since it is essentially a "gamification" of a therapeutic exercise (Fleming et al., 2017; Johnson et al., 2016; Sardi et al., 2017).

Summary and Conclusions

The virtual rendering of two VFE treatment components was largely successful due to its basis in two broadly applicable approaches. First, the task was based in computational motor neuroscience concepts that generally examine motor learning in tasks with redundancy. Given that airflow and vocal intensity do not uniquely determine performance, but rather an infinite number of combinations of the two can lead to good performance, voicing is an example of a motor skill with redundancy. Hence, methods developed in limb motor control could be adapted to quantify the process of learning in this task. Second, the VFE protocol was fractionated into its treatment components using the RTSS framework, allowing an evaluation of which treatment components were most amenable to virtual rendering, that is, Exercises 1 and 4. The RTSS concept of a treatment target was necessary to identify which execution variables should be measured (i.e., mean airflow and vocal intensity) and how they should be combined to achieve the target's solution manifold (i.e., the ratio of vocal intensity divided by mean airflow within a specific flow and loudness range). The RTSS concept of a mechanism of action helped conceptualize where the virtual variability metrics (i.e., costs, SCI) fit into the treatment protocol (i.e., they relate how changes in the target resulted from practice). Because the task was based on two broadly applicable approaches, this study could be a useful model for creating therapeutically meaningful virtual tasks in other rehabilitation fields heavily reliant on motor learning (e.g., speech, upper extremity training, gait training). It is critical to explicitly connect the elements of a virtual task and its underlying treatment theory (i.e., how the ingredients are hypothesized to directly affect the target). When these connections are absent or left to speculation, fully quantitative renderings will provide increased measurement without improved knowledge regarding what improved patient outcomes and why.

Acknowledgments

This work was supported by the Voice Health Institute and the National Institutes of Health under Grants R01-SD045639, R01-HD081346, and R01-HD087089 (PI: Dagmar Sternad); R-33-DC011588 and P50-DC015446 (PI: Robert Hillman); and F31-DC014412 (PI: Jarrad Van Stan). 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. The authors would like to thank James Kobler for making the three-dimensional printed attachments for the Phonatory Aerodynamic System flow heads as well as Andrew J. Ortiz and Daryush D. Mehta for signal processing advice.

References

Abe, M. O., & Sternad, D. (2013). Directionality in distribution and temporal structure of variability in skill acquisition. *Frontiers in Human Neuroscience, 7,* 225. https://doi.org/10.3389/fnhum.2013.00225

Ajemian, R., D'Ausilio, A., Moorman, H., & Bizzi, E. (2013). A theory for how sensorimotor skills are learned and retained in noisy and nonstationary neural circuits. *Proceedings of the National Academy of Sciences, 110*(52), E5078–E5087. https://doi.org/10.1073/pnas.1320116110

Andrade, P. A., Wistbacka, G., Larsson, H., Södersten, M., Hammarberg, B., Simberg, S., Švec, J. G., & Granqvist, S. (2016). The flow and pressure relationships in different tubes commonly used for semi-occluded vocal tract exercises. *Journal of Voice, 30*(1), 36–41. https://doi.org/10.1016/j.jvoice.2015.02.004

Angadi, V., Croake, D., & Stemple, J. (2019). Effects of vocal function exercises: A systematic review. *Journal of Voice, 33*(1), 124.E113–124.E134. https://doi.org/10.1016/j.jvoice.2017.08.031

Barbado Murillo, D., Caballero Sánchez, C., Moreside, J., Vera-García, F. J., & Moreno, F. J. (2017). Can the structure of motor variability predict learning rate. *Journal of Experimental Psychology: Human Perception and Performance, 43*(3), 596–607. https://doi.org/10.1037/xhp0000303

Berg, E. E., Hapner, E., Klein, A., & Johns, M. M., III (2008). Voice therapy improves quality of life in age-related dysphonia: A case-control study. *Journal of Voice, 22*(1), 70–74. https://doi.org/10.1016/j.jvoice.2006.09.002

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

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*(5), 709–728. https://doi.org/10.1080/02687038.2012.676888

Cardis, M., Casadio, M., & Ranganathan, R. (2017). High variability impairs motor learning regardless of whether it affects task performance. *Journal of Neurophysiology, 119*(1), 39–48. https://doi.org/10.1152/jn.00158.2017

Chen, S. H., Liu, H., Xu, Y., & Larson, C. R. (2007). Voice F0 responses to pitch-shifted voice feedback during English speech. *The Journal of the Acoustical Society of America, 121*(2), 1157–1163. https://doi.org/10.1121/1.2404624

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

Cohen, R., & Sternad, D. (2009). Variability in motor learning: Relocating, channeling and reducing noise. *Experimental Brain Research, 193*(1), 69–83. https://doi.org/10.1007/s00221-008-1596-1

Colton, R. H., Casper, J. K., & Leonard, R. J. (2006). *Understanding voice problems: A physiological perspective for diagnosis and treatment*. Lippincott Williams & Wilkins.

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

Cusumano, J. P., & Cesari, P. (2006). Body-goal variability mapping in an aiming task. *Biological Cybernetics, 94*(5), 367–379. https://doi.org/10.1007/s00422-006-0052-1

Dienes, Z., & Berry, D. (1997). Implicit learning: Below the subjective threshold. *Psychonomic Bulletin & Review, 4*(1), 3–23. https://doi.org/10.3758/BF03210769

Faisal, A. A., Selen, L. P., & Wolpert, D. M. (2008). Noise in the nervous system. *Nature Reviews Neuroscience, 9*(4), 292–303. https://doi.org/10.1038/nrn2258

Fleming, T. M., Bavin, L., Stasiak, K., Hermansson-Webb, E., Merry, S. N., Cheek, C., Mathijs, L., Lau, H. M., Pollmuller, P., & Hetrick, S. (2017). Serious games and gamification for mental health: Current status and promising directions. *Frontiers in Psychiatry, 7,* 215. https://doi.org/10.3389/fpsyt.2016.00215

Gillespie, A. I., & Gartner-Schmidt, J. (2016). Immediate effect of stimulability assessment on acoustic, aerodynamic, and patient-perceptual measures of voice. *Journal of Voice, 30*(4), 507.E509–507.E514. https://doi.org/10.1016/j.jvoice.2015.06.004

Gillivan-Murphy, P., Drinnan, M. J., O'Dwyer, T. P., Ridha, H., & Carding, P. (2006). The effectiveness of a voice treatment approach for teachers with self-reported voice problems. *Journal of Voice, 20*(3), 423–431. https://doi.org/10.1016/j.jvoice.2005.08.002

Guenther, F. H., Ghosh, S. S., & Tourville, J. A. (2006). Neural modeling and imaging of the cortical interactions underlying syllable production. *Brain and Language, 96*(3), 280–301. https://doi.org/10.1016/j.bandl.2005.06.001

Hart, T., Dijkers, M. P., Whyte, J., Turkstra, L. S., Zanca, J. M., Packel, A., Van Stan, J. H., Ferraro, M., & Chen, C. (2019). A theory-driven system for the specification of rehabilitation treatments. *Archives of Physical Medicine and Rehabilitation, 100*(1), 172–180. https://doi.org/10.1016/j.apmr.2018.09.109

He, K., Liang, Y., Abdollahi, F., Bittmann, M. F., Kording, K., & Wei, K. (2016). The statistical determinants of the speed of motor learning. *PLOS Computational Biology, 12*(9), Article e1005023. https://doi.org/10.1371/journal.pcbi.1005023

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*(2), 373–392. https://doi.org/10.1044/jshr.3202.373

Hillman, R. E., Stepp, C. E., Van Stan, J. H., Zañartu, M., & Mehta, D. D. (2020). An updated theoretical framework for vocal hyperfunction. *American Journal of Speech-Language Pathology, 29*(4), 2254–2260. https://doi.org/10.1044/2020_AJSLP-20-00104

Hirano, M., & McCormick, K. R. (1986a). Clinical examination of voice. *The Journal of the Acoustical Society of America*.

Hirano, M., & McCormick, K. R. (1986b). Clinical examination of voice by Minoru Hirano. *The Journal of the Acoustical Society of America, 80*(4), 1273. https://doi.org/10.1121/1.393788

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

Holmberg, E. B., Hillman, R. E., & Perkell, J. S. (1988). Glottal airflow and transglottal air pressure measurements for male and female speakers in soft, normal, and loud voice. *The Journal of the Acoustical Society of America, 84*(2), 511–529. https://doi.org/10.1121/1.396829

Holmberg, E. B., Hillman, R. E., & Perkell, J. S. (1989). Glottal airflow and transglottal air pressure measurements for male and female speakers in low, normal, and high pitch. *Journal of Voice, 3*(4), 294–305. https://doi.org/10.1016/S0892-1997(89)80051-7

Jacobson, B. H., Johnson, A., Grywalski, C., Silbergleit, A., Jacobson, G., Benninger, M. S., & Newman, C. W. (1997). The Voice Handicap Index (VHI): Development and validation. *American Journal of Speech-Language Pathology, 6*(3), 66–70. https://doi.org/10.1044/1058-0360.0603.66

Jie, L.-J., Kleynen, M., Meijer, K., Beurskens, A., & Braun, S. (2018). The effects of implicit and explicit motor learning in gait rehabilitation of people after stroke: Protocol for a randomized controlled trial. *JMIR Research Protocols, 7*(5), Article e142. https://doi.org/10.2196/resprot.9595

Johnson, D., Deterding, S., Kuhn, K.-A., Staneva, A., Stoyanov, S., & Hides, L. (2016). Gamification for health and wellbeing: A systematic review of the literature. *Internet Interventions, 6,* 89–106. https://doi.org/10.1016/j.invent.2016.10.002

Kal, E., Prosée, R., Winters, M., & Van Der Kamp, J. (2018). Does implicit motor learning lead to greater automatization of motor skills compared to explicit motor learning? A systematic review. *PLOS ONE, 13*(9), Article e0203591. https://doi.org/10.1371/journal.pone.0203591

Kandel, E. R., Schwartz, J. H., & Jessell, T. M. (2000). *Principles of Neural Science* (4th ed.). McGraw-Hill.

Kaneko, M., Hirano, S., Tateya, I., Kishimoto, Y., Hiwatashi, N., Fujiu-Kurachi, M., & Ito, J. (2015). Multidimensional analysis on the effect of vocal function exercises on aged vocal fold atrophy. *Journal of Voice, 29*(5), 638–644. https://doi.org/10.1016/j.jvoice.2014.10.017

Kapsner-Smith, M. R., Hunter, E. J., Kirkham, K., Cox, K., & Titze, I. R. (2015). A randomized controlled trial of two semi-occluded vocal tract voice therapy protocols. *Journal of Speech, Language, and Hearing Research, 58*(3), 535–549. https://doi.org/10.1044/2015_JSLHR-S-13-0231

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

Kridgen, S., Hillman, R. E., Stadelman-Cohen, T., Zeitels, S., Burns, J. A., Hron, T., Krusemark, C., Muise, J., & Van Stan, J. H. (2020). Patient-reported factors associated with the onset of hyperfunctional voice disorders. *Annals of Otology, Rhinology & Laryngology*. Advance online publication. https://doi.org/10.1177/0003489420956379

Larson, C. R., Burnett, T. A., Kiran, S., & Hain, T. C. (2000). Effects of pitch-shift velocity on voice F0 responses. *The Journal of the Acoustical Society of America, 107*(1), 559–564. https://doi.org/10.1121/1.428323

Lã, F. M., Wistbacka, G., Andrade, P. A., & Granqvist, S. (2017). Real-time visual feedback of airflow in voice training: Aerodynamic properties of two flow ball devices. *Journal of Voice, 31*(3), 390.E391–390.E398. https://doi.org/10.1016/j.jvoice.2016.09.024

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*(3), 277–298. https://doi.org/10.1044/1058-0360(2008/025)

Masters, R. S. (1992). Knowledge, knerves and know-how: The role of explicit versus implicit knowledge in the breakdown of a complex motor skill under pressure. *British Journal of Psychology, 83*(3), 343–358. https://doi.org/10.1111/j.2044-8295.1992.tb02446.x

Mehler, D. M. A., Reichenbach, A., Klein, J., & Diedrichsen, J. (2017). Minimizing endpoint variability through reinforcement learning during reaching movements involving shoulder, elbow and wrist. *PLOS ONE, 12*(7), Article e0180803. https://doi.org/10.1371/journal.pone.0180803

Müller, H., & Sternad, D. (2004). Decomposition of variability in the execution of goal-oriented tasks: Three components of skill improvement. *Journal of Experimental Psychology: Human Perception and Performance, 30*(1), 212–233. https://doi.org/10.1037/0096-1523.30.1.212

Müller, H., & Sternad, D. (2009). Motor learning: Changes in the structure of variability in a redundant task. In D. Sternad (Ed.), *Progress in motor control* (pp. 439–456). Springer. https://doi.org/10.1007/978-0-387-77064-2_23

Nguyen, D. D., & Kenny, D. T. (2009). Randomized controlled trial of vocal function exercises on muscle tension dysphonia in Vietnamese female teachers. *Journal of Otolaryngology—Head & Neck Surgery, 38*(2), 261–278.

Pasa, G., Oates, J., & Dacakis, G. (2007). The relative effectiveness of vocal hygiene training and vocal function exercises in preventing voice disorders in primary school teachers. *Logopedics Phoniatrics Vocology, 32*(3), 128–140. https://doi.org/10.1080/14015430701207774

Patel, R. R., Pickering, J., Stemple, J., & Donohue, K. D. (2012). A case report in changes in phonatory physiology following voice therapy: Application of high-speed imaging. *Journal of Voice, 26*(6), 734–741. https://doi.org/10.1016/j.jvoice.2012.01.001

Pedrosa, V., Pontes, A., Pontes, P., Behlau, M., & Peccin, S. M. (2016). The effectiveness of the comprehensive voice rehabilitation program compared with the vocal function exercises method in behavioral dysphonia: A randomized clinical trial. *Journal of Voice, 30*(3), 377.E311–377.E319. https://doi.org/10.1016/j.jvoice.2015.03.013

Peng, C. K., Havlin, S., Stanley, H. E., & Goldberger, A. L. (1995). Quantification of scaling exponents and crossover phenomena in nonstationary heartbeat time series. *Chaos: An Interdisciplinary Journal of Nonlinear Science, 5*(1), 82–87. https://doi.org/10.1063/1.166141

Ramig, L. O., & Verdolini, K. (1998). Treatment efficacy: Voice disorders. *Journal of Speech, Language, and Hearing Research, 41*(1), S101–S116. https://doi.org/10.1044/jslhr.4101.s101

Rosen, C. A., Lee, A. S., Osborne, J., Zullo, T., & Murry, T. (2004). Development and validation of the Voice Handicap Index-10. *The Laryngoscope, 114*(9), 1549–1556. https://doi.org/10.1097/00005537-200409000-00009

Roy, N., Gray, S. D., Simon, M., Dove, H., Corbin-Lewis, K., & Stemple, J. C. (2001). An evaluation of the effects of two treatment approaches for teachers with voice disorders: A prospective randomized clinical trial. *Journal of Speech, Language, and Hearing Research, 44*(2), 286–296. https://doi.org/10.1044/1092-4388(2001/023)

Sardi, L., Idri, A., & Fernández-Alemán, J. L. (2017). A systematic review of gamification in e-health. *Journal of Biomedical Informatics, 71,* 31–48. https://doi.org/10.1016/j.jbi.2017.05.011

Sauder, C., Roy, N., Tanner, K., Houtz, D. R., & Smith, M. E. (2010). Vocal function exercises for presbylaryngis: A multidimensional assessment of treatment outcomes. *Annals of Otology, Rhinology & Laryngology, 119*(7), 460–467. https://doi.org/10.1177/000348941011900706

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

Scholz, J. P., Schöner, G., & Latash, M. L. (2000). Identifying the control structure of multijoint coordination during pistol shooting. *Experimental Brain Research, 135*(3), 382–404. https://doi.org/10.1007/s002210000540

Singh, P., Jana, S., Ghosal, A., & Murthy, A. (2016). Exploration of joint redundancy but not task space variability facilitates supervised motor learning. *Proceedings of the National Academy of Sciences, 113*(50), 14414–14419. https://doi.org/10.1073/pnas.1613383113

Stemple, J. C. (2005). A holistic approach to voice therapy. *Seminars in Speech and Language, 26*(2), 131–137. https://doi.org/10.1055/s-2005-871209

Stemple, J. C., Lee, L., D'Amico, B., & Pickup, B. (1994). Efficacy of vocal function exercises as a method of improving voice production. *Journal of Voice, 8*(3), 271–278. https://doi.org/10.1016/S0892-1997(05)80299-1

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

Sternad, D. (2018). It's not (only) the mean that matters: Variability, noise and exploration in skill learning. *Current Opinion in Behavioral Sciences, 20,* 183–195. https://doi.org/10.1016/j.cobeha.2018.01.004

Sternad, D., Huber, M. E., & Kuznetsov, N. (2014). Acquisition of novel and complex motor skills: Stable solutions where intrinsic noise matters less. In M. F. Levin (Ed.), *Progress in motor control* (pp. 101–124). Springer. https://doi.org/10.1007/978-1-4939-1338-1_8

Story, B. H., Titze, I. R., & Hoffman, E. A. (1998). Vocal tract area functions for an adult female speaker based on volumetric imaging. *The Journal of the Acoustical Society of America, 104*(1), 471–487. https://doi.org/10.1121/1.423298

Tanaka, S., & Gould, W. J. (1983). Relationships between vocal intensity and noninvasively obtained aerodynamic parameters in normal subjects. *The Journal of the Acoustical Society of America, 73*(4), 1316–1321. https://doi.org/10.1121/1.389235

Tanner, K., Sauder, C., Thibeault, S. L., Dromey, C., & Smith, M. E. (2010). Vocal fold bowing in elderly male monozygotic twins: A case study. *Journal of Voice, 24*(4), 470–476. https://doi.org/10.1016/j.jvoice.2008.10.010

Tay, E. Y. L., Phyland, D. J., & Oates, J. (2012). The effect of vocal function exercises on the voices of aging community choral singers. *Journal of Voice, 26*(5), 672.E619–672.E627. https://doi.org/10.1016/j.jvoice.2011.12.014

Teixeira, L. C., & Behlau, M. (2015). Comparison between vocal function exercises and voice amplification. *Journal of Voice, 29*(6), 718–726. https://doi.org/10.1016/j.jvoice.2014.12.012

Van Stan, J. H., Dijkers, M. P., Whyte, J., Hart, T., Turkstra, L. S., Zanca, J. M., & Chen, C. (2019). The Rehabilitation Treatment Specification System: Implications for improvements in research design, reporting, replication, and synthesis. *Archives of Physical Medicine and Rehabilitation, 100*(1), 146–155. https://doi.org/10.1016/j.apmr.2018.09.112

Van Stan, J. H., Park, S.-W., Jarvis, M., Mehta, D. D., Hillman, R. E., & Sternad, D. (2017). Measuring vocal motor skill with a virtual voice-controlled slingshot. *The Journal of the Acoustical Society of America, 142*(3), 1199–1212. https://doi.org/10.1121/1.5000233

Wu, H. G., Miyamoto, Y. R., Castro, L. N. G., Ölveczky, B. P., & Smith, M. A. (2014). Temporal structure of motor variability is dynamically regulated and predicts motor learning ability. *Nature Neuroscience, 17*(2), 312–321. https://doi.org/10.1038/nn.3616

Zarate, J. M., Wood, S., & Zatorre, R. J. (2010). Neural networks involved in voluntary and involuntary vocal pitch regulation in experienced singers. *Neuropsychologia, 48*(2), 607–618. https://doi.org/10.1016/j.neuropsychologia.2009.10.025

Ziegler, A., Verdolini Abbott, K., Johns, M., Klein, A., & Hapner, E. R. (2014). Preliminary data on two voice therapy interventions in the treatment of presbyphonia. *The Laryngoscope, 124*(8), 1869–1876. https://doi.org/10.1002/lary.24548