Measuring vocal motor skill with a virtual voice-controlled slingshot

Jarrad H. Van Stan[a)]
Center for Laryngeal Surgery and Voice Rehabilitation, Massachusetts General Hospital, Boston, Massachusetts 02114, USA

Se-Woong Park
Department of Biology, Northeastern University, Boston, Massachusetts 02115, USA

Matthew Jarvis
Newark, Delaware 19711, USA

Daryush D. Mehta[b)] and Robert E. Hillman[b)]
Center for Laryngeal Surgery and Voice Rehabilitation, Massachusetts General Hospital, Boston, Massachusetts 02114, USA

Dagmar Sternad
Departments of Biology, Electrical and Computer Engineering, and Physics, Northeastern University, Boston, Massachusetts 02115, USA

(Received 9 March 2017; revised 8 August 2017; accepted 9 August 2017; published online 5 September 2017)

Successful voice training (e.g., singing lessons) and vocal rehabilitation (e.g., therapy for a voice disorder) involve learning complex, vocal behaviors. However, there are no metrics describing how humans learn new vocal skills or predicting how long the improved behavior will persist post-therapy. To develop measures capable of describing and predicting vocal motor learning, a theory-based paradigm from limb motor control inspired the development of a virtual task where subjects throw projectiles at a target via modifications in vocal pitch and loudness. Ten subjects with healthy voices practiced this complex vocal task for five days. The many-to-one mapping between the execution variables pitch and loudness and resulting target error was evaluated using an analysis that quantified distributional properties of variability: Tolerance, noise, covariation costs (TNC costs). Lag-1 autocorrelation (AC1) and detrended-fluctuation-analysis scaling index (SCI) analyzed temporal aspects of variability. Vocal data replicated limb-based findings: TNC costs were positively correlated with error; AC1 and SCI were modulated in relation to the task's solution manifold. The data suggests that vocal and limb motor learning are similar in how the learner navigates the solution space. Future work calls for investigating the game's potential to improve voice disorder diagnosis and treatment. © *2017 Acoustical Society of America*. [http://dx.doi.org/10.1121/1.5000233]

[ZZ] Pages: 1199–1212

I. INTRODUCTION

Successful voice therapy depends upon patients with voice disorders learning or re-learning vocal motor behaviors—e.g., vocal loudness (Ramig *et al.*, 1995; Schalling *et al.*, 2013; Van Stan *et al.*, 2017), vocal efficiency (Titze, 1992; Titze, 2006), vocal endurance (Buekers, 1998; Schneider and Bigenzahn, 2005; Schneider *et al.*, 2006; Stemple *et al.*, 1994), and voice quality (Kempster *et al.*, 2009; Kreiman *et al.*, 1993; Verdolini-Marston *et al.*, 1995)—to reduce voice-related impairments in daily life (Hogikyan and Sethuraman, 1999; Jacobson *et al.*, 1997). Therefore, "learning" new motor patterns—defined as a relatively permanent change in motor behavior (Schmidt and Lee, 2011)—is a critical part of the voice therapy process. In the same way, singing lessons depend upon vocalists establishing new complex behaviors such as improved voice quality on high notes or improved transitions across register changes (e.g., chest voice into head voice). However, little is known regarding how humans learn new vocal motor skills as research into properties of the cortico-bulbar sensorimotor system typically use well-learned (i.e., habituated) vocal behaviors—e.g., sustained vowels or glissandos (Burnett *et al.*, 1998; Larson *et al.*, 2000; Zarate *et al.*, 2010) and syllables or speech (Chen *et al.*, 2007; Guenther *et al.*, 2006; Tourville *et al.*, 2008; Xu *et al.*, 2004). Also, clinical voice treatment designs focus on average differences between isolated time points—e.g., before versus after surgery/voice therapy (Holmberg *et al.*, 2003; Ramig and Verdolini, 1998; Roy *et al.*, 2003; Roy *et al.*, 2002)—which do not take into account how behavior change occurred. While these clinical investigations improve the field's diagnostic capabilities and provide empirical support for the effectiveness of voice treatments, they rarely offer theoretical insights into how patients learn improved behaviors or how long the measured improvements will last after discontinuing

[a)] Also at: Department of Surgery, Harvard Medical School, Boston, MA 02115, USA. Electronic mail: jvanstan@mgh.harvard.edu

[b)] Also at: Department of Surgery, Harvard Medical School, Boston, MA 02115, USA.

therapy—i.e., carryover or retention. However, there is a growing number of studies in the voice field that apply motor learning principles in the hopes of maximizing carryover or retention through fine-tuning practice and feedback variables (Schalling *et al.*, 2013; Steinhauer and Grayhack, 2000; Van Stan *et al.*, 2015; Wong *et al.*, 2011). These approaches typically measure a patient's overall performance (e.g., accuracy or error), and not how the vocal motor system improved performance (e.g., how did the subject modify various aspects of phonation to minimize error?). Additionally, motor learning studies have asked subjects to pay attention solely to the results of performance (external focus) or the execution of performance (internal focus) (Wulf *et al.*, 1998); these studies neglect to quantify the relationship between how changes in execution are related to improvements in skill. If a method was developed that could examine how humans learn new vocal motor skills, this could produce clinically important measures capable of estimating or predicting a patient's likelihood of improvement, long-term retention, or disorder recurrence after discharge from therapy. However, it is important to note that retention of newly learned motor skills is reliant upon many variables in addition to motor performance (e.g., attention, motivation, cognition, medical history, and baseline skill level) (Marinelli *et al.*, 2017; Studenka *et al.*, 2017).

Theories from the field of motor control and learning frequently attempt to quantify how the central nervous system (CNS) controls and learns new movements. Therefore, these theories may guide the development of assessment approaches that can describe how a patient improved his/her vocal motor behavior and estimate the new behavior's degree of permanence. For example, a model of *redundant motor tasks*—tasks with infinitely many ways to achieve success—has potential to offer insights into how people establish new movements (Cusumano and Cesari, 2006; Kudo *et al.*, 2000; Martin *et al.*, 2001; Müller and Sternad, 2003; Müller and Sternad, 2009; Scholz *et al.*, 2000). Redundant motor tasks can be described by how *execution* variables relate to *result* variables: if there are more execution variables that map into fewer result variables, the task has redundancy. Such redundancy produces an infinite number of combinations of execution variables that can achieve a desired result. It is important to note that redundancy in motor performance can arise from two sources: (1) that of motor equivalence, where multiple bodily configurations map onto a singular outcome, i.e., finger position in space (Kelso *et al.*, 1998) and (2) that of task equivalence, where multiple task-specific variables map onto a singular result in task performance, i.e., getting a perfect score in darts, where the bull's eye is an area and the dart can hit it with many different orientations (Müller and Sternad, 2003). Therefore, the term "execution variable" can refer to many types of variables, from biomechanical (joint angle positions, velocities, etc.) to physics-based (release velocity, release angle, etc.). The most frequently cited example of redundant motor performance is Nicolai Bernstein's observation that even expert blacksmiths exhibited slightly different arm kinematics (multiple execution variables represented by joint angles) during each swing of the hammer, yet consistently hit their desired end point on the anvil (singular result) (Bernstein, 1967). This example of motor equivalence also has task equivalence, since the anvil is not a single point and can be hit with many different orientations. In order to systematically study such motor tasks, virtual environments have been developed where the physics of the task is mathematically modeled so that multiple execution variables fully determine the result or error (John and Cusumano, 2007; Müller and Sternad, 2004). For example, in a virtual throwing task, the result (minimum ball distance from a target) can be fully determined by the user's angle and velocity when releasing the ball (Müller and Sternad, 2009). Therefore, the relationship between execution variables and the result creates a mathematical null space or *solution manifold*—defining those executions that all lead to the desired result. Using this approach, investigators can quantify how subjects learn new motor skills by relating practice-based performance improvements (reduction in error) to changes in variability of the execution variables.

All vocal behaviors targeted in voice therapy are redundant by nature. In other words, patients must learn how to covary multiple *execution* aspects of phonation on one level—e.g., variables such as pitch, loudness, or vocal fold ab/adduction—to achieve a desired *result* defined at another level—e.g., improved messa di voce (increasing and subsequently decreasing loudness whilst maintaining the same pitch), vocal efficiency (decreased input for the same or more output) (Titze, 1992), modification of resonance or timbre. However, current approaches to developing virtual environments for voice-related actions have been confined to one descriptive level. For example, the software program VISI-PITCH (KayPENTAX, Montvale, NJ) allows users to manipulate objects by modifying either pitch, loudness, or both simultaneously; i.e., changes in pitch and/or loudness are results totally determined by themselves. Another example is the PROSODIC MARIONETTE (Patel *et al.*, 2012), which displays multiple voiced features in real-time (e.g., loudness, pitch, word duration) to improve prosody, but the mathematical relationship between all features and overall prosody is not quantified. In general, most quantitative voice therapy tools (i.e., biofeedback tools) have simply displayed a number representing the desired target behavior (e.g., jitter, shimmer, cepstral peak prominence, electromyography) and asked a patient to modify this number as directed (e.g., increase, decrease, or stay in a desired range) (Ferrand, 1995; Ma *et al.*, 2013; van Leer *et al.*, 2016; Wong *et al.*, 2011). Therefore, the resulting data provide minimal insight into how vocal performance is achieved and what accounts for any improvements.

An area of extensive research in motor control is dedicated to characterizing the structure of the ever-present variability in movements—"structure" refers to the distributional (e.g., Gaussian, anisotropy) or temporal (e.g., Brownian motion, pink noise) characteristics of variability. The human sensorimotor system exhibits variability—sometimes referred to as noise—at multiple time scales, at all levels of function (e.g., the cellular physiology of neuronal activation or the accuracy of throwing a ball) and at all levels of skill (e.g., even expert performance produces trial-to-trial fluctuations) (Ajemian *et al.*, 2013; Faisal *et al.*, 2008; Sternad *et al.*, 2014). Therefore, it is believed that investigations into distributional and temporal variability of motor performance will provide unique insights into a subject's sensorimotor function, as well as how the sensorimotor system establishes new skills. More specifically, variability of measured execution and result variables are hypothesized to shed light on control strategies of the CNS.

Analyses of temporal variability in execution variables during well-learned movements—e.g., walking (Dingwell *et al.*, 2010) or grasping (Rácz and Valero-Cuevas, 2013)—have revealed selective control by the CNS in error-relevant or error-irrelevant directions on the task's solution manifold. Furthermore, when applying this approach to learning novel movements—e.g., throwing tasks (Abe and Sternad, 2013), two-arm pointing (Domkin *et al.*, 2005), or posture/balance (Asaka *et al.*, 2008)—temporal variability amongst execution variables in the early stages of practice appears to have minimal directional preference. However, later in practice, the CNS starts to selectively channel temporal variability/noise into error-irrelevant directions (parallel to the solution manifold, where variability does not affect error) and variability in error-relevant directions (orthogonal to the solution manifold where variability can obviously affect error) is lowered.

Studying the structure of vocal motor variability has the potential to provide new diagnostic and therapeutic insights in the field of voice disorders. Vocal biomarkers derived from variability analyses are hypothesized to detect degradation in the complexity or quality of an individual's motor coordination. Many areas outside the field of voice disorders have attained useful biomarkers of neurological and psychological dysfunction from variability-based measures in the voice signal using cross-correlation and/or detrended fluctuation analyses (Helfer *et al.*, 2014; Horwitz *et al.*, 2013; Williamson *et al.*, 2014; Williamson *et al.*, 2015). Other studies have demonstrated potential to discriminate pathological and normal voices when applying stability-based measures (closely related to variability) to sustained vowels (Herzel *et al.*, 1994; Little *et al.*, 2007; Zhang and Jiang, 2008; Zhang *et al.*, 2004). However, none of these studies addressed the question of learning and therefore cannot offer insights as to how vocal skills improve with practice.

Sternad and colleagues (Cohen and Sternad, 2009) developed three costs that assess how variability amongst execution variables directly affects resulting performance. Tolerance cost (T-cost) evaluates sensitivity of the result space to the distribution of execution variables, noise cost (N-cost) evaluates cost to performance per stochastic variability, and covariation cost (C-cost) evaluates the cost to performance per suboptimal covariation between execution variables. Across a range of studies (Abe and Sternad, 2013; Chu *et al.*, 2016; Cohen and Sternad, 2009; Müller and Sternad, 2004), the lowest to highest cost to performance has been T-cost, C-cost, and N-cost, respectively. Also, using pairwise correlations between each cost and mean error, the greatest to least contributor to reduced error has been T-cost, C-cost, and N-cost, respectively. Finally, these three costs have been shown to evolve over different time scales; i.e., T-cost reached asymptote at a significantly faster rate compared to N-cost and C-cost. In summary, it appears that subjects first improve performance exponentially through finding an error-tolerant space in the solution manifold (T-cost). Further improvement proceeds at a slower time scale by exploiting the covariation between execution variables in the result space (C-cost) and, to a lesser extent, reducing stochastic dispersion (N-cost) (Abe and Sternad, 2013; Chu *et al.*, 2016; Cohen and Sternad, 2009; Müller and Sternad, 2004).

To the authors' knowledge, there are no studies to date that have used a vocal motor task to analyze how execution changes over time to produce practice-based improvements. Since the cortico-bulbar (head/neck control) and cortico-spinal (core and limb control) sensorimotor systems differ with regard to several anatomical/physiological factors (e.g., bilateral/unilateral cortical input, prevalence of a gamma neural system, and interconnection with respiration) (Brandon *et al.*, 2003; Kandel *et al.*, 2000; Simonyan and Horwitz, 2011), generalization of limb-based findings to vocal motor learning is a non-trivial endeavor and requires empirical study.

Overall, the specific purpose of the following study is to investigate how changes in execution variability relate to improvements in skilled vocal performance. This study has two aims: In aim 1, we develop a voice-controlled virtual environment mimicking the motor learning approach used with redundant limb-based movements. More specifically, a video game was developed where the user controls a virtual slingshot with two vocal execution variables (related to fundamental frequency and vocal intensity) to hit a target with a projectile. In aim 2, we assess if variability in this vocal task evolves by the same changes in distributional and temporal variability empirically demonstrated in limb-based motor learning. Aim 2 has multiple hypotheses based on replicating empirical findings from limb movements. For distributional variability, it is expected that (a) the TNC-costs will be rank-ordered with T-cost the lowest, then C-cost, and the highest values for N-cost; (b) T-cost will be the strongest contributor to reducing error, with C-cost second, and N-cost last; (c) T-cost will decrease at a faster time scale than C-cost and N-cost (i.e., T-cost will reach asymptotic performance in the shortest time). For temporal variability, it is expected that during asymptotic performance (i.e., when the vocal skill is well learned), the subjects will display directional sensitivity to the solution manifold. More specifically, autocorrelation and detrended fluctuation analysis will reveal tight control in the direction orthogonal to the solution manifold (error-relevant direction) and lack of control in the direction parallel to the solution manifold (error-irrelevant direction).

The long-term goal of this line of research is to investigate if objective measures of variability derived from a virtual environment will provide theoretical and empirical insights into vocal motor control/learning—ultimately for improved assessment/treatment of behaviorally based (e.g., muscle tension dysphonia, vocal fold nodules, polyps) or neurologically based voice disorders.

II. METHODS

A. Participants

The goal of this study was to acquire data from ten adult subjects (five male and five female) to obtain adequate power for repeated-measure group-based statistics. Fourteen subjects (nine male and five female, mean age = 25 years, range = 19–37 years) performed the experimental task after providing written informed consent in accordance with the Institutional Review Board of the Massachusetts General Hospital. Four male subjects did not complete the entire voice video game protocol due to a technical problem during practice (signal saturation, two subjects), allergies associated with vocal deterioration (one subject), and early termination due to multiple cancellations (one subject). The study enrolled only professional singers to maximize the likelihood of all subjects attaining asymptotic performance within 5 days of practice. Professional singers were considered any student enrolled in a vocal performance degree at music conservatories in the Boston area, or any person who reported their primary income was from singing. The singers reported no history of voice disorders and were judged to have normal voice quality by a speech-language pathologist with specialization in voice.

B. Experimental setup

Figure 1(A) shows an illustration of the experimental set-up. Each subject's voice was recorded using a miniature accelerometer (model BU-27135; Knowles Corp., Itasca, IL) placed on the anterior neck using double-sided tape [model 2181, 3 M, Maplewood, MN, see Fig. 1(A) inset]. An interface circuit supplies power to the accelerometer and delivers the neck-skin acceleration signal to custom C# software running on a MacBook Pro (Apple, Cupertino, CA) with WINDOWS 8.1 (Microsoft, Seattle, WA). Each day, subjects were seated in front of the computer screen and the accelerometer was affixed to approximately the same spot between their thyroid cartilage and the superior aspect of the sternum. Measurements (in millimeters) were taken from anatomical landmarks (e.g., wrinkles on the neck or superior border of the sternum) to the top or bottom of the accelerometer to minimize placement variability.

Since the game was played via recording neck skin acceleration, it is important to note that the amplitude of neck-skin acceleration (ACC_{amp}) is not exactly the same as the acoustic sound pressure level (SPL); e.g., full lip occlusion with the voiced nasal /m/ can elicit the same ACC_{amp} as the vowel /a/, but the SPL will be significantly different between the two degrees of mouth opening. Therefore, all subjects played the game using an /m/ to minimize perceptual discrepancies in loudness between auditory input (SPL) and measured ACC_{amp}. It has been established that SPL and ACC_{amp} demonstrate a strong, significant linear correlation within a subject when the supraglottal track is not time varying (i.e., in a static position) (Cheyne *et al.*, 2003; Fryd *et al.*, 2016; Švec *et al.*, 2005). Therefore, due to the experimental constraints (closed lips and no variation in accelerometer placement each day), it is reasonable to assume that changes in ACC_{amp} closely represent the changes in SPL. Also, the fundamental frequency (f_0), which is obtained from the acceleration signal and used for the virtual environment, is an execution variable known to be highly correlated between neck-skin acceleration and acoustics (Coleman, 1988; Mehta *et al.*, 2016; Sugimoto and Hiki, 1960).

The software processed the acceleration signal (recorded with a 10 kHz low-pass filter, 22 050 Hz sampling rate, and 16-bit quantization) every 30 ms to produce estimates of f_0 and ACC_{amp}. The software computed ACC_{amp} in decibels (dB) with the reference intensity at full scale of the sound card. A fast Fourier transform was used to obtain f_0 estimates from the first peak that is $\geq$30% of the highest peak in the spectrum. Frequency detection errors occurred in less than one percent of all trials. For each trial, a simple voice activity detection algorithm was implemented to determine when voicing began and ended, as well as to provide the exact starting/ending values of f_0 and ACC_{amp}.

C. Voice-controlled virtual environment

Figure 1(B) illustrates the vocal task—displayed on the laptop screen (resolution 2560 × 1440 pixels; pix)—which consists of throwing a ball using a slingshot to hit a target

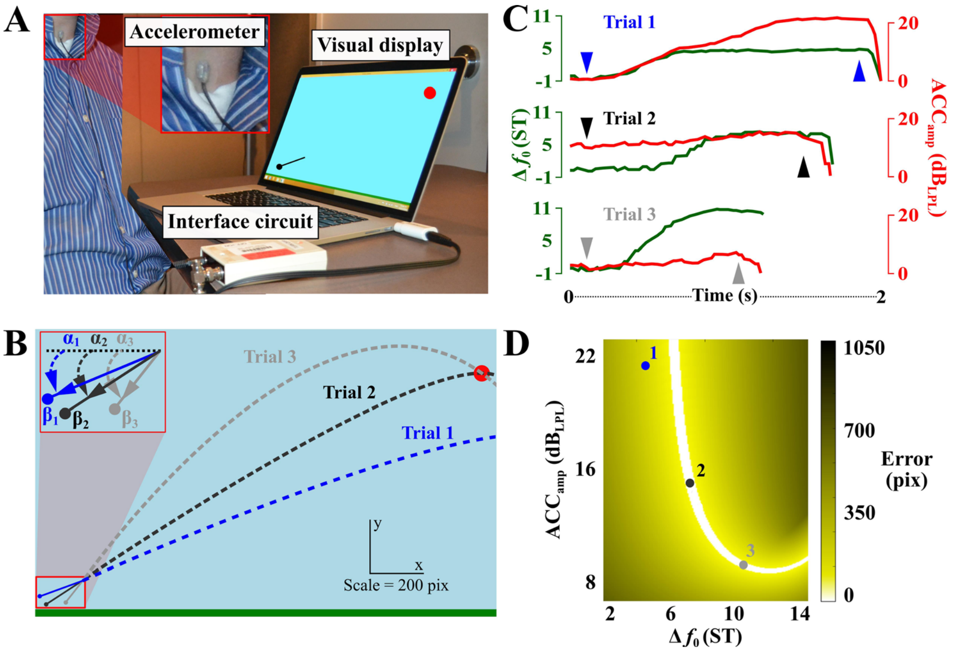

FIG. 1. (Color online) (A) The virtual vocal throwing task. Subjects threw a ball at a target using neck-skin acceleration amplitude above their lowest phonation level (ACC_{amp}, in dB_{LPL}) and changes in fundamental frequency [Δf_0, in semitones (st)]. Though the sling is pictured, it was not visible during practice. (B) Visual display showing the sling and ball trajectories for three exemplary trials. α and β represent angle (in radians) and stretch (in pixels). (C) The three plots show frame-based values of Δf_0 and ACC_{amp} for the three trials. Also, arrows pointing down and up show the start and release values for the slingshot, respectively. Notice that in all three throws, there are moments where one variable is stationary and another is changing. (D) The execution space indicates the amount of error in pixels (pix) for combinations of slingshot release values of Δf_0 and ACC_{amp} (values shown for three trials).

circle. In this two-dimensional (2D) virtual environment, the slingshot's release angle α (in radians) and stretch β (in pixels) fully determine the trajectory of the ball and whether the ball hits or misses the target. The ball follows a simple, frictionless trajectory as defined by Newton's laws of motion, where the initial ball velocities in the horizontal (v_{ix}) and vertical (v_{iy}) directions are determined by α and β, respectively, according to the following relations:

$$v_{ix} = \beta \cos \alpha, \tag{1}$$

$$v_{iy} = \beta \sin \alpha. \tag{2}$$

The 2D position of the ball (x, y) is determined by the following functions of time (t), initial velocities (v_{ix}, v_{iy}), and gravity (g):

$$x(t) = v_{ix}t, \tag{3}$$

$$y(t) = v_{iy}t + \frac{gt^2}{2}. \tag{4}$$

Table I lists all the modifiable parameters related to the 2D virtual environment.

The angle α of the slingshot is controlled via the change in f_0 (Δf_0), in semitones (ST), between the starting and ending f_0 values. The stretch β of the slingshot is controlled via ACC_{amp} during voicing and measured in dB above the lowest phonation level, denoted dB_{LPL}.

The slingshot and ball appear on the screen once the subject begins voicing. At this initial f_0, the slingshot is parallel to the ground (i.e., at 0°) and the subject must increase his/her f_0 to pull the slingshot downward, which aims the ball trajectory upward. The slingshot can also be angled upward to aim the ball trajectory downward with a decrease in f_0 from the initial f_0. A 90° range (−45° to 45°) for Δf_0 was set at 18 ST for all subjects. The range of slingshot stretch was fixed for all subjects between their lowest phonation level (0 dB_{LPL}) at minimal stretch and 22.5 dB_{LPL} at maximal stretch.

TABLE I. Parameters of the virtual environment and specific values used for all subjects.

Parameter	Value used	Description
Field		
Width	1600 px	Virtual environment width (horizontal dimension)
Height	900 px	Virtual environment height (vertical dimension)
Slingshot position		
X	150 px	X and Y coordinates (from the left and bottom of the screen, respectively) of the center of the sling.
Y	75 px	
Target[a]		
X	1150 px	X and Y coordinates of the center of the target circle. Ø represents the diameter of the target circle.
Y	600 px	
Ø	60 px	
Obstacle[b]		
X	1000 px	X coordinate represents the location of an obstacle (i.e., a wall) from the left of the screen. Y_t represents the height (i.e. the "top") of the obstacle minus Y_b (i.e., the "bottom").
Y_t	20 px	
Y_b	20 px	
Gravity		
	-0.0175 px/s^2	Force of gravity on the ball

[a] During exploratory testing, the target was placed at X = 500 px and Y = 700 px.

[b] The virtual environment can contain a wall, but this was not used for the current study.

To begin a trial, the voice signal must have had six consecutive frames (180-ms of voicing) above a f_0 cut-off of 65-Hz and a subject-specific minimum ACC_{amp} threshold. The minimum ACC_{amp} threshold was based on each subject's average vocal intensity over ten trials during which s/he was asked to phonate as softly as possible without voice breaks or whispering; i.e., their lowest phonation level (dB_{LPL}). The ACC_{amp} and f_0 values contained in the sixth voiced frame were used as the start values for each trial. The ball is released once the subject stops voicing; the voiced signal needs to decrease below the ACC_{amp} and/or the f_0 minimum thresholds, at which time the tenth-to-last voiced frame supplies the release ACC_{amp} and f_0 values. Figure 1(C) shows example Δf_0 and ACC_{amp} trajectories of the three throws in Fig. 1(B), with downward pointing triangles indicating slingshot "start" times and upward pointing triangles indicating slingshot "release" times. A video illustrating the vocal slingshot task in action during five trials in the voice-controlled 2D virtual environment is given in Mm. 1.

Mm. 1. Video showing a subject performing five test trials of the vocal slingshot task in the virtual environment. The first four throws miss the target, and the fifth throw hits the middle of the target. When the ball hits the target's center, the ball sticks until the subject begins another practice trial. Note the volume (ACC_{amp}) and pitch (Δf_0) bars on the top of the screen (as well as the slingshot) are not seen by the study subjects during their five days of practice in the study. This is a file of type "mpg" (1.6 MB).

To evaluate trial error, the minimum distance between the ball's trajectory and the center of the target is calculated. The target is a circle with a 60-pixel diameter; i.e., a "hit" occurs when the projectile passes within ± 30 pixels of the circle's center [Fig. 1(B)]. An explicit example of the vocal task's many-to-one mapping or redundancy can be seen in Fig. 1(B), where two exemplary trajectories (trials 2 and 3) show different strategies for hitting the target. The trajectory of trial 2 can be described as a "direct shot," where the user released the ball at approximately a 35° angle ($\Delta f_0 = 7.02$ ST) with a relatively large slingshot stretch of 198 pixels ($\text{ACC}_{\text{amp}} = 14.85$ dB_{LPL}). The trajectory of trial 3 can be described as a "finesse shot," where the user released the ball at a steeper angle of approximately 52° ($\Delta f_0 = 10.35$ ST) with a relatively small slingshot stretch of 125 pixels ($\text{ACC}_{\text{amp}} = 9.28$ dB_{LPL}). The trajectory of trial 1 missed the target and the user released the ball at an angle of approximately 24° (4.72 ST) with a very large slingshot stretch of 276 pixels [$\text{ACC}_{\text{amp}} = 20.57$ dB_{LPL}; stretch for trial 1 is not to scale in Fig. 1(B)].

Since the virtual environment allows a direct mathematical mapping between the two execution variables (Δf_0 and ACC_{amp}) and the resulting error, this error (minimum distance between trajectory and center of the target) can be portrayed with a color code on a 2D execution space [Fig. 1(D)]. Note that the solution manifold, the set of all zero-error solutions, is depicted by white. Despite the simplicity of the task, the solution manifold has a highly nonlinear U-shape. The task is to arrive at this manifold with minimal error. The three dots represent the user's Δf_0 and ACC_{amp} at ball release for the three trials shown in Figs. 1(B) and 1(C). Most importantly, dots 2 and 3 land on the white solution manifold, signifying that both trials hit the target despite being significantly different in execution space. Dot 1 lands on a non-white part of the execution space signifying non-zero error for that trial.

D. Experimental protocol

The overall design of the study called for five days of practice with the voice-controlled virtual environment. The first day began with a verbal description of how to control the virtual slingshot with user-controlled pitch and loudness. Then, the subjects completed 100 trials of exploratory testing with the virtual environment, in which they could see the slingshot moving in real time with pitch and loudness changes (see Mm. 1). Throughout the five days of subsequent practice the slingshot was made invisible so that subjects only could see the ball trajectory after release. This was done to mimic the natural conditions of vocal learning, where vision is not routinely considered part of the real-time feedback loop; feedback constituents are traditionally auditory and proprioceptive/somatosensory (Tourville and Guenther, 2011). Furthermore, three blocks of 100 practice trials were completed each day, for all five days of practice (1500 practice trials total). The target was in a different location for the exploratory testing phase from the one used in the five days of experimental practice. During exploratory testing, the target was higher to the left of the screen (closer to the slingshot) compared to the target placement used for practice [shown in Fig. 1(B)]. Subjects were permitted to have one to three days between practice sessions.

E. Analysis of distributional variability

Figure 2 illustrates the tolerance, noise, and covariation cost (TNC costs) calculations for a given data set of 100 trials by one subject in the execution space (Cohen and Sternad, 2009); day 1 of practice is on the top row and day 5 of practice is on the bottom row. In each panel, the gray data points represent the veridical distributions and the darker (color online) data show the transformed data that optimize each cost.

T-cost is the cost to overall performance for not finding the most error-tolerant area of the execution space. T-cost is estimated by generating an optimized data set in which the mean release angle (i.e., Δf_0) and the mean release stretch (i.e., ACC_{amp}) were shifted in execution space to the location yielding the best overall result. The dispersion in execution space is preserved during this process. More specifically, the data set was shifted on a grid of 1500 × 1500 possible center points and the boundaries of this grid were determined by the limits of the task. The angles tested as centers were limited to those between 0° and 90° (which was equivalent to 0–18 ST). The velocity/stretch values tested as centers were limited to those between 0 and 300 pixels of stretch (which was equivalent to 0–22.5 dB_{LPL}). This range was determined by the settings established for all subjects individually and there were no other options outside of these ranges. The optimization procedure shifted the dataset through every possible center point and evaluated its mean result at each location. 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 and the algebraic difference between the two mean error values defined T-cost.

N-cost is the cost to overall performance due to non-optimal stochastic variability in 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

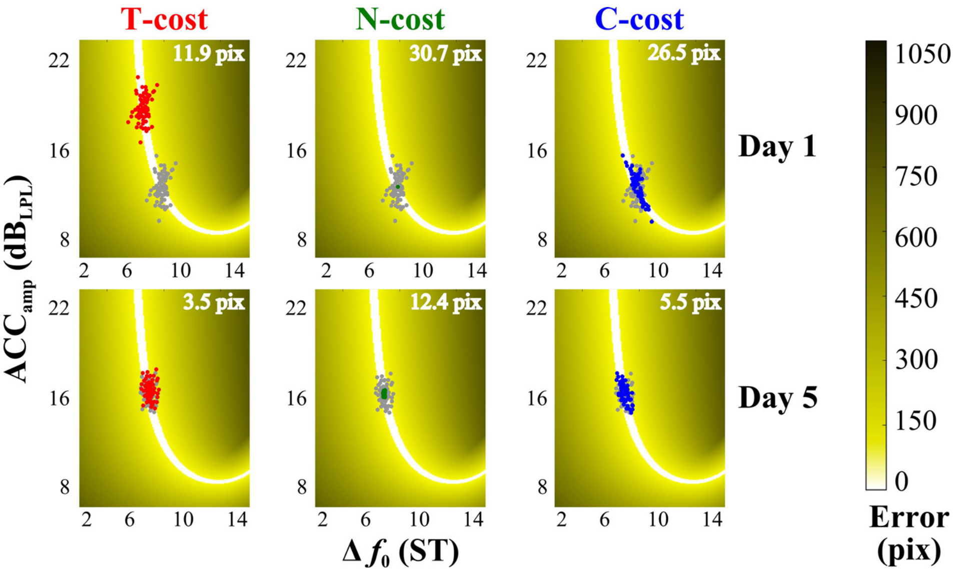

FIG. 2. (Color online) Exemplary and corresponding virtual sets of two practice blocks from one subject, used to illustrate the tolerance noise covariation cost analysis method (T-cost, N-cost, and C-cost, respectively). The left column shows data optimized in terms of T-cost; the middle column shows data optimized in terms of N-cost, and the right column shows data optimized in terms of C-cost. Gray circles represent actual throws made by the subject's phonatory gesture, and red, green, or blue circles represent surrogate data with one component idealized (tolerance, noise, or covariation, respectively). The panels in the top and bottom rows show data from the first block of practice on day 1 and the last block of practice on day 5, respectively.

possible mean error, while leaving overall mean Δf_0 and $\mathrm{ACC}_{\mathrm{amp}}$ unchanged. Though one would initially expect all data sets to be optimally reduced to a single point (the mean Δf_0 and $\mathrm{ACC}_{\mathrm{amp}}$), note that each data set is evaluated in terms of its result (minimum distance from the target). Therefore, the mean of a data set may not always fall directly on the solution manifold and a small distribution might achieve better results on average. In the numerical procedure, the data set was shrunk in 100 steps to converge onto the mean, where the radial distance for every data point was divided into 100 steps. Subsequently, all data points were scaled towards the mean at 1% intervals and the mean error was evaluated at each interval. The algebraic difference between the means of the interval that produced the lowest mean error (optimized data set) and the original data set defined N-cost.

C-cost is the overall cost to performance due to insufficient alignment with the direction of the solution manifold; i.e., not exploiting the redundancy in the execution space. C-cost is estimated by generating an optimized data set where the Δf_0 and $\mathrm{ACC}_{\mathrm{amp}}$ values were not modified, but the individual Δf_0–$\mathrm{ACC}_{\mathrm{amp}}$ trial-by-trial pairings were recombined to achieve the lowest possible mean error via a greedy hill climbing algorithm using a pairwise matching procedure (Russell and Norvig, 2002). Specifically, all Δf_0–$\mathrm{ACC}_{\mathrm{amp}}$ pairs were rank-ordered from lowest error (best) to highest error (worst); $i = 1, 2, 3, \ldots, 100$ since all data sets were of 100 trials. Subsequently, the worst performing $\Delta f_{0\ (i=100)}$ was paired with $\mathrm{ACC}_{\mathrm{amp}\ (i=99)}$ and $\mathrm{ACC}_{\mathrm{amp}\ (i=100)}$ was paired with $\Delta f_{0\ (i=99)}$; the mean result of the new $\mathrm{error}_{(i=100)}$ and the $\mathrm{error}_{(i=99)}$ were compared to the original mean error of the two trials. If the error improved over the original, the swap was accepted. As a next step, $\mathrm{ACC}_{\mathrm{amp}\ (i=100)}$ was swapped with $\mathrm{ACC}_{\mathrm{amp}\ (i=98)}$ and the resulting mean error of the two trials was evaluated. If the mean result improved, the swap was accepted. This continued until $\Delta f_{0\ (i=100)}$ was compared with $\mathrm{ACC}_{\mathrm{amp}\ (i=1)}$; i.e., $\mathrm{ACC}_{\mathrm{amp}\ (i=100)}$ was swapped with $\mathrm{ACC}_{\mathrm{amp}\ (i=1)}$. After this sequence of 99 comparisons, the same sequence was repeated with $\mathrm{ACC}_{\mathrm{amp}\ (i=99)}$; therefore the batch consisted of 4950 comparisons. The batch of procedures was repeated on the improved set until no further swaps could be made. The algebraic difference between the mean error of this optimized data set and the original data set defined C-cost.

An example of data from a single subject in Fig. 2 shows representative changes in all three cost measures due to improved vocal performance. For example, the optimal data transformation shown for T-cost in the upper left panel of Fig. 2 expresses that the subject could have improved her performance by 11.9 pix, if tolerance were optimized. Similarly, for N-cost, the result could be improved by 30.7 pix if the noise were reduced optimally. Finally, for C-cost, the result could be improved by 26.5 pix if the covariation between execution variables was optimal. The subject's data set exhibits decreased costs to performance in all three metrics late in practice (day 5) compared to early in practice (Day 1).

F. Analysis of directionality in execution space

The analysis of directional variability in the execution space was replicated from (Abe and Sternad, 2013) and is illustrated in Fig. 3. To investigate the temporal structure in Δf_0 and $\mathrm{ACC}_{\mathrm{amp}}$, the two axes of the execution space had to be normalized since they are in terms of two different units. To this end, the data for each (nonoverlapping) 100-trial block per subject were transformed into z-scores according to the mean and standard deviation of Δf_0 and $\mathrm{ACC}_{\mathrm{amp}}$ for that block.

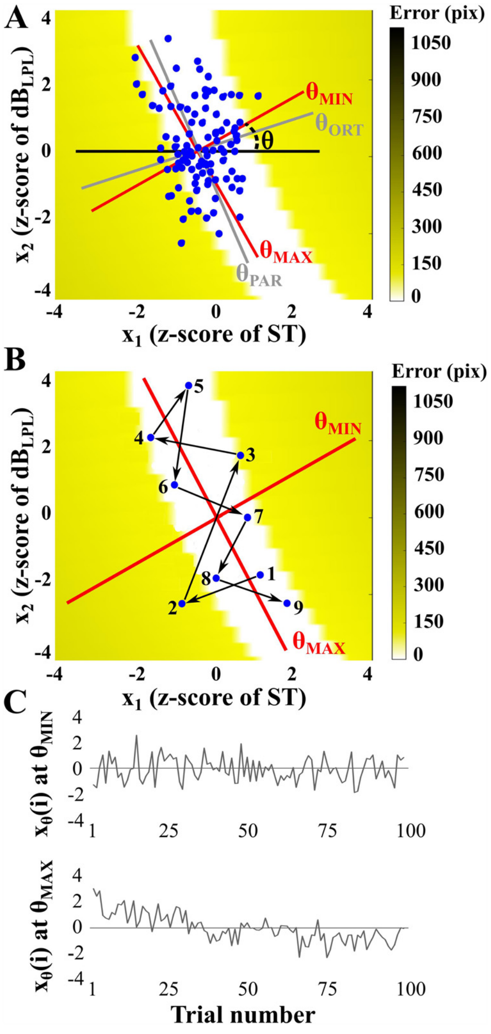

FIG. 3. (Color online) (A) Execution space and rotation axis used to analyze temporal structure in various directions. The execution space was normalized using the individual subject's mean and standard deviation. The gray lines show the directions/rotations parallel (θ_{PAR}) or orthogonal (θ_{ORT}) to the solution manifold and the red lines show the directions/rotations where the lag-1 autocorrelation coefficient (AC1) was lowest (θ_{MIN}) and highest (θ_{MAX}). Number of trials = 100. (B) Example using nine consecutive trials to illustrate trial-by-trial dynamics in the two-dimensional execution space. (C) The upper time series shows $x_\theta(i)$ for the rotation angle θ with the minimum AC1 value and the lower time series shows $x_\theta(i)$ for the rotation angle θ with the maximum AC1 value.

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 dataset using the following equation:

$$x_\theta(i) = x_1(i)\cos\theta + x_2(i)\sin\theta, \qquad (5)$$

where i is the trial index, $x_\theta(i)$ denotes the new time series after projection onto the line, and x_1 and x_2 denote the zscore of Δf_0 and $\mathrm{ACC}_{\mathrm{amp}}$, respectively. It has been shown that variability analysis is highly sensitive to the coordinate system and without normalization, the execution space has no metric (Sternad *et al.*, 2010). The angle θ of this line was zero when parallel to the x axis [Δf_0 direction, black line in Fig. 3(A)] and 0.5π rad when parallel to the y-axis ($\mathrm{ACC}_{\mathrm{amp}}$ direction). The center of the data was defined by the mean of Δf_0 and $\mathrm{ACC}_{\mathrm{amp}}$ for each block of 100 trials for each individual. This line was then rotated through $0 < \theta < \pi$ rad, 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 autocorrelation and detrended fluctuation analysis. 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} [gray lines in Fig. 3(A)].

G. Analysis of temporal variability

The temporal structure of $x_\theta(i)$ obtained for all rotation angles θ was evaluated by autocorrelation and detrended fluctuation analysis (DFA). These two analysis methods have been chosen because they provide statistical quantifications of temporal persistence/anti-persistence on short-term (autocorrelation) and long-term (DFA) time scales. More specifically, persistence is statistically defined when future fluctuations are likely to be in the same direction as current fluctuations; anti-persistence is defined when currently observed fluctuation are in the opposite direction of future fluctuations (Brenner *et al.*, 2013; Collins and De Luca, 1993; Collins and De Luca, 1994; John *et al.*, 2016). From the autocorrelation analysis, the lag-1 autocorrelation coefficient (AC1) was reported to evaluate trial-to-trial variability since AC1 provides a correlation between the signal and the signal shifted by 1 trial. Temporal structure beyond lag-1 was evaluated using DFA. The DFA method was chosen since it is a modification of the root-mean square analysis of a random walk that is relatively insensitive to non-stationarities 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 scaling index (SCI) was obtained from the slope of a linear regression (Peng *et al.*, 1995). This SCI has traditionally been used to estimate the fractal dimensions of a time series, which provides an indication of the statistical properties of the fluctuations contained in the time series (Duarte and Sternad, 2008; Feder, 1988). Sets of 100 trials were used in the analysis of directionality in execution space.

Temporal variability is classified as either uncorrelated *white noise* ($\mathrm{AC1} = 0$ and $\mathrm{SCI} = 0.5$); *anti-persistence* denoting stable dynamic behavior and error correction ($\mathrm{AC1} < 0$ and $\mathrm{SCI} < 0.5$), or *persistence* denoting unstable dynamic behavior and lack of error correction ($\mathrm{AC1} > 0$ and $\mathrm{SCI} > 0.5$) (Collins and De Luca, 1993; Collins and De Luca, 1994; Dingwell and Cusumano, 2010; Dingwell *et al.*, 2010). It was hypothesized that the lowest value of AC1 or SCI (labeled θ_{MIN}) would be coincident with the direction orthogonal to the solution manifold (θ_{ORT}), as this is the most error-relevant direction. The rotation angle in execution space with the highest values of AC1 or SCI (labeled θ_{MAX}) will be coincident with the direction parallel to the solution manifold (θ_{PAR}), which is the least error-relevant direction. Figure 3(A) represents the angles θ_{MIN} and θ_{MAX} in execution space, and Fig. 3(B) shows hypothetical data (nine consecutive trials) to clearly show trial-by-trial fluctuations that produce negative AC1 along the θ_{MIN} axis (trial-to-trial changes alternate left/right) and positive AC1 along the θ_{MAX} axis (trial-to-trial changes persist up/down). Figure 3(C) shows time series of execution variables, $x_\theta(i)$, at two different rotation angles—specifically, the signals at those angles with minimum AC1 (θ_{MIN}) and maximum AC1 (θ_{MAX}).

H. Statistical analysis

Performance improvement across practice was evaluated by fitting exponential functions to the error (minimum distance between the ball trajectory and target), where $y = ae^{-bx} + c$. Fits were performed for each participant and for the entire group's mean error, calculated from absolute values of 50 trials. Similarly, exponential functions were fitted to T-cost, N-cost, and C-cost over time, evaluated for blocks of 50 trials. The exponential functions assessed the different time scales of change between the costs and error via the time constant $1/b$. To further demonstrate significant improvement over practice, the first 50 trials from all practice days were analyzed with repeated measure analyses of variance (RM-ANOVA) with respect to mean error, T-cost, N-cost, and C-cost. Significant main effects were followed up with one-tailed paired t-tests (*post hoc*) since the hypothesized direction of change was in one direction (reduction of all metrics over time). All *post hoc* comparisons were in reference to day 1 (i.e., four total comparisons per metric) since human performance improvement is exponential (most reduction in error metrics occur very early in practice). Bonferroni p-value corrections were not done because this is a preliminary study where the focus was on finding statistically robust ($p < 0.05$) medium-to-large effect sizes (as determined via Cohen's d). The individual contribution of each cost towards error reduction was evaluated through Spearman's rank correlation coefficient ρ (T-cost, N-cost, or C-cost versus mean error) and resulting ρ values were tested for significant differences using three paired t-tests (T-cost versus N-cost, T-cost versus C-cost, and C-cost versus N-cost).

The subjects' sensitivity to directions in the execution space was assessed using all data from the last day (day 5) of practice. This data set consisted of 3 blocks (100 practice trials per block) from all ten subjects, where θ_{MIN} and θ_{MAX} were calculated for each block using AC1 or SCI (total practice blocks $= 30$). Two paired t-tests (one with data from AC1 and the other from SCI) were used to assess the difference between the 30-paired values of θ_{MIN} and θ_{MAX}. If subjects developed a sensitivity to how trial-to-trial fluctuations in the execution space related to performance error, the direction with minimum AC1 and SCI (θ_{MIN}) will be significantly lower than the direction with maximum AC1 and SCI

(θ_{MAX}). This will be interpreted as increased CNS control 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 (Cohen, 1988). All statistics were calculated using SPSS software (version 22.0, IBM, Armonk, NY). All variables were confirmed to have normal distributions via one-sample Kolmogorov-Smirnov tests.

III. RESULTS

A. Performance improvement

Before analyzing variability as it relates to practice, one must first demonstrate that the group of subjects significantly improved their performance. Figure 4(A) shows the progression of error across practice for all ten individuals who completed 5 days of practice. Only the exponential fits are shown for visual clarity. The fit across all subjects and the mean error per 50 trial block is shown in the inset of Fig. 4(A). One subject (in gray) started out with small error and maintained his relatively high performance throughout all days of practice. Even though this subject did not demonstrate a learning trend, he was included in all analyses. All other subjects had statistically significant exponential fits with r-values (the nonlinear correlation coefficient between the exponential fit and data) ranging from 0.45 to 0.77. An exponential function was also fitted across the group using overall mean error, which resulted in a significant r-value of 0.83 [see inset of Fig. 4(A)]. The subjects began practice at a mean error of 58.57 (15.23) pixels during their first block of trials (n = 100) and decreased it to 44.30 (12.46) pixels during their first block of trials on their last practice day (recall that a "hit" is defined as an error of 30 pixels or less). Therefore, even after 5 days of practice (1500 trials), the group of singers was not hitting the target on average; i.e., the task was challenging to learn. The time scale $1/b$ was 4.878, indicating that the time constant was approximately 244 individual trials (i.e., slightly less than one day of practice). Furthermore, the results of a RM-ANOVA indicated a significant change in error across practice days—Greenhouse-Geisser $F(2.07, 18.61) = 5.708$, $\eta^2 = 0.39$, $p = 0.011$—and *post hoc* t-tests demonstrated a statistically significant decrease in error for practice days 2–5 compared to day 1 ($p = 0.020$, $p < 0.001$, $p = 0.005$, $p = 0.001$, respectively) with large effect sizes (mean $d = 1.24$ across the four comparisons).

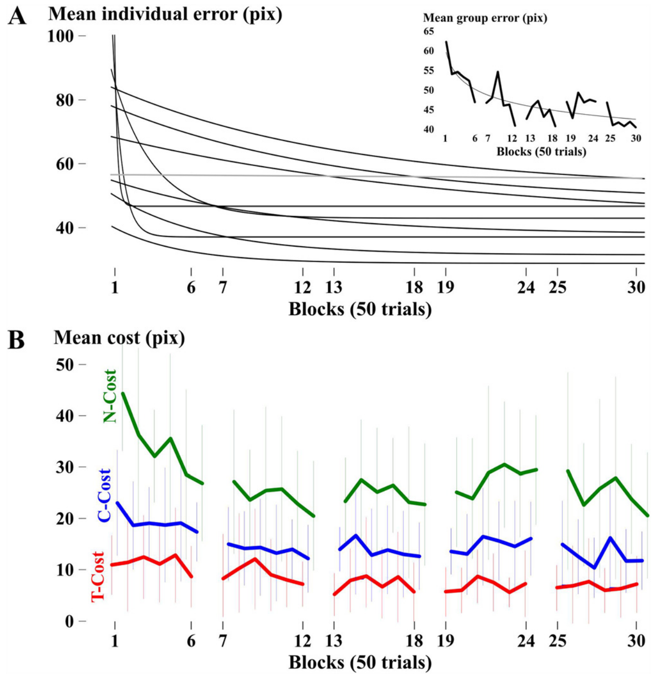

FIG. 4. (Color online) (A) Exponential fits to each individual's mean error computed for blocks of 50 trials, where the gray line represents the subject who did not show improvement across practice. The inset shows the mean error across all ten subjects with spacing between lines representing new days (six trial blocks per day for five days). (B) Means and standard deviations of T-cost (red), C-cost (blue), and N-cost (green) per trial block for all subjects.

B. Distributional variability—TNC-costs

Figure 2 showed exemplary data early in practice (day 1) and late in practice (day 5) from one subject. Comparing days 1 and 5, it can be seen that the distribution of execution variables moves towards a more error-tolerant location on the solution manifold (T-cost), the dispersion of the data cloud decreased (N-cost), and the anisotropy of the data changed to align with the solution manifold (C-cost). Figure 4(B) shows how the group averages of these TNC-costs decreased with practice. As hypothesized, N-cost was the highest cost to performance followed by C-cost, and finally T-cost. The results of the RM-ANOVA using N-cost and C-cost demonstrated a significantly decreased cost to performance across all practice days. More specifically, N-cost decreased across all practice days—Greenhouse-Geisser $F(2.64, 23.75) = 6.215$, $\eta^2 = 0.41$, $p = 0.001$—and *post hoc* paired t-tests demonstrated a significant decrease in N-cost for practice days 2–5 compared to day 1 ($p = 0.003$, $p < 0.001$, $p = 0.001$, $p = 0.007$, respectively) with large effect sizes (mean $d = 1.55$ across the four comparisons). C-cost decreased across all practice days—Greenhouse-Geisser $F(2.48, 22.35) = 3.83$, $\eta^2 = 0.30$, $p = 0.03$—and *post hoc* paired t-test demonstrated a significant decrease in C-cost for practice days 2–5 compared to day 1 ($p = 0.014$, $p = 0.003$, $p = 0.013$, $p = 0.037$, respectively) with large effect sizes (mean $d = 0.85$ across the four comparisons). T-cost changes, on the other hand, did not attain a significant main effect in the RM-ANOVA—Greenhouse-Geisser $F(1.94, 17.47) = 2.48$, $\eta^2 = 0.22$, $p = 0.11$—but *post hoc* paired t-tests demonstrated a significant decrease in T-cost for practice days 3–5 compared to day 1 ($p = 0.008$, 0.028, 0.023, respectively). The lack of a significant main effect is likely due to insufficient statistical power since the associated mean effect size across all four comparisons indicated a moderate-to-large decrease in T-cost ($d = 0.66$).

Exponential fit functions demonstrated significant r-values of 0.80 for T-cost, 0.84 for N-cost, and 0.82 for C-cost. N-cost and C-cost evolved at a faster time scale than T-cost. More specifically, the b-coefficients were 0.095 for T-cost, corresponding to a time constant of 526 trials, or slightly less than 2

TABLE II. Spearman correlations (ρ) between error and each of the TNC costs. Subject ID denotes sex of each individual: Male (M) and female (F).

Subject	T-cost	N-cost	C-cost
F3	0.825[a]	0.453[b]	0.348
F4	0.655[a]	0.079	0.025
F5	0.839[a]	0.388[b]	0.299
F6	0.679[a]	0.632[a]	0.608[a]
F7	0.485[c]	0.828[a]	0.583[a]
M2	0.586[a]	0.593[a]	0.747[a]
M3	0.382[b]	0.315	0.479[c]
M6	0.528[c]	0.697[a]	0.558[a]
M8	0.531[c]	0.624[a]	0.306
M9	0.457[b]	0.560[a]	0.453[b]
Mean (SD)	0.60 (0.15)	0.52 (0.22)	0.44 (0.21)

[a] $p < 0.001$.
[b] $p < 0.05$.
[c] $p < 0.01$.

days of practice. The time scale for C-cost was 0.246, corresponding to 203 trials, or slightly less than 1 day of practice. The time scale for N-cost was 0.428, equivalent to 117 trials, or approximately a half day of practice. These profiles were correlated with the error profiles. Table II displays the resulting rho values from Spearman's correlations between error and T-, N-, or C-cost, respectively. Qualitatively, the ranking from highest to lowest correlation across participants was different than hypothesized. However, paired t-tests demonstrated no significant difference between mean rho values when comparing between costs at a group level.

C. Temporal variability—AC1 and SCI at θ_{MIN} or θ_{MAX}

Figure 3(A) shows that, for the exemplary data set, θ_{MAX} aligns at an angle approximately parallel to the solution manifold (error-irrelevant direction θ), and θ_{MIN} forms an angle approximately orthogonal to the solution manifold (error-relevant direction θ). Additionally, Fig. 3(C) displays exemplary time series for the signal at minimum and maximum AC1 to illustrate how the two time series can be different. The signal at θ_{MIN} shows anti-persistent temporal structure fluctuating around the mean. In contrast, the signal at θ_{MAX} shows persistent temporal structure with a noticeable drift in the first half of the trials.

Figure 5 summarizes the results of the temporal analyses with respect to direction θ. According to paired t-tests, θ_{MIN} was significantly lower than θ_{MAX} in both AC1 and SCI analyses with very large effect sizes; AC1: $t(29) = 13.65$, $p < 0.001$, $d = 2.92$; SCI: $t(29) = 12.06$, $p < 0.001$, $d = 2.62$. Therefore, both temporal correlation measures demonstrated significantly different trial-by-trial dynamics between θ_{MIN} and θ_{MAX}; i.e., uncorrelated, stable dynamics at θ_{MIN} and persistent dynamics at θ_{MAX}, indicative of selective control depending on the direction in solution space.

IV. DISCUSSION

A voice-controlled video game has been described and was used to investigate whether theoretically based findings from limb motor learning were applicable to vocal motor learning. The relationship between motor performance and distributional variability were quantitatively examined, as well as the time scales over which error and TNC-costs evolved. Finally, temporal variability in the two-dimensional execution space was evaluated to determine if subjects' trial-to-trial behavior were sensitive to error-relevant directions in the solution manifold.

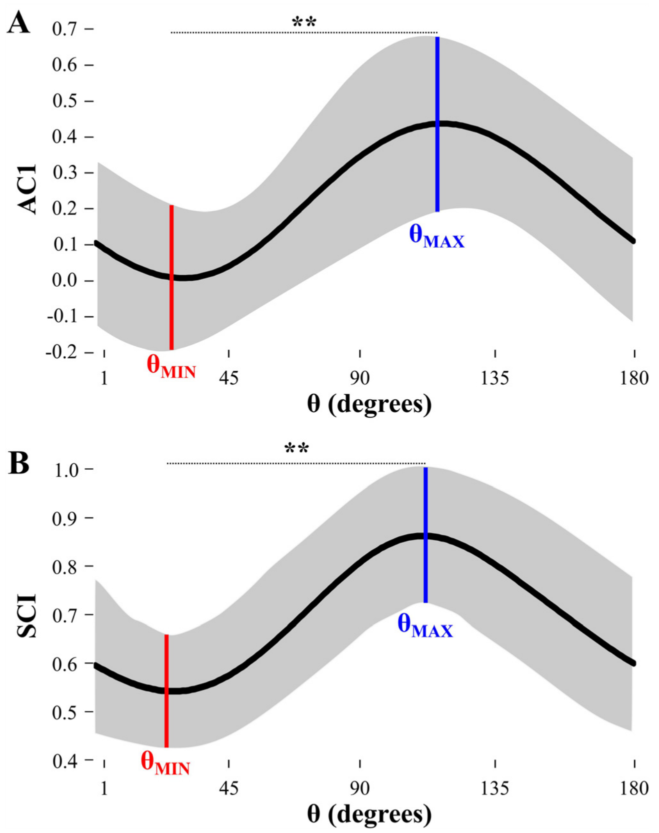

FIG. 5. (Color online) Analysis of temporal variability and directionality in execution space. (A) Lag-1 autocorrelation coefficient (AC1) and (B) detrended fluctuation analysis scaling index (SCI) of $x_\theta(i)$ as a function of rotation angle θ across all subjects' last day of practice (three blocks of 100 trials each). Solid horizontal black curves show the mean measure across all 30 practice blocks in each group, and shaded areas represent ±1 standard deviation around the mean. θ is indicated for which AC1 and SCI are minimum (θ_{MIN}) and maximum (θ_{MAX}), with their difference statistically significant (**$p < 0.01$).

Before interpreting the variability results, it will be helpful to first examine the final skilled vocal behavior exhibited by the group of ten professional singers. Figure 6 provides a visualization of how each subject performed in the execution space during his/her final block of 100 trials on practice day 5. A gray ellipse represents each subject's final block of practice and was centered at the final practice block's mean ACC_{amp} and Δf_0. Overall, it qualitatively appears that they all performed a relatively similar vocal gesture. On average, the subjects produced most harmonic intervals between a diminished and augmented fifth (mean ± 1 standard deviation; 6.98 ± 0.49 ST) at over two times louder than their lowest phonation level ($14.57\ dB_{LPL} \pm 1.74$ dB).

In order to interpret how difficult the task was (and how skilled the final behavior became), published perturbation studies were consulted for estimates of minimal controllable limits. These limits were approximately 0.20–0.25 ST for f_0

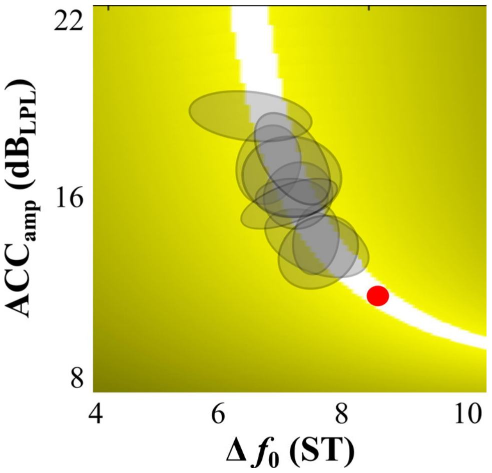

FIG. 6. (Color online) Summary of each subject's error in the execution space for their final block (100 trials) of practice on day 5, represented by gray ellipses. The aspect ratio of the execution space was scaled according to the estimated controllable limits such that a Δf_0 of 0.2 ST and $\mathrm{ACC_{amp}}$ of 0.5 $\mathrm{dB_{LPL}}$ represent the same length in the execution space (the red circle below the ellipses is a scale marker).

(Burnett and Larson, 2002; Hain *et al.*, 2000; Liu and Larson, 2007) and 0.5–0.75 dB for sound pressure level (Bauer *et al.*, 2006; Larson *et al.*, 2007). Studies on just noticeable differences provided even lower estimated limits, but they were not used since the derivations were based on perceptual judgments not connected to motor performance (Braida *et al.*, 1984; Fletcher and Munson, 1933; Klatt, 1973; Nikjeh *et al.*, 2008; Tervaniemi *et al.*, 2005). To show how these estimated minimal controllable limits related to game play, a circle (axes diameters = 0.2 ST and 0.5 dB) was placed on Fig. 6 (below the gray ellipses) where it could fit inside the solution manifold; i.e., trial-to-trial variability would be present, but the ball would still hit the target for every throw. First, the proximity of the standard deviation of Δf_0 and $\mathrm{ACC_{amp}}$ across all individual subjects (0.37 ST and 0.73 dB, respectively) to the estimated controllable limits (0.2 ST and 0.5 dB, respectively) speaks to the resulting elite vocal performance of the group as a whole and the high difficulty of the virtual task. Second, it is realistic to assume that trial-by-trial variations reflect volitional vocal control (or lack of control) in relation to error correction, since the standard deviation of Δf_0 and $\mathrm{ACC_{amp}}$ exhibited by the study subjects are above these estimated controllable limits. And finally, the subjects found a location on the solution manifold (gray ellipses in Fig. 6) that was reasonably tolerant to variability, i.e., the subjects could realistically reduce their variability and stay on the zero error, or white, solution manifold.

A. Distributional variability—TNC-costs

In general, the relationship between costs and error is similar to findings in limb movement studies. However, two results were different than what was expected based on previous limb work. T-cost exhibited the slowest time scale (which it typically demonstrates the fastest time scale of all cost metrics in limb studies) and C-cost and N-cost exhibited higher positive correlations with error than observed in limb research. These two results may be due to the baseline skill level of the ten subjects; i.e., they were all professional singers highly competent at manipulating pitch and loudness before practicing the game. Thus, the expert singers may have been able to find an error-tolerant space on the solution manifold very early in practice. Therefore, T-cost appeared to evolve at longer temporal periods than C-cost and N-cost. Also, the large group-based correlation coefficients between N-cost or C-cost and error indicated that subjects were past the stage of establishing a new vocal skill, and instead, were in the later stages of skill learning where they fine-tuned their motor performance. It has been noted in previous limb-based experiments that correlations between C-cost or N-cost and error increased during improvement of expert performance. Strong positive N-cost correlations with error have been noted in subjects who performed well during early practice (Abe and Sternad, 2013) and in subjects who had professional sports experience (Cohen and Sternad, 2009).

B. Temporal variability—Directionality in execution space

As seen in Fig. 5 (and supported via statistical results), trial-by-trial dynamics demonstrated clear modulations across different directions in the execution space, suggesting that the subjects were sensitive to the direction of the solution manifold. Furthermore, θ_{MIN} for AC1 was around zero instead of negative and θ_{MIN} for SCI was around 0.5 instead of <0.5; i.e., both minima should be characterized as white noise instead of anti-persistence/error correction. One would have expected to find negative AC1 values or SCI values <0.5 at θ_{MIN} since, theoretically, low values would represent error-correction—and anti-persistent structure has been found at θ_{MIN} in studies using well-learned behaviors (e.g., walking or pointing) (Dingwell *et al.*, 2010; van Beers *et al.*, 2013). This apparent deviation was also reported in a previous study using a similar task and explained through a computational learning model. This model suggested that additive noise sources (e.g., perceptual and motor noise) could obscure corrective errors from feedback and result in positive AC1 and SCI at θ_{MIN} (Abe and Sternad, 2013). Furthermore, the mapping from execution to error is highly nonlinear; therefore, perceiving the direct relation between error and execution is difficult and the resulting time series of execution variables may not demonstrate strong anti-persistent/error-correction patterns. Last, the significantly lower AC1 and SCI at θ_{MIN} compared to θ_{MAX} support the hypothesis that subjects became sensitive to the directionality of the underlying non-linear solution manifold and compensated on a trial-by-trial basis in only those directions that were closely related to a change the error.

V. CONCLUSION AND FUTURE DIRECTIONS

The results of this investigation suggest that vocal motor learning shares similar features to limb motor learning—especially how variability is modified over time to reduce the overall error. This is particularly significant as replicating findings from motor control/learning studies based on limb movements (outside of perturbation and adaptation paradigms) are not often successful when applied to bulbar movements/skills (e.g., speech, swallowing, voice); for reviews see

Bislick *et al.*, 2012; Maas *et al.*, 2008. To date, the study of variability in vocal motor behavior has focused primarily on decreases or increases in relation to performance improvement or normal versus pathological. In contrast, the present study demonstrates that the vocal motor system may not be primarily concerned with simply decreasing variability, but selectively channeling temporal and distributional variability according to the task demands. Furthermore, selective control of variability may even be indicative of late-stage skill learning or habituation.

This paradigm warrants future work to investigate the potential for quantification of improvements in voice training and novel clinical capabilities in the field of voice disorder rehabilitation—the ultimate goal being to improve assessment and treatment of patients with voice disorders. For example, the TNC-costs approach enabled novel characterizations of upper-limb control and learning in children with dystonia (Chu *et al.*, 2016) and a virtual throwing task has been used in a group of patients with Parkinson's disease (Pendt *et al.*, 2012). Consequently, it seems feasible that this approach may be applied to patients with neurologic disorders associated with vocal deterioration (e.g., Parkinson's disease, multiple sclerosis, amyotrophic lateral sclerosis). New voice-specific diagnostic features (i.e., vocal biomarkers) could be derived from the game through analysis of the Δf_0–ACC$_{\text{amp}}$ trajectory, TNC-costs, or directional variability in execution space. These features could also become therapeutically important if they correlated with behavioral retention or longer-term learning. For example, the degree of correlation between N- or C-cost and error, or directional variability analysis could help quantify if a patient is implicitly aligning his/her performance with the underlying solution manifold. Also, since the current investigation included only subjects with high/expert baseline levels of vocal skill (i.e., professional singers), future investigations should test this paradigm in those with novice levels of baseline skill.

The voice-controlled slingshot task may help patients improve their specific pathological vocal features that are subtle and otherwise difficult to volitionally vary. Also, just as singers practice vocalises and pianists practice scales, the slingshot task could be a platform for patients to improve basic vocal skills or for experts to improve vocal technique. Patients who demonstrate minimal kinesthetic/proprioceptive awareness may find indirect benefits (i.e., learning to learn) through using the visual feedback to help establish a complex two-dimensional pitch-loudness vocal skill. Additionally, patients with Parkinson's disease may benefit from playing the game in its current state, since the vocal execution variables are related to known targets for therapy (pitch and loudness) (Ramig *et al.*, 2001). This is also true of vocal function exercises since they hypothetically focus on rebalancing and strengthening the phonatory system through prolonged soft voicing at specific pitches (Stemple *et al.*, 1994). It could additionally be hypothesized that adherence to therapy or therapeutic outcomes would be improved with the addition of a "game" to treatment regimens. The features used to control the virtual slingshot could vary according to patient diagnosis; e.g., for patients with vocal hyperfunction, vocal intensity could be replaced with a measure of aerodynamic importance such as subglottal pressure or glottal airflow (Fryd *et al.*, 2016; Zañartu *et al.*, 2014) or a ratio of aerodynamics to acoustics (i.e., a vocal efficiency type measure). Even more specifically, future work that has potential for clinical adoption (i.e., voice therapy) should include the addition of voice quality measures (e.g., cepstral peak prominence, spectral tilt, noise-to-harmonics ratio) or laryngeal biomechanics (e.g., open/closed quotient, maximum flow declination rate, AC Flow). Finally, the acquisition/retention paradigm in motor learning may be applied to this video game to study the effects of varying feedback and practice variables (e.g., concurrent/terminal feedback, frequency of feedback, massed/distributed practice) (Schmidt and Lee, 2011).

ACKNOWLEDGMENTS

This work was supported by the National Institutes of Health through Grant Nos. R01-HD045639, R01-HD081346, and R01-HD087089 (awarded to D.S.); R33-DC011588 and P50-DC015446 (awarded to R.H.); F31-DC014412 (awarded to J.V.S.); and the Voice Health Institute. The article's contents are solely the responsibility of the authors and do not necessarily represent the official views of the NIH. The authors would like to thank Harold Cheyne for developing the accelerometer circuit.

Abe, M. O., and Sternad, D. (**2013**). "Directionality in distribution and temporal structure of variability in skill acquisition," Front. Human Neurosci. **7**, 225.

Ajemian, R., D'Ausilio, A., Moorman, H., and Bizzi, E. (**2013**). "A theory for how sensorimotor skills are learned and retained in noisy and nonstationary neural circuits," Proc. Natl. Acad. Sci. U.S.A. **110**(52), E5078–E5087.

Asaka, T., Wang, Y., Fukushima, J., and Latash, M. L. (**2008**). "Learning effects on muscle modes and multi-mode postural synergies," Exp. Brain Res. **184**(3), 323–338.

Bauer, J. J., Mittal, J., Larson, C. R., and Hain, T. C. (**2006**). "Vocal responses to unanticipated perturbations in voice loudness feedback: An automatic mechanism for stabilizing voice amplitude," J. Acoust. Soc. Am. **119**(4), 2363–2371.

Bernstein, N. A. (**1967**). The Co-ordination and Regulation of Movements (Pergamon, London).

Bislick, L. P., Weir, P. C., Spencer, K., Kendall, D., and Yorkston, K. M. (**2012**). "Do principles of motor learning enhance retention and transfer of speech skills? A systematic review," Aphasiology **26**(5), 709–728.

Braida, L. D., Lim, J. S., Berliner, J. E., Durlach, N. I., Rabinowitz, W. M., and Purks, S. R. (**1984**). "Intensity perception. XIII. Perceptual anchor model of context-coding," J. Acoust. Soc. Am. **76**(3), 722–731.

Brandon, C. A., Rosen, C., Georgelis, G., Horton, M. J., Mooney, M. P., and Sciote, J. J. (**2003**). "Staining of human thyroarytenoid muscle with myosin antibodies reveals some unique extrafusal fibers, but no muscle spindles," J. Voice **17**(2), 245–254.

Brenner, E., Cañal-Bruland, R., and van Beers, R. J. (**2013**). "How the required precision influences the way we intercept a moving object," Exp. Brain Res. **230**(2), 207–218.

Buekers, R. (**1998**). "Are voice endurance tests able to assess vocal fatigue?," Clin. Otolaryngol. Allied Sci. **23**(6), 533–538.

Burnett, T. A., Freedland, M. B., Larson, C. R., and Hain, T. C. (**1998**). "Voice F0 responses to manipulations in pitch feedback," J. Acoust. Soc. Am. **103**(6), 3153–3161.

Burnett, T. A., and Larson, C. R. (**2002**). "Early pitch-shift response is active in both steady and dynamic voice pitch control," J. Acoust. Soc. Am. **112**(3), 1058–1063.

Chen, S. H., Liu, H., Xu, Y., and Larson, C. R. (**2007**). "Voice F0 responses to pitch-shifted voice feedback during English speech," J. Acoust. Soc. Am. **121**(2), 1157–1163.

Cheyne, H. A., Hanson, H. M., Genereux, R. P., Stevens, K. N., and Hillman, R. E. (**2003**). "Development and testing of a portable vocal accumulator," J. Speech Lang. Hear. Res. **46**(6), 1457–1467.

Chu, V. W., Park, S.-W., Sanger, T., and Sternad, D. (**2016**). "Children with dystonia can learn a novel motor skill: Strategies that are tolerant to high variability," IEEE Trans. Neural Syst. Rehab. Eng. **24**(8), 847–858.

Cohen, J. (**1988**). *Statistical Power Analysis for the Behavioral Sciences*, 2nd ed. (Earlbaum, Hillsdale).

Cohen, R., and Sternad, D. (**2009**). "Variability in motor learning: Relocating, channeling and reducing noise," Exp. Brain Res. **193**(1), 69–83.

Coleman, R. F. (**1988**). "Comparison of microphone and neck-mounted accelerometer monitoring of the performing voice," J. Voice **2**(3), 200–205.

Collins, J. J., and De Luca, C. J. (**1993**). "Open-loop and closed-loop control of posture: A random-walk analysis of center-of-pressure trajectories," Exp. Brain Res. **95**(2), 308–318.

Collins, J. J., and De Luca, C. J. (**1994**). "Random walking during quiet standing," Phys. Rev. Lett. **73**(5), 764.

Cusumano, J. P., and Cesari, P. (**2006**). "Body-goal variability mapping in an aiming task," Biol. Cybern. **94**(5), 367–379.

Dingwell, J. B., and Cusumano, J. P. (**2010**). "Re-interpreting detrended fluctuation analyses of stride-to-stride variability in human walking," Gait Posture **32**(3), 348–353.

Dingwell, J. B., John, J., and Cusumano, J. P. (**2010**). "Do humans optimally exploit redundancy to control step variability in walking?," PLoS Comput. Biol. **6**(7), e1000856.

Domkin, D., Laczko, J., Djupsjöbacka, M., Jaric, S., and Latash, M. L. (**2005**). "Joint angle variability in 3D bimanual pointing: Uncontrolled manifold analysis," Exp. Brain Res. **163**(1), 44–57.

Duarte, M., and Sternad, D. (**2008**). "Complexity of human postural control in young and older adults during prolonged standing," Exp. Brain Res. **191**(3), 265–276.

Faisal, A. A., Selen, L. P., and Wolpert, D. M. (**2008**). "Noise in the nervous system," Nat. Rev. Neurosci. **9**(4), 292–303.

Feder, J. (**1988**). *Fractals* (Plenum, New York).

Ferrand, C. T. (**1995**). "Effects of practice with and without knowledge of results on jitter and shimmer levels in normally speaking women," J. Voice **9**(4), 419–423.

Fletcher, H., and Munson, W. A. (**1933**). "Loudness, its definition, measurement and calculation," Bell Syst. Tech. J. **12**(4), 377–430.

Fryd, A. S., Van Stan, J. H., Hillman, R. E., and Mehta, D. D. (**2016**). "Estimating subglottal pressure from neck-surface acceleration during normal voice production," J. Speech Lang. Hear Res. **59**(6), 1335–1345.

Guenther, F. H., Ghosh, S. S., and Tourville, J. A. (**2006**). "Neural modeling and imaging of the cortical interactions underlying syllable production," Brain Lang. **96**(3), 280–301.

Hain, T. C., Burnett, T. A., Kiran, S., Larson, C. R., Singh, S., and Kenney, M. K. (**2000**). "Instructing subjects to make a voluntary response reveals the presence of two components to the audio-vocal reflex," Exp. Brain Res. **130**(2), 133–141.

Helfer, B. S., Quatieri, T. F., Williamson, J. R., Keyes, L., Evans, B., Greene, W. N., Vian, T., Lacirignola, J., Shenk, T. E., and Talavage, T. M. (**2014**). "Articulatory dynamics and coordination in classifying cognitive change with preclinical mTBI," in *INTERSPEECH*, pp. 485–489.

Herzel, H., Berry, D., Titze, I. R., and Saleh, M. (**1994**). "Analysis of vocal disorders with methods from nonlinear dynamics," J. Speech Lang. Hear. Res. **37**(5), 1008–1019.

Hogikyan, N. D., and Sethuraman, G. (**1999**). "Validation of an instrument to measure voice-related quality of life (V-RQOL)," J. Voice **13**(4), 557–569.

Holmberg, E. B., Doyle, P., Perkell, J. S., Hammarberg, B., and Hillman, R. E. (**2003**). "Aerodynamic and acoustic voice measurements of patients with vocal nodules: Variation in baseline and changes across voice therapy," J. Voice **17**(3), 269–282.

Horwitz, R., Quatieri, T. F., Helfer, B. S., Yu, B., Williamson, J. R., and Mundt, J. (**2013**). "On the relative importance of vocal source, system, and prosody in human depression," in *IEEE International Conference on Body Sensor Networks*, pp. 1–6.

Jacobson, B. H., Johnson, A., Grywalski, C., Silbergleit, A., Jacobson, G., Benninger, M. S., and Newman, C. W. (**1997**). "The Voice Handicap Index (VHI): Development and validation," Am. J. Speech-Lang. Pathol. **6**(3), 66–70.

John, J., and Cusumano, J. P. (**2007**). "Inter-trial dynamics of repeated skilled movements," in *ASME 2007 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference* (American Society of Mechanical Engineers, New York), pp. 707–716.

John, J., Dingwell, J. B., and Cusumano, J. P. (**2016**). "Error correction and the structure of inter-trial fluctuations in a redundant movement task," PLoS Comput. Biol. **12**(9), e1005118.

Kandel, E. R., Schwartz, J. H., and Jessell, T. M. (**2000**). *Principles of Neural Science* (McGraw-Hill, New York).

Kelso, J., Fuchs, A., Lancaster, R., Holroyd, T., Cheyne, D., and Weinberg, H. (**1998**). "Dynamic cortical activity in the human brain reveals motor equivalence," Nature **392**(6678), 814–818.

Kempster, G. B., Gerratt, B. R., Verdolini Abbott, K., Barkmeier-Kraemer, J., and Hillman, R. E. (**2009**). "Consensus auditory-perceptual evaluation of voice: Development of a standardized clinical protocol," Am. J. Speech-Lang. Pathol. **18**(2), 124–132.

Klatt, D. H. (**1973**). "Discrimination of fundamental frequency contours in synthetic speech: Implications for models of pitch perception," J. Acoust. Soc. Am. **53**(1), 8–16.

Kreiman, J., Gerratt, B. R., Kempster, G. B., Erman, A., and Berke, G. S. (**1993**). "Perceptual evaluation of voice quality: Review, tutorial, and a framework for future research," J. Speech Hear. Res. **36**(1), 21–40.

Kudo, K., Tsutsui, S., Ishikura, T., Ito, T., and Yamamoto, Y. (**2000**). "Compensatory coordination of release parameters in a throwing task," J. Motor Behav. **32**(4), 337–345.

Larson, C. R., Burnett, T. A., Kiran, S., and Hain, T. C. (**2000**). "Effects of pitch-shift velocity on voice F∼ 0 responses," J. Acoust. Soc. Am. **107**, 559–564.

Larson, C. R., Sun, J., and Hain, T. C. (**2007**). "Effects of simultaneous perturbations of voice pitch and loudness feedback on voice F0 and amplitude control," J. Acoust. Soc. Am. **121**(5), 2862–2872.

Little, M. A., McSharry, P. E., Roberts, S. J., Costello, D. A., and Moroz, I. M. (**2007**). "Exploiting nonlinear recurrence and fractal scaling properties for voice disorder detection," BioMed. Eng. Online **6**(1), 23.

Liu, H., and Larson, C. R. (**2007**). "Effects of perturbation magnitude and voice F0 level on the pitch-shift reflex," J. Acoust. Soc. Am. **122**(6), 3671–3677.

Ma, E. P.-M., Yiu, G. K.-Y., and Yiu, E. M.-L. (**2013**). "The effects of self-controlled feedback on learning of a 'relaxed phonation task'," J. Voice **27**(6), 723–728.

Maas, E., Robin, D. A., Austermann Hula, S. N., Freedman, S. E., Wulf, G., Ballard, K. J., and Schmidt, R. A. (**2008**). "Principles of motor learning in treatment of motor speech disorders," Am. J. Speech-Lang. Pathol. **17**(3), 277–298.

Marinelli, L., Quartarone, A., Hallett, M., Frazzitta, G., and Ghilardi, M. F. (**2017**). "The many facets of motor learning and their relevance for Parkinson's disease," Clin. Neurophysiol. **128**(7), 1127.

Martin, T. A., Greger, B. E., Norris, S. A., and Thach, W. T. (**2001**). "Throwing accuracy in the vertical direction during prism adaptation: Not simply timing of ball release," J. Neurophysiol. **85**(5), 2298–2302.

Mehta, D. D., Van Stan, J. H., and Hillman, R. E. (**2016**). "Relationships between vocal function measures derived from an acoustic microphone and a subglottal neck-surface accelerometer," IEEE/ACM Trans. Audio Speech Lang. Process. **24**(4), 659–668.

Müller, H., and Sternad, D. (**2003**). "A randomization method for the calculation of covariation in multiple nonlinear relations: Illustrated with the example of goal-directed movements," Biol. Cybern. **89**(1), 22–33.

Müller, H., and Sternad, D. (**2004**). "Decomposition of variability in the execution of goal-oriented tasks: Three components of skill improvement," J. Exp. Psychol.: Human Percept. Perform. **30**(1), 212.

Müller, H., and Sternad, D. (**2009**). "Motor learning: Changes in the structure of variability in a redundant task," in *Progress in Motor Control* (Springer, Berlin), pp. 439–456.

Nikjeh, D. A., Lister, J. J., and Frisch, S. A. (**2008**). "Hearing of note: An electrophysiologic and psychoacoustic comparison of pitch discrimination between vocal and instrumental musicians," Psychophysiology **45**(6), 994–1007.

Patel, R., Brumberg, J., Veilleux, N., and Shattuck-Hufnagel, S. (**2012**). "Prosodic Marionette," Communication Analysis and Design Laboratory, Northeastern University.

Pendt, L. K., Maurer, H., and Müller, H. (**2012**). "The influence of movement initiation deficits on the quantification of retention in Parkinson's disease," Front. Hum. Neurosci. **6**, 226.

Peng, C. K., Havlin, S., Stanley, H. E., and Goldberger, A. L. (**1995**). "Quantification of scaling exponents and crossover phenomena in nonstationary heartbeat time series," Chaos Interdisc. J. Nonlin. Sci. **5**(1), 82–87.

Rácz, K., and Valero-Cuevas, F. (**2013**). "Spatio-temporal analysis reveals active control of both task-relevant and task-irrelevant variables," Front. Comput. Neurosci. **7**, 155.

Ramig, L. O., Countryman, S., Thompson, L. L., and Horii, Y. (**1995**). "Comparison of two forms of intensive speech treatment for Parkinson disease," J. Speech Lang. Hear. Res. **38**(6), 1232–1251.

Ramig, L. O., Sapir, S., Fox, C., and Countryman, S. (**2001**). "Changes in vocal loudness following intensive voice treatment (LSVT®) in individuals with Parkinson's disease: A comparison with untreated patients and normal age-matched controls," Move. Disord. **16**(1), 79–83.

Ramig, L. O., and Verdolini, K. (**1998**). "Treatment efficacy voice disorders," J. Speech Lang. Hear. Res. **41**(1), S101–S116.

Roy, N., Weinrich, B., Gray, S. D., Tanner, K., Stemple, J. C., and Sapienza, C. M. (**2003**). "Three treatments for teachers with voice disorders: A randomized clinical trial," J. Speech Lang. Hear. Res. **46**(3), 670–688.

Roy, N., Weinrich, B., Gray, S. D., Tanner, K., Toledo, S. W., Dove, H., Corbin-Lewis, K., and Stemple, J. C. (**2002**). "Voice amplification versus vocal hygiene instruction for teachers with voice disorders: A treatment outcomes study," J. Speech Lang. Hear. Res. **45**(4), 625–638.

Russell, S., and Norvig, P. (**2002**). *Artificial Intelligence* (Prentice Hall, New York).

Schalling, E., Gustafsson, J., Ternstrom, S., Bulukin Wilen, F., and Sodersten, M. (**2013**). "Effects of tactile biofeedback by a portable voice accumulator on voice sound level in speakers with Parkinson's disease," J. Voice **27**(6), 729–737.

Schmidt, R. A., and Lee, T. D. (**2011**). *Motor Control and Learning: A Behavioral Emphasis* (Human Kinetics, Champaign, IL).

Schneider, B., and Bigenzahn, W. (**2005**). "How we do it: Voice therapy to improve vocal constitution and endurance in female student teachers," Clin. Otolaryngol. **30**(1), 66–71.

Schneider, B., Enne, R., Cecon, M., Diendorfer-Radner, G., Wittels, P., Bigenzahn, W., and Johannes, B. (**2006**). "Effects of vocal constitution and autonomic stress-related reactivity on vocal endurance in female student teachers," J. Voice **20**(2), 242–250.

Scholz, J. P., Schöner, G., and Latash, M. L. (**2000**). "Identifying the control structure of multijoint coordination during pistol shooting," Exp. Brain Res. **135**(3), 382–404.

Simonyan, K., and Horwitz, B. (**2011**). "Laryngeal motor cortex and control of speech in humans," Neuroscientist **17**(2), 197–208.

Steinhauer, K., and Grayhack, J. P. (**2000**). "The role of knowledge of results in performance and learning of a voice motor task," J. Voice **14**(2), 137–145.

Stemple, J. C., Lee, L., D'Amico, B., and Pickup, B. (**1994**). "Efficacy of vocal function exercises as a method of improving voice production," J. Voice **8**(3), 271–278.

Sternad, D., Huber, M. E., and Kuznetsov, N. (**2014**). "Acquisition of novel and complex motor skills: Stable solutions where intrinsic noise matters less," in *Progress in Motor Control* (Springer, Berlin), pp. 101–124.

Sternad, D., Park, S.-W., Muller, H., and Hogan, N. (**2010**). "Coordinate dependence of variability analysis," PLoS Comput. Biol. **6**(4), e1000751.

Studenka, B. E., Dorsch, T. E., Ferguson, N. L., Olsen, C. S., and Gordin, R. D. (**2017**). "Nonlinear assessment of motor variability during practice and competition for individuals with different motivational orientations," Learn. Motiv. **58**, 16–26.

Sugimoto, T., and Hiki, S. (**1960**). "Extraction of the pitch of a voice from the vibration of the outer skin of the trachea," J. Acoust. Soc. Jpn. **1**(4), 291–293.

Švec, J. G., Titze, I. R., and Popolo, P. S. (**2005**). "Estimation of sound pressure levels of voiced speech from skin vibration of the neck," J. Acoust. Soc. Am. **117**(3), 1386–1394.

Tervaniemi, M., Just, V., Koelsch, S., Widmann, A., and Schröger, E. (**2005**). "Pitch discrimination accuracy in musicians vs nonmusicians: An event-related potential and behavioral study," Exp. Brain Res. **161**(1), 1–10.

Titze, I. R. (**1992**). "Vocal efficiency," J. Voice **6**(2), 135–138.

Titze, I. R. (**2006**). "Voice training and therapy with a semi-occluded vocal tract: Rationale and scientific underpinnings," J. Speech Lang. Hear. Res. **49**(2), 448–459.

Tourville, J. A., and Guenther, F. H. (**2011**). "The DIVA model: A neural theory of speech acquisition and production," Lang. Cogn. Process. **26**(7), 952–981.

Tourville, J. A., Reilly, K. J., and Guenther, F. H. (**2008**). "Neural mechanisms underlying auditory feedback control of speech," Neuroimage **39**(3), 1429–1443.

van Beers, R. J., Brenner, E., and Smeets, J. B. (**2013**). "Random walk of motor planning in task-irrelevant dimensions," J. Neurophysiol. **109**(4), 969–977.

van Leer, E., Pfister, R. C., and Zhou, X. (**2016**). "An iOS-based cepstral peak prominence application: Feasibility for patient practice of resonant voice.," J. Voice **31**(1), 131.e9–131.e16.

Van Stan, J. H., Mehta, D. D., and Hillman, R. E. (**2015**). "The effect of voice ambulatory biofeedback on the daily performance and retention of a modified vocal motor behavior in participants with normal voices," J. Speech Lang. Hear. Res. **58**(3), 713–721.

Van Stan, J. H., Mehta, D. D., Petit, R., Sternad, D., Muise, J., Burns, J. A., and Hillman, R. E. (**2017**). "Integration of motor learning principles into real-time ambulatory voice biofeedback and example implementation via a clinical case study with vocal fold nodules," Am. J. Speech-Lang. Pathol. **26**(1), 1–10.

Verdolini-Marston, K., Katherine Burke, M., Lessac, A., Glaze, L., and Caldwell, E. (**1995**). "Preliminary study of two methods of treatment for laryngeal nodules," J. Voice **9**(1), 74–85.

Williamson, J. R., Quatieri, T. F., Helfer, B. S., Ciccarelli, G., and Mehta, D. D. (**2014**). "Vocal and facial biomarkers of depression based on motor incoordination and timing," in *Proceedings of the 4th International Workshop on Audio/Visual Emotion Challenge*, pp. 65–72.

Williamson, J. R., Quatieri, T. F., Helfer, B. S., Ciccarelli, G., and Mehta, D. D. (**2015**). "Segment-dependent dynamics in predicting Parkinson's disease," in *Proceedings of INTERSPEECH*, pp. 1–5.

Wong, A. Y.-H., Ma, E. P.-M., and Yiu, E. M.-L. (**2011**). "Effects of practice variability on learning of relaxed phonation in vocally hyperfunctional speakers," J. Voice **25**(3), e103–e113.

Wulf, G., Höß, M., and Prinz, W. (**1998**). "Instructions for motor learning: Differential effects of internal versus external focus of attention," J. Motor Behav. **30**(2), 169–179.

Xu, Y., Larson, C. R., Bauer, J. J., and Hain, T. C. (**2004**). "Compensation for pitch-shifted auditory feedback during the production of Mandarin tone sequences," J. Acoust. Soc. Am. **116**(2), 1168–1178.

Zañartu, M., Galindo, G. E., Erath, B. D., Peterson, S. D., Wodicka, G. R., and Hillman, R. E. (**2014**). "Modeling the effects of a posterior glottal opening on vocal fold dynamics with implications for vocal hyperfunctional," J. Acoust. Soc. Am. **136**(6), 3262–3271.

Zarate, J. M., Wood, S., and Zatorre, R. J. (**2010**). "Neural networks involved in voluntary and involuntary vocal pitch regulation in experienced singers," Neuropsychologia **48**(2), 607–618.

Zhang, Y., and Jiang, J. J. (**2008**). "Acoustic analyses of sustained and running voices from patients with laryngeal pathologies," J. Voice **22**(1), 1–9.

Zhang, Y., McGilligan, C., Zhou, L., Vig, M., and Jiang, J. J. (**2004**). "Nonlinear dynamic analysis of voices before and after surgical excision of vocal polyps," J. Acoust. Soc. Am. **115**(5), 2270–2277.