Research Article

Efficacy of Conversation Training Therapy for Patients With Benign Vocal Fold Lesions and Muscle Tension Dysphonia Compared to Historical Matched Control Patients

Amanda I. Gillespie,^a Jonathan Yabes,^b Clark A. Rosen,^c and Jackie L. Gartner-Schmidt^d

Purpose: Conversation training therapy (CTT) is the 1st voice therapy approach to eliminate the traditional therapeutic hierarchy and use patient-driven conversation as the sole therapeutic stimulus. The purpose of this investigation was to determine the efficacy of CTT compared to standard-of-care voice therapy approaches for the treatment of patients with voice disorders.

Method: A prospective study of CTT treatment outcomes in adults with dysphonia due to primary muscle tension dysphonia or benign vocal fold lesions compared to age, gender, and diagnosis historical matched control (HMC) patients was used. The primary outcome was change in Voice Handicap Index–10 (VHI-10); secondary outcomes included acoustic, aerodynamic, and auditory-perceptual outcomes. Data were collected before treatment (baseline), at the start of each therapy session, 1 week after the final therapy session (short-term follow-up), and 3 months after the final therapy session (long-term follow-up).

Results: For the CTT group, statistically significant improvements were observed for VHI-10. Though statistically

significant improvements were observed for the VHI-10 for the HMC group, the CTT group saw significantly greater improvement in VHI-10. Furthermore, equivalent gains were observed following only 2 sessions of CTT compared to 4–8 sessions of traditional therapy. Significant improvements in the CTT group were observed for cepstral peak prominence in a vowel, fundamental frequency, Cepstral Spectral Index of Dysphonia in a vowel and connected speech, vocal intensity, average airflow in speech in a reading passage, number of breaths and duration of reading passage, and auditory-perceptual measurement of overall voice severity.

Conclusions: Results support the hypothesis that training voice techniques in the context of spontaneous conversational speech improves patient perception of voice handicap and acoustic, aerodynamic, and auditory-perceptual voice outcomes both immediately following treatment and at long-term follow-up. CTT participants also demonstrated significantly larger decreases in VHI-10 compared to HMC participants who received standard-of-care, nonconversational, hierarchical-based voice therapy.

Voice therapy is a patient-centered treatment method used to modify behaviors that cause or contribute to voice disorders. Voice therapy is the most common intervention for many of the nearly 88 million people

in the United States who experience voice disorders (Roy, Merrill, Gray, & Smith, 2005; U.S. Census Bureau, 2005). Benign vocal fold lesions and primary muscle tension dysphonia (MTD) constitute over 60% of voice clinic case-loads (Angsuwarangsee & Morrison, 2002; Cohen, Kim, Roy, Asche, & Courey, 2012; Mozzanica et al., 2016; Roy, 2003; Van Houtte, Van Lierde, D'Haeseleer, & Claeys, 2010), and over 80% of patients with lesions and nearly 100% of those with MTD are treated exclusively with voice therapy (Rosen et al., 2012; Verdolini, Rosen, & Branski, 2005). Hierarchical, non-conversation-based voice therapy is the standard of care for most speech-language pathologists (SLPs) in clinical practice (e.g., Casper & Murry,

^aDepartment of Otolaryngology, Emory University, Atlanta, GA

^bDepartment of Biostatistics, University of Pittsburgh, PA

^cDepartment of Otolaryngology, University of California, San Francisco

^dDepartment of Otolaryngology, University of Pittsburgh, PA

Correspondence to Amanda I. Gillespie: Amanda.i.gillespie@emory.edu

Editor-in-Chief: Bharath Chandrasekaran

Editor: Jack Jiang

Received March 26, 2019

Revision received July 8, 2019

Accepted August 8, 2019

https://doi.org/10.1044/2019_JSLHR-S-19-0136

Disclosure: A. I. Gillespie and J. L. Gartner-Schmidt receive royalties from MedBridge, Inc., for educational videos on voice therapy topics, including those discussed in this article. All other authors have declared that no competing interests existed at the time of publication.

2000; Chen, Hsiao, Hsiao, Chung, & Chiang, 2007; Niebudek-Bogusz et al., 2008). A voice therapy hierarchy is an approach that builds in complexity from such components as breathing exercises, to single sounds and short phrases, to conversation last, if at all. Most published therapy programs follow such a hierarchy and address conversational voice only minimally (Bassiouny, 1998; Carding, Horsley, & Docherty, 1999; Casper & Murry, 2000; Chen et al., 2007; Holmberg, Hillman, Hammarberg, Södersten, & Doyle, 2001; Niebudek-Bogusz et al., 2008; Rodriguez-Parra, Adrián, & Casado, 2011; Schneider & Bigenzahn, 2005; Sellars, Carding, Deary, MacKenzie, & Wilson, 2002; van Leer & Connor, 2012). However, research on patient perceptions of voice therapy indicates a clear need to target conversational voice use in treatment (Iwarsson, 2015; Iwarsson, Morris, & Balling, 2017; Ohlsson, 2016; Ziegler, Dastolfo, Hersan, Rosen, & Gartner-Schmidt, 2014). The purpose of the current investigation is to determine the efficacy of a novel approach to voice therapy, conversation training therapy (CTT; Gartner-Schmidt et al., 2016). CTT does not rely on a therapeutic hierarchy and uses patient conversation as the sole stimulus in all sessions. The theoretical rationale for a focus on conversational speech throughout voice therapy is in line with several theories of motor learning that can be applied to voice therapy. The following section outlines motor learning principles potentially disrupted by traditional therapeutic hierarchies and explains how such an effect may contribute to patient attrition and poor long-term treatment outcomes.

Motor Learning in Voice Therapy

Motor learning is the process of changing a behavior based on practice or experience (Schmidt & Lee, 2005). Hierarchical approaches to voice therapy are discordant with principles of motor learning theory, which may prevent patients from learning and retaining the voice techniques being trained and may lengthen time spent in treatment, as seen in other motor learning tasks (Schmidt & Lee, 2005). First, traditional hierarchical treatment does not agree with the part-versus-whole practice of motor learning. In voice therapy, parts of a complex task (i.e., single sounds) are practiced in isolation, instead of the whole task (i.e., connected speech). While this “single sounds” approach may improve immediate performance, it is detrimental to learning and, by extension, long-term retention of the skill (Titze & Verdolini Abbott, 2012). Per Titze and Verdolini Abbott (2012), practicing a segment of a task when the end goal is production of the whole task together has “questionable value,” such as voiced communication, wherein the parts (i.e., breathing, resonance, articulation) cannot occur in isolation. By definition, voiced communication is a parallel processing system. In summary, a practice approach focused on conversation involves simultaneous activation of all parts of the skill and thus may enhance learning and retention. Second, as cognitive effort is necessary for motor learning to occur, traditional approaches of practicing vocal tasks in isolation may hinder long-term skill retention

(Verdolini & Lee, 2004). Having patients become aware of the sound and feel of their voiced communication while also having to think of what to say in conversation is inherently difficult and thus should improve learning.

The third principle, which impacts motor learning, is that of contextual relevance, which states that, for a skill to be learned and transferred to novel situations (i.e., spontaneous conversation), the tasks trained in the learning of that skill must closely resemble real-life tasks (Schmidt & Lee, 2005). In standard-of-care voice therapy, conversation is addressed last, if at all (Bassiouny, 1998; Carding et al., 1999; Casper & Murry, 2000; Chen et al., 2007; Holmberg et al., 2001; Niebudek-Bogusz et al., 2008; Rodriguez-Parra et al., 2011; Schneider & Bigenzahn, 2005; Sellars et al., 2002; van Leer & Connor, 2012), which is in conflict with this motor learning principle. Likewise, because voice production in conversation is the product of multiple levels of motor activities, consciously focusing on each level alone during training impedes learning the levels as a whole. Verbal instructions and teaching “parts of the motor movement” are of limited value, whereas experiential learning is the best method of teaching. Therefore, the closer the voice training can mirror conversational speech, the better the skill will be learned, retained, and transferred across speaking conditions.

The fourth motor learning principle important to changing vocal behaviors is associated with the schedule of practice. According to schema theory, learning occurs best when practice is varied across conditions and foils (Titze & Verdolini Abbott, 2012). In the case of voice practice, this variability could translate to practice in different environments, noise levels, and conversation partners, as well as variability in the stimuli practiced. Because patient-generated conversation is, by definition, variable, practicing voice techniques in conversation should enhance vocal motor skill learning, retention, and transfer (Ohlsson, 2016).

Finally, voice therapy hierarchies stand in opposition to the neuroplasticity principles of salience and specificity via training tasks that are irrelevant to the patient and to the end goal of communication. For example, lip trills and maximum phonation time are not explicitly related to the act of communicating and, therefore, are potentially detrimental to skill learning, which may deter adherence to voice therapy practice and attrition (Kleim & Jones, 2008). Because standard-of-care voice therapy promotes hierarchical, noncontextual treatment and encourages blocked, not variable, practice, it stands to question if the rampant issues with attrition and disorder relapse may be related to the construction of the therapy.

Problems With Attrition and Relapse in Voice Therapy

The use of a hierarchical treatment approach necessitates a protracted time in treatment. Most studies utilizing such an approach require more than eight to 12 voice therapy sessions, with one report of 24 sessions, each lasting 45–90 min in length necessary to achieve goals (Casper &

Murry, 2000; Chen et al., 2007; Niebudek-Bogusz et al., 2008; Rodriguez-Parra et al., 2011; Schindler et al., 2012; Sellars et al., 2002). This length of time in treatment puts a tremendous burden on patient resources, including time off from work, travel to the treatment center, cost of treatment, and, perhaps most importantly, length of time without a functional voice for communicative activities of daily living. It may not be surprising, therefore, that attrition rates for behavioral voice therapy are estimated to be more than 65% (Hapner, Portone-Maira, & Johns, 2009; Roy, Bless, Heisey, & Ford, 1997), and at least one study found voice problem relapse rates at nearly 70% (Roy et al., 1997). Not only does patient attrition potentially hurt the patient, it is also costly to health care institutions due to uncollected billing revenue and uncompensated SLP time (Litts, Gartner-Schmidt, Clary, & Gillespie, 2015). One reason for attrition and relapse could be that patients' voice needs are not being met in therapy. Ziegler et al. (2014) found that transfer of voice therapy techniques to "real life" conversation was the most useful part of voice therapy, as well as the most difficult aspect of treatment. Similarly, another study of patients' perceptions of voice therapy found that patients thought nonspeech voice therapy exercises were difficult to perform outside therapy (van Leer & Connor, 2010). SLPs also think that generalization is a difficult aspect of voice therapy (Ohlsson, 2016). These data combined indicate that the structure of standard-of-care, hierarchical voice therapy is not serving the communicative needs of patients and may contribute to costly relapse and dropout rates.

In summary, healthy, balanced phonation in conversation is the goal of voice therapy. Both patients and SLPs report transfer of voice techniques to conversation as the most challenging part of therapy. Incorporating the motor learning theories of a focus on whole practice, variable practice, and contextual relevance during treatment may improve skill learning and long-term retention compared to standard-of-care approaches to voice therapy for patients with benign vocal fold lesions and MTD. The limitations of traditional strategies may contribute to unnecessarily protracted times in treatment, high attrition rates, and lack of long-term retention, leading to substantial voice problem relapse rates (Hapner et al., 2009; Rodriguez-Parra et al., 2011; Titze & Verdolini Abbott, 2012; Ziegler et al., 2014).

Development of CTT

Drs. Gartner-Schmidt and Gillespie created the original CTT approach with therapeutic skills that included a focus on sensory feedback (Titze & Verdolini Abbott, 2012), clinician language avoiding specific biomechanical instructions, patient goal setting (Titze & Verdolini Abbott, 2012), practice across multiple real-life contexts (Titze & Verdolini Abbott, 2012), and "whole" (e.g., conversation) versus "part" (e.g., single phonemes) training (Fontana, Mazzardo, Furtado, & Gallagher, 2009; Titze & Verdolini Abbott, 2012). CTT is based on experiential learning processes (i.e., exploratory learning) and increasing awareness of the

sensations of voice in conversational speech as opposed to a prescriptive exercise-based theoretical construct (e.g., do *X* number of reps of a sound twice a day; Verdolini & Lee, 2004; Verdolini-Marston & Balota, 1994). Subsequently, CTT was formulated on the use of the "clear speech technique" (Picheny, Durlach, & Braida, 1985, 1986), an intelligibility strategy for speaking to listeners with hearing loss and/or in adverse listening conditions, such as when ambient noise is high; the "clear speech technique" has had many positive applications to patients with Parkinson's disease (Ferguson, 2004; Picheny et al., 1985). When speakers use clear speech, sound and phrase duration, pauses, intensity, and intonation increase, vowels are less reduced, speaking rate decreases, and a wider range of fundamental frequencies are produced (Bradlow, Kraus, & Hayes, 2003; Picheny et al., 1986; Smiljanić & Bradlow, 2008). Data show that immediate aerodynamic, auditory, and patient-perceptual voice improvements occur in people with voice disorders when asked to produce "clear speech" (Gillespie & Gartner-Schmidt, 2016). In addition, "clear speech" requires the speaker to increase articulatory precision with the production of crisp, clear consonants and to focus on the kinesthetic feedback of the oral sensations from the consonants. This focus on the external effect (e.g., sensations) of a gesture (e.g., reduced muscle tension, reduced vocal fold impact stress during phonation) enhances learning (Schmidt & Lee, 2005; Verdolini Abbott, 2008a; Wulf & Weigelt, 1997). These effects are similar to those achieved with the well-known voice techniques of resonant voice and flow phonation, which result in an ideal laryngeal configuration for voice production (Berry et al., 2001; Gartner-Schmidt, 2013; Peterson, Verdolini-Marston, Barkmeier, & Hoffman, 1994; Schmidt & Lee, 2005; Verdolini, 2000; Verdolini, Druker, Palmer, & Samawi, 1998; Verdolini Abbott, 2008a, 2008b, 2011). A major difference is that "clear speech" trains these effects immediately in connected speech; therefore, "clear speech" effectively combines techniques that are well grounded in the voice literature into one skill that can only be achieved in connected speech—the ultimate goal of voice therapy. All CTT skills are, therefore, trained in the context of patient-led conversation (Gartner-Schmidt et al., 2016).

The current investigation aimed to determine the efficacy of CTT by comparing acoustic, aerodynamic, auditory-perceptual, and patient-reported outcome measures before and after CTT treatment. Next, patient-reported outcome measures before and after CTT treatment were compared with patient-reported outcomes from age-, gender-, and diagnosis-matched historical controls (voice therapy patients) treated with standard-of-care, hierarchical, non-conversation-based voice therapies.

Method

Study Design

This was a single-arm, single-center, prospective efficacy study with a 1:1 matched historical control (ClinicalTrials.gov No. CT02441348) conducted at the University of

Pittsburgh Voice Center. The University of Pittsburgh Institutional Review Board approved the protocol.

CTT Participants

Recruited participants were male and female patients. The inclusion criteria were as follows: diagnosed with only benign vocal fold lesions or primary MTD by a multidisciplinary team of a laryngologist and an SLP, deemed amenable to voice therapy (i.e., not surgical candidates, without other laryngologic diagnoses) per voice therapy stimulability testing (Gillespie & Gartner-Schmidt, 2016; Rosen et al., 2012), nonsmoking (to avoid potential deleterious cumulative effects of smoking on laryngeal health and voice quality), age of 16–60 years (to avoid enrolling patients whose voices may change during the course of treatment and follow-up due to hormonal changes associated with puberty or aging voice), normal hearing (determined by pure-tone audiometry), baseline Voice Handicap Index–10 (VHI-10) of > 11 (Arffa, Krishna, Gartner-Schmidt, & Rosen, 2012), and a willingness to attend all therapeutic intervention and follow-up sessions. The exclusion criteria were as follows: no prior history of voice surgery, no history of voice therapy in the last year, no history of pulmonary disease (per patient report), and no history of other serious chronic medical condition that may affect voice (per patient report). Participants attended four 45-min sessions of voice therapy. Four sessions were chosen based on our pilot data, which demonstrated results in an average of three sessions, and published data that the mode number of voice therapy sessions for patients who completed voice therapy was four sessions.

Screening

All potential patient participants completed the VHI-10, a validated 10-question assessment of a patient's perception of the handicapping effects of voice on their life (Rosen, Lee, Osborne, Zullo, & Murry, 2004). Next, patients underwent instrumental acoustic and aerodynamic testing in a sound-treated voice laboratory. All patient instructions were scripted and standardized for repeatability across individuals. For acoustic analyses, patients were seated upright with a head-mounted microphone (SHURE Beta 54) placed at a 45° angle from his or her mouth and were instructed to read the six sentences from the Consensus Auditory-Perceptual Evaluation of Voice (CAPE-V) protocol (Kempster, Gerratt, Verdolini Abbott, Barkmeier-Kraemer, & Hillman, 2009), which were recorded with the Analysis of Dysphonia in Speech and Voice program in the Computerized Speech Lab (PENTAX) and saved for future analyses. Aerodynamic assessment was then completed using the Phonatory Aerodynamic System 6600 (PENTAX). This system's hardware consists of a pneumotach coupled to a face mask and an external microphone. The speaker phonates into the face mask, and expired air flows to the pneumotach, which consists of a stainless steel mesh screen with pressure transducers on either side. The system calculates the pressure difference on either

side of the screen to determine airflow rate. The microphone is positioned at the end of the pneumotach and internally calibrated per system specifications to represent a microphone situated 15 cm from the speaker's mouth. The patient sat with the face mask held snugly over his or her nose and mouth. Once in place, the patient read the first four sentences of the Rainbow Passage to capture phonatory aerodynamic functioning (Fairbanks, 1960; Gartner-Schmidt et al., 2015; Roy, 2003).

Stimulability Testing

Patients were deemed amenable to voice therapy using a standardized stimulability test developed by the research team and conducted as routine part of the medical speech evaluation (Gillespie & Gartner-Schmidt, 2016). Patients were assessed for stimulability to resonance, flow, and clear speech prompts. Patients were referred for study inclusion if they demonstrated one quantitative change in acoustics or aerodynamics or one qualitative change (i.e., patient noted a change to the sound or feel of voice) following each of the stimulability prompts (resonance, clear speech, or flow phonation). Furthermore, the evaluating SLP also noted if a qualitative change was observed in the patient's voice following the trials. The use of stimulability testing for voice therapy ensured that patients enrolled in the investigation had the baseline ability to, at the very least, hear or feel voice change with common therapeutic prompts, which data have shown indicates a greater likelihood of success in treatment (Bonilha & Dawson, 2012; Dejonckere & Lebacqz, 2001).

Laryngeal Examination

A fellowship-trained laryngologist performed laryngeal exams using flexible or rigid laryngoscopy with stroboscopy, per standard published protocols (Roehm & Rosen, 2004; Rosen & Murry, 2000; Young & Rosen, 2011). All exams were recorded and saved. If a lesion was present, the laryngologist and the SLP made a determination about lesion severity requiring surgery based on a published existing nomenclature paradigm, as well as whether they passed the standardized voice therapy stimulability testing (Akbulut et al., 2016; Gillespie & Gartner-Schmidt, 2016; Rosen et al., 2012). Diagnoses of primary MTD were made based on a patient's self-perceived voice problem in the absence of laryngoscopic structural or physiological deficit following laryngoscopy (Roy, 2003; Verdolini et al., 2005). Once inclusion/exclusion criteria were met, patients were consented for and enrolled in the study.

SLP Assignment

Participants were randomly allocated to an SLP trained in CTT at the same voice center. Each participant was randomized to a primary and secondary SLP to accommodate scheduling preferences of the participant. If the primary SLP could not accommodate four consecutive sessions with the participant, then all sessions were scheduled with the secondary SLP. All therapies for a given participant were completed by the same SLP.

Outcomes

The primary outcome was the VHI-10 score (Rosen et al., 2004). Secondary outcomes included acoustic, aerodynamic, and auditory-perceptual outcomes and functional gains for occupational voice demands. Outcomes were collected at the start of each therapy session, 1 week posttherapy follow-up, and 3 months posttherapy follow-up by a trained research coordinator. The time frame for data collection was specifically selected immediately prior to a therapy session to capture vocal learning that may have occurred as a result of the previous session and to avoid capturing performance improvements gained in a session. Acoustic analyses included the Cepstral Spectral Index of Dysphonia (CSID), a multifactorial estimate of dysphonia severity that correlates with the visual analog scale for overall voice severity used in the CAPE-V (Awan, Roy, Jetté, Meltzner, & Hillman, 2010), cepstral peak prominence (CPP) and its standard deviation, CPP fundamental frequency (CPP F_0) and its standard deviation, and average vocal intensity in dB SPL. Aerodynamic outcomes included average airflow in speech, and average number of breaths, the average duration of a standard reading passage (the Rainbow Passage; Fairbanks, 1960). Auditory-perceptual outcomes included overall voice severity as determined by the CAPE-V score provided by three blinded raters, each SLP specialized in voice disorders (Kempster et al., 2009). Each rater listened via headsets, with computer volume at a set level that was not adjusted for the duration of the ratings. Raters listened to all six CAPE-V sentences, which were randomized by speaker and time point, with a different randomization scheme for each rater. Temporal and occupational outcomes included time to treatment success as measured by weekly change in VHI-10 and participant weekly response to a yes/no question (*Can you do what you need to do with your voice?*), missed work hours due to voice problems as determined by weekly work logs, and skill retention at long-term follow-up as determined by maintenance of VHI-10, acoustic, aerodynamic, and auditory-perceptual outcomes. Finally, at each session and follow-up time point, participants were asked to rate their percent normal vocal function (0% = *not at all normal*, 100% = *completely normal*) and vocal effort (0 = *no effort*, 100 = *maximum effort*) on two visual analog scales. Participants were compensated for participation.

Historical Matched Control Patients

Historical matched controls (HMCs) were selected from a database of patients treated with standard-of-care voice therapy at the same tertiary care, urban, academic voice center. Those who underwent voice therapy between 2010 and 2012 and met the same inclusion/exclusion criteria as the CTT cohort were included in the pool of potential HMCs. Using a 1:1 ratio, patients were matched on baseline variables, including age, gender, and diagnosis.

Standard-of-Care Voice Therapy

HMC patients received standard-of-care voice therapy. This intervention was defined as using the general class of resonant voice therapies, Lessac-Madsen resonant voice program (Verdolini Abbott, 2008b), flow phonation (Gartner-Schmidt, 2014), and various semi-occluded vocal tract postures to elicit improved glottal closure and heighten anterior vibration sensations with optimal muscular balance.

HMC Outcomes

VHI-10 scores for HMC patients were collected from the medical record at the clinical visit immediately prior to therapy initiation and the first posttreatment clinical visit. The timing of the posttreatment visit ranged from 1 week to 1 month.

CTT Voice Therapy

Each CTT participant completed one 45-min voice therapy session per week for four consecutive weeks. Sessions occurred at intervals of no less than 5 days and no more than 10 days apart. Each therapy session was video-recorded for fidelity checks by the Co-Investigator (J. G. S.). These checks ensured SLP accuracy in delivering each CTT component, homework instructions, and maintaining conversation as the sole therapeutic stimulus. There are six components to CTT: clear speech, auditory/kinesthetic awareness, rapport building, negative practice, embedded basic training gestures, and prosody and pauses (see Table 1). Detailed descriptions of the components of CTT are outlined in the article describing the development of CTT (Gartner-Schmidt et al., 2016).

Although the first component is “clear speech,” after that, the order of the component parts of CTT is based on the participant’s individual needs, and the SLP was free to choose whichever component they deemed most useful for the participant. All six components of CTT were covered before discharge. Homework for the CTT program increased in complexity based on the components taught in each weekly session. After the first session, participants were encouraged to increase awareness of voice production and specifically the sound and feel of voice when speaking during the following week. In addition, participants were required to practice their CTT voice seven times a day for 1 min, followed by 30 s of negative practice, and then finished off by a return to 1 min of their CTT voice. Practice was to be variable—at different times and in varying conditions; conversational speech is intrinsically variable by nature.

Data Management

Research Electronic Data Capture was used to manage all participant data after initial screening through study completion.

Table 1. Name and brief description of each conversation training therapy component.

Clear speech	The participant is asked to use crisp, clear consonants and precise articulation while beginning a conversation on a topic of their choosing (Picheny et al., 1986). Care is taken to not overarticulate or become robotic or monotone.
Rapport building	Beginning immediately with a conversation of the participant's choice builds the therapist–patient relationship from the outset. This relationship is critical to success in treatment.
Auditory/kinesthetic awareness	Participants were asked to attend to changes in both the sound and the feel of voice while producing clear speech. Participants noted how all consonants feel differently as they are spoken and were encouraged to explore and increase attention to sensory information. The SLP routinely questioned the participant, which, thereby, required the participant to discover and increase awareness of specific sensations. This experiential learning is also important in motor learning (Verdolini & Lee, 2004).
Negative practice and labeling	The participant speaks with the voice production that motivated treatment seeking (e.g., hoarseness, fatigue; Iwarsson, 2015; Ohlsson, 2016). The participant labels or names and describes the two different voices in his or her own words, which promotes patient participation in the therapeutic process and requires less patient reliance on SLP feedback (Bjork, 1994). Labeling the two different voices (a meta-therapy skill) also necessitates synthesizing the participant's sensations and perceptions into “one package” (label) instead of having the patient remember a “laundry list” of sensations (Helou, 2017). Provides mastery experience and improves learning and self-efficacy for voice change (Bandura, 1977).
Embedded basic training gesture	As originally coined by Verdolini Abbott (2008a), the participant intermittently sustains consonants in conversation and focuses on the tactile voice goal of feeling sounds of speech and the auditory goal of hearing vocal clarity in consonants in conversation (Verdolini Abbott, 2008a).
Prosody, pauses, and projection	Patterns of stress and intonation are trained by asking participants to vary the intonation/inflection and stress/emphasis of words in conversational speech. Loud voice was achieved via instruction to increase the sensation of consonantal energy in the head/face, slow the rate, and increase the mouth opening.

Note. SLP = speech-language pathologist.

Power and Sample Size

Sample size calculations were based on a reduction in VHI-10 of 5, which is clinically meaningful (Gartner-Schmidt & Rosen, 2011). This assumption translated to an effect size of 0.52, assuming an *SD* of 9.6 (Rosen et al., 2004). To calculate the sample size needed when using historical controls, we used a formula proposed by Zhang, Cao, and Ahn (2010) for continuous outcomes. Assuming a conservative dropout rate of 15%, a sample size of 59 participants per group would achieve at least 80% power to detect a 5-point difference in VHI-10 between CTT and HMC participants at $\alpha = .05$. This sample size also achieves 80% power to detect at least a 4.8-point change from baseline controlling for an experiment-wise error rate of 0.05 for five time points.

Statistical Analyses

Distributions of baseline characteristics overall, by diagnosis and by gender, were evaluated. Means and standard deviations or medians and quartiles were used to summarize continuous variables, whereas frequencies and percentages were used for categorical variables.

Therapy Session 1 remained a pretreatment time point since patients had not yet received any therapy when outcomes were collected at the beginning of the session. In unadjusted analyses, VHI-10 change from baseline was calculated, and paired *t* tests were used to assess whether VHI-10 improved significantly at each time point (Session 2, Session 3, Session 4, 1 week posttherapy completion, and 3 months posttherapy completion) from baseline. Testing

VHI-10 improvement at 1 week (short term) and 3 months (long term) from the end of treatment allowed investigation into long-term effects of CTT. To exploit the longitudinal data collection and account for missing data, a linear mixed model for raw VHI-10 scores with a patient-specific random intercept and categorical time points as predictor was fitted. To correct for multiple testing on five follow-up time points, Sidak-adjusted *p* values were also calculated. For secondary outcomes, a similar analyses approach was used, except for temporal outcome (binary) in which McNemar's test for paired proportions was performed. To assess the efficacy of CTT relative to standard-of-care therapy, VHI-10 changes from baseline to 1 week after the last therapy session (short term) were compared between treatment groups (CTT vs. HMC). Unadjusted comparisons used a paired *t* test on the VHI-10 change scores. Adjusted analyses used a linear mixed model for VHI-10 raw scores with group, time point (baseline vs. follow-up), and Group $\times$ Time Point interaction as primary covariates. The Group $\times$ Time Point interaction coefficient captured the effect of CTT. Adjustments for the matching variables (age at diagnosis, gender, and diagnosis type) and inclusion of a random patient intercept were used to account for within-subject correlations over time.

For auditory-perceptual analyses, each CTT participant at each time point was evaluated by three independent SLP raters to generate the overall severity scores. The sequence at which each rater performed the ratings across all CTT participants was random. At each time point, correlations between ratings of the three raters were examined using scatter plots and the Pearson correlation coefficient.

Rating scores were summarized at each time point stratified by the rater using means and standard deviations or medians and quartiles. Changes from baseline to each follow-up time point were also examined.

To test for significant changes in ratings over time, linear mixed models with time as the primary covariate and random participant effects were used to account for the within-subject correlations over time. Raters were added to the model to adjust for potential rater effects. An examination of whether changes in ratings over time varied across raters by testing for the interaction between time point and rater was also conducted.

Statistical analyses were performed with Stata 15.0 (StataCorp) and R Open 3.5.1 (Microsoft). All tests were two-sided, and a p value of $\leq .05$ was considered significant.

Results

CTT Participant Recruitment and Demographics

Participant flow throughout the trial is depicted in Figure 1. Sixty-eight total potential participants were recruited from the clinic for study participation. Of these, 60 met eligibility criteria, and 49 returned for the first voice therapy session. One patient dropped out after the first therapy session. The remaining 48 participants completed all four therapy sessions. Forty-seven completed the 1-week follow-up, and 43 participants completed the 3-month follow-up. Participants had an average age of 39 years and were mostly women (76.7%), diagnosed with MTD (61.7%), White (79.7%), and nearly all non-Hispanic (96.7%; see Table 2).

Primary Outcome

The primary outcome, change in VHI-10 from baseline, was assessed before the start of Therapy Session 2, Therapy Session 3, Therapy Session 4, 1 week posttreatment completion, and 3 months posttreatment completion follow-up time points for the CTT patient group. Univariate analyses at each time point show a significant decrease in VHI-10 across all time points after the first session (see Table 3). Specifically, a linear mixed model estimated that VHI-10 at baseline had a mean of 20.7 with a 95% confidence interval (CI) of [19.0, 22.3], which significantly decreased over the therapy sessions (mean change [95% CI]: before Therapy Session 2, -4.3 [-5.6 , -2.9]; before Therapy Session 3, -6.4 [-7.8 , -5.0]; before Therapy Session 4, -9.2 [-10.6 , -7.8]) and after therapy completion (after 1 week, -10.2 [-11.6 , -8.8]; after 3 months, -12.7 [-14.2 , -11.3]), all $ps < .001$ with and without Sidak adjustment for multiple testing (see Table 4 and Figure 2). Statistically significant decreases in VHI-10 started after only one CTT session.

HMC Patient Demographics

To compare CTT with standard-of-care therapy, 48 HMC patients were found after age, gender, and diagnosis matching with the CTT participants. Over 50% (5/9) of the

SLPs who provided the standard-of-care voice therapy to the HMC patients were the same SLPs as those who provided CTT treatment for the current study. Table 5 shows that matching was successful with age, gender, and diagnosis distribution similar between the CTT and HMC groups. The baseline VHI-10 scores were also similar (CTT: $M = 21.1$, $SD = 5.9$; HMC: $M = 20.4$, $SD = 7.0$; $p = .58$). Short-term follow-up data for CTT and HMC were respectively captured at 1 week and from 1 week to 1 month following the final therapy session. A linear mixed model was fitted to compare changes in VHI-10 from baseline to short-term follow-up between groups (see Table 6). While VHI-10 decreased in both the CTT and HMC patients (mean change [95% CI]: -10.4 [-13.1 , -7.8] vs. -5.8 [-8.4 , -3.1], both $ps < .001$), the reduction in CTT was greater by nearly 5 points (effect estimate [95% CI]: -4.7 [-8.4 , -0.9], $p = .016$; see Figure 3).

Age, Gender, and Diagnosis Interactions

Potential interaction effects of age, gender, and diagnosis (vocal fold lesions or MTD) on VHI-10 change were assessed. No significant age, gender, or diagnosis interaction effect was found, indicating there was no significant difference in VHI-10 change by age, gender, or presence of vocal fold lesions or MTD ($p > .05$ for all comparisons).

Secondary Outcomes

Change from baseline in acoustic, aerodynamic, and auditory-perceptual analyses of voice and patient perception of percent normal vocal function and vocal effort was also assessed before the start of Therapy Session 2, Therapy Session 3, Therapy Session 4, 1-week follow-up, and 3-month follow-up time points.

Acoustic Analyses

Acoustic measures for males and females were analyzed by gender. For females, there was a significant increase in CPP in a sustained /a/ vowel from baseline ($M = 10.3$, $SD = 1.9$) to before Therapy Session 2 ($M = 11.2$, $SD = 1.9$, Sidak $p = .005$), before Therapy Session 3 ($M = 11.2$, $SD = 1.8$, Sidak $p = .035$), and 1-week follow-up ($M = 11.1$, $SD = 1.4$, Sidak $p = .028$). A significant increase in CPP F_0 , measured in the sentence “we were away a year ago,” was observed from baseline ($M = 181.5$ Hz, $SD = 24.3$) to before Therapy Session 3 ($M = 201.4$ Hz, $SD = 22.2$), before Therapy Session 4 ($M = 203.1$ Hz, $SD = 24.1$), and 1-week follow-up ($M = 208.7$ Hz, $SD = 20.4$), Sidak $p < .001$ for all comparisons. CPP F_0 increased to 190.2 at long-term follow-up, but this change was not significant. The standard deviation of CPP F_0 —a theorized measure of intonational variety and one of the components of CTT—increased significantly from baseline ($M = 28.7$ Hz, $SD = 12.2$) to 1-week follow-up ($M = 37.8$, $SD = 12.5$, Sidak $p = .012$). This measure was also initially significant at Session 4 and 3-month follow-up; however, after Sidak adjustment, the significance went away.

Figure 1. CONSORT (Consolidated Standards of Reporting Trials) diagram of participant flow from inclusion through final data collection. CTT = conversation training therapy; VHI-10 = Voice Handicap Index–10.

For males, an increase in CPP of sustained /a/ vowel was observed from baseline ($M = 12.2$ dB, $SD = 2.0$) to before Therapy Session 2 ($M = 14.2$ dB, $SD = 1.5$, $p = .016$), Therapy Session 4 ($M = 14.1$ dB, $SD = 2.0$, $p = .026$), and 1-week follow-up ($M = 14.4$ dB, $SD = 1.4$, $p = .013$) but did not achieve significance after Sidak correction. No significant differences were observed for CPP F_0 standard deviation.

For both genders, a significant improvement in CSID in the sustained vowel /a/ was observed from baseline ($M = 21.7$, $SD = 19.3$) to before Voice Therapy 2 ($M = 11.2$, $SD = 17.2$, Sidak $p = .003$), before Voice Therapy 3 ($M = 11.4$, $SD = 16.0$, Sidak $p = .001$), before Voice Therapy 4 ($M = 10.6$, $SD = 17.2$, Sidak $p = .003$), and 1-week

follow-up ($M = 10.2$, $SD = 18.0$, Sidak $p = .003$). The significance originally observed at 3-month follow-up went away after Sidak adjustment. A significant improvement in CSID in the sentence “we were away a year ago” was observed from baseline ($M = 3.7$, $SD = 16.1$) to before Voice Therapy 4 ($M = -3.2$, $SD = 11.8$, Sidak $p = .02$) and 1-week follow-up ($M = -3.2$, $SD = 9.3$, Sidak $p < .004$). CSID values at 3-month follow-up decreased ($M = 2.6$), but this decrease was not statistically significant. Average dB SPL significantly decreased from baseline ($M = 76.6$ dB, $SD = 3.8$) to before Therapy Session 2 ($M = 72.0$, $SD = 2.6$), before Therapy Session 3 ($M = 72.4$, $SD = 3.1$), before Therapy Session 4 ($M = 71.3$, $SD = 9.8$), 1-week follow-up

Table 2. Demographics of the conversation training therapy group.

Demographic	Total (<i>N</i> = 60)	Gender		<i>p</i>	Diagnosis		<i>p</i>
		Female (<i>n</i> = 46)	Male (<i>n</i> = 14)		Lesion (<i>n</i> = 23)	MTD (<i>n</i> = 37)	
Female, <i>n</i> (%)	46 (76.7)	46 (100.0)	0 (0.0)		22 (95.7)	24 (64.9)	.02
Lesion diagnosis, <i>n</i> (%)	23 (38.3)	22 (47.8)	1 (7.1)	.02	23 (100.0)	0 (0.0)	
Age (years), <i>M</i> (<i>SD</i>)	38.55 (11.88)	39.35 (12.48)	35.93 (9.61)	.29	37.0 (11.9)	39.5 (11.9)	.42
Ethnicity, <i>n</i> (%)				> .99			.14
Hispanic	1 (1.7)	1 (2.2)	0 (0.0)		1 (4.3)	0 (0.0)	
Not Hispanic	58 (96.7)	44 (95.7)	14 (100.0)		21 (91.3)	37 (100.0)	
Unknown	1 (1.7)	1 (2.2)	0 (0.0)		1 (4.3)	0 (0.0)	
Race, <i>n</i> (%)				.697			.04
Asian	1/59 (1.7)	1/45 (2.2)	0 (0.0)		1/22 (4.5)	0 (0.0)	
Black	7/59 (11.9)	6/45 (13.3)	1 (7.1)		5/22 (22.7)	2 (5.4)	
White	47/59 (79.7)	34/45 (75.6)	13 (92.9)		16/22 (72.7)	31 (83.8)	
More than one race	4 (6.8)	4 (8.9)	0 (0.0)		0 (0.0)	4 (10.8)	
Laryngoscopy type, <i>n</i> (%)				.75			.02
Flexible	38/59 (64.4)	29 (63.0)	9/13 (69.2)		10 (43.5)	28/36 (77.8)	
Rigid	21/59 (35.6)	17 (37.0)	4/13 (30.8)		13 (56.5)	8/36 (22.2)	

Note. MTD = muscle tension dysphonia.

($M = 73.1$, $SD = 2.4$), and 3-month follow-up ($M = 73.6$, $SD = 2.7$), Sidak $p < .01$ for all comparisons.

Aerodynamic Analyses

Due to known differences in average airflow in speech between males and females, this measure was analyzed by gender. For females, there was a significant increase in average airflow in speech from baseline ($M = 173.2$ ml/s, $SD = 60.9$) to before Therapy Session 2 ($M = 210.3$ ml/s, $SD = 55.1$), before Therapy Session 3 ($M = 228.5$ ml/s, $SD = 64.2$), before Therapy Session 4 ($M = 228.3$ ml/s, $SD = 74.9$), 1-week follow-up ($M = 238.5$ ml/s, $SD = 85.1$), and 3-month follow-up ($M = 232.9$ ml/s, $SD = 71.4$), Sidak $p < .05$ for all comparisons. For males, a significant improvement (increase) in average airflow in speech was observed from baseline ($M = 192.5$, $SD = 65.8$) to before Therapy Session 2 ($M = 231.7$, $SD = 55.6$), before Therapy Session 3 ($M = 235.8$, $SD = 54.2$), before Therapy Session 4 ($M = 233.3$, $SD = 53.7$), 1-week follow-up ($M = 240.8$, $SD = 61.1$), and 3-month follow-up ($M = 245.5$, $SD = 67.3$), Sidak $p < .05$ for all comparisons.

Across genders, the number of breaths taken during the Rainbow Passage reading—representative of the pauses for the breath replenishment tenet of CTT—demonstrated significant improvement from baseline ($M = 4.8$, $SD = 1.8$) to before Therapy Session 2 ($M = 6.1$, $SD = 2.0$), before Therapy Session 3 ($M = 6.9$, $SD = 2.0$), before Therapy Session 4 ($M = 6.3$, $SD = 1.8$), 1-week follow-up ($M = 6.8$, $SD = 2.3$), and 3-month follow-up ($M = 6.3$, $SD = 2.2$), Sidak $p < .001$ for all comparisons. Likewise, duration of reading passage time also significantly increased from baseline ($M = 23.9$ s, $SD = 3.1$ s) to before Therapy Session 2 ($M = 26.7$, $SD = 3.7$), before Therapy Session 3 ($M = 27.7$, $SD = 4.4$), before Therapy Session 4 ($M = 26.9$, $SD = 3.8$), 1-week follow-up ($M = 28.2$, $SD = 5.4$), and 3-month follow-up ($M = 26.8$, $SD = 4.4$), Sidak $p < .001$ for all comparisons.

Auditory-Perceptual Analyses

In a linear mixed model, the Rater $\times$ Time Point interaction effect was not significant ($p = .62$), indicating that changes over time in auditory-perceptual ratings did not vary significantly between raters. A model with a main effect of time adjusted for rater demonstrated a significant reduction in auditory-perceptual mean ratings of overall voice severity before Session 3 (estimate [95% CI]: -3.6 [-6.4 , -0.8], $p = .012$), Session 4 (-5.3 [-8.1 , -2.5], $p < .001$), after 1 week (-5.9 [-8.7 , -3.0], $p < .001$), and after 3 months (-3.0 [-5.9 , -0.1], $p = .041$).

Percent Normal Function

A significant increase in participant-rated percent normal vocal function was observed from baseline ($M = 63.0\%$, $SD = 16.7\%$) to before Therapy Session 3 ($M = 72.4\%$, $SD = 18.6\%$, Sidak $p = 0.013$), before Therapy Session 4 ($M = 76.8\%$, $SD = 21.6\%$, Sidak $p < .001$), 1-week follow-up ($M = 81.2\%$, $SD = 16.7\%$, Sidak $p < .001$), and 3-month follow-up ($M = 81.1\%$, $SD = 21.1\%$, Sidak $p < .001$). Relatedly, participants were asked to answer the yes/no statement “I can do what I need to do with my voice.” At baseline, only 31.2% of patients responded in the affirmative, which improved significantly before Session 3 (56.2%, Sidak $p = .03$), before Session 4 (66.7%), after 1 week (78.7%), and after 3 months (83.3%), Sidak $p < .001$ for the last three time points.

Vocal Effort

A significant decrease in participant-rated vocal effort (rated on a scale of 0–100, with 100 = *most effort* and 0 = *no effort*) was observed from baseline ($M = 56.6$, $SD = 28.6$) to before Therapy Session 4 ($M = 35.4$, $SD = 24.7$), 1-week follow-up ($M = 28.3$, $SD = 21.5$), and 3-month follow-up ($M = 26.5$, $SD = 23.2$), Sidak $p < .001$ for all comparisons.

Table 3. Means, standard deviations (SD), confidence intervals (CI), and *p* values for all measured outcomes at all time points, stratified by gender.

Outcome	Follow-up time point	<i>N</i>	Baseline <i>M</i> (SD) ^a	Follow-up <i>M</i> (SD)	Change <i>M</i> [95% CI]	<i>p</i> ^b	<i>p</i> ^c
Males							
VHI	Session 2	48	21.1 (5.9)	16.8 (6.4)	-4.4 [-5.8, -3.0]	< .001	< .001
	Session 3	47	21.3 (5.9)	14.7 (6.7)	-6.6 [-8.1, -5.0]	< .001	< .001
	Session 4	48	21.1 (5.9)	11.8 (6.9)	-9.3 [-11.1, -7.5]	< .001	< .001
	1 week	47	21.0 (5.9)	10.7 (7.0)	-10.3 [-12.1, -8.5]	< .001	< .001
	3 months	42	21.4 (6.0)	8.4 (7.8)	-13.0 [-15.0, -11.0]	< .001	< .001
Sustained Vowel–CPP	Session 2	46	10.7 (2.1)	12.0 (2.2)	1.2 [0.7, 1.8]	< .001	< .001
	Session 3	48	10.8 (2.2)	11.8 (2.0)	1.0 [0.4, 1.6]	.002	.011
	Session 4	48	10.8 (2.2)	11.7 (2.2)	0.9 [0.3, 1.5]	.004	.022
	1 week	46	10.8 (2.2)	12.0 (2.0)	1.2 [0.6, 1.8]	< .001	< .001
	3 months	42	10.9 (2.2)	11.5 (2.3)	0.6 [0.0, 1.1]	.045	.206
Sustained Vowel–CPP SD	Session 2	46	1.2 (0.8)	0.9 (0.6)	-0.4 [-0.6, -0.1]	.005	.023
	Session 3	48	1.2 (0.9)	1.1 (1.4)	-0.1 [-0.6, 0.3]	.559	.983
	Session 4	48	1.2 (0.9)	0.8 (0.5)	-0.4 [-0.7, -0.2]	.002	.011
	1 week	46	1.3 (0.9)	0.8 (0.6)	-0.4 [-0.7, -0.1]	.004	.019
	3 months	42	1.3 (0.9)	0.7 (0.4)	-0.5 [-0.8, -0.2]	< .001	.002
Sustained Vowel–CSID	Session 2	45	21.7 (19.3)	11.1 (17.2)	-10.6 [-16.4, -4.7]	< .001	.003
	Session 3	48	20.6 (19.8)	11.4 (16.0)	-9.2 [-14.7, -3.7]	.001	.007
	Session 4	48	20.6 (19.8)	10.6 (17.2)	-10.0 [-15.4, -4.6]	< .001	.003
	1 week	46	21.6 (19.5)	10.2 (18.0)	-11.4 [-17.6, -5.1]	< .001	.003
	3 months	42	19.8 (19.4)	12.6 (17.0)	-7.2 [-13.1, -1.3]	.018	.085
Sentence–CPP <i>F</i> ₀	Session 2	46	168.6 (32.5)	177.3 (45.6)	8.6 [-1.7, 19.0]	.098	.404
	Session 3	48	167.7 (32.9)	183.4 (37.5)	15.8 [9.8, 21.7]	< .001	< .001
	Session 4	48	167.7 (32.9)	182.1 (46.5)	14.5 [6.3, 22.6]	< .001	.004
	1 week	46	169.1 (32.8)	189.3 (38.6)	20.2 [13.2, 27.3]	< .001	< .001
	3 months	42	167.5 (34.2)	173.5 (59.1)	6.0 [-8.5, 20.5]	.407	.927
Sentence–CPP <i>F</i> ₀ SD	Session 2	46	32.3 (15.2)	31.9 (15.4)	-0.4 [-6.5, 5.7]	.886	> .99
	Session 3	48	31.6 (15.4)	33.9 (11.2)	2.3 [-2.5, 7.0]	.343	.877
	Session 4	48	31.6 (15.4)	35.8 (13.5)	4.2 [-1.0, 9.4]	.109	.439
	1 week	46	31.4 (15.6)	37.2 (13.4)	5.8 [0.2, 11.4]	.041	.191
	3 months	42	31.3 (15.9)	37.6 (14.8)	6.4 [0.9, 11.9]	.024	.114
Sentence–CSID	Session 2	46	4.1 (16.4)	-0.8 (16.5)	-4.8 [-10.8, 1.1]	.108	.435
	Session 3	48	3.7 (16.1)	0.0 (15.2)	-3.7 [-9.9, 2.6]	.243	.751
	Session 4	48	3.7 (16.1)	-3.2 (11.8)	-6.9 [-11.5, -2.3]	.004	.02
	1 week	46	4.1 (16.3)	-3.2 (9.3)	-7.3 [-11.4, -3.2]	< .001	.004
	3 months	42	4.3 (16.5)	2.6 (15.3)	-1.6 [-6.5, 3.2]	.497	.968
No. of breaths	Session 2	48	4.8 (1.8)	6.1 (2.0)	1.3 [0.8, 1.8]	< .001	< .001
	Session 3	47	4.8 (1.8)	6.9 (2.0)	2.0 [1.6, 2.5]	< .001	< .001
	Session 4	48	4.8 (1.8)	6.3 (1.8)	1.5 [1.1, 2.0]	< .001	< .001
	1 week	46	4.8 (1.8)	6.8 (2.3)	2.0 [1.4, 2.5]	< .001	< .001
	3 months	42	4.6 (1.7)	6.3 (2.2)	1.6 [1.0, 2.3]	< .001	< .001
Duration	Session 2	48	23.9 (3.1)	26.7 (3.7)	2.7 [1.7, 3.8]	< .001	< .001
	Session 3	47	24.0 (3.1)	27.7 (4.4)	3.6 [2.3, 4.9]	< .001	< .001
	Session 4	48	23.9 (3.1)	26.9 (3.8)	2.9 [1.9, 4.0]	< .001	< .001
	1 week	46	23.8 (3.0)	28.2 (5.4)	4.4 [3.0, 5.8]	< .001	< .001
	3 months	42	23.8 (3.0)	26.8 (4.4)	3.0 [1.8, 4.3]	< .001	< .001
Average flow	Session 2	46	178.3 (62.1)	215.9 (55.4)	37.6 [19.0, 56.3]	< .001	< .001
	Session 3	46	178.7 (61.8)	230.4 (61.2)	51.7 [38.0, 65.4]	< .001	< .001
	Session 4	47	177.7 (61.5)	229.6 (69.5)	51.9 [34.3, 69.6]	< .001	< .001
	1 week	46	176.7 (61.7)	239.1 (78.9)	62.4 [42.5, 82.3]	< .001	< .001
	3 months	42	176.7 (59.2)	236.2 (69.7)	59.5 [41.1, 78.0]	< .001	< .001
dB SPL	Session 2	46	76.6 (3.8)	72.0 (2.6)	-4.6 [-5.8, -3.4]	< .001	< .001
	Session 3	46	76.7 (3.7)	72.4 (3.1)	-4.3 [-5.4, -3.3]	< .001	< .001
	Session 4	47	76.6 (3.7)	71.3 (9.8)	-5.3 [-8.3, -2.2]	.001	.005
	1 week	46	76.5 (3.7)	73.1 (2.4)	-3.5 [-4.4, -2.5]	< .001	< .001
	3 months	42	76.2 (3.8)	73.6 (2.7)	-2.6 [-3.7, -1.5]	< .001	< .001
<i>F</i> ₀	Session 2	46	162.6 (34.1)	177.3 (38.2)	14.7 [11.1, 18.4]	< .001	< .001
	Session 3	46	162.7 (34.1)	179.0 (36.1)	16.2 [12.7, 19.8]	< .001	< .001
	Session 4	47	162.8 (33.7)	178.9 (41.0)	16.1 [11.0, 21.1]	< .001	< .001
	1 week	46	164.7 (33.4)	182.8 (36.2)	18.1 [14.0, 22.2]	< .001	< .001
	3 months	42	164.0 (35.0)	180.3 (37.8)	16.3 [12.1, 20.6]	< .001	< .001
Vocal effort (past week)	Session 2	46	58.3 (28.8)	55.1 (25.3)	-3.2 [-10.0, 3.7]	.357	.89
	Session 3	45	57.5 (28.7)	46.5 (25.3)	-11.0 [-19.6, -2.4]	.014	.067
	Session 4	47	58.6 (28.6)	35.4 (24.7)	-23.3 [-32.6, -13.9]	< .001	< .001
	1 week	46	57.7 (28.2)	28.3 (21.5)	-29.4 [-37.9, -20.9]	< .001	< .001
	3 months	39	59.3 (28.1)	26.5 (23.2)	-32.8 [-42.2, -23.5]	< .001	< .001

(table continues)

Table 3. (Continued).

Outcome	Follow-up time point	N	Baseline M (SD) ^a	Follow-up M (SD)	Change M [95% CI]	p ^b	p ^c
PNF	Session 2	44	61.9 (17.0)	65.2 (21.4)	3.3 [-3.3, 9.9]	.316	.85
	Session 3	43	63.0 (16.7)	72.4 (18.6)	9.5 [3.4, 15.6]	.003	.016
	Session 4	46	62.1 (16.6)	76.8 (21.6)	14.7 [7.6, 21.8]	< .001	< .001
	1 week	45	62.4 (16.7)	81.2 (16.7)	18.8 [12.3, 25.2]	< .001	< .001
	3 months	39	62.4 (16.6)	81.1 (21.1)	18.6 [10.6, 26.6]	< .001	< .001
PSS score	1 week	46	17.6 (7.4)	16.1 (7.1)	-1.5 [-3.2, 0.2]	.083	.353
	3 months	41	17.4 (7.3)	14.0 (7.2)	-3.4 [-5.6, -1.3]	.002	.012
Temporal outcome ^d	Session 2	48	15 (31.2%)	17 (35.4%)	4.2%	.724	.998
	Session 3	48	15 (31.2%)	27 (56.2%)	25.0%	.006	.029
	Session 4	48	15 (31.2%)	32 (66.7%)	35.4%	< .001	.002
	1 week	47	15 (31.9%)	37 (78.7%)	46.8%	< .001	< .001
	3 months	42	12 (28.6%)	35 (83.3%)	54.8%	< .001	< .001
Females							
VHI	Session 2	36	20.7 (5.9)	16.4 (6.3)	-4.3 [-5.8, -2.7]	< .001	< .001
	Session 3	35	20.9 (5.9)	14.7 (6.9)	-6.2 [-8.1, -4.3]	< .001	< .001
	Session 4	36	20.7 (5.9)	11.6 (7.4)	-9.1 [-11.4, -6.9]	< .001	< .001
	1 week	35	20.6 (5.9)	10.3 (6.9)	-10.3 [-12.3, -8.3]	< .001	< .001
	3 months	31	21.3 (6.2)	8.0 (8.1)	-13.3 [-15.5, -11.0]	< .001	< .001
Sustained Vowel–CPP	Session 2	35	10.3 (1.9)	11.2 (1.9)	1.0 [0.4, 1.6]	.001	.005
	Session 3	36	10.3 (1.9)	11.2 (1.8)	0.9 [0.3, 1.5]	.007	.035
	Session 4	36	10.3 (1.9)	10.9 (1.6)	0.6 [-0.0, 1.3]	.059	.263
	1 week	34	10.2 (1.9)	11.1 (1.4)	1.0 [0.3, 1.6]	.006	.028
	3 months	31	10.3 (2.0)	10.6 (1.7)	0.3 [-0.3, 0.8]	.339	.873
Sustained Vowel–CPP SD	Session 2	35	1.1 (0.8)	0.8 (0.6)	-0.3 [-0.6, 0.0]	.059	.26
	Session 3	36	1.1 (0.8)	1.1 (1.6)	-0.0 [-0.6, 0.6]	.992	> .99
	Session 4	36	1.1 (0.8)	0.8 (0.5)	-0.3 [-0.6, -0.0]	.042	.193
	1 week	34	1.1 (0.9)	0.9 (0.6)	-0.2 [-0.6, 0.1]	.133	.51
	3 months	31	1.2 (0.9)	0.7 (0.5)	-0.4 [-0.8, -0.1]	.011	.053
Sustained Vowel–CSID	Session 2	34	25.7 (16.9)	17.0 (13.6)	-8.7 [-14.7, -2.8]	.005	.027
	Session 3	36	25.0 (17.3)	15.9 (13.5)	-9.1 [-14.7, -3.5]	.002	.011
	Session 4	36	25.0 (17.3)	17.0 (13.1)	-8.0 [-13.3, -2.7]	.004	.021
	1 week	34	26.6 (16.3)	16.1 (16.1)	-10.5 [-17.5, -3.4]	.005	.024
	3 months	31	23.4 (17.5)	18.4 (14.8)	-5.1 [-11.1, 0.9]	.093	.387
Sentence–CPP F ₀	Session 2	35	181.5 (24.3)	192.5 (40.6)	10.9 [-1.8, 23.7]	.091	.379
	Session 3	36	181.5 (24.0)	201.4 (22.2)	19.9 [13.6, 26.2]	< .001	< .001
	Session 4	36	181.5 (24.0)	203.1 (24.1)	21.6 [15.5, 27.7]	< .001	< .001
	1 week	34	184.3 (21.5)	208.7 (20.4)	24.4 [16.2, 32.7]	< .001	< .001
	3 months	31	182.4 (25.0)	190.2 (55.4)	7.9 [-9.9, 25.6]	.372	.903
Sentence–CPP F ₀ SD	Session 2	35	29.4 (12.3)	31.3 (14.4)	1.9 [-4.5, 8.3]	.551	.982
	Session 3	36	29.1 (12.2)	33.9 (10.5)	4.7 [-0.1, 9.6]	.056	.25
	Session 4	36	29.1 (12.2)	36.1 (10.0)	6.9 [1.2, 12.6]	.019	.091
	1 week	34	28.7 (12.2)	37.8 (12.5)	9.2 [3.5, 14.8]	.002	.012
	3 months	31	28.3 (12.0)	36.4 (13.0)	8.0 [1.8, 14.2]	.013	.063
Sentence–CSID	Session 2	35	6.0 (15.0)	2.0 (17.4)	-4.1 [-11.0, 2.9]	.246	.756
	Session 3	36	5.7 (14.9)	1.3 (9.9)	-4.4 [-8.8, 0.1]	.055	.248
	Session 4	36	5.7 (14.9)	0.2 (9.9)	-5.4 [-10.7, -0.2]	.044	.2
	1 week	34	6.3 (15.0)	-1.4 (9.3)	-7.8 [-12.3, -3.2]	.002	.008
	3 months	31	6.2 (15.2)	5.5 (16.1)	-0.8 [-5.7, 4.2]	.752	.999
No. of breaths	Session 2	36	4.9 (1.6)	6.2 (2.0)	1.3 [0.7, 1.9]	< .001	< .001
	Session 3	35	4.9 (1.7)	6.9 (2.1)	2.1 [1.6, 2.5]	< .001	< .001
	Session 4	36	4.9 (1.6)	6.5 (1.8)	1.6 [1.2, 2.1]	< .001	< .001
	1 week	34	4.9 (1.7)	7.0 (2.5)	2.1 [1.5, 2.7]	< .001	< .001
	3 months	31	4.5 (1.5)	6.3 (2.1)	1.7 [1.0, 2.5]	< .001	< .001
Duration	Session 2	36	24.2 (3.3)	27.3 (3.8)	3.2 [2.0, 4.4]	< .001	< .001
	Session 3	35	24.3 (3.2)	28.4 (4.6)	4.2 [2.6, 5.7]	< .001	< .001
	Session 4	36	24.2 (3.3)	27.7 (3.8)	3.5 [2.4, 4.6]	< .001	< .001
	1 week	34	24.0 (3.1)	29.2 (5.6)	5.2 [3.6, 6.8]	< .001	< .001
	3 months	31	23.9 (3.2)	27.4 (4.6)	3.5 [2.0, 4.9]	< .001	< .001
Average flow	Session 2	34	173.2 (60.9)	210.3 (55.1)	37.1 [12.7, 61.5]	.004	.02
	Session 3	34	173.8 (60.6)	228.5 (64.2)	54.7 [37.2, 72.2]	< .001	< .001
	Session 4	35	172.6 (60.2)	228.3 (74.9)	55.7 [32.7, 78.7]	< .001	< .001
	1 week	34	171.2 (60.2)	238.5 (85.1)	67.4 [41.4, 93.3]	< .001	< .001
	3 months	31	170.3 (55.3)	232.9 (71.4)	62.6 [39.4, 85.7]	< .001	< .001

(table continues)

Table 3. (Continued).

Outcome	Follow-up time point	N	Baseline M (SD) ^a	Follow-up M (SD)	Change M [95% CI]	p ^b	p ^c
dB SPL	Session 2	34	76.8 (3.8)	72.3 (2.3)	-4.6 [-5.9, -3.2]	< .001	< .001
	Session 3	34	77.0 (3.7)	73.0 (2.5)	-4.0 [-5.2, -2.8]	< .001	< .001
	Session 4	35	76.8 (3.8)	73.1 (2.8)	-3.7 [-5.0, -2.4]	< .001	< .001
	1 week	34	76.7 (3.8)	73.5 (2.5)	-3.3 [-4.4, -2.1]	< .001	< .001
	3 months	31	76.3 (3.8)	73.9 (2.9)	-2.4 [-3.7, -1.1]	< .001	.004
F ₀	Session 2	34	178.3 (23.6)	195.8 (23.3)	17.5 [13.5, 21.5]	< .001	< .001
	Session 3	34	178.4 (23.5)	196.3 (22.0)	17.9 [13.8, 22.0]	< .001	< .001
	Session 4	35	178.1 (23.3)	196.6 (30.1)	18.5 [12.2, 24.8]	< .001	< .001
	1 week	34	181.2 (19.9)	201.6 (17.4)	20.4 [15.5, 25.3]	< .001	< .001
	3 months	31	180.3 (23.7)	198.2 (23.5)	17.8 [13.6, 22.1]	< .001	< .001
Vocal effort (past week)	Session 2	35	59.9 (28.8)	56.6 (26.3)	-3.2 [-11.7, 5.3]	.445	.947
	Session 3	35	59.9 (28.8)	47.9 (25.4)	-11.9 [-22.4, -1.5]	.026	.123
	Session 4	35	59.9 (28.8)	36.2 (24.6)	-23.7 [-34.7, -12.7]	< .001	< .001
	1 week	34	58.7 (28.3)	29.4 (21.1)	-29.3 [-38.9, -19.6]	< .001	< .001
	3 months	29	62.4 (27.4)	27.6 (24.9)	-34.9 [-45.6, -24.2]	< .001	< .001
PNF	Session 2	34	61.6 (18.2)	65.8 (21.5)	4.1 [-3.6, 11.8]	.283	.811
	Session 3	34	62.6 (17.7)	71.0 (19.7)	8.4 [0.9, 15.9]	.029	.138
	Session 4	35	61.9 (18.0)	79.1 (19.4)	17.3 [10.1, 24.4]	< .001	< .001
	1 week	34	62.3 (18.1)	81.3 (17.2)	19.0 [11.0, 27.0]	< .001	< .001
	3 months	30	61.9 (17.7)	78.6 (22.7)	16.8 [6.9, 26.6]	.002	.008
PSS score	1 week	34	17.1 (7.6)	15.8 (7.0)	-1.3 [-3.5, 0.9]	.245	.755
	3 months	30	17.2 (7.8)	13.4 (6.8)	-3.9 [-6.6, -1.1]	.007	.036
Temporal outcome ^d	Session 2	36	11 (30.6%)	13 (36.1%)	5.6%	.683	.997
	Session 3	36	11 (30.6%)	19 (52.8%)	22.2%	.027	.127
	Session 4	36	11 (30.6%)	24 (66.7%)	36.1%	.004	.018
	1 week	35	11 (31.4%)	27 (77.1%)	45.7%	< .001	.002
	3 months	31	8 (25.8%)	25 (80.6%)	54.8%	< .001	.001

Note. VHI = Voice Handicap Index; CPP = cepstral peak prominence; CSID = Cepstral Spectral Index of Dysphonia; F₀ = fundamental frequency; PNF = percent normal function; PSS = Perceived Stress Scale.

^aVariability in baseline means is due to recalculation of means for each time point comparison to account for patient attrition. ^bp value calculated using a paired *t* test. ^cSidak-corrected *p* value. PSS score was corrected for two follow-up time points. All other variables were corrected for five follow-up time points. ^dThe table shows *n* (%); *p* values were calculated using McNemar's test.

Occupational Voice

The number of hours of work missed due to voice problems was assessed by patient report from baseline to the start of each therapy and follow-up session. The majority (70%–84%) of patients did not miss any work hours due to voice problems at baseline; therefore, no treatment effect was observed.

Number of Sessions

Participants in the CTT group were prospectively assigned to attendance at four voice therapy sessions.

Table 4. Linear mixed-model estimates for the Voice Handicap Index–10 change for each conversation training therapy session.

Time point	Estimate [95% CI]	p ^a	p ^b
Baseline	20.7 [19.0, 22.3]	< .001	< .001
Session 2 vs. Baseline	-4.3 [-5.6, -2.9]	< .001	< .001
Session 3 vs. Baseline	-6.4 [-7.8, -5.0]	< .001	< .001
Session 4 vs. Baseline	-9.2 [-10.6, -7.8]	< .001	< .001
1 week vs. Baseline	-10.2 [-11.6, -8.8]	< .001	< .001
3 months vs. Baseline	-12.7 [-14.2, -11.3]	< .001	< .001

^ap value from a linear mixed-model Wald test. ^bSidak-corrected *p* value for five follow-up time points.

Forty-eight percent of HMC patients completed four sessions of voice therapy, 27% completed five sessions, 8% completed six sessions, 2% completed seven sessions, and 15% completed eight sessions.

Discussion

We compared the efficacy of a novel voice therapy approach, CTT to HMC patients diagnosed with benign vocal fold lesions and MTD treated with standard-of-care voice therapy. Results support the efficacy of CTT in these populations. Participants treated with CTT showed greater improvement in perceived voice handicap than those treated with standard-of-care therapy. Most importantly, participants in the CTT cohort achieved equivalent gains in terms of mean VHI-10 decrease following only two sessions of CTT compared to those observed at therapy completion (four to eight sessions) with standard-of-care therapy.

CTT represents a shift in the current prevailing voice therapy paradigm, demonstrating success with a program that does not include any prescriptive exercises but allows the patient to select the stimuli through their own narrative. In doing so, SLPs fine tune and guide treatment to the patient's personal, occupational, vocational, and vocal

Figure 2. Graphical representation of mean Voice Handicap Index–10 (VHI-10) scores at each time point for conversation training therapy participants. CI = confidence interval.

deficits. Progress is determined by the patient's own vocal discoveries in their conversational speech. Participants experienced a significant decrease in their perception of voice handicap following only one therapy session, a clinically meaningful decrease (-5) by two sessions, and an average VHI-10 score of 2 *SDs* from the normal mean at 1-week follow-up (Arffa et al., 2012). Even though the HMC group showed significant improvement with standard-of-care voice therapy, VHI-10 change scores were larger for the CTT group, and the CTT group's change scores reached within 2 *SDs* of the norm for individuals with nonvoice disorder 1 week following therapy whereas the HMC group did not. Notably, patients' perception of voice handicap demonstrated continued improvement (decrease) 3 months following the end of treatment. This change trend was also observed for secondary acoustic, aerodynamic, and auditory-perceptual outcomes. Table 7 displays each measure that demonstrated a statistically significant improvement and the corresponding data collection time point where the improvement was observed. Results support our pilot data, which demonstrated that patients were discharged having met therapeutic goals after an average of three (range: 2–4) CTT sessions (Gartner-Schmidt et al., 2016).

Congruence between patient perception of voice handicap and objective, instrumental voice measures is uncommon. Past findings from this author group revealed that only aerodynamic outcomes correlated with change in VHI-10 and only for patients with unilateral vocal fold immobility (Gillespie, Gooding, Rosen, & Gartner-Schmidt, 2014). Other authors have found a similar lack of correlation between patient-reported outcomes and aerodynamic measures (Cheng & Woo, 2010). However, those studies used single-sentence or consonant–vowel repetitions, whereas our aerodynamic outcomes included connected speech of moderate duration (> 20 s). Thus, changes in connected speech may be more linked to a patient's report of vocal handicap. Likewise, we and others have questioned the reliability and sensitivity to change of acoustic outcomes (Carding et al., 2004; Gillespie, Dastolfo, Magid, & Gartner-Schmidt, 2014; Gillespie, Gooding, et al., 2014; Smits, Marres, & de Jong, 2012). Instead of examining primarily frequency as in the aforementioned studies, in the current study, we used time- and frequency-based acoustic outcomes (i.e., cepstral analyses). Similar to the aerodynamic outcomes, past acoustic investigations have been conducted on sustained phoneme analyses, whereas the current study analyzed acoustic measures in not only vowels but also connected speech, a more ecologically valid task for speakers with voice problems. Lastly, to our knowledge, this is the first voice therapy study to use a stimulability protocol as part of its inclusion criteria, which likely led to optimal patients being included in the trial.

A central hypothesis of the current investigation was that training voice in patient-generated conversational speech would encourage increased vocal motor learning compared to hierarchical treatments. Multiple results support this hypothesis. First, improvements were observed earliest in outcomes that assessed sustained vowels and later in connected speech. Because the tested specific phrases are not part of the therapy program, this pattern of improvement potentially demonstrates generalization, a benchmark of learning, of the skill to novel tasks, that is, connected, natural speech. Second, a hallmark of learning is that skill acquisition occurs with time, whereas performance tends to improve immediately. Thus, the continued improvement 3 months posttherapy termination might demonstrate ongoing skill acquisition as opposed to shorter

Table 5. Characteristics of matched samples.

Characteristic	CTT (<i>n</i> = 48)	Historical control (<i>n</i> = 48)	<i>p</i> ^a
Female, <i>n</i> (%)	36 (75.0)	36 (75.0)	.99
Disease diagnosis, <i>n</i> (%)			.99
Lesion	18 (37.5)	18 (37.5)	
MTD	30 (62.5)	30 (62.5)	
Age (years), <i>M</i> (<i>SD</i>)	37.2 (12.0)	35.9 (11.6)	.587
Baseline VHI-10, <i>M</i> (<i>SD</i>)	21.1 (5.9)	20.4 (7.0)	.583

Note. CTT = conversation training therapy; MTD = muscle tension dysphonia; VHI-10 = Voice Handicap Index–10.

^a*p* value calculated using a *t* test for age or a chi-square test for categorical variables.

Table 6. Linear mixed-model estimates for Voice Handicap Index–10 in conversation training therapy (CTT) compared to historical matched controls.

Variable	Estimate [95% CI]	<i>p</i> ^a
CTT	0.6 [–2.0, 3.3]	.641
Time point	–5.8 [–8.4, –3.1]	< .001
CTT × Time Point	–4.7 [–8.4, –0.9]	.016
MTD	–0.5 [–3.3, 2.3]	.72
Male	1.7 [–1.4, 4.8]	.287
Age	0.1 [–0.0, 0.2]	.175

Note. MTD = muscle tension dysphonia.

^a*p* value from a linear mixed-model Wald test.

term outcomes, which may represent vocal performance, but not learning. Finally, in standard-of-care voice therapy, transfer to speech occurs at the end of the therapeutic process, if at all. By targeting transfer in the first therapy and throughout, in CTT, patients are given longer periods of time to concentrate on transfer while being supported by the SLP during the process (Ohlsson, 2016).

Many of the findings in this study support the acoustic, aerodynamic, and perceptual outcomes found in the clear speech literature (Picheny et al., 1985, 1986). To the best of our knowledge, this is the first study to replicate findings in patients with MTD and benign vocal fold lesions. In contrast to other studies that have used clear speech, participants in this study demonstrated a significant decrease

in vocal intensity as a result of treatment. We hypothesized that participants treated with CTT would result in an increase in intensity, as reflected in the clear speech literature (Gillespie & Gartner-Schmidt, 2016; Picheny et al., 1986). This result may be explained via examination of the aerodynamic data. Airflow in speech significantly increased; if a subsequent increase in estimated subglottal pressure did not occur, then a decrease in laryngeal resistance and, potentially, intensity would be decreased. Subglottal pressure data were not collected to confirm this hypothesis but would be worth exploring in future investigations. Furthermore, participants experienced a significant decrease in vocal effort with treatment. This decrease in effort, coupled with an increase in airflow, may also explain the decrease in vocal intensity. Finally, clear speech was only one component of CTT, and an exact replication of findings from other studies that used only clear speech cannot be expected.

Another unexpected finding was that of lack of significant improvement in the standard deviation of the fundamental frequency for males. Prosody—variations in frequency and intensity—was a trained component of CTT, and we hypothesized that this training would be reflected in an increase in the standard deviation of the fundamental, which was observed only for female participants. The gender difference may simply be due to more females than males enrolled in the trial, and therefore, we were not powered to detect significant gender differences for all outcomes tested. Alternatively, perhaps the nature of the acoustic tasks measured (single-sentence reading) did not warrant a change in prosody for all participants, and therefore, the measure only improved for females. On the other hand, number of breaths and duration of time of the reading passage both increased, which we hypothesize is representative of training pauses in speech for breath replenishment as a component of CTT.

Potentially related to the increase in fundamental frequency standard deviation for females is the significant decrease in participant-rated vocal effort. Past findings have demonstrated that vocal fatigue—a proxy for effort—decreases with increased frequency variability (McCabe & Titze, 2002). Participants in this study also had significant increases in average phonatory airflow, which may also mitigate vocal effort.

Future Directions

In order to further ascertain the impact of CTT on learning, we plan to study longer term outcomes of CTT by prospectively following a patient group for 1–2 years posttreatment. Furthermore, we plan to prospectively, systematically compare CTT to non-conversation-based voice therapy approaches in a large sample of patients with voice disorders in order to compare the treatment methodologies on generalization and learning outcomes. We plan to investigate the characteristics that indicate a patient is most appropriate for CTT intervention versus more traditional approaches and how to best determine CTT candidacy. Finally, we will develop and test a pediatric CTT program.

Figure 3. Graphical representation of adjusted mean changes in Voice Handicap Index–10 (VHI-10) for the conversation training therapy (CTT) and historical matched control (HMC) groups.

Table 7. Significant improvement by time point.

Outcome measure	Session 2 ^a	Session 3	Session 4	Short term	Long term
Patient perception outcomes					
VHI-10	x	x	x	x	x
% Normal vocal function		x	x	x	x
"I can do what I need to do"			x	x	x
Vocal effort			x	x	x
Acoustic outcomes					
CPP F ₀ SD			x	x	x
CSID /a/	x	x			
CSID sentence			x	x	
dB SPL	x	x	x	x	x
Aerodynamic outcomes					
Average airflow in speech	x	x	x	x	x
No. of breaths in speech	x	x	x	x	x
Duration	x	x	x	x	x
Auditory-perceptual outcome					
Overall voice severity		x	x	x	x

Note. VHI-10 = Voice Handicap Index-10; CPP F₀ = cepstral peak prominence fundamental frequency; CSID = Cepstral Spectral Index of Dysphonia.

^aMeasure was collected prior to the start of each session, representing change induced by the previous session.

Limitations

This study has three noteworthy limitations. First, this study was conducted at a single center with SLPs who contributed to the development of the final CTT program. In turn, positive CTT bias could have occurred. How results would differ with the inclusion of other SLPs with different practice patterns and patient populations remains unknown. Future studies exploring the efficacy of CTT are planned and will include multiple voice centers. However, the majority of the therapists were the same for both the HMC and CTT groups, which mitigates a therapist effect on the treatment outcomes. Second, despite presenting with self-perceived voice handicaps, the acoustic and auditory-perceptual baseline results indicate that many of the participants in this study had mild voice disorders (Lewandowski et al., 2018). Thus, how a more severely dysphonic population than the one studied here would respond to the treatment is unknown. Finally, a loss of power due to attrition at the 3-month time point was experienced. Despite multiple attempts to contact these participants, the reason for six participants being lost to the 3-month follow-up was not ascertained and remains unknown. This loss of power likely affected our ability to determine if the numeric, descriptive improvements in outcomes observed at the 3-month time point were significant.

Conclusions

CTT is the first voice therapy approach to eliminate the traditional therapeutic hierarchy and use patient-driven conversation as the sole therapeutic stimulus. Results support the hypothesis that training voice techniques in the context of spontaneous conversational speech improves patient perception of voice handicap as well as acoustic, aerodynamic, and auditory-perceptual voice outcomes

both immediately following treatment and at long-term follow-up.

Acknowledgments

This study was supported by National Institute on Deafness and Other Communication Disorders Grant R03 DC015305 (Principal Investigator: Gillespie; Co-Investigators: Yabes, Rosen, and Gartner-Schmidt) and the University of Pittsburgh Competitive Medical Research Fund Award (Principal Investigator: Gillespie). The authors would like to acknowledge Tina Harrison, Maurice Goodwin, Diana Becker, Hayley Buxton, Libby Smith, and Kwonho Jeong, for their assistance with the project, as well as the University of Pittsburgh Voice Center's Speech-Language Pathology team of Rita Hersan, Tracey Thomas, Christina Dastolfo-Hromack, and Ali Lewandowski.

References

- Akbulut, S., Gartner-Schmidt, J. L., Gillespie, A. I., Young, V. N., Smith, L. J., & Rosen, C. A. (2016). Voice outcomes following treatment of benign midmembranous vocal fold lesions using a nomenclature paradigm. *The Laryngoscope*, 126(2), 415–420. <https://doi.org/10.1002/lary.25488>
- Angsuwarangsee, T., & Morrison, M. (2002). Extrinsic laryngeal muscular tension in patients with voice disorders. *Journal of Voice*, 16(3), 333–343. [https://doi.org/10.1016/S0892-1997\(02\)00105-4](https://doi.org/10.1016/S0892-1997(02)00105-4)
- Arffa, R. E., Krishna, P., Gartner-Schmidt, J., & Rosen, C. A. (2012). Normative values for the Voice Handicap Index-10. *Journal of Voice*, 26(4), 462–465. <https://doi.org/10.1016/j.jvoice.2011.04.006>
- Awan, S. N., Roy, N., Jetté, M. E., Meltzner, G. S., & Hillman, R. E. (2010). Quantifying dysphonia severity using a spectral/cepstral-based acoustic index: Comparisons with auditory-perceptual judgements from the CAPE-V. *Clinical Linguistics & Phonetics*, 24(9), 742–758. <https://doi.org/10.3109/02699206.2010.492446>

- Bandura, A.** (1977). Toward a unifying theory of behavioral change. *Psychological Review*, 84, 191–215. <https://doi.org/10.1037/0033-295X.84.2.191>
- Bassiouny, S.** (1998). Efficacy of the accent method of voice therapy. *Folia Phoniatrica et Logopaedica*, 50(3), 146–164. <https://doi.org/10.1159/000021458>
- Berry, D. A., Verdolini, K., Montequin, D. W., Hess, M. M., Chan, R. W., & Titze, I. R.** (2001). A quantitative output-cost ratio in voice production. *Journal of Speech, Language, and Hearing Research*, 44(1), 29–37. [https://doi.org/10.1044/1092-4388\(2001/003\)](https://doi.org/10.1044/1092-4388(2001/003))
- Bjork, R. A.** (1994). Memory and metamemory considerations in the training of human beings. In J. Metcalfe & A. P. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 185–205). Cambridge, MA: MIT Press.
- Bonilha, H. S., & Dawson, A. E.** (2012). Creating a mastery experience during the voice evaluation. *Journal of Voice*, 26(5), 665.e1–665.e7. <https://doi.org/10.1016/j.jvoice.2011.09.004>
- Bradlow, A. R., Kraus, N., & Hayes, E.** (2003). Speaking clearly for children with learning disabilities. *Journal of Speech, Language, and Hearing Research*, 46(1), 80–97. [https://doi.org/10.1044/1092-4388\(2003/007\)](https://doi.org/10.1044/1092-4388(2003/007))
- Carding, P. N., Horsley, I. A., & Docherty, G. J.** (1999). A study of the effectiveness of voice therapy in the treatment of 45 patients with nonorganic dysphonia. *Journal of Voice*, 13(1), 72–104. [https://doi.org/10.1016/S0892-1997\(99\)80063-0](https://doi.org/10.1016/S0892-1997(99)80063-0)
- Carding, P. N., Steen, I. N., Webb, A., Mackenzie, K., Deary, I. J., & Wilson, J. A.** (2004). The reliability and sensitivity to change of acoustic measures of voice quality. *Clinical Otolaryngology & Allied Sciences*, 29(5), 538–544. <https://doi.org/10.1111/j.1365-2273.2004.00846.x>
- Casper, J. K., & Murry, T.** (2000). Voice therapy methods in dysphonia. *Otolaryngologic Clinics of North America*, 33(5), 983–1002. [https://doi.org/10.1016/S0030-6665\(05\)70259-0](https://doi.org/10.1016/S0030-6665(05)70259-0)
- Chen, S. H., Hsiao, T. Y., Hsiao, L. C., Chung, Y. M., & Chiang, S. C.** (2007). Outcome of resonant voice therapy for female teachers with voice disorders: Perceptual, physiological, acoustic, aerodynamic, and functional measurements. *Journal of Voice*, 21(4), 415–425. <https://doi.org/10.1016/j.jvoice.2006.02.001>
- Cheng, J., & Woo, P.** (2010). Correlation between the Voice Handicap Index and voice laboratory measurements after phonosurgery. *Ear, Nose & Throat Journal*, 89(4), 183–188. <https://doi.org/10.1177/014556131008900411>
- Cohen, S. M., Kim, J., Roy, N., Asche, C., & Courey, M.** (2012). Direct health care costs of laryngeal diseases and disorders. *The Laryngoscope*, 122(7), 1582–1588. <https://doi.org/10.1002/lary.23189>
- Dejonckere, P. H., & Lebacqz, J.** (2001). Plasticity of voice quality: A prognostic factor for outcome of voice therapy? *Journal of Voice*, 15(2), 251–256. [https://doi.org/10.1016/S0892-1997\(01\)00025-X](https://doi.org/10.1016/S0892-1997(01)00025-X)
- Fairbanks, G.** (1960). *Voice and articulation drill book* (2nd ed.). New York, NY: Harper & Row.
- Ferguson, S.** (2004). Talker differences in clear and conversational speech: Vowel intelligibility for normal hearing listeners. *The Journal of the Acoustical Society of America*, 116, 2365–2373. <https://doi.org/10.1121/1.1788730>
- Fontana, F. E., Mazzardo, O., Furtado, O., Jr., & Gallagher, J. D.** (2009). Whole and part practice: A meta-analysis. *Perceptual and Motor Skills*, 109(2), 517–530. <https://doi.org/10.2466/pms.109.2.517-530>
- Gartner-Schmidt, J. L.** (2013). Flow phonation. In A. Behrman (Ed.), *The complete voice therapy workbook*. San Diego, CA: Plural.
- Gartner-Schmidt, J. L.** (2014). Flow phonation. In J. Stemple & E. Hapner (Eds.), *Voice therapy clinical case studies* (4th ed.). San Diego, CA: Plural.
- Gartner-Schmidt, J. L., Gherson, S., Hapner, E. R., Muckala, J., Roth, D., Schneider, S., & Gillespie, A. I.** (2016). The development of conversation training therapy: A concept paper. *Journal of Voice*, 30(5), 563–573. <https://doi.org/10.1016/j.jvoice.2015.06.007>
- Gartner-Schmidt, J. L., Hirai, R., Dastolfo, C., Rosen, C. A., Yu, L., & Gillespie, A. I.** (2015). Phonatory aerodynamics in connected speech. *The Laryngoscope*, 125(12), 2764–2771. <https://doi.org/10.1002/lary.25458>
- Gartner-Schmidt, J. L., & Rosen, C.** (2011). Treatment success for age-related vocal fold atrophy. *The Laryngoscope*, 121(3), 585–589. <https://doi.org/10.1002/lary.21122>
- Gillespie, A. I., Dastolfo, C., Magid, N., & Gartner-Schmidt, J.** (2014). Acoustic analysis of four common voice diagnoses: Moving toward disorder-specific assessment. *Journal of Voice*, 28(5), 582–588. <https://doi.org/10.1016/j.jvoice.2014.02.002>
- 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.e9–507.e14. <https://doi.org/10.1016/j.jvoice.2015.06.004>
- Gillespie, A. I., Gooding, W., Rosen, C., & Gartner-Schmidt, J.** (2014). Correlation of VHI-10 to voice laboratory measurements across five common voice disorders. *Journal of Voice*, 28(4), 440–448. <https://doi.org/10.1016/j.jvoice.2013.10.023>
- Hapner, E., Portone-Maira, C., & Johns, M. M., III.** (2009). A study of voice therapy dropout. *Journal of Voice*, 23(3), 337–340. <https://doi.org/10.1016/j.jvoice.2007.10.009>
- Helou, L.** (2017). Crafting the dialogue: Meta-therapy in transgender voice and communication training. *Perspectives of the ASHA Special Interest Groups*, 2(10), 83–91. <https://doi.org/10.1044/persp2.SIG10.83>
- Holmberg, E. B., Hillman, R. E., Hammarberg, B., Södersten, M., & Doyle, P.** (2001). Efficacy of a behaviorally based voice therapy protocol for vocal nodules. *Journal of Voice*, 15(3), 395–412. [https://doi.org/10.1016/S0892-1997\(01\)00041-8](https://doi.org/10.1016/S0892-1997(01)00041-8)
- Iwarsson, J.** (2015). Facilitating behavioral learning and habit change in voice therapy—Theoretic premises and practical strategies. *Logopedics Phoniatrics Vocology*, 40(4), 179–186. <https://doi.org/10.3109/14015439.2014.936498>
- Iwarsson, J., Morris, D. J., & Balling, L. W.** (2017). Cognitive load in voice therapy carry-over exercises. *Journal of Speech, Language, and Hearing Research*, 60(1), 1–12. https://doi.org/10.1044/2016_JSLHR-S-15-0235
- 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\(2008/08-0017\)](https://doi.org/10.1044/1058-0360(2008/08-0017))
- Klein, J. A., & Jones, T. A.** (2008). Principles of experience-dependent neural plasticity: Implications for rehabilitation after brain damage. *Journal of Speech, Language, and Hearing Research*, 51, 225–239. [https://doi.org/10.1044/1092-4388\(2008/018\)](https://doi.org/10.1044/1092-4388(2008/018))
- Lewandowski, A., Gillespie, A. I., Kridgen, S., Jeong, K., Yu, L., & Gartner-Schmidt, J.** (2018). Adult normative data for phonatory aerodynamics in connected speech. *The Laryngoscope*, 128(4), 909–914. <https://doi.org/10.1002/lary.26922>
- Litts, J. K., Gartner-Schmidt, J. L., Clary, M. S., & Gillespie, A. I.** (2015). Impact of laryngologist and speech pathologist

- coassessment on outcomes and billing revenue. *The Laryngoscope*, 125(9), 2139–2142. <https://doi.org/10.1002/lary.25349>
- McCabe, D. J., & Titze, I. R. (2002). Chant therapy for treating vocal fatigue among public school teachers: A preliminary study. *American Journal of Speech-Language Pathology*, 11, 356–369. [https://doi.org/10.1044/1058-0360\(2002\)040](https://doi.org/10.1044/1058-0360(2002)040)
- Mozzanica, F., Ginocchio, D., Barillari, R., Barozzi, S., Maruzzi, P., Ottaviani, F., & Schindler, A. (2016). Prevalence and voice characteristics of laryngeal pathology in an Italian voice therapy-seeking population. *Journal of Voice*, 30(6), 774.e13–74.e21. <https://doi.org/10.1016/j.jvoice.2015.11.018>
- Niebudek-Bogusz, E., Sznurowska-Przygocka, B., Fiszer, M., Kotylo, P., Sinkiewicz, A., Modrzewska, M., & Sliwiska-Kowalska, M. (2008). The effectiveness of voice therapy for teachers with dysphonia. *Folia Phoniatrica et Logopaedica*, 60(3), 134–141. <https://doi.org/10.1159/000120290>
- Ohlsson, A. C. (2016). Verbal Instruction Model (VIM) in voice therapy. *Logopedics Phoniatrics Vocology*, 41(1), 41–46. <https://doi.org/10.3109/14015439.2014.949303>
- Peterson, K. L., Verdolini-Marston, K., Barkmeier, J., & Hoffman, H. T. (1994). Comparison of aerodynamic and electroglottographic parameters in evaluating clinically relevant voicing patterns. *Annals of Otolaryngology, Rhinology & Laryngology*, 103(5), 335–346. <https://doi.org/10.1177/000348949410300501>
- Picheny, M., Durlach, N., & Braid, L. (1985). Speaking clearly for the hard of hearing. I: Intelligibility differences between clear and conversational speech. *Journal of Speech and Hearing Research*, 28(1), 96–103. <https://doi.org/10.1044/jshr.2801.96>
- Picheny, M., Durlach, N., & Braid, L. (1986). Speaking clearly for the hard of hearing. II: Acoustic characteristics of clear and conversational speech. *Journal of Speech and Hearing Research*, 29(4), 434–446. <https://doi.org/10.1044/jshr.2904.434>
- Rodriguez-Parra, M. J., Adrián, J. A., & Casado, J. C. (2011). Comparing voice-therapy and vocal-hygiene treatments in dysphonia using a limited multidimensional evaluation protocol. *Journal of Communication Disorders*, 44(6), 615–630. <https://doi.org/10.1016/j.jcomdis.2011.07.003>
- Roehm, P., & Rosen, C. (2004). Dynamic voice assessment using flexible laryngoscopy—How I do it: A targeted problem and its solution. *American Journal of Otolaryngology*, 25(2), 138–141. <https://doi.org/10.1016/j.amjoto.2003.09.008>
- Rosen, C. A., Gartner-Schmidt, J., Hathaway, B., Simpson, C. B., Postma, G. N., Courey, M., & Sataloff, R. T. (2012). A nomenclature paradigm for benign midmembranous vocal fold lesions. *The Laryngoscope*, 122(6), 1335–1341. <https://doi.org/10.1002/lary.22421>
- 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>
- Rosen, C. A., & Murry, T. (2000). Diagnostic laryngeal endoscopy. *Otolaryngologic Clinics of North America*, 33(4), 751–757. [https://doi.org/10.1016/S0030-6665\(05\)70241-3](https://doi.org/10.1016/S0030-6665(05)70241-3)
- Roy, N. (2003). Functional dysphonia. *Current Opinion in Otolaryngology & Head and Neck Surgery*, 11(3), 144–148.
- Roy, N., Bless, D. M., Heisey, D., & Ford, C. N. (1997). Manual circumlaryngeal therapy for functional dysphonia: An evaluation of short- and long-term treatment outcomes. *Journal of Voice*, 11(3), 321–331. [https://doi.org/10.1016/S0892-1997\(97\)80011-2](https://doi.org/10.1016/S0892-1997(97)80011-2)
- Roy, N., Merrill, R. M., Gray, S. D., & Smith, E. M. (2005). Voice disorders in the general population: Prevalence, risk factors, and occupational impact. *The Laryngoscope*, 115(11), 1988–1995. <https://doi.org/10.1097/01.mlg.0000179174.32345.41>
- Schindler, A., Mozzanica, F., Ginocchio, D., Maruzzi, P., Atac, M., & Ottaviani, F. (2012). Vocal improvement after voice therapy in the treatment of benign vocal fold lesions. *Acta Otorhinolaryngologica Italica*, 32(5), 304–308.
- Schmidt, R. A., & Lee, T. D. (2005). *Motor control and learning. A behavioral emphasis* (4th ed.). Champaign, IL: Human Kinetics Publishers.
- Schneider, B., & Bigenzahn, W. (2005). How we do it: Voice therapy to improve vocal constitution in female student teachers. *Clinical Otolaryngology*, 30, 66–71. <https://doi.org/10.1111/j.1365-2273.2004.00937.x>
- Sellars, C., Carding, P. N., Deary, I. J., MacKenzie, K., & Wilson, J. A. (2002). Characterization of effective primary voice therapy for dysphonia. *The Journal of Laryngology & Otology*, 116(12), 1014–1018. <https://doi.org/10.1258/002221502761698757>
- Smiljanić, R., & Bradlow, A. R. (2008). Temporal organization of English clear and conversational speech. *The Journal of the Acoustical Society of America*, 124(5), 3171–3182. <https://doi.org/10.1121/1.2990712>
- Smits, R., Marres, H., & de Jong, F. (2012). The relation of vocal fold lesions and voice quality to voice handicap and psychosomatic well-being. *Journal of Voice*, 26(4), 466–470. <https://doi.org/10.1016/j.jvoice.2011.04.005>
- Titze, I. R., & Verdolini Abbott, K. (2012). *Vocology: The science and practice of voice habilitation*. Salt Lake City, UT: National Center for Voice and Speech.
- U.S. Census Bureau. (2005). *Age and sex distribution in 2005*. Retrieved from <http://www.census.gov/population/pop-profile/dynamic/AgeSex.pdf>
- Van Houtte, E., Van Lierde, K., D'Haeseleer, E., & Claeys, S. (2010). The prevalence of laryngeal pathology in a treatment-seeking population with dysphonia. *The Laryngoscope*, 120(2), 306–312. <https://doi.org/10.1002/lary.20696>
- van Leer, E., & Connor, N. P. (2010). Patient perceptions of voice therapy adherence. *Journal of Voice*, 24(4), 458–469. <https://doi.org/10.1016/j.jvoice.2008.12.009>
- van Leer, E., & Connor, N. P. (2012). Use of portable digital media players increases patient motivation and practice in voice therapy. *Journal of Voice*, 26(4), 447–453. <https://doi.org/10.1016/j.jvoice.2011.05.006>
- Verdolini, K. (2000). Resonant voice therapy. In J. Stemple (Ed.), *Voice therapy: Clinical studies* (2nd ed., pp. 46–61). San Diego, CA: Singular.
- Verdolini, K., Druker, D. G., Palmer, P. M., & Samawi, H. (1998). Laryngeal adduction in resonant voice. *Journal of Voice*, 12(3), 315–327. [https://doi.org/10.1016/S0892-1997\(98\)80021-0](https://doi.org/10.1016/S0892-1997(98)80021-0)
- Verdolini, K., & Lee, T. D. (2004). Optimizing motor learning in speech interventions: Theory and practice. In C. M. Sapienza & J. Casper (Eds.), *For clinicians by clinicians: Vocal rehabilitation in medical speech-language pathology* (pp. 403–446). Austin, TX: Pro-Ed.
- Verdolini, K., Rosen, C. A., & Branski, R. C. (Eds.) (2005). *Classification manual for voice disorders–I*. Mahwah, NJ: Erlbaum.
- Verdolini Abbott, K. (2008a). *Lessac-Madsen resonant voice therapy*. San Diego, CA: Plural.
- Verdolini Abbott, K. (2008b). *Lessac-Madsen resonant voice therapy* (2nd ed.). San Diego, CA: Plural.
- Verdolini Abbott, K. (2011). *Clinician manual: Casper-Stone confidential voice therapy*. Kankakee, IL: MultiVoiceDimensions.
- Verdolini-Marston, K., & Balota, D. (1994). Role of elaborative and perceptual integrative processes in perceptual-motor performance. *Journal of Experimental Psychology: Learning*,

-
- Memory, and Cognition*, 20, 739–749. <https://doi.org/10.1037/0278-7393.20.3.739>
- Wulf, G., & Weigelt, C.** (1997). Instructions about physical principles in learning a complex motor skill: To tell or not to tell. *Research Quarterly for Exercise and Sport*, 68(4), 362–367. <https://doi.org/10.1080/02701367.1997.10608018>
- Young, V. N., & Rosen, C.** (2011). Videostroboscopy: USA perspective. In E. P. Ma & E. M. Yiu (Eds.), *Handbook of voice assessments* (pp. 99–112). San Diego, CA: Plural.
- Zhang, S., Cao, J., & Ahn, C.** (2010). Calculating sample size in trials using historical controls. *Clinical Trials*, 7(4), 343–353. <https://doi.org/10.1177/1740774510373629>
- Ziegler, A., Dastolfo, C., Hersan, R., Rosen, C., & Gartner-Schmidt, J.** (2014). Perceptions of voice therapy from patients diagnosed with primary muscle tension dysphonia and benign mid-membranous vocal fold lesions. *Journal of Voice*, 28(6), 742–752. <https://doi.org/10.1016/j.jvoice.2014.02.007>