Connecting Auditory-Perceptual Prompts Used in Voice Therapy to Anatomy and Physiology: Application to the Estill Voice Model and the Rehabilitation Treatment Specification System

[*]Elizabeth U. Grillo, [†,‡]Jeremy Wolfberg, [§]Karen Perta, [†,‡,¶]Jarrad Van Stan, and [‖]Kimberly Steinhauer, *West Chester, ‖Pittsburgh, Pennsylvania, †‡¶Boston, Massachusetts, and §Elmhurst, Illinois

Summary: Purpose. This clinical tutorial will present the concept of applying auditory-perceptual prompts (implicit instruction) typically used in voice therapy to the anatomy and physiology of the voice production system (explicit instruction) via the Estill Voice Model (EVM) and the Rehabilitation Treatment Specification System (RTSS).

Methods. EVM offers an integrated implicit-explicit instructional approach to voice training allowing for isolated practice of vocal structures (explicit) that interact to produce functional voice qualities (implicit), such as modal speech and louder projected voice qualities. In EVM, voice quality is correlated with the specific anatomy and physiologic adjustments via 13 Estill Figures and Options (eg, Larynx Figure has three options: High, Mid, and Low). RTSS provides a framework to connect client change in functioning (ie, target) with clinician action (ie, ingredients). Mechanisms of action connect the target to the ingredients by hypothesizing how the treatment is expected to work.

Results. Evidence is provided for connecting auditory-perpetual voice prompts with the anatomy and physiology of voice and supporting an integrated implicit-explicit approach to voice therapy. The concept of linking commonly used implicit auditory-perceptual prompts used in voice therapy (eg, humming, loud "aahh") to explicit anatomy and physiology training (eg, 13 Estill Figures and Options) is demonstrated using EVM and the RTSS framework with case studies and video examples.

Conclusions. Clinicians may choose to use anatomy and physiology of voice to define and provide explicit instruction for typically used implicit auditory-perceptual prompts. Future research is warranted to test the concept applied to voice therapy models in the literature across prevention and treatment of voice disorders.

Key Words: Voice–Voice therapy–Rehabilitation Treatment Specification System–Estill Voice Model.

INTRODUCTION

Behavioral voice therapy, either by itself or in combination with surgical and/or pharmacological intervention from a laryngologist, is the main form of treatment for people with voice problems. Historically, voice therapy approaches facilitated by speech-language pathologists (SLP) have focused on auditory-perceptual prompts (eg, humming, loud voice, twang, etc) provided by the SLP without specific client training in the anatomy and physiology that created the prompt.[1–8] Recently, the literature has defined auditory-perceptual prompts as implicit instruction and references to anatomy and physiology as explicit instruction.[9–11] An implicit instruction may involve imagery (eg, visualize a constant line with no stops), analogy (eg, "make your voice like a calm sea"), reference to a common emotion (eg, laugh, cry, hum), or wholistic gesture (eg, yawn). Directives that emphasize the auditory-perceptual elements of a task (eg, use clear speech or a higher pitch) may also be considered implicit in nature. In contrast, an explicit instruction provides detailed biomechanical information regarding the movement pattern being executed (eg, feel how the larynx lowers during a yawn). Explicit instruction may not include directives related to general, referred bodily sensations that lack reference to the specific anatomical structures involved (eg, feel your voice in your face).

In contrast to traditional speech-language pathology voice therapy approaches, the Estill Voice Model (EVM), created by Jo Estill over 40 years ago, facilitates both implicit and explicit instruction in an integrated approach to voice training. From EVM's inception, it has connected implicit auditory-perceptual voice qualities (eg, opera, sob, belt, nasal twang, etc) with explicit definitions and training of the voice qualities by anatomy (ie, 13 Figures) and physiology (ie, Figure Options). In the Figures for Voice Control, first level of the model, there are 13 Figures related to the major anatomical structures of the voice production system.[12] Each Figure has physiological options

Accepted for publication June 25, 2024.

From the *Department of Communication Sciences and Disorders, West Chester University of Pennsylvania, West Chester, Pennsylvania; †Center for Laryngeal Surgery and Voice Rehabilitation at Massachusetts General Hospital, Boston, Massachusetts; ‡MGH Institute of Health Professions, Boston, Massachusetts; §Department of Communication Sciences and Disorders, Elmhurst University, Elmhurst, Illinois; ¶Department of Surgery at Harvard Medical School, Boston, Massachusetts; and the ‖Estill Voice International, Pittsburgh, Pennsylvania.

Address correspondence and reprint requests to Elizabeth U. Grillo, Department of Communication Sciences and Disorders at West Chester University of Pennsylvania, 201 Carter Drive, Suite 413, West Chester, PA 19383.
E-mail: egrillo@wcupa.edu

Journal of Voice, Vol xx, No xx, pp. xxx–xxx
0892-1997

https://doi.org/10.1016/j.jvoice.2024.06.025

within the given anatomical structure as well as corresponding hand gestures to represent the hypothesized movement associated with voice production (eg, False Vocal Folds (FVF) Figure with Constrict, Mid, and Retract options). In the second level, there are Figure Combinations for six Estill Voice Qualities.[12] Each Estill Quality is defined by the options of the 13 Figures. For example, Estill Oral Twang is operationalized with Aryepiglottic Sphincter (AES) Narrow and Velum High as two of the most relevant Estill Figure Options. EVM's origins began with the performing voice with recent application to the prevention and treatment of clients with speaking voice problems.[9,10,13,14] For example, narrowing AES Figure for "ring" in opera and belt has been applied to increase power in hypofunctional voices, and Thin to Stiff phonation patterns for the True Vocal Folds (TVF): Body-Cover Figure has been applied to optimize contact for hyperfunctional voices. Tables 1 and 2 contain all 13

TABLE 1.
Estill Figures and Options with Summary of Action for Larynx Structures

Figures and Options	Summary of Action
Larynx Structures	
True Vocal Folds (TVF): Onset/Offset	Start and stop sound at the level of the TVF TVF closure precedes/ends breath flow
Glottal 	
Aspirate-gradual 	Breath flow precedes/follows movement of TVF gradually
Aspirate-abrupt 	Breath flow precedes/follows movement of TVF abruptly
Smooth 	TVF closure coordinated with breath flow simultaneously
True Vocal Folds (TVF): Body-Cover	Source of TVF oscillation based on contribution of mass and glottal shape Periodic cycles of thick mass and adducted TVF
Thick 	
Thin 	Periodic cycles of thin mass and adducted TVF
Stiff 	Periodic cycles of thinner mass and variably abducted TVF
Slack 	Aperiodic or periodic pulsed cycles of TVF
False Vocal Folds (FVF)	Movement of FVF toward or away from midline above the TVF Gentle adduction of FVF over TVF
Constrict 	
Mid 	FVF abducted midway over TVF
Retract 	FVF abducted wide away from midline with clear view of TVF
Thyroid	Cartilage that acts as protective shield and anterior attachment for TVF Thyroid cartilage vertically aligned atop cricoid and trachea
Vertical 	
Tilt 	Thyroid cartilage tilted forward and decreasing space of the cricothyroid visor
Cricoid	Most inferior cartilage providing base for arytenoids and posterior attachment of TVF Cricoid vertically aligned with thyroid cartilage atop the trachea
Vertical 	
Tilt 	Cricoid tilts anteriorly to increase the space of the cricothyroid visor
Aryepiglottic Sphincter (AES)	Area of epilaryngeal tube above the FVF and TVF AES relaxes to increase epilaryngeal space
Wide 	
Narrow 	AES tightens to decrease anterior-posterior area of epilarynx

TABLE 2.
Estill Figures and Options with Summary of Action for Vocal Tract and Support Structures

Figures and Options	Summary of Action
Vocal Tract Structures	
Larynx	Movement of entire complex of laryngeal cartilages to vary length of vocal tract
High 	Highest position of larynx as in the initiation of a swallow
Mid 	Larynx rests midway as in relaxed breathing
Low 	Lowest position of larynx as in the initiation of a yawn
Velum	Simplified action of velopharyngeal port that influences coupling of nasal to oral cavities
High 	Velum raises high to seal against posterior pharyngeal wall totally opening the oropharynx
Mid 	Velum remains in mid position between tongue and posterior pharyngeal wall opening both the naso and oropharynx
Low 	Velum lowers to tongue totally opening the nasopharynx
Tongue	Independent placement of tongue dorsum
High 	Dorsum lifted to level of upper molars
Mid 	Dorsum positioned between upper and lower molars
Low 	Dorsum pressed down to level of lower molars
Compress 	Dorsum moves forward while tongue tip moves back
Jaw	Structure that influences oral and pharyngeal spaces
Forward	Lower jaw moves forward as in underbite
Mid	Upper and lower jaw aligned, relaxed
Back	Lower jaw pulled slightly back as in an overbite
Drop	Jaw drops down and mouth opens
Lips	Most anterior structure of vocal tract that influences length
Protrude	Lips in forward relaxed "pucker" position
Mid	Lips relaxed
Spread	Corner of lips pulled laterally to shorten vocal tract
Support Structures	
Head & Neck (H&N)	Velar, sub-occipital, and sternocleidomastoid muscles that lend support to voice production
Relax 	Postural muscles of head and neck and framework for vocal tract relaxed
Anchor 	Postural muscles of head and neck are engaged to stabilize the framework of vocal tract
Torso	Latissimus dorsi, pectoralis major, and quadratus lumborum muscles that lend to stability and management of breath flow
Relax 	Muscles of shoulder and back lightly relaxed
Anchor 	Muscles of shoulder and back engage with respiratory movements remaining free

Figures with corresponding options and hand gestures organized by larynx, vocal tract, and support structures. Jaw and Lips are the only Figures without hand gestures because the physiological options are visible.

There are potential benefits to both the clinician and client for connecting auditory-perceptual voice prompts (implicit) to the anatomy and physiology of the voice system (explicit). For clinicians, understanding the anatomy and physiology that underlies the auditory-perceptual changes might better equip clinicians to change their client's voice quality. It is possible that understanding biomechanics allows clinicians to model nuanced variations of the same voice facilitator (eg, humming) and adjust the esthetics of the target voice quality, providing clients with

more options for their new voice. Using relevant aspects of the anatomy and physiology with hand gestures that physically represent what is hypothesized to be changing in the voice system, the clinician can implement these changes to help clients replicate the mutually agreed-upon, therapeutic voice target and increase their motivation to practice.[9,10,15,16] For clients, understanding the anatomical and physiological changes hypothesized to be responsible for the new voices learned in voice therapy provides a clear rationale for why the auditory-perceptual prompt works. With the rationale, the client may buy-in to the therapeutic process and stay the course. By using anatomy and physiology to guide treatment, the client is given more choices for the desired voice targets, placing them in the driver's seat to determine what voice to use to meet a specific communication need. With autonomy to make voice decisions, the client is in control, a powerful motivator for continued practice.

Clinical research on the efficacy and effectiveness of rehabilitation interventions, including voice therapy, has primarily focused on comparing outcome measures collected before and after an intervention is delivered, with little to no focus on describing what a clinician is having a client do or how clinician actions lead to changes in their client's functioning. This hinders the field's ability to replicate and reliably determine the critical clinician action and connected changes in client functioning. The Rehabilitation Treatment Specification System (RTSS) has been described as a proposed solution to this problem.[17–20]

Within the RTSS, a single target, ingredient(s), and mechanism(s) of action (MoA) are defined as a treatment component.[17] A target is a change in client function achieved through specific clinician actions (ie, ingredients). For example, the client will increase habitual use of resonant voice quality as the target. Ingredients are the actions that the clinician takes to achieve a target; for example, providing opportunities to practice voicing in a bottom-up speech hierarchy, providing feedback on the voicing practice, and discussing how the client can be capable and motivated to perform the practice. Both targets and ingredients are always observable and measurable. For example, the target of increasing habitual use of resonant voice quality could be measured by the client's and/or clinician's perceptual judgment on a scale from 0%–100% accuracy after a 10-minute conversation. Ingredients are further defined by dose (ie, quantitative measure). For example, the ingredient of practice voicing within sentences could be measured by the number of sentence repetitions during the session. MoAs are the connection between the ingredients and the singular target. MoAs can be known and measurable or hypothesized and not directly measurable during routine clinical care. For example, learning by doing and changes in vocal fold kinematics (barely abducted/adducted[21]) are MoAs that connect a target of increased forward resonance with ingredients of providing opportunities for the client to practice voicing in a bottom-up speech hierarchy given clinician feedback.

To categorize treatment components that share similar MoAs, the RTSS also defines three broad treatment groups: 1) Organ Functions treatment components that share an MoA of changing or replacing the function of organ(s), 2) Skills and Habits treatment components that share an MoA of learning by doing through practice, and 3) Representations treatment components that share an MoA of changing mental representations, such as cognition and volitional behavior.[17,19] The RTSS also emphasizes the importance of volition, defined as client's effort needed to engage in and adhere to a behavior.[20] Figure 1 provides example treatment components within each of the three treatment groups using terminology typically used when applying the RTSS framework. Previous work, called the RTSS-Voice,[22,23] has operationalized target and ingredients within voice therapy, mapped the RTSS onto the concept of meta-therapy,[24] and applied the RTSS and RTSS-Voice to 59 standard-of-care voice therapy sessions.[25]

One goal of applying the RTSS framework is to determine the targets, ingredients, and MoAs that are shared generally, across speech-language pathology therapy models and more specifically, across voice therapy models described in the literature. Subsequently, investigations can assess the efficacy and effectiveness of these shared targets, ingredients, and MoAs with recommendations for which ones should be present in every voice therapy session regardless of the specific model (eg, Resonant Voice Therapy described by Stemple and Verdolini Abbott[6–8]). A potential concept that may be important for testing is connecting auditory-perceptual prompts (eg, ask the client to hum for resonant voice or say "aaahhh" for loud voice) to anatomy and physiology of the voice production system (eg, TVF: Body-Cover Thin for resonant voice or TVF: Body-Cover Thick for loud voice). The concept of connecting auditory-perceptual prompts to anatomy and physiology may be used as targets, ingredients, or MoAs depending upon the plan created by the clinician. The literature supports the use of anatomy and physiology to guide treatment in speech disorders,[26,27] voice,[9–16] and dysphagia.[28,29] Speech therapy has historically provided explicit, anatomically based instructions for highly visible structures such as the tongue (eg, tongue tip to alveolar ridge for /s/). EVM Figures with hand gestures and RTSS MoAs guide clinicians to help clients understand and produce physiologic changes in the voice that are less visible (eg, TVF: Body-Cover Thin, FVF Retract, AES Narrow, etc).

The purpose of this clinical tutorial is to present the concept of applying auditory-perceptual prompts used in voice therapy to anatomy and physiology of the voice production system via EVM and RTSS. An in-depth explanation of both EVM and RTSS is outside the scope of this article. For more information, please see the following references for EVM[12,30–32] and RTSS.[17–20,22,23–25] This clinical tutorial will address two primary aims. First,

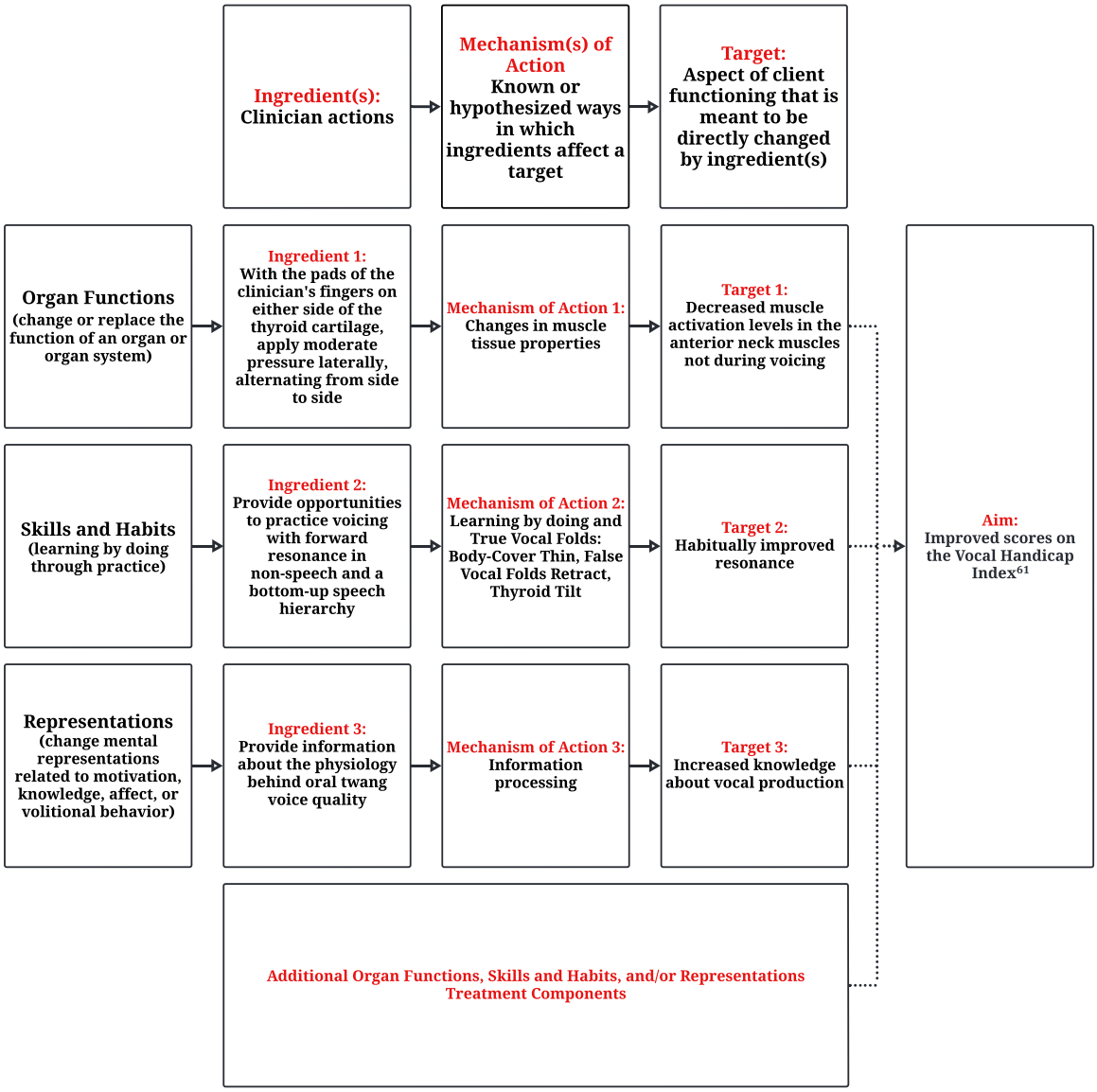

FIGURE 1. Rehabilitation Treatment Specification System (RTSS) examples of Organ Functions, Skills and Habits, Representations, ingredients, mechanism(s) of action (MoA), and targets in voice therapy. To illustrate how Aims are not directly connected to treatment components, a dotted line was placed between the example treatment components and the Aim. An additional treatment component box was also added to illustrate that more treatment components may be introduced to achieve the broad Aim of the intervention.[61]

evidence will be presented for connecting auditory-perceptual prompts to anatomy and physiology of voice production in EVM and for using an integrated implicit-explicit instructional approach to voice training. Second, the concept will be applied to voice therapy using EVM and RTSS with case studies and video demonstrations.

EVIDENCE

It is already common clinical practice to connect auditory-perceptual aspects of voice quality with physiologic findings on laryngoscopy to inform the assessment and treatment of voice disorders. For example, muscle tension dysphonia (MTD) is often characterized by excess supraglottic activity in the FVFs (ie, FVF Figure Constrict) and may be treated with semi-occluded vocal tract exercises (SOVTs), which have been shown to have a widening effect on the supraglottis (ie, FVF Figure Retract).[33,34] Vocal fold paralysis is commonly characterized by weak, breathy voice quality (ie, TVF: Body-Cover Figure Stiff) and may be treated with injection augmentation to increase TVF contact.[35] The 13 EVM Figures and their corresponding options can be used to characterize the auditory-perceptual aspects of voice quality in a physiologic way. Selecting Figure Options to describe vocal quality is a method of hypothesizing what is happening in the mechanism as that voice is produced. When available, laryngoscopy can be used to confirm many Figure Options.

Varying levels of evidence exist to support the 13 EVM Figures and their options. For instance, the three options Glottal, Aspirate, and Smooth (ie, simultaneous voicing with airflow) of TVF: Onset/Offset Figure are prevalent in trained singers, whereas the Glottal and Aspirate options are commonly found in the untrained speaking voice.[36] The four options Thick, Thin, Stiff, and Slack in the TVF: Body-Cover Figure have been successfully differentiated using acoustic and aerodynamic measures.[15] Similarly, following explicit instructions for release of FVF constriction, TVF: Body-Cover Thick, and Larynx Low resulted in optimal acoustic output for vocally healthy participants.[37] The FVF Figure includes the option Constrict, which is a common feature of laryngeal hyperfunction,[33] and Retract, which decreases mediolateral compression of the FVFs in a manner similar to that observed during SOVTs.[34] In terms of the Thyroid cartilage Figure (ie, options Vertical and Tilt), laryngeal tilting has been observed as a component of voice quality that is independent of pitch change.[38] The Larynx Figure represents a continuum from low to high laryngeal position. Changes in vertical larynx position are a component of pitch change[39] as well as a known characteristic of different singing voice qualities, with operatic vs contemporary singing styles associated with lower and higher laryngeal positions, respectively.[38,40–42] The AES Figure and its option Narrow are associated with the production of twang voice quality and an acoustic intensity boost in the range from 2500–3000 Hz or 2800–4300 Hz.[43,44] During the production of twang voice quality, narrowing activity has also been shown to occur in surrounding supralaryngeal structures, including the pharyngeal walls in the lateral dimension and tongue base in the anteroposterior dimension.[42] The narrowed vocal tract adjustment associated with twang quality is a separate and distinct vocal tract gesture from nasality, which is created with a lowered velum and open velopharyngeal port.[42,45] The Velum Figure and its Options (ie, High, Mid, and Low) address the phenomenon of nasality, and even untrained singers can learn to control the extremes of velopharyngeal activity across vowels[46] and speaking tasks of increasing complexity.[11] The six Estill Qualities of speech, falsetto, sob, twang, opera, and belt were distinguished by vocal fold oscillation,[47] and the "*irrintzi* (a folkloric shout emitted in a single breath used by the Basque people)" was analyzed acoustically and defined by EVM to include TVF: Body-Cover Thick, Thyroid Vertical, Cricoid Tilt, Larynx High, AES Narrow, FVF Retract, Tongue High, Velum High, and Anchored Head & Neck and Torso.[48]

Related to evidence for linking auditory-perceptual voice prompts to anatomy and physiology in voice therapy, recent literature suggests positive outcomes for clients with voice problems and healthy voice users. An integrated implicit-explicit instructional approach using EVM for clients with voice problems has been described in the literature.[9,10] The basic protocol was used with 30 clients with voice disorders and included five steps that are not linear, but instead are integrated; 1) implicit clinician modeling of voice prompts followed by client imitation, 2) clinician trains explicit anatomy and physiology of voice, 3) gestures are used to cue for physical changes within the anatomy and physiology of the voice production system, 4) client practices using anatomy and physiology to guide successful productions, and 5) "client becomes their own clinician."[9] The most common Figures and Options that were trained for the 30 clients with voice disorders were FVF Retract, Thyroid Tilt, AES Narrow, and Head & Neck Anchor.[9] In addition, sob voice quality (ie, FVF Retract, AES Wide, Larynx Low, and TVF: Body-Cover Thin or Stiff) has been used to treat clients with mild MTD resulting in positive treatment outcomes as measured by the Consensus Auditory-Perceptual Evaluation of Voice (CAPE-V) and smoothed cepstral peak prominence.[14] Madill et al[14] primarily used participant imitation of a clinician implicit model of a clear, quiet, and effortless /ŋ/. Explicit instruction was provided only if imitation did not elicit the target sob quality. A study examining practice structure and knowledge of results (KR) for hypophonic voice users, novice vocalists, and expert vocalists demonstrated, when learning twang (ie, AES Narrow), a blocked practice schedule enhanced short-term effects of motor acquisition.[49] Random practice and 55% KR degraded performance during short-term acquisition but enhanced long-term performance effects of motor learning. Steinhauer and Eichhorn[49] facilitated twang in an integrated approach while providing auditory-perceptual prompts (eg, quack like a duck, taunting child, wicked witch, etc) linked with the explicit anatomy and physiology training of the back or body of the tongue touching the top molars and FVF Retract. Participants were also provided visual feedback on the level of twanginess in their voice via the "Twangometer" with increased dB spectral intensity between 2.0 and 3.5 kHz.

In vocally untrained participants learning speech nasalance tasks of increasing difficulty, results indicated that, as compared to implicit or explicit instruction alone, integrated instruction (ie, combining auditory-perceptual and biomechanical directives) resulted in larger learning gains and less performance variability.[11] Compared to implicit instruction alone, explicit instruction resulted in lower learning gains on more complex tasks (ie, phrase level) but slightly improved learning outcomes on simple tasks (ie, vowels and syllables). Conversely, compared to explicit instruction only, implicit instruction alone yielded larger learning gains on complex tasks (ie, connected speech). While a few studies in the speech literature have examined the effects of biomechanical attention and/or instruction on speech-motor learning,[50–53] no other studies to date have included an integrated group. Perhaps implicit and explicit processes, as two synergistic components of a functional whole, can work in conjunction across the motor learning trajectory from simple to complex tasks.

The Global Voice Prevention and Therapy Model with Estill Figures and Qualities was successful in using an integrated implicit-explicit approach with 82 vocally healthy

student teachers.[16] The student teachers learned four new voice qualities to meet daily vocal needs of teaching: 1) new resonant voice in connected speech, 2) falsetto for quiet, confidential talking, 3) oral twang for talking over noise, and 4) belt for healthy yelling. Each of the voices used implicit auditory-perceptual prompts connected with explicit training on the anatomical and physiological options for each of the qualities. For example, oral twang was facilitated with the implicit prompt "beep-beep" producing a piercing robotic-like sound. As the client produced the beep-beep, the clinician provided visuals with hand gestures, videos and pictures of the larynx, and narrowband spectrograms demonstrating AES Narrow while maintaining a Mid to High Velum. The student teachers contrasted both Wide and Narrow options of AES.

In sum, the literature has indicated positive evidence for using EVM's 13 Figures and Options to represent voice qualities in singing and speaking and positive outcomes for using an integrated approach to voice therapy linking both implicit and explicit training. Future research is warranted to compare the integrated approach with implicit only and explicit only across varying types of voice disorders. The following section describes linking auditory-perceptual prompts (implicit) with the anatomy and physiology of the voice production system (explicit) using EVM in the RTSS framework with case studies and video examples.

APPLICATION

The concept of connecting auditory-perceptual voice prompts to anatomy and physiology is embedded in the core principles of EVM with Jo Estill's central question, "how am I doing this?" as she sang for audiences around the world. Her question sparked a career move from performance to voice research and teaching. As she studied the anatomy and physiology of the voice production system and the principles of power, source, and filter, she began to formulate answers to her question which led to the creation of EVM. From the beginning, EVM used implicit auditory-perceptual qualities in singing (eg, falsetto, opera, belt, twang) and applied Jo's central question (eg, how am I producing falsetto, opera, belt, and twang). The question was answered by explicit references to 13 anatomical structures of the voice production system called Figures with physiological options (Tables 1 and 2). For example, the AES Figure is an anatomical larynx structure that connects the right and left arytenoid cartilages to the epiglottis. AES Figure has two options, Wide and Narrow. Wide AES is produced with a large gap between the arytenoids and the epiglottis represented by an open or wide hand gesture between thumb and fingers that mimics what is hypothesized to be occurring in the voice system (Table 1). Wide AES corresponds to an implicit auditory-perceptual prompt of a yawn-sigh and the sob voice quality. In contrast, Narrow AES is produced with the arytenoids and epiglottis close together represented by a narrow or closed hand gesture between thumb and fingers that mimics the physiological voice pattern (Table 1). Narrow AES corresponds to an implicit auditory-perceptual prompt of a bright, piercing, or extreme forward resonant voice heard in twang. Only recently have the terms implicit, explicit, and an integrated implicit-explicit approach been used to describe the varying instructional approaches for eliciting target speech/voice behaviors.[9,10] Even before these recent terms, EVM's core principle connected various auditory-perceptual voice prompts (implicit) with the corresponding anatomy and physiology of the voice system (explicit). In EVM training, the implicit prompts for voice quality are tied to the isolated practice of the explicit anatomical Figure with Options for an integrated implicit-explicit approach.

EVM was originally created from Jo's central question related to singing, but the model also has applications to injured voices, speaking voice modifications or enhancements, and prevention and treatment of speaking voice disorders. After all, voice shares the same anatomical structures (Figures) and physiological options regardless of the type of utterance. EVM Figures and Options present anatomy and physiology concepts into simple, manageable exercises accompanied by hand gestures that represent the internal workings of the voice system that are not easily visualized. Most importantly, the anatomy information is immediately mapped to what the body is doing to explain the why and the how. Related to singing, EVM has been used to study and train voice quality control from classical to contemporary commercial singers.[54] Related to speaking, the literature has indicated success in training AES Narrow to achieve twang for patients with hypophonia,[13] facilitating voice therapy for patients with voice disorders, [9,10,14] teaching four new voices for vocally healthy student teachers,[16] training Velum and its extreme options of Low and High,[11] distinguishing TVF: Body-Cover Figure and its four options by acoustic and aerodynamic measures,[15] and release of FVF constriction, TVF: Body-Cover Thick, and Larynx Low for optimal acoustic output.[37]

EVM is taught in Estill Foundations, and level 1 and 2 courses are offered through Estill Voice Training (EVT). EVT is an American Speech-Language-Hearing Association continuing education provider that applies the principles of EVM to voice training and treatment. Clinicians, researchers, students, and educators may consider attending an EVT course to learn more about EVM. An alternative would be to explore the existing literature on EVM with careful consideration of how to apply the model in practice. EVM can be applied to any auditory-perceptual voice prompt or quality used in singing, speaking, and speech-language pathology behavioral voice therapy approaches. For example, resonant voice, loud voice, breathy voice, to name a few, can all be defined and trained through the Estill Figures and Options.

When someone connects auditory-perceptual prompts (implicit) to anatomy and physiology of voice (explicit), they are attempting to learn how to use their voice in a new

way through practice, which primarily maps onto the Skills and Habits treatment group in the RTSS framework (Figure 1). For example, in an integrated implicit-explicit approach, the implicit prompt of humming in resonant voice can be paired with explicit instruction of TVF: Body-Cover Figure with the Thin depth of contact option. In addition to its application to Skills and Habits treatment components, auditory-perceptual prompts tied to anatomy and physiology may be used as part of Representations treatment components. For example, to increase a client's knowledge about a particular topic as a *pedagogical* target (eg, defining twang voice quality), the clinician may use the Estill Figures and Options as an *informational* ingredient (eg, providing information about AES Narrow to define twang).

In terms of targets, TVF:Onset/Offset Figure maps onto the *glottal onset* RTSS-Voice target and the Estill Quality of twang coincides with the *resonance* RTSS-Voice target.[23] For ingredients, training with Estill's 13 Figures primarily involves the RTSS-Voice ingredient of *opportunities to practice voicing*.[23] The Figures of Head & Neck and Torso Anchor align with the *opportunities to practice alignment/posture* RTSS-Voice ingredient.[23] Additionally, the Estill Figures and Qualities may be used as *volition* ingredients. Volition ingredients are introduced to increase a client's Capability, Opportunity, and Motivation to perform a Behavior.[22,55] For example, the use of an integrated implicit-explicit approach could be introduced as a volition ingredient. When this approach is used, a clinician theorizes that adding explicit instruction to the implicit model will increase the client's capability for accurate performance of the voicing practice. A clinician could also theorize that knowledge about the Estill Figures may contribute to motivating the client to practice. For example, if the client knows more about the physiologic rationale behind the vocal exercises being practiced, this could motivate them to adhere to practice recommendations.

Overall, most Estill Figures appear to align with the mechanism(s) of action (MoA) (ie, the hypothesized or observed ways the ingredients affect a target). While Estill Figure Options can be observed through laryngeal imaging, clinicians do not always complete laryngeal imaging before and after voice therapy and often conclude that voice therapy was successful without a post-therapy laryngoscopy. Rather, they often rely on client reports and auditory perception to measure change in client functioning (ie, the RTSS target). While the primary MoA for a Skills and Habits treatment component is learning by doing in the RTSS-Voice, the Estill Figures further describe what is occurring or hypothesized to be occurring within the laryngeal, vocal tract, and support structures to lead to the desired change in client functioning. This is consistent with EVM's goal of connecting perception with anatomy and physiology.

Clinical examples of how to apply the concept of connecting auditory-perceptual voice prompts used in voice therapy to anatomy and physiology will be provided in the next section highlighting implicit training and an integrated implicit-explicit approach with EVM. The examples will be linked to Skills and Habits and Representations treatment components of the RTSS through targets, ingredients, and MoAs. Of note, goals will be included within the case study descriptions, but the concepts of short-term and long-term goals are not synonymous with terms defined within RTSS (eg, target, ingredient, MoA, etc). Goals define a desired behavior under a certain condition measured by an acceptable criterion within a given period of time, while the RTSS attempts to make a direct and causal connection between clinician actions (ingredients) and a subsequent change in client functioning (target).

The following hypothetical cases were created to represent what is typically seen in the real world with clients who report voice problems demonstrating the use of an integrated implicit-explicit approach within the RTSS framework. In the three cases, note that there are hypothetical, simulated video demonstrations with two of the authors playing the role of client and clinician. All cases have a presumed normal neurological system. Research into the application of the integrated implicit-explicit approach for specific clinical populations is still ongoing. As some aspects of motor learning have been shown to differ in populations with neurological disorders, such as Parkinson's disease,[56–58] this approach may not be suitable for patients with all types of voice disorders.

Case 1—Mary

Mary Smith is a 50-year-old cisgender female who was diagnosed with prenodular swelling and MTD. She is an administrative assistant in a doctor's office and reports that her voice progressively worsens throughout the day. By the end of the work week, she frequently loses her voice. Auditory-perceptual evaluation using the CAPE-V[59] revealed moderate strain and mild roughness. Mary was stimulable for improved voice quality with the humming facilitator and cues to focus on forward resonance.

Table 3 provides an overview of one of the treatment components used during Mary's course of voice therapy. An *increase in forward resonance* (Skills and Habits) target was addressed by providing the client with *opportunities to practice voicing* (ingredient) with increased forward resonance in a bottom-up speech hierarchy. Additionally, the ingredients *provide volition ingredients* and *provide feedback* were also introduced, as these ingredients, when present within a Skills and Habits treatment component, tell the client what to practice (volition) and how to modify their practice error (feedback). The MoA was *learning by doing*.[17] EVM was part of the MoA and connected the auditory-perceptual prompt (humming) with anatomy and physiology of the voice system (Estill Figures) to facilitate increased forward resonance.

In the case of Mary, the clinician decided to begin the treatment component from an implicit approach, requiring that the client imitate the clinician's model with no explicit training. Even with the implicit approach, the clinician is

TABLE 3.
The Humming Prompt Linked to the Anatomy and Physiology of the Voice System through Ingredients and Mechanisms of Action of the Rehabilitation Treatment Specific System for One Treatment Component Used with Mary Smith

Target: Client Function Directly Changed by Ingredients	Ingredient(s): Clinician Actions	Mechanism(s) of Action: Ways in Which the Treatment is Expected to Work	Dose: Amount of Ingredients
Increase forward resonance as measured perceptually by clinician and client judgment on a scale from 0%–100% accuracy.	Provide opportunities to practice voicing (a) Resonant voice (b) In a bottom-up speech hierarchy of hum, facilitator phrase (hum-hum, my mom makes muffins), and automatic speech (counting 1–5) and continuum of loudness difficulty from quiet resonant voice work, breathy resonant voice work, to louder resonant voice work.	Learning by doing: Clinician hypothesizes that "humming quietly" links to the following anatomy and physiology of voice: True Vocal Folds (TVF): Onset/Offset Smooth, TVF: Body-Cover Thin, Thyroid Tilt, False Vocal Folds (FVF) Retract, Head & Neck (H&N) Anchor See Client Mary - quiet hum video example in Appendix A	Practice -Number of repetitions until producing increased resonance in 90% of trials.
	Provide feedback (a) Client and clinician delivered on a scale from 0%–100% accuracy after practice trials. (b) Verbal	Clinician hypothesizes that "breathy humming" links to the following anatomy and physiology of voice: TVF: Onset/Offset Aspirate, TVF: Body-Cover Stiff, Thyroid Tilt, FVF Retract, H&N Anchor See Client Mary - breathy hum video example in Appendix A	Feedback -Client and clinician rating occurs after 10 trials
	Provide volition ingredients (1) Increase motivation to practice by discussing how the voices being introduced and practiced are different options that the client can choose from to meet their daily vocal needs. (2) Increase capability to accurately practice by providing a clinician implicit model of a forward resonant voice (quiet hum, breathy hum, and louder hum) within hum, facilitator phrase, and automatic speech. (3) If the client does not meet 90% accuracy with imitating the implicit model in (2), then the clinician may use an integrated implicit-explicit approach to increase the client's capability to accurately practice. This would include (a) reviewing anatomy and physiology of the relevant Estill Figure(s) by showing a video or picture, in this example, False Vocal Folds (FVF). (b) reviewing the Estill hand gestures for each relevant Figure with Options, in this example, FVF. (c) showing the change in the harmonics of a	Clinicians hypothesizes that "louder humming" links to the following anatomy and physiology of voice: TVF Onset/Offset Smooth, TVF: Body-Cover Thick, Thyroid Tilt, FVF Retract, H&N Anchor	Volition -As needed, until client practices correctly.

TABLE 3 (*Continued*)

Target: Client Function Directly Changed by Ingredients	Ingredient(s): Clinician Actions	Mechanism(s) of Action: Ways in Which the Treatment is Expected to Work	Dose: Amount of Ingredients
	narrowband spectrogram for the relevant Figures and Options, in this example, FVF. (d) modeling the relevant Estill Figures and Options, in this example FVF, in non-speech (grunting, breathing, and laughing) and speech tasks (vowels, oh-you, moh-you). (e) asking the client to imitate the clinician's model as outlined in (d). After client completes (e), then return to opportunities to practice voicing first described in this column.	In this Client Mary - louder hum video example in Appendix A, the client has difficulty imitating the louder hum from the implicit model. The clinician explicitly trains FVF Figure.	

still using anatomy and physiology within the MoA to guide the implicit model of the "hum" due to many options of production (Table 3). For example, the hum may be quiet or breathy, which is consistent with previous research that showed that voicing with forward resonance leads to barely abducted/adducted TVFs[21] with a hypothesized connection to the Estill Figure of TVF: Body-Cover with the options of Thin or Stiff. However, if the resonant voice is carried over into something louder, then it is hypothesized that the TVF: Body-Cover option would be Thick. Defining the hum by anatomy and physiology of the system (ie, Estill Figures) reduces confusion for both the clinician and client and connects directly to what may be changing in the voice system. For example, the quiet hum may be modeled with TVF: Body-Cover Thin, whereas the breathy hum may be modeled with TVF: Body-Cover Stiff. The client can decide if the quiet or breathy hum is preferred for one-on-one conversation or talking confidentially, giving them the freedom of choosing the voice that meets a specific vocal demand. In addition, if a client has difficulty imitating the resonant voice from the clinician's implicit model (eg, quiet hum, breathy hum, or louder hum), then the clinician has the option of adding an integrated approach by continuing to use the implicit model of the hum with the addition of explicit training of the relevant anatomy and physiology (ie, Estill Figures and Options).

For Mary, who can correctly imitate from the clinician's implicit model for both the quiet and breathy hum as outlined in the following goal, from the clinician's implicit model of resonant voice, the client will produce the resonant voice in humming, facilitator phrase, and automatic speech with 90% accuracy as determined by the clinician and client judgment on a scale from 0%–100%. However, as Mary attempts to imitate the louder hum, she cannot produce it with 90% accuracy. Mary's production is effortful and strained. The clinician hypothesizes that Mary is inappropriately recruiting FVF Constrict when attempting to produce the louder hum. The clinician decides to move to an integrated implicit-explicit training model continuing to use the louder hum prompt with the addition of specific anatomy and physiology training of FVF Figure as outlined in the following goal, the client will produce FVF Constrict, Mid, and Retract in non-speech and speech tasks with 90% accuracy as determined by client and clinician judgment on a scale from 0%–100%. If the clinician stayed with the implicit-only approach, then there is a possibility that Mary would not have correctly produced the louder hum staying only with the breathy and quiet hum. With only the quiet and breathy hum, Mary may wonder "how can I use this voice when I need to get louder when talking with a group of friends?" "I don't see how this will work for me." After two sessions, Mary may not return to voice therapy; therefore, the use of an implicit-only approach led to a failed treatment outcome.

Case 2—Jennifer

Jennifer Paul, a 25-year-old cisgender female, was diagnosed with MTD. She is a full-time caregiver for older adults. She complains of extreme effort in her throat when she talks. Jennifer is stimulable to a breathy voice and prefers the breathy voice over other voice options that were trialed during stimulability. From the beginning of the treatment component for Jennifer, the clinician implemented an integrated implicit-explicit approach. The clinician hypothesized that the TVF: Body-Cover Figure may be most responsible for production of the breathy

TABLE 4.
The Breathy Voice Quality Linked to the Anatomy and Physiology of the Voice System through Targets, Ingredients, and Mechanisms of Action of the Rehabilitation Treatment Specific System for One Treatment Component Used with Jennifer Paul

Target: A Change in Client Function	Ingredient(s): Actions That a Clinician Takes	Mechanism(s) of Action: Direct Effects That the Ingredients Have on a Target	Dose: Quantitative Variations of the Ingredients
Skills and Habits: Increase use of a breathy voice quality as measured perceptually by clinician and client judgment on a scale from 0%–100% accuracy.	Provide opportunities to practice voicing (a) Breathy voice (b) In a bottom-up speech hierarchy (ie, vowels, phrases, sentences).	Learning by doing: Clinician hypothesizes that breathy voice links to the following anatomy and physiology: True Vocal folds (TVF): Body-Cover Stiff	Practice -Number of repetitions until producing increased breathy voice quality in 90% of trials
	Provide feedback (a) Client and clinician delivered on a scale from 0%–100% accuracy after practice trials in the speech hierarchy. (b) Verbal	See Client Jennifer's video example in Appendix A.	Feedback -Client and clinician rating occurs after trials
	Provide volition ingredients (1) The clinician will facilitate an integrated implicit-explicit approach to increase the client's capability to accurately practice. This would include (a) reviewing anatomy and physiology of the relevant Estill Figure by showing a video or picture, in this example, True Vocal Folds (TVF): Body-Cover. (b) reviewing the Estill hand gestures for each relevant Figure with Options, in this example, TVF: Body-Cover. (c) showing the change in the harmonics of a narrowband spectrogram for the relevant Figures and Options, in this example, TVF: Body-Cover. (d) modeling the relevant Estill Figures and Options in speech tasks, in this example, TVF: Body-Cover. (e) asking the client to imitate the clinician's model as outlined in (c) and (d). (2) Increase capability to accurately practice by providing a clinician implicit model integrated with explicit training of a breathy voice quality in a bottom-up speech hierarchy.		Volition -As needed, until client practices correctly.

voice. Beginning of treatment focused on training all the options of TVF: Body-Cover Figure in the following goal, client will produce all options of TVF: Body-Cover Figure (Thick, Thin, Stiff, and Slack) in vowels as measured perceptually by clinician and client judgment on a scale from 0%–100% accuracy. Training included linking the auditory-perceptual voice prompt heard in each option of TVF: Body-Cover Figure with the hypothesized anatomical and physiological changes in body-cover dynamics represented in hand gestures and visual feedback on a narrowband spectrogram. After Jennifer achieved the goal of producing all options for TVF: Body-Cover Figure, the following goal was implemented, the client will produce breathy voice (ie, defined as TVF: Body-Cover Stiff) following the implicit model from the clinician and explicit training of the Figure in a bottom-up speech hierarchy with 90% accuracy as determined by clinician and client judgment on a scale from 0%–100%. Recall that Jennifer is stimulable to breathy

TABLE 5.
The Loud Voice Prompt Linked to the Anatomy and Physiology of the Voice System through Mechanisms of Action of the Rehabilitation Treatment Specific System for One Treatment Component Used with Sally Brooke

Target: A Change in Client Function	Ingredient(s): Actions That a Clinician Takes	Mechanism(s) of Action: Direct Effects that the Ingredients Have on a Target	Dose: Quantitative Variations of the Ingredients
Skills and Habits: Increase loudness as measured perceptually by clinician and client judgment on a scale from 0%–100% accuracy.	Provide opportunities to practice voicing (a) Loud voice (b) In a bottom-up speech hierarchy (ie, vowels, facilitator phrase, automatic speech, phrases, sentences, memorized speech, specific spontaneous speech acts, monologue, and conversation).	Learning by doing: Clinician hypothesizes that loud voice links to the following anatomy and physiology of voice via Belt: True Vocal Folds (TVF): Onset/Offset Glottal, TVF: Body-Cover Thick, Aryepiglottic Sphincter (AES) Narrow, Cricoid Tilt, Head & Neck (H&N) Anchor, Torso Anchor See Client Sally loud voice (belt) video example in Appendix A for vowels, facilitator phrase, and phrases	Practice -Number of repetitions until producing increased loudness in 90% of trials
	Provide feedback (a) Client and clinician delivered on a scale from 0%–100% accuracy after practice trials. (b) Verbal	Clinician hypothesizes that loud voice links to the following anatomy and physiology of voice via Twang: TVF: Body-Cover Thin or Thick, AES Narrow, Velum Mid or High See Client Sally loud voice (twang) video example in Appendix A for sentences and memorized speech acts	Feedback -Client and clinician rating after trials
	Provide volition ingredients (1) Increase motivation to practice by discussing how the voices being introduced and practiced are different options that the client can choose from to meet their daily vocal needs. (2) Increase capability to accurately practice by providing a clinician implicit model of a loud voice in a bottom-up speech hierarchy.	Clinician hypothesizes that loud voice links to the following anatomy and physiology of voice via: TVF: Onset/Offset Glottal, TVF: Body-Cover	Volition -As needed, until client practices correctly.

TABLE 5 (*Continued*)

Target: A Change in Client Function	Ingredient(s): Actions That a Clinician Takes	Mechanism(s) of Action: Direct Effects that the Ingredients Have on a Target	Dose: Quantitative Variations of the Ingredients
		Thick See Client Sally loud voice video example in Appendix A for memorized speech acts and specific spontaneous speech acts.	

voice and prefers breathy voice over other "new" options; therefore, the clinician hypothesized that TVF: Body-Cover Stiff is primarily responsible for the breathy voice and uses that option in a bottom-up speech hierarchy.

The target is *increasing use of the breathy voice quality* (Skills and Habits) with the ingredient of *opportunities to practice voicing* (Table 4). Just as in Case 1, the ingredients *provide volition ingredients* and *provide feedback* were also

TABLE 6.
Twang Quality Linked to the Anatomy and Physiology of the Voice System through Representations Targets, Ingredients, and Mechanisms of Action of the Rehabilitation Treatment Specific System for One Treatment Component Used with Sally Brooke

Target: A Change in Patient Function	Ingredient(s): Actions That a Clinician Takes	Mechanism(s) of Action: Direct Effects That the Ingredients Have on a Target	Dose: Quantitative Variations of the Ingredients
Representations: Increase the amount of knowledge as measured by the client answering verbal questions from the clinician with 90% accuracy.	Provide information (1) Clinician will provide information about the relevant Figures for twang quality. This would include (a) reviewing anatomy and physiology of the relevant Estill Figure by showing a video, picture, or hand gestures, in this example, False Vocal Folds (FVF), Aryepiglottic Sphincter (AES), and Velum. (b) reviewing the Estill hand gestures for each relevant Figure with Options, in this example, FVF, AES, and Velum. (c) showing the change in the harmonics of a narrowband spectrogram for the relevant Figures with Options, in this example, AES. (d) discussing how the relevant Figures can be combined to produce different Estill qualities, in this example, Twang quality is produced with FVF Retract, AES Narrow, and Velum Mid or High.	Information processing in the brain. See Client Sally - increase knowledge of twang video example in Appendix A.	Information -Number of repetitions until client answers clinician's questions with 90% accuracy.

introduced. The MoA was learning by doing as the Estill Figures connected the auditory-perceptual prompt (breathy voice) with anatomy and physiology of the voice system (TVF: Body-Cover Stiff).

In the video example, Jennifer produces all options of TVF: Body-Cover and carries over TVF: Body-Cover Stiff into a speech hierarchy. It is also apparent that Jennifer consistently produces an "effortful" or "husky" production during Figure option training and breathy voice (TVF: Body-Cover Stiff) in speech. The clinician hypothesizes that the auditory-perceptual terms of "effortful" and "husky" relate back to her underlying diagnosis of MTD with a connection to the anatomy and physiology of FVF Figure Constrict. Jennifer may be inappropriately recruiting constricted FVFs when practicing. If the clinician has access to perform flexible laryngoscopy, then the clinician can test that hypothesis for accuracy. If scoping is not an option, the clinician can still test the hypothesis by pursuing an integrated implicit-explicit training of FVF Figure (Constrict, Mid, Retract) as the next logical ingredient. Jennifer's ability to replace the "effortful" or "husky" voice quality (FVF Constrict) with an "open" voice quality (FVF Retract) provides support for the clinician's hypothesis. The plan would be to have Jennifer produce breathy voice (TVF: Body-Cover Stiff) with an open sound (FVF Retract) in a speech hierarchy.

Case 3—Sally

Sally Brooke, a 35-year-old cisgender female teacher, works with first graders in an elementary school. She needs to talk loudly over the noise of the classroom. Sally is stimulable for loud voice. The perceptual term of loud voice suggests many definitions. Loud voice may be achieved with greater adduction of the TVFs,[30,31] cricoid tilt,[30,31] twang produced with a narrowed epilarynx via the AES,[13,30,31,60] to name a few options. In Table 5, the clinician begins treatment with implicit modeling of the loud voice. Because of the variation in options for being loud, the clinician must be clear on the type of loud voice that will be modeled for the client. EVM was part of the MoA and connected the auditory-perceptual prompt (loud voice) with anatomy and physiology of the voice system (Estill Figures) offering many options for production of the loud voice to meet the client's needs, demonstrating the benefit of using the Estill Figures to inform the clinician's modeling of the prompt. The next likely additional ingredient in the treatment component may involve an integrated approach with the implicit model linked to explicit training of anatomy and physiology of the loud voice(s).

Table 5 provides an overview of one of Sally's treatment targets to *increase loudness* (Skills and Habits) by the ingredient of *providing opportunities to practice voicing*. *Volition* and *feedback* ingredients were also provided. The MoA was *learning by doing* as the clinician implicitly modeled the prompt and the client imitated the model in the following goal, after clinician implicit model of the loud voice, the client will imitate with 90% accuracy at each step of the hierarchy as determined by clinician and client judgment on a scale from 0%–100%. In Table 5, three different loud voice options are modeled in different parts of the speech hierarchy.

As therapy continues, Sally determines that both belt and twang are the best options for producing a loud voice in the classroom. Belt will be used for healthy yelling in phrases to get the students' attention (eg, eyes up, phones down, walking feet, attention up here) and twang will be used to talk over noise of the classroom. At this point in the session, outlined in Table 6, the clinician decided to offer the treatment target of *increase amount of knowledge* (Pedagogical Target-Representations) with the ingredient of *providing information* about how twang quality is produced. The following goal is facilitated, after the clinician provides information about how twang is produced, the client will verbally answer questions posed by the clinician with 90% accuracy.

CONCLUSIONS

The RTSS framework offers the ability to link clinician action with client change in function, providing the opportunity to assess the efficacy and effectiveness of voice therapy concepts through targets, ingredients, and MoAs. An example of a concept involves the connection between auditory-perceptual prompts used in voice therapy with anatomy and physiology of the voice production system. EVM links the voice qualities of speech, falsetto, sob, twang, belt, and opera to the voice production system through 13 Figures with specific options. Each Figure represents an anatomical structure (eg, TVF: Body-Cover, Thyroid, AES, etc) that is physiologically manipulated to produce different options (eg, TVF: Body-Cover Stiff, Thyroid Tilt, AES Narrow). Using the Figures, clinicians may define and instruct typical auditory-perceptual prompts used in voice therapy (eg, humming, loud "aaahhh," beep-beep or quack-quack, breathy "aaahhh") with changes in anatomy (13 Estill Figures) and physiology (Figure Options) for an integrated implicit-explicit instructional approach to voice therapy. The case studies illustrated the concept applied to the RTSS framework.

In the case of Mary, the clinician addressed the ingredients from an implicit approach, modeling the target voices with the expectation that Mary would imitate the model. Even with beginning implicitly, the clinician still used anatomy and physiology (Estill Figures) to guide the MoA of the auditory-perceptual model allowing for more nuanced voice targets. When Mary struggled with imitating the implicit model by recruiting constricted FVFs, then the clinician adapted the ingredients to include an integrated implicit-explicit approach linking the louder resonant voice in "humming" to training of the FVF Figure. With Jennifer, voice therapy initiated with an integrated implicit-explicit approach to training the options of TVF: Body-Cover Figure with the ultimate plan of linking breathy voice with TVF: Body-Cover Stiff in speech tasks. With

Sally, a loud voice was implicitly modeled by the clinician with MoAs that defined loud voice in different ways based on the Estill Figure Options. The clinician's knowledge of the hypothesized changes in the anatomy and physiology of the voice system informed the different voice models; therefore, offering more voice options to meet Sally's various loud voice needs. As therapy progressed, the clinician provided information to Sally on how twang is produced in the voice system using the relevant Estill Figures and Options.

The concept of connecting auditory-perceptual prompts used in voice therapy with anatomy and physiology of the voice production was applied through targets, ingredients, and MoAs of the RTSS framework. Clinicians may link commonly used voice therapy prompts to the Estill Figures and Options allowing for targets defined by anatomy and physiology, an integrated implicit-explicit approach to ingredients, and more nuanced MoAs. Future research should investigate the concept applied to current voice therapy models in the literature for the prevention and treatment of voice disorders.

Data availability

Data sharing is not applicable to this article as no datasets were generated or analyzed during the current study.

Declaration of Competing Interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Elizabeth U. Grillo reports financial support was provided by the National Institutes of Health. Elizabeth U. Grillo reports a relationship with the National Institutes of Health that includes funding grants. The other authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The concept of applying anatomy and physiology to auditory-perceptual prompts used in voice therapy is in the Global Voice Prevention and Therapy Model (GVPTM) with Estill Figures and Qualities. Investigation of the GVPTM with Estill Figures and Qualities has been supported by two National Institute on Deafness and Other Communication Disorders of the National Institutes of Health grants, R15DC014566 and R15DC019954.

Appendix A. Supporting Information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.jvoice.2024.06.025.

References

1. Barrichelo VM, Behlau M. Perceptual identification and acoustic measures of the resonant voice based on "Lessac's Y-Buzz"-a preliminary study with actors. *J Voice.* 2007;21:46–53.
2. Belsky MA, Shelly S, Rothenberger SD, et al. Phonation resistance training exercises (PhoRTE) with and without expiratory muscle strength training (EMST) for patients with presbyphonia: a non-inferiority randomized clinical trial. *J Voice.* 2023;37:398–409.
3. Boone DR, McFarlane SC, Von Berg SL, et al. *The Voice and Voice Therapy*. Boston: Allyn & Bacon; 2005.
4. Gartner-Schmidt J, Gherson S, Hapner ER, et al. The development of conversation training therapy: a concept paper. *J Voice.* 2016;30:563–573.
5. Lessac A. *The Use and Training of the Human Voice: A Practical Approach to Speech and Voice Dynamics*. 3rd ed. New York: Mayfield Publishing Company; 1997.
6. Stemple JC, Roy N, Klaben BK. *Clinical Voice Pathology: Theory and Management*. San Diego, CA: Plural Publishing; 2018.
7. Titze IR, Verdolini Abbott K. *Vocology: The Science and Practice of Voice Habilitation*. Salt Lake City, UT: National Center for Voice and Speech; 2012.
8. Verdolini Abbott K. *Lessac-Madsen Resonant Voice Therapy: Clinician Manual*. San Diego: Plural Publishing; 2008.
9. Tellis C. Integrated implicit-explicit learning approach to voice therapy. *Perspect ASHA Spec Interest Groups.* 2014;24:111–118.
10. Tellis C. New concepts in motor learning and training related to voice rehabilitation. *Perspect ASHA Spec Interest Groups.* 2018;3:56–67.
11. Perta K, Bae Y, Vuolo J, et al. The role of instructions in motor learning of oral versus nasalized speech targets. *J Speech Lang Hear Res.* 2023;66:4398–4413.
12. Steinhauer K, McDonald MM, Estill J. *The Estill Voice Model: Theory & Translation*. Pittsburgh, PA: Estill Voice International; 2017.
13. Lombard LE, Steinhauer KM. A novel treatment for hypophonic voice: twang therapy. *J Voice.* 2007;21:294–299.
14. Madill C, Chacon A, Kirby E, et al. Active ingredients of voice therapy for muscle tension voice disorders: a retrospective data audit. *J Clin Med.* 2021;10:4135.
15. Barone NA, Ludlow CL, Tellis CM. Acoustic and aerodynamic comparisons of voice qualities produced after voice training. *J Voice.* 2021;35:157.e11.
16. Grillo EU. A nonrandomized trial for student teachers of an in-person and telepractice Global Voice Prevention and Therapy Model with Estill Voice Training assessed by the VoiceEvalU8 app. *Am J Speech Lang Pathol.* 2021;30:566–583.
17. Hart T, Dijkers MP, Whyte J, et al. A theory-driven system for the specification of rehabilitation treatments. *Arch Phys Med Rehabil.* 2019;100:172–180.
18. Zanca JM, Turkstra LS, Chen C, et al. Advancing rehabilitation practice through improved specification of interventions. *Arch Phys Med Rehabil.* 2019;100:164–171.
19. Van Stan JH, Dijkers MP, Whyte J, et al. The rehabilitation treatment specification system: implications for improvements in research design, reporting, replication, and synthesis. *Arch Phys Med Rehabil.* 2019;100:146–155.
20. Whyte J, Dijkers MP, Hart T, et al. The importance of voluntary behavior in rehabilitation treatment and outcomes. *Arch Phys Med Rehabil.* 2019;100:156–163.
21. Verdolini K, Druker DG, Palmer PM, et al. Laryngeal adduction in resonant voice. *J Voice.* 1998;12:315–327.
22. Van Stan JH, Whyte J, Duffy JR, et al. Rehabilitation treatment specification system: methodology to identify and describe unique targets and ingredients. *Arch Phys Med Rehabil.* 2021;102:521–531.
23. Van Stan JH, Whyte J, Duffy JR, et al. Voice therapy according to the rehabilitation treatment specification system: expert consensus ingredients and targets. *Am J Speech Lang Pathol.* 2021;30:2169–2201.

24. Helou LB, Gartner-Schmidt JL, Hapner ER, et al. Mapping metatherapy in voice interventions onto the rehabilitation treatment specification system. *Semin Speech Lang.* 2021;42:5–18.
25. Wolfberg J, Whyte J, Doyle P, et al. Rehabilitation Treatment Specification System: application to everyday clinical care. *Am J Speech Lang Pathol.* 2024;33:814–830. https://doi.org/10.1044/2023_ajslp-23-00283.
26. Preston J, Brick N, Landi N. Ultrasound biofeedback treatment for persisting childhood apraxia of speech. *Am J Speech Lang Pathol.* 2013;22:627–643.
27. Sugden E, Lloyd S, Lam J, et al. Systematic review of ultrasound biofeedback in intervention for speech sound disorders. *Int J Lang Commun Disord.* 2019;54:705–728.
28. Macrae P, Anderson C, Taylor-Kamara I, et al. The effects of feedback on volitional manipulation of airway protection during swallowing. *J Motor Behav.* 2014;46:133–139.
29. Azola S, Sunday K, Humbert I. Kinematic biofeedback improves accuracy of learning a swallowing maneuver and accuracy of clinician cues during training. *Dysphagia.* 2017;32:115–122.
30. Klimek MM, Obert KB, Steinhauer K. *The Estill Voice Training System, Level One: Compulsory Figures for Voice Control*. Pittsburgh, PA: Think Voice International, LLC; 2010.
31. Klimek MM, Obert KB, Steinhauer K. *The Estill Voice Training System, Level Two: Figure Combinations for Six Voice Qualities*. Pittsburgh, PA: Think Voice International, LLC; 2010.
32. Steinhauer K, McDonald Klimek M. Vocal traditions: Estill Voice Training®. *Voice Speech Rev.* 2019;13:354–359. https://doi.org/10.1080/23268263.2019.1605707.
33. Dabirmoghaddam P, Aghajanzadeh M, Erfanian R, et al. Comparative study of increased supraglottic activity in normal individuals and those with muscle tension dysphonia (MTD). *J Voice.* 2019;35:554–558.
34. Meerschman I, Van Lierde K, D'haeseleer E, et al. Immediate and short-term effects of straw phonation in air or water on vocal fold vibration and supraglottic activity of adult patients with voice disorders visualized with strobovideolaryngoscopy: a pilot study. *J Voice.* 2024;38:392–403.
35. Haddad R, Ismail S, Khalaf M, et al. Lipoinjection for unilateral vocal fold paralysis treatment: a systematic review and meta-analysis. *Laryngoscope.* 2022;132:1630–1640.
36. Steinhauer K, Grayhack JP, Smiley-Oyen AL, et al. The relationship among voice onset, voice quality, and fundamental frequency: a dynamical perspective. *J Voice.* 2004;18:432–442. https://doi.org/10.1016/J.JVOICE.2004.01.006.
37. Madill C, Nguyen DD. Impact of instructed laryngeal manipulation on acoustic measures of voice–preliminary results. *J Voice.* 2023;37:143.e1.
38. Aaen M, McGlashan J, Sadolin C. Investigating laryngeal "tilt" on same-pitch phonation – preliminary findings of vocal mode metal and density parameters as alternatives to cricothyroid-thyroarytenoid "mix". *J Voice.* 2019;33:806.e9–806.e21.
39. Hong YT, Hong KH, Jun JP, et al. The effects of dynamic laryngeal movements on pitch control. *Am J Otolaryngol.* 2015;36:660–665.
40. Echternach M, Popeil L, Traser L, et al. Vocal tract shapes in different singing functions used in musical theater singing: a pilot study. *J Voice.* 2014;28:653.e1–653.e7.
41. Sundberg J, Gramming P, LoVetri J. Comparison of pharynx, source, formant, and pressure characteristics in operatic and musical theater singing. *J Voice.* 1993;7:301–310.
42. Perta K, Bae Y, Obert K. A pilot investigation of twang quality using magnetic resonance imaging. *Logop Phoniatr Vocol.* 2021;46:77–85.
43. Titze I, Story B. Acoustic interaction of the voice source with the lower vocal tract. *J Acoust Soc Am.* 1997;101:2234–2243.
44. Yanagisawa R, Estill J, Kmucha ST, et al. The contribution of aryepiglottic constriction to "ringing" voice quality—a videolaryngoscopic study. *J Voice.* 1989;3:342–350.
45. Titze IR. Acoustic interpretation of resonant voice. *J Voice.* 2001;28:147–155.
46. Steinhauer K, Grayhack JP, Smiley-Oyen AL, et al. The relationship among voice onset, voice quality, and fundamental frequency: a dynamical perspective. *J Voice.* 2004;18:432–442. https://doi.org/10.1016/J.JVOICE.2004.01.006.
47. Fric M, Dobrovolna´ A, Amarante Andrad P. Comparison of laryngoscopic, glottal and vibratory parameters among Estill qualities – case study. *Biomed Signal Process Control.* 2024;37:1–19.
48. Martínez-Arellano A, Campo A, Del Rio B, et al. Describing the acoustic and vocal production characteristics of the irrintzi: feasibility of its use for the treatment of voice disorders. *J Speech Lang Hear Res.* 2022;65:3789–3797.
49. Steinhauer K, Eichhorn K. Effect of practice structure and feedback frequency on voice motor learning in older adults. *J Voice.* 2023;18:432–442. https://doi.org/10.1016/j.jvoice.2023.04.006.
50. Freedman S, Maas E, Caligiuri, et al. Internal versus external: oral motor performance as a function of attentional focus. *J Speech Lang Hear Res.* 2007;50:131–136.
51. Lisman A, Sadagopan N. Focus of attention and speech motor performance. *J Commun Disord.* 2013;46:281–293.
52. Lo E, Wong A, Tse A, et al. Effects of error experience on learning to lower speech nasalance level. *Am J Speech Lang Pathol.* 2019;28:448–456.
53. Wong A, Tse A, Ma E, et al. Effects of error experience when learning to simulate hypernasality. *J Speech Lang Hear Res.* 2013;56:1764–1773.
54. Fantini M, Fussi F, Crosetti E, et al. Estill Voice Training and voice quality control in contemporary commercial singing: an exploratory study. *Logop Phoniatr Vocol.* 2017;42:146–152. https://doi.org/10.1080/14015439.2016.1237543.
55. Michie S, van Stralen MM, West R. The behaviour change wheel: a new method for characterising and designing behaviour change interventions. *Implement Sci.* 2011;6:42. https://doi.org/10.1186/1748-5908-6-42.
56. Whitfield J, Goberman A. Speech motor sequence learning: acquisition and retention in Parkinson disease and normal aging. *J Speech Lang Hear Res.* 2017;60:1477–1492.
57. Whitfield J, Goberman A. Speech motor sequence learning: effect of Parkinson disease and normal aging on dual-task performance. *J Speech Lang Hear Res.* 2017;60:1752–1765.
58. Wu T, Chan P, Hallett M. Effective connectivity of neural networks in automatic movements in Parkinson's disease. *NeuroImage.* 2010;49:2581–2587.
59. Kempster GB, Gerratt BR, Abbott KV, et al. Consensus auditory-perceptual evaluation of voice: development of a standardized clinical protocol. *Am J Speech Lang Pathol.* 2009;18:124–132.
60. Lott J. The use of the twang technique in voice therapy. *Perspect Voice Voice Disord.* 2014;24:119–123. https://doi.org/10.1044/VVD24.3.119.
61. Jacobson BH, Johnson A, Grywalski C, et al. The Voice Handicap Index (VHI): development and validation. *Am J Speech Lang Pathol.* 1997;6:66–70. https://doi.org/10.1044/1058-0360.0603.66.