Real-Time Resonance Biofeedback for Gender-Affirming Voice Training: Usability Testing of the *TruVox* Web-Based Application[☆]

[*]Tara McAllister, [*]Collin Eagen, [†]Victoria S. McKenna, [‡]Yi Shan, [‡]Peter Traver, [*]Jayce Garner, [§]Daphna Harel, [¶]Tae Hong Park, and [||]Vesna D. Novak, *‡§*New York, New York,* †*Orlando, Florida,* ¶*West Lafayette, Indiana, and* ||*Cincinnati, Ohio*

SUMMARY: Objectives. Gender-affirming voice training (GAVT) can reduce voice dysphoria for some transgender/ gender-diverse individuals, but access is often limited by factors such as high cost and a shortage of well-trained providers. Interactive software could increase the accessibility of GAVT, but existing applications are limited in functionality, particularly in the area of resonance. This paper reports the design and usability testing of a web-based software module that provides visual-acoustic biofeedback for vocal tract resonance. This preliminary study focuses on the use of the software for voice feminization.

Materials and Methods. The resonance module, part of a larger project to develop free GAVT software called *TruVox*, represents resonant frequencies of the vocal tract as peaks in a real-time linear predictive coding spectrum. It features built-in resonance targets for English vowel sounds, a routine for practicing resonance modification in utterances of increasing length, and video tutorials explaining how to use the website. This paper reports the results of both online and in-lab usability testing of the software. After an unstructured (online) or structured (in-person) interaction with the software, participants completed the system usability scale (SUS) and provided qualitative feedback. Participants in the in-lab study additionally completed the intrinsic motivation inventory and produced vowel targets in one-syllable words for acoustic measurement at the start and end of the session.

Results. Participants who completed in-lab testing rated the software with an average SUS score of 75.25, indicating acceptable-to-good usability. In their qualitative feedback, users reported that the visual-acoustic biofeedback helped them understand and control their vocal resonance; they also identified a number of areas where usability could be improved. While little acoustic change was expected in light of the short duration of the session, 8/10 participants showed a trend in the expected direction for the vowel that was practiced most extensively.

Conclusion. The results of this usability testing suggest that our web-based biofeedback tool has the potential to enhance resonance training in the GAVT context. Future goals to improve the app include refining our automated resonance tracking algorithm for more interpretable feedback, allowing users to set individualized resonance targets based on their own voice at baseline, and adding targets for nonbinary and transmasculine users. Clinical trials to measure acoustic changes after an extended period of interaction with the software are also planned.

Keywords: Transgender–Gender-diverse–Gender-affirming voice training–Resonance-biofeedback.

INTRODUCTION

Gender-affirming voice training

Transgender and gender-diverse individuals (henceforth referred to as "trans people"), defined here as those whose gender identity differs from their sex assigned at birth, are estimated to represent approximately 1.6% of the U.S. population and up to 5% of young adults.[1] Compared with the general population, trans individuals experience a lower average quality of life.[2] While this can in large part be attributed to systemic discrimination, it also reflects elevated rates of mental health concerns linked to gender dysphoria.[3] Voice dysphoria—a feeling of discomfort arising when a person's voice does not align with their gender identity[4]—can be a significant contributor. Although some trans people seek surgical interventions to alleviate voice-gender incongruence, these procedures are invasive, costly, and limited in the effect they produce on the voice.[5,6] Gender-affirming voice training (GAVT), commonly administered by speech-language pathologists (SLPs), offers a nonsurgical approach for reducing this incongruence.[7] However, GAVT delivered by SLPs is subject to barriers, most notably the high cost of treatment, which often is not covered by insurance.[8,9] In addition, most SLPs receive limited training in the area of GAVT, with the result that few report feeling confident in their ability to deliver

Accepted for publication October 3, 2025.

☆ Research reported in this publication was supported by the National Institute of Deafness and Other Communication Disorders of the National Institutes of Health under award number R21DC021537 (McAllister and Novak, multi-PI), by seed funding from the Music and Audio Research Laboratory at NYU, and by an NYU Steinhardt Dean's Innovation Grant.

From the *Department of Communicative Sciences and Disorders, New York University, New York, New York; †School of Communication Sciences and Disorders, University of Central Florida, Orlando, Florida; ‡Department of Music and Performing Arts Professions, New York University, New York, New York; §Department of Applied Statistics, Social Science, and Humanities, New York University, New York, New York; ¶Department of Music, Purdue University, West Lafayette, Indiana; and the ||Department of Electrical and Computer Engineering, University of Cincinnati, Cincinnati, Ohio.

Address correspondence and reprint requests to Tara McAllister, Department of Communicative Sciences and Disorders, New York University, 665 Broadway, Suite 600, New York, New York 10012. E-mail: tkm214@nyu.edu

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

https://doi.org/10.1016/j.jvoice.2025.10.004

GAVT services.[10–12] Comprehensive preparation would ideally include not only technical expertise in voice, phonetics, and speech acoustics, but also familiarity with gender diversity and the ways communication relates to social identity. A final barrier is the small number of SLPs who themselves identify as gender-diverse, which can lead to a disconnect between stakeholder priorities and clinical practice.[6]

A promising avenue to address these limitations is the development of interactive software that can support users in completing voice training exercises independently or in tandem with clinician-guided sessions. Survey studies have highlighted a strong interest in tools that incorporate visual-acoustic biofeedback.[13] Biofeedback leverages instrumentation to provide real-time visual information about physiological processes that typically fall below the level of conscious control.[14] In biofeedback for speech, a visual target or model of the desired pronunciation is displayed, and users attempt to adjust their vocal output to align with that target while observing real-time feedback of their own performance. Several existing GAVT apps (eg, Voice Tools, EvaF.app, and Christella VoiceUp) feature biofeedback based on fundamental frequency (f_o), which corresponds to the rate of vocal fold vibration and correlates with perceived vocal pitch. However, we are aware of no currently available software providing real-time feedback on vocal tract resonance.[13] This paper presents a new software module designed to fill that gap by delivering real-time resonance biofeedback in the context of GAVT. The module is part of a broader project to develop a comprehensive set of free web-based tools for GAVT called *TruVox*.[15] The software is freely available at https://truvox.bitslabstart.com/transvoice.

Resonance in GAVT

Vocal resonance plays a significant role in how listeners perceive speaker gender, making it a critical component of GAVT.[16] The resonant frequencies of the vocal tract, or formants, differ in relative position for contrasting sounds such as /i/ and /u/. They also correlate with vocal tract length, with longer vocal tracts yielding lower resonant frequencies. Because male vocal tracts are longer on average than female vocal tracts, women's formant frequencies (in studies of speakers who are presumed to be cisgender) tend to be roughly 18% higher.[17] A variety of terms have been used to describe gendered differences in resonance. The higher resonant frequencies commonly linked to feminine-perceived voices are sometimes described as "brighter" or "smaller," while the lower resonant frequencies more often associated with masculine-perceived voices can be described as "darker" or "larger."[18] In this study, we follow a common convention in speech pathology by adopting the terms "brighter" and "darker."[19] In gender-affirming resonance training, speakers attempt to behaviorally modify their resonance in a way that simulates a longer or shorter vocal tract.[20] While trans people who pursue GAVT generally regard resonance as an important part of voice modification, they often express limited understanding of what resonance entails or how it can be intentionally modified.[13]

A prior laboratory study[21] provided a proof-of-concept demonstration that visual-acoustic biofeedback could be leveraged to modify vocal tract resonance in the GAVT context. Twelve trans women were familiarized with a real-time display of resonant frequencies. They were cued to produce the words *bad*, *bod*, and *bud* while attempting to adjust their vocal tract configuration to match a target resonance shifted above their baseline formant frequencies. Vowels produced with visual feedback cueing an upward shift showed significantly higher resonant frequencies than unshifted productions, and these higher formants were associated with higher femininity ratings from blinded listeners on Amazon's Mechanical Turk crowdsourcing platform. However, the Windows-based software used to provide biofeedback in that study is expensive and has limited accessibility to learners without specialized training.

The current study presents the design and initial testing of a browser-based application intended to support resonance training in GAVT. We begin by describing the technical features of the resonance module and outlining the clinical training tasks embedded in it. We then report results of user acceptance testing in which trans women evaluated the system, beginning with an initial online study (n = 6) followed by an in-lab study (n = 10). Our decision to focus on voice feminization in this initial study was motivated by several considerations. It is known that hormone replacement therapy affects voice characteristics in trans men but not in trans women.[22] Thus, while behavioral modification of voice is of interest to gender-diverse individuals of all identities, it may be a particularly high priority for trans women. The high level of interest in GAVT among transfeminine individuals has also been documented in previous survey and chart review studies.[13,23] However, the expansion of this research to encompass transmasculine and nonbinary perspectives represents an important next step.

SOFTWARE PROTOTYPE

Real-time linear predictive coding display

In GAVT resonance training with biofeedback, users view a live visual display of formant frequencies as they speak. This can be displayed as a real-time spectrogram[24] or a real-time linear predictive coding (LPC) spectrum.[25] The present study focuses on the real-time LPC implementation because the scrolling motion of a real-time spectrogram (which has time on the x-axis) can be challenging for users to monitor. LPC spectra are well-suited for visualizing sounds with prominent formant structure, including vowels and vowel-like consonants such as English /ɹ/ and /l/ sounds.[26] The current version of the resonance module is limited to vowel targets. However, vowel-like consonants are also relevant in GAVT[27] and could be added to the software in the future.

FIGURE 1. Screenshot of the *TruVox* resonance module.

Figure 1 displays the LPC spectrum of the *TruVox* resonance module, which is implemented in JavaScript. At the time of writing, it runs only in the Chrome browser. The LPC spectrum has frequency on the x-axis and amplitude on the y-axis; the peaks in the spectrum represent resonant frequencies of the vocal tract. As the user speaks into the device microphone, the display representing the spectral envelope is updated in real time. White lines overlaid on the spectrum represent local maxima identified by a peak-picking algorithm, which flags peaks that exceed a predefined amplitude threshold.

Target Slider

Visible in Figure 1 is a vertical line superimposed over the spectrum. This is an adjustable-frequency slider that serves as the target used in training: participants view their wave while sustaining a vowel sound and modify their vocal tract configuration in an effort to shift their formant peak to align with the slider. Within the framework of TruVox, our emphasis is on the second resonant frequency. In reality, all formant frequencies are relevant to the perception of speaker gender, not just the second formant. However, we judged that it could be challenging for users to monitor multiple frequency peaks simultaneously, and we chose to target a single formant as a means of reducing cognitive load. We selected the second formant because it has been characterized as the most important predictor of perceived vocal femininity.[28–30] While speech scientists refer to the second formant as F2, in our software (and in the rest of this paper), we refer to it as R2 (for "resonance 2"), because the word *formant* is unfamiliar to most individuals without phonetic training.

The target for R2 is always defined in relation to a specific vowel, since R2 differs across vowels—in fact, the magnitude of difference in R2 between vowels can be much larger than the difference between brighter and darker voices. While it is a goal for future work to set R2 in a customized manner based on the user's resonance at baseline, formant estimation in our prototype software was insufficiently stable to yield a confident estimate of baseline R2. (This point is discussed further under "Limitations and Future Directions".) Instead, R2 targets were defined from a reference sample of adult speakers of American English.[31] The speakers were identified as men and women; it was not specified if they were cis- or transgender. Based on pilot testing by author C. Eagen, an SLP with experience providing GAVT, the target was set to 1.5 standard deviations below the mean R2 frequency reported for women in Clopper et al.[31] Means and standard deviations were averaged across the dialect regions represented in the reference sample because it was judged that the additional complexity needed for dialect-specific targets could reduce the usability of the software.

When a vowel is selected, the R2 slider automatically moves to a target frequency based on the reference sample. Users can manually adjust the slider to different frequency values (eg, if they find the built-in target too challenging to attain) and click “save custom target” to store the current slider position locally using cookies. Buttons are provided to toggle between the reference target and the user-defined custom value.

Navigation and help dialog

From any view within the software, users can access three main pages: “Tutorial,” “Setup,” and “Practice.” Additionally, a “Help Navigating the App” button is available. On the Setup page, this help dialog offers extra guidance to explain the available options (eg, “Some vowels, like 'ee' or 'eh', are relatively easier to say with bright resonance, and some vowels, like ‘ah’ or ‘oo’, are more challenging”). For the “Practice” page, the help dialog provides some contextualizing text (eg, “Here you will see the wave and the adjustable slider”) and tips (eg, “Keep in mind that everyone’s voice is different, and the target we suggest as a starting point may not be a perfect fit for you. You should feel free to adjust the slider location to a value that feels comfortable for you.”).

Tutorial

Even trained professionals like SLPs may have a limited understanding of what resonance represents and how it can be controlled, and learners without previous vocal or phonetic training are expected to have minimal knowledge of this topic. Therefore, we chose to create a series of brief video tutorials in which author C. Eagen, drawing on his experience as a trained vocalist and SLP, explains concepts related to resonance in a voiceover while recording a video of the real-time LPC display. The topics covered in the tutorial are summarized in Table 1, and the text of the tutorials can be found in a supplement to this paper.

Setup page

On this page, users can select one or more vowel sounds to target during practice. Because many users may not be familiar with phonetic symbols, the vowels are represented using key words: *seed, hid, head, sad, bud, good, god, lawn,* and *spoon*. These targets are arranged from highest to lowest R2 frequency, based on reference data from

TABLE 1.
Description of Elements Covered in the *TruVox* Resonance Tutorial Videos

Module	Content
What is resonance?	This module explains that resonance is shaped by the size and configuration of the vocal tract above the vocal folds. The narrator sustains different vowel sounds while the corresponding LPC spectrum is displayed. The narrator highlights the second peak, R2, as the primary focus for resonance training and introduces the adjustable slider used to set a target R2 value.
Bright and dark resonance	This module makes the following points using still images of the LPC spectrum: • R2 is further to the right (higher frequency) in brighter/more feminine-perceived voices and further to the left (lower frequency) in darker/more masculine-perceived voices. • However, the position of R2 varies depending on the vowel, and the differences in R2 across vowels are often greater than those between brighter and darker voice qualities.
Modifying resonance	The narrator sustains an “uh” vowel and demonstrates shifting R2 to both a brighter and a darker target frequency.
Pitch versus *resonance*	This module features a split-screen display, with a real-time pitch tracker on one side and the real-time LPC spectrum on the other. The narrator illustrates how f_o and resonance can be adjusted independently of one another.
Strategies for resonance	The narrator demonstrates three articulatory adjustments intended to shift R2 toward a higher frequency: • Retraction of the lips; • Forward placement of the tongue in the oral cavity; • Narrowing the pharyngeal space.
Resonance chains	The narrator demonstrates resonance modification in utterances of increasing length, described in detail below.

Abbreviation: LPC, linear predictive coding.

ARTICLE IN PRESS

American English women reported in Clopper et al.[31] The setup page also includes a prompt asking whether the user's goal is a more feminine, more masculine, or nonbinary voice; however, only the voice feminization option is currently functional. Users can also choose whether the words presented will be blocked by vowels for more consistent practice, or presented in a randomized order to make practice more challenging.

Practice page

As illustrated in Figure 2, this view displays the real-time LPC spectrum alongside an adjustable slider, which defaults to a vowel-specific target frequency as previously described. A text banner prompts users to connect an external microphone (as opposed to using their device's built-in microphone) to enhance the signal-to-noise ratio and minimize the risk of vocal strain. The target vowel is shown both as a text label (eg, "eh") and within a sample word (eg, "text"). In practice mode, users are instructed to sustain the target vowel and visually assess whether their second peak, R2, approaches the slider's position. If it does not, they are encouraged to apply the strategies introduced in the tutorial video to better approximate the target. Once satisfied with their performance on the isolated vowel, users may click a button to "try a different vowel" or proceed to practice the same vowel in a more complex context, as described in the next section.

Resonance chains

Practicing a vowel in isolation is of limited functional utility, as natural speech involves rapid transitions between different sounds. However, using a real-time biofeedback display during connected speech poses a challenge, since the visual target must shift dynamically as the speaker produces different vowels. To address this issue, the TruVox app introduces "resonance chains"—structured sequences ranging from one to five syllables, all containing the same vowel. For instance, the steps in a resonance chain targeting the vowel "uh" might include: (1) *some*, (2) *someone*, (3) *someone was*, (4) *someone was up*, and (5) *someone's mum was up*. As demonstrated in the tutorial,

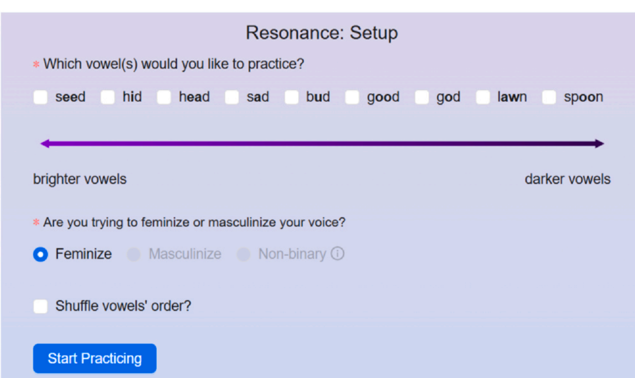

FIGURE 2. Screenshot of the setup page in the TruVox resonance module.

learners are encouraged to move through the chain step by step, elongating the target vowel in each syllable and visually monitoring whether their production aligns with the R2 target. This results in a deliberately exaggerated, chanted style of speech. Once the learner is able to approximate the target frequency at the full chain level, the tutorial recommends repeating the chain with more natural prosody and intonation. The software contains a total of 27 chains (three chains each for nine vowels). Some of the chains are loaded with nasal sounds, which are favored in voice practice by some SLPs because they encourage ease of vocal fold vibration.[32]

MATERIALS AND METHODS

Ethics approval

Study procedures were reviewed and approved by the Institutional Review Board of the University of Cincinnati, protocol #2023-0831; the activities described in this paper were carried out at New York University under a reliance agreement with the University of Cincinnati (NYU IRB protocol FY2024-8952). Participants were recruited through social media posts and targeted outreach to SLPs and researchers who work in GAVT and could disseminate the study flyer to individuals who might be interested in participating. The study flyer described the inclusionary criteria, the time requirement for the study and the activities that would take place, and the rate of compensation ($20 per hour). The flyer also described the eligibility requirements for the study, which specified that participants must be at least 18 years of age, must consider English their native language or one of their native languages, must have been assigned male at birth and identify as transgender, must have normal or corrected-to-normal hearing and vision, must have no significant history of speech, language, or voice disorder, and must have no prior history of voice feminization surgery. Prior to completion of any study activities, participants provided written documentation of informed consent using an electronic consent form implemented in Research Electronic Data Capture (REDCap) software.[33,34]

Online user testing

Participants

Participants in the online usability study had an average age of 28 years (SD = 7.9; range = 18–40 years; one participant did not report their age). Five of the six self-identified as trans women, and a final participant selected "I prefer not to answer." Five participants identified as White and one identified as multiracial (Black and White); no participant reported Hispanic or Latine ethnicity. With respect to voice training experience, two participants reported no previous experience with GAVT. Others reported some GAVT experience (eg, "10 sessions professional vocal therapy"; "multiple gender-affirming voice therapy sessions"; "voice training…for about two months"). Three participants reported training or experience as a singer.

Protocol

Online usability testers completed the testing process independently using their own device. After completing the consent form and background questionnaires, participants were instructed to complete the following steps: (a) watch all six video tutorials; (b) read the "Help navigating the app" dialog; (c) review the Setup page; (d) navigate to the Practice page and interact with the wave while producing sustained vowel sounds; (e) interact with the Practice wave while producing phrase chains. At the end of their participation, participants were prompted to complete the system usability scale (SUS)[35] and completed a questionnaire asking them to comment on strengths and weaknesses of each part of the software (tutorial, setup, help dialog, practice for sustained vowels, and practice with resonance chains). All data capture was carried out using the REDCap system[33,34] hosted by NYU.

In-person user testing

Participants

Following the completion of online user testing and the integration of modifications suggested by those testers, ten different trans women were recruited to participate in a single-session usability evaluation conducted in a laboratory setting. Before scheduling a session, participants were required to complete a REDCap screening survey intended to verify that they met all of the inclusionary criteria detailed above. As in the online phase, all participants provided written informed consent and completed demographic and voice training history forms. A pure-tone hearing screening was administered, assessing detection of tones at 500, 1000, 2000, and 4000 Hz at 20-dB HL.[36] (No hearing screening was conducted for the remote cohort because online self-administration on heterogeneous hardware is unreliable, and because hearing acuity was not central to the study.) Three participants did not detect the 4000-Hz tone in one or both ears, suggesting possible high-frequency hearing loss. However, they were retained in the study, as the R2 frequency targeted in our resonance module generally falls at or below 3000 Hz.[31] The average participant age was 45.4 years (SD = 17; range = 27–72 years). This range included both younger and older adults, which is important given that gender-affirming voice services are also needed by older trans people. Nine participants identified as White, and one identified as Asian. No participant reported Hispanic or Latine ethnicity. Three participants reported no previous experience with GAVT, four reported individual practice with videos or software tools, and three reported some formal GAVT experience (eg, "I began voice training in October 2023 but did not proceed past a few lessons"; "I have just started working with a voice coach.").

Prior to engaging with the software, participants were recorded reading a list of nine words representing the stressed monophthong vowels of English. Each list contained one word per vowel (eg, "seed" for /i/) that was repeated four times in a pseudorandom order. Audio was captured with Praat software[37] using a Shure SM48 tabletop microphone positioned roughly five inches from the participant's mouth. A sampling rate of 44100 Hz and 16-bit encoding were used.

Protocol

The in-person user testing session began with a one-minute vocal warm-up focused on resonance, including humming and sustained phonation. Participants then viewed the tutorial videos titled "What is resonance?" and "Bright and dark resonance." Following this, the experimenter conducted a brief comprehension check in which participants were asked to sustain three vowels and point to the second resonant frequency (R2) on the display. They were also asked to indicate the direction in which they would shift R2 to produce a brighter resonance.

Next, participants viewed the tutorial videos "Modifying resonance" and "Pitch versus resonance." They were then given one minute to interact freely with the resonance feedback display and were encouraged to explore how changes in resonance differ from changes in pitch.

Following this unstructured exploration, participants watched the tutorial titled "Strategies for resonance" and proceeded to a more extended phase of practice using the resonance feedback display with guidance from the first author. The experimenter selected three target vowels ("eh," which is intrinsically fairly bright, "uh," which falls in the middle of the bright-dark range, and "aw", which is intrinsically quite dark) and asked participants to sustain each one. If a participant's neutral R2 peak was already close to the target slider's reference value, the experimenter asked whether they wished to brighten their resonance further. If so, the target was adjusted upward and saved as a custom value. When the participant's R2 peak was below the target, the experimenter provided coaching on strategies (eg, lip retraction, pharyngeal narrowing) to help them approach the target frequency. This guided practice lasted approximately ten minutes.

Participants then viewed the "Resonance chains" tutorial video and engaged in a final phase of guided practice using the resonance chains functionality. If they were not yet ready to practice beyond single words, they were encouraged to complete a chain for comprehension only and then returned to practicing at the vowel or single-word level.

Like the online testers, the in-lab participants completed the SUS and provided qualitative feedback on each aspect of the app. In addition, in-lab participants completed the intrinsic motivation inventory (IMI), a multidimensional self-report instrument to measure participants' subjective experience during an experimental interaction.[38] The version of the IMI adopted here included subscales for interest/enjoyment (eg, "I found the exercise very interesting"), perceived competence (eg, "I think I am pretty good at the exercise"), effort/importance (eg, "I put a lot of effort into the exercise"), and pressure/tension (eg, "I felt very tense during the exercise"). Scores range from 1 to 7; they are averaged across the two items for each subscale that were included in the present version. For the final subscale, pressure/tension, lower scores are interpreted as better; for all others, higher scores indicate a more positive subjective experience. Lastly, participants were re-recorded producing the list of words representing the stressed monophthongs of English. These productions were acoustically measured as described in Section 4.3.

RESULTS

Online user testing

SUS scores

Average SUS scores provided by the online usability testers are shown in Table 2. Across the six participants, the average overall SUS score was 73.75 out of 100 (standard deviation = 15.1, range = 55–100). The questions associated with the lowest usability ratings were "I would imagine that most people would learn to use this website very quickly" and "I needed to learn a lot of things before I could get going with this website." This was consistent with our expectations, since resonance is a challenging concept and users were required to view a series of videos (roughly ten minutes total watch time) before embarking on resonance practice. SUS scores are compared with benchmarks of 70 ("acceptable") and 80 ("good").[39] While the scores provided by online testers indicated room for improvement, the average score exceeded 70, which we treated as the threshold for proceeding to the next phase of testing.

Qualitative feedback

Because there is potential for redundancy between the feedback provided by our two groups of testers, online and in-person, here, we emphasize aspects of online tester feedback that were used to refine the software between the two testing time points, or that were raised by online but not in-person testers.

Like the in-person testers described in more detail below, the online testers generally praised the **tutorial videos** as clear and easy to follow, but they also offered some suggestions for improvement. For example, they suggested exploring methods to make the videos more interactive, such as interleaving informational content with opportunities to practice interacting with the software. In response to this feedback, we added an option to launch the biofeedback wave on the same screen as the tutorial, so users can more easily go back and forth between the two.

Online testers raised two points about the tutorials that helped illuminate ways in which users' preferences might diverge from the perspectives built into the app. One user characterized the training approach embodied in the tutorials as "a Resonant Oral Therapy approach" that is "pretty unpopular in a lot of informal trans voice spaces," adding, "If someone had previous negative exposure to an SLP or had spent a lot of time in these informal spaces, they might be a bit hesitant." We found this critique unexpected because we had specifically avoided adopting cues we associate with resonant voice therapy, such as "feel a buzz in your lips." This is because, while we regard resonant voice therapy as an effective means to promote healthy phonation,[40] we are not aware of evidence that it will reliably raise resonant frequencies. For an extended critique of how resonance modification is commonly handled by the SLP profession from the perspective of a trans voice educator, see Huff.[6] While we have some differences of opinion relative to that author, we are in broad agreement with her assertion that cues for resonance modification in SLP-guided GAVT are often unhelpfully vague. For this reason, we only included resonance-shaping cues that we regard as specific strategies to shorten the effective length of the vocal tract or the oral or pharyngeal cavity.

TABLE 2.
Usability Ratings for Each SUS Item in Both the Online and the In-Person Testing Cohorts

Item	Rating (online)	Rating (In-person)	Polarity
I think that I would like to use this website frequently.	4.0	4.5	Pos
I found the website unnecessarily complex.	2.3	1.7	Neg
I thought the website was easy to use.	4.0	3.7	Pos
I think that I would need the support of a technical person to be able to use this website.	1.8	1.7	Neg
I found the various functions in this website to be well-integrated.	4.5	3.3	Pos
I thought there was too much inconsistency in this website.	2.2	1.8	Neg
I would imagine that most people would learn to use this website very quickly.	3.3	3.3	Pos
I found the website very cumbersome to use.	1.3	1.8	Neg
I felt very confident using the website.	3.8	4.2	Pos
I needed to learn a lot of things before I could get going with this website.	2.5	1.9	Neg

Abbreviation: SUS, system usability scale.
Each item has a minimum score of 1 and a maximum score of 5. Item polarity indicates whether greater usability is associated with a high score (Pos) or a low score (Neg)

The tester's feedback was a helpful reminder that our resonance module still reflects the way GAVT is practiced by the SLP profession, and that it will be important to continue to integrate perspectives from non-SLP stakeholders as we refine our offering.

A second illuminating comment came from a user who identified as Black, who noted, "I'd recommend avoiding terms like 'lightness' and 'darkness,' as they can carry racial connotations, especially when discussing voice characteristics." While we did not act on this suggestion immediately, since the bright versus dark metaphor is embedded throughout the tutorials and the software interface, it served as a useful indication that terminology widely adopted in SLP circles can be received differently by stakeholders from other backgrounds. We return to this point in our discussion of limitations of the present study.

The *Setup* component was generally described as quick and easy to use, and users appreciated the flexibility to adjust and save custom targets and to choose from a selection of target vowels and words. One user noted that she was uncertain if her microphone was detected effectively and suggested adding a microphone check component at the setup stage.

Online testers found the *practice component with sustained vowels* to be reasonably clear and easy to use. They reported that having a well-defined goal combined with real-time feedback helped them make adjustments to their resonance. One user reported having a significant "aha!" moment after some practice. Like the in-person testers described in more detail below, online testers expressed a desire for more tips on how to alter their resonance when they were struggling.

Some users reported difficulty getting the R2 peak or the white lines to appear on the wave graph. However, these users mentioned that they were using their laptop's built-in microphone, which is not recommended. Our immediate response was to add a prominent message encouraging users to use an external microphone in the interest of avoiding vocal strain. In the long term, we are interested in exploring methods for automatic monitoring of voice quality, with the possibility of notifying users of any changes suggestive of possible vocal strain.

Another suggestion was to show a "shadow" display of the full formant structure for a target vowel, as opposed to a slider representing the location of a single peak, R2. Users also expressed interest in a function that would allow them to record their peaks and scroll back to compare them to the target. Integrating these two pieces of feedback, we added a feature that allows users to take a snapshot of their wave (including all formant peaks), which can be superimposed over the live LPC spectrum as a point of comparison. The snapshot feature is represented in Figure 3.

Online testers also had largely positive feedback about the *practice component with resonance chains*, describing this element as "cool and helpful." However, some users found the quick-changing formants to be slightly overwhelming and the waveform hard to track while saying the whole phrase. Suggestions were made for automated detection to show where the second peak was at various points, as well as the possibility to record and get averages of resonance.

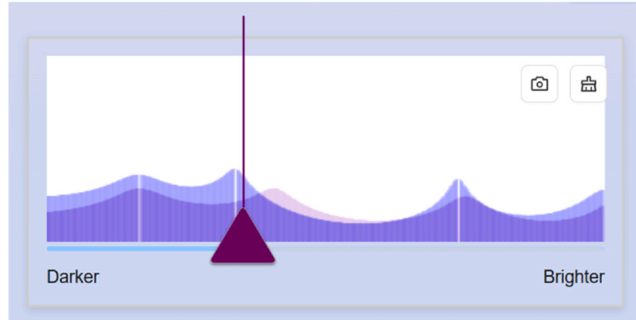

FIGURE 3. Screenshot of the snapshot feature in the *TruVox* resonance module. The darker trace is a freeze-frame created at the time the user hit the camera button; the lighter trace is the live wave, which moves in response to changes in input.

In-person user testing

Results

SUS scores for the in-person testers are shown in Table 2. Across the ten participants, the average overall SUS score was 75.25 out of 100 (standard deviation = 10.7, range = 57.5–87.5). This falls in the acceptable-to-good range and represents a small numerical improvement relative to the average SUS score from the online testing phase. Like the online testers, the in-lab testers gave the software a relatively low usability rating for the item "I would imagine that most people would learn to use this website very quickly." Unlike the online testers, the other lowest-rated item for the in-lab testers was "I found the various functions in this website to be well integrated."

IMI scores are shown in Table 3. These scores suggest that in-lab participants had a positive experience with the

TABLE 3.
Average Participant Ratings for Each IMI Item

Item	Rating	Polarity
I put a lot of effort into the exercise.	6.2	Pos
I think I am pretty good at the exercise.	5.2	Pos
I found the exercise very interesting.	6.7	Pos
I tried as hard as I could during the exercise.	6.5	Pos
I felt very tense during the exercise.	2.7	Neg
I'm satisfied with my performance in the exercise.	5.2	Pos
I enjoyed the exercise very much.	6.1	Pos
I felt pressured during the exercise.	2.2	Neg

Abbreviation: *IMI, intrinsic motivation inventory.*
Each item has a minimum score of 1 and a maximum score of 7. Item polarity indicates whether greater motivation is associated with a high score (Pos) or a low score (Neg).

TABLE 4.
Examples of Positive and Critical Feedback Based on Users' Experience Interacting With the *TruVox* Resonance Module in the Lab Setting

Module	Positive feedback	Critical feedback
Tutorials	"I appreciate that they are short and to the point. They were clear and focused"; "Great delivery by the speaker"; "the person who narrated them had the vocal skill to illustrate the points"	"It might have been useful to make the exercises interactive ie give the user the chance to mimic the videos"; "Sometimes I spaced out while I was listening...Maybe something more repetitive and more instructive ('do this now do this') could have been more helpful?"
"Help Navigating the App"	"clear and effective"; "really helpful"	"I got lost a bit. Example images and formatting might help."
Setup	"The setup was pretty simple"	"The setup is end user friendly but seems a little scattered. I spent a brief period to find the logic of the layout and sequence of the setup."
Practice with Wave	"I absolutely loved this and was fascinated by it. Adding a visual component to voice training helped me put together what I was missing in the training that I already had been doing and having a visual marker to 'hit' during practice made a world of difference."	"I think the user interface needs some work. As someone who sees a lot of numbers and graphs, I could follow it but I think it could be more accessible for folks with less quantitative backgrounds. Maybe change colors as you get to the target?"
Practice with Resonance Chains	"The phase chains were awesome because it reinforced the practice and ultimately got me to a place where I could speak the chain phrases more normally"	" A bit challenging to observe the resonances when other vowels and consonants were pulling the peaks around."

software with respect to interest/enjoyment and effort/importance; both of these scales had an average score of 6.4 out of 7. The scale reporting self-perception of competence was more mixed, with an average score of 5.3 out of 7. Lastly, the average score of 2.45 on the pressure/tension subscale (where lower scores are better) indicated that participants experienced some degree of tension while interacting with the software.

Qualitative feedback

Summary impressions of user feedback are provided in this section. Illustrative quotes from users can be found in Table 4.

Feedback on the *tutorial videos* was overwhelmingly positive, with participants specifically praising the narrator's delivery of the content. Two participants pointed out that the tutorials appeared in a different order on the website than their presentation during usability testing; they will be reordered in response to this feedback. One participant suggested making the tutorials more interactive—for example, by prompting users to pause and experiment with the strategies demonstrated in each video. Although we had modified the layout in response to online user feedback so that the interactive wave display is available to the user on the same page as the tutorial videos, the repeated occurrence of this request indicates that we should explore further strategies to integrate the tutorial videos with practice with the wave.

Users had divided opinions about the *help dialog*. Five users provided mainly positive feedback, while four users characterized the text as overly dense and provided suggestions for reorganization. An updated version of the help page is currently in development.

As in online testing, the majority of users expressed satisfaction with the *setup component*, describing it as simple and easy to use. One user dissented, reporting that she had some difficulty understanding the logic and sequence of the setup. Given that responses on the SUS instrument suggested that users did not perceive the functions of the software as very well-integrated, we will explore options to make the setup process more intuitive.

For the *practice component with sustained vowels*, several users expressed appreciation for the core resonance visualization, which, to our knowledge, is not featured in any other GAVT tools. Users also offered suggestions for enhancement, including an option to record and play back their own voice along with the corresponding wave movement. One user recommended integrating prerecorded model utterances that demonstrate the target resonance for various vowels, which we are in the process of implementing at this time. Another user suggested that the practice routine could feature pop-up text reminding users of the resonance-modification strategies introduced in the tutorial. Finally, one user pointed out that the software is more challenging to use for speakers with rough vocal quality (eg, due to age or damage to the vocal folds), since the peaks of the wave are less clearly visible.

For the *practice component with resonance chains*, five users shared primarily positive feedback. Four users offered suggestions related to the difficulty of tracking resonant peaks for target sounds when alternating with other

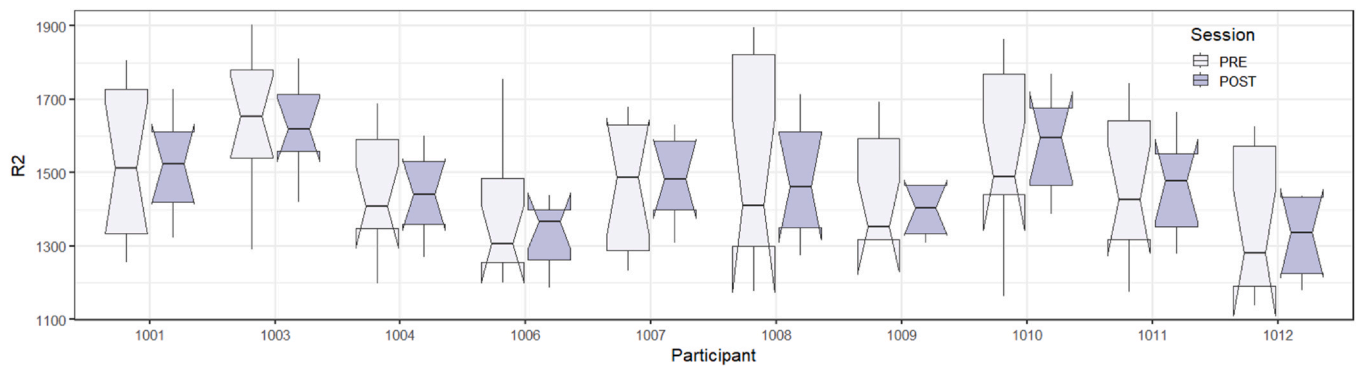

FIGURE 4. Boxplots showing the second resonant frequency (R2) in vowels from single-word utterances produced by in-lab user testers before and after their interaction with the *TruVox* resonance module. A higher R2 is generally associated with a more feminine-perceived voice.

sounds. Recommendations included providing video models for each target chain, adding supplementary visuals and verbal explanations, or using an alternative display with a "fading trailing edge" to help smooth transitions between sounds. One user also observed that some chains included a mix of vowel sounds and recommended further review to correct possible errors; this correction has been implemented.

A final open-ended text field invited participants to share any additional feedback about the software. One participant noted that the resonance tool's compatibility with only the Chrome browser poses a limitation, particularly for users who prefer open-source software. Two participants suggested expanding the software to support more advanced levels of practice. Their comments included a request for an "option to save and display multiple vowels for more complex strings" and an "option of showing two vowel targets with mixed vowel chains—although maybe that gets to be cluttered." While the second participant rightly noted that such complexity might hinder usability, this feature could still be valuable for advanced users. Finally, two users highlighted the need to refine the user interface for broader accessibility. One user (quoted in Table 4) mentioned that her high level of comfort with technology made the tool accessible to her, but she could see less tech-savvy users struggling to understand the wave display. This is a particularly important observation, given that many in-lab testers had backgrounds in technology and engineering—likely a result of outreach through a Discord channel that attracted participants with similar professional profiles.

Acoustic measurement

Using Praat software for acoustic analysis,[37] a trained research assistant measured the formant frequencies of vowels produced in monosyllabic words (*seed, hid, head, sad, bud, good, god, lawn,* and *spoon*) at the start and end of the session. First, the research assistant generated Praat TextGrids demarcating each word in the recorded probe list for each participant. Then the .WAV audio file and TextGrid for each probe list were passed to the Montreal Forced Aligner[41] to identify phoneme boundaries within words. The vowel intervals identified in each textgrid were extracted for acoustic measurement using FastTrack, which generates an ensemble of possible formant tracks across multiple LPC settings and then selects the setting that produces the smoothest overall trajectory.[42] More specifically, FastTrack compares different LPC orders, which are set indirectly in Praat using the Maximum Formant setting (eg, 5 formants at 5000 Hz, 5100 Hz, 5200 Hz, etc). The setting is selected that yields the lowest residual error for a 4th-order polynomial fitted to the first three formants. The selected setting was then used to extract formant values in 5 time bins across the duration of each vowel. No inferential statistics were conducted on the acoustically measured data due to the preliminary nature of this research. However, the measured values were plotted to allow visual inspection of trends across participants, vowels, and time points. Figure 4 depicts R2 measures for the vowel "uh," which participants spent the most time practicing in their structured interaction with the software. In the boxplots, the center line represents the median R2 frequency across four repetitions, and the notch represents a 95% confidence interval around the median value. Out of the ten participants, eight showed a numerical trend toward higher values of R2 at the post training time point (where a higher R2 is associated with more feminine-perceived voices).[28–30,43] While the pilot nature of this investigation does not allow strong conclusions, it does suggest that in the short term, participants have some success in altering their resonant frequencies using the real-time biofeedback display.

Discussion

While the findings of this study are preliminary, they provide a positive indicator that our web-based biofeedback tool has the potential to facilitate GAVT resonance training in the transfeminine context. Participants who completed in-lab testing rated the system with an average SUS score of 75.25, which falls within the range typically interpreted as indicating acceptable-to-good usability. In their qualitative feedback, users noted that the visual-acoustic biofeedback was helpful for increasing their awareness and control of vocal resonance. Substantial acoustic changes were not anticipated, given the very brief duration of training provided; a typical course of SLP-guided GAVT features an average of 12 sessions 45-60 min in duration.[44] Nevertheless, eight out of ten participants exhibited a shift in the expected direction for the vowel that received the most focused practice, suggesting initial responsiveness to the training paradigm. Thus, we view the results of this study as clearing the path for further use and testing of the software. Our immediate next steps are focused on implementing suggestions provided by testers, such as providing improved integration between the content of the tutorial videos and the setup and practice routines. We then intend to measure software usage and changes in voice characteristics in the context of a structured clinical trial.

Limitations and future directions

Increasing representation

A limitation of this study is its exclusive focus on transfeminine participants. An important goal for future work is to add targets and instructions customized for users who identify as transmasculine or nonbinary. In principle, the resonance module's core practice mechanism can be applied equally well to darkening resonance as to brightening it. However, additional pilot testing is needed to establish recommended targets for trans men, as transmasculine voices are influenced by gender-affirming hormone therapy[45] and thus involve different considerations from those of transfeminine voices. For nonbinary users, the same principles could be applied, depending on whether their goal is to shift their voice toward brighter or darker resonance. In general, to flexibly accommodate the needs of users with differing identities, we would like to shift away from our current reliance on reference values from talkers who are presumed to be cisgender. In the long term, we aim to allow users to set R2 targets in a customized fashion. Specifically, we would like to measure the user's R2 at baseline for a range of vowels, then allow them to set a custom target representing their desired degree of shift (eg, "moderately brighter"; "slightly darker") relative to that baseline. However, the success of this customized goal-setting will depend on refining the core formant-tracking algorithm of the software, as we describe in more detail in the next section.

In addition to the question of representation of different gender identities, it is important to consider the representation of individuals from different demographic groups and language backgrounds. Our participants were overwhelmingly White and non-Hispanic/Latine. One participant who identified as Black provided helpful feedback that the "bright" versus "dark" terminology we adopted for resonance targets could have unwelcome connotations for users of color. This example highlights the importance of integrating perspectives of individuals representing different racial and ethnic identities as we continue to develop and test the software. We also acknowledge that the current software is configured only to target the vowels of American English, and it does not accommodate dialectal variation within the American English context. (Recall that we considered adopting dialect-specific vowel targets but judged that the added complexity in the user experience outweighed the gains in target precision.) The same basic interaction could be applied across any language or language variety, particularly if we enable the functionality for participants to set targets relative to their own baseline formants. We hope to have the opportunity to integrate targets for other languages in the future development of the software.

Refining the formant-tracking algorithm

Feedback from our testers made it clear that users would appreciate a biofeedback display that is easier to understand than the current real-time LPC. The multiple peaks and rapidly changing nature of the LPC can be challenging for users to understand and utilize, and our testers (most of whom had a high level of comfort with technology) pointed out that this could present a significant barrier for users who are less tech-savvy. We agree that a simplified visualization of resonance represents an important goal, but it is unlikely to be achieved without substantial changes to the core technology. For example, consider the suggestion we received from one user to make the wave change color as R2 approaches the target frequency. We agree that such a modification could help users better interpret the feedback display, but we have not pursued it at this time because the naive peak-picking algorithm implemented in the current version of the software is insufficiently stable to support this functionality. The LPC algorithm suffers from a known problem of detecting spurious peaks[46,47] or missing peaks that are present in the signal. The magnitude of measurement error introduced by spurious or missed peaks is often large. As a consequence, any visualization that hinges on automatic detection of peaks can create flickering behavior that may confuse users more than it helps them. (In the current implementation of the LPC wave, automatic peak detection is used to superimpose the white lines seen in Figure 1. These lines do indeed exhibit flickering behavior, and users are explicitly instructed to pay more attention to the overall shape of the wave envelope than to the specific presence or absence of a white line.)

State-of-the-art formant detection algorithms, such as KARMA[48] and DeepFormants,[49] have been demonstrated to achieve more reliable detection of formant frequencies across talkers with different vocal tract sizes. However, these algorithms have not been implemented in a real-time context. Members of our team are exploring the possibility of real-time implementation of contemporary formant algorithms, which could unlock alternative visualizations that could make the software more user-friendly. In addition, as noted above, the ability to set individualized targets depends on our ability to accurately measure a user's formants at baseline. Even if it proves intractable to implement state-of-the-art formant detection algorithms in real time, they could still potentially be integrated into an initial setup routine to measure users' baseline resonant frequencies in order to set custom targets.

Measuring outcomes in the clinical context

This work serves as a foundation for clinical trials in which individuals pursuing voice feminization make use of the software over a period of multiple days or weeks. In a preliminary feasibility study,[50] six trans women used the *TruVox* suite of biofeedback tools[15] to work on both pitch and resonance over a three-week period. They were familiarized with *TruVox* in four virtual sessions with an SLP and were instructed to complete daily home practice using the software. While the results were suggestive of improvement in self-perception of voice, as measured by the Utah Gender Presentation Scale for Communication and Self-Efficacy Scale for Voice Modification in Trans Women,[51] the feasibility study is limited by the small sample size and lack of a control condition. Our next step will be to use single-case experimental methodology (eg, multiple-baseline across-subjects design) to measure training-related changes with a control condition to establish internal validity. In addition to self-perception of voice, we aim to obtain acoustic measurements and perceptual ratings of trans women's voices before and after practice with our software.

One caution that has been raised in previous studies of various types of biofeedback[25,52,53] is that gains made in the training setting do not automatically generalize to contexts in which enhanced feedback is not available. If generalization is not immediately achieved, it may be encouraged through a period of follow-up treatment in which task difficulty is adjusted in an adaptive fashion, including expanded opportunities for practice without biofeedback.[54] It will also be important to measure participants' voice satisfaction, voice acoustics, etc at different time points (eg, immediately after training, six weeks after training, and six months after training) in order to understand the durability of any training-related voice changes.

Conclusion

In summary, this study provides early evidence supporting the usability of a browser-based biofeedback tool targeting resonance in the GAVT context. By combining real-time acoustic feedback with structured practice and tutorial support, the software helps users build awareness and control of vocal resonance, which is a critical but challenging component of perceived voice gender. While further technical refinements and broader representation in testing are needed, the usability scores, qualitative feedback, and preliminary acoustic trends observed here point to the promise of this tool. The current version of the software is being made publicly available while the research team pursues its ongoing iterative development and clinical validation.

Declaration of Competing Interest

The authors declare the following financial interests/personal relationships that may be considered as potential competing interests: Tara McAllister is a member of Sonority Labs LLC, a for-profit entity that was created to explore the possibility of commercialization of staRt, an iOS/web application for visual-acoustic biofeedback. The staRt app is related to the TruVox resonance tool reported here. However, the TruVox code base has not been licensed to Sonority Labs LLC, and the authors do not plan to pursue software commercialization in the GAVT context. 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.

Declaration of Generative AI and AI-assisted technologies in the writing process

During the preparation of this work, the authors used GPT-4o for assistance with LaTeX formatting and for minor rephrasing. After using this tool/service, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

Acknowledgments

The authors are grateful to Erin Doty, Yi-No Elizabeth Luan, and Renee Gustin for their contributions to the development of resonance chains, and to Will Haack and Helen Carey for their contributions to software development. A preliminary version of a portion of this work was presented as a four-page proceedings paper at Interspeech 2025. The present paper substantially extends that work by including additional data from an online testing study, acoustic analyses of resonance before and after training, and an expanded discussion.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.jvoice.2025.10.004.

References

1. Brown A. *About 5% of young adults in the u.s. say their gender is different from their sex assigned at birth*. Tech. Rep. Pew Research Institute; 2022.
2. Nobili A, Glazebrook C, Arcelus J. Quality of life of treatment-seeking transgender adults: a systematic review and meta-analysis. *Rev Endocrine Metabol Disord*. 2018;19:199–220.

3. Dhejne C, Van Vlerken R, Heylens G, Arcelus J. Mental health and gender dysphoria: a review of the literature. *Int Rev Psychiatry.* 2016;28:44–57.
4. Hancock AB, Krissinger J, Owen K. Voice perceptions and quality of life of transgender people. *J Voice.* 2011;25:553–558.
5. Song TE, Jiang N. Transgender phonosurgery: a systematic review and meta-analysis. *Otolaryngol Head Neck Surg.* 2017;156:803–808.
6. Huff AZ. Modern responses to traditional pitfalls in gender affirming behavioral voice modification. *Otolaryngol Clin N Am.* 2022;55:727–738.
7. Chadwick KA, Coleman R, Andreadis K, Pitti M, Rameau A. Outcomes of gender-affirming voice and communication modification for transgender individuals. *Laryngoscope.* 2022;132:1615–1621.
8. Gelfer MP. Voice treatment for the male-to-female transgendered client. *Am J Speech Lang Pathol.* 1999;8:201–208.
9. DeVore EK, Gadkaree SK, Richburg K, et al. Coverage for gender-affirming voice surgery and therapy for transgender individuals. *Laryngoscope.* 2021;131:E896–E902.
10. Matthews J-J, Olszewski A, Petereit J. Knowledge, training, and attitudes of students and speech-language pathologists about providing communication services to individuals who are transgender. *Am J Speech Lang Pathol.* 2020;29:597–610.
11. Sawyer J, Perry JL, Dobbins-Scaramelli A. A survey of the awareness of speech services among transgender and transsexual individuals and speech-language pathologists. *Int J Transgenderism.* 2014;15:146–163.
12. Hancock A, Haskin G. Speech-language pathologists' knowledge and attitudes regarding lesbian, gay, bisexual, transgender, and queer (lgbtq) populations. *Am J Speech Lang Pathol.* 2015;24:206–221.
13. Bush EJ, Krueger BI, Cody M, Clapp JD, Novak VD. Considerations for voice and communication training software for transgender and nonbinary people. *J Voice.* 2024;38:1251.e1–1251.e20.
14. Volin RA. A relationship between stimulability and the efficacy of visual biofeedback in the training of a respiratory control task. *Am J Speech Lang Pathol.* 1998;7:81–90.
15. Wang X, Weese S, McAllister T, McKenna V, Novak V. *Design and Usability Evaluation of a Web-Based Pitch Control Training App for Transgender Women.* In: Proceedings of the International Conference on Human-Computer Interaction, 224–234.Springer; 2024.
16. Leung Y, Oates J, Chan SP. Voice, articulation, and prosody contribute to listener perceptions of speaker gender: a systematic review and meta-analysis. *J Speech Lang Hearing Res.* 2018;61:266–297.
17. Hillenbrand JM, Clark MJ. The role of f0 and formant frequencies in distinguishing the voices of men and women. *Attention Perception Psychophys.* 2009;71:1150–1166.
18. Kowalchuk T. W*hat trans* folks talk about when they talk about voice: Learning about voice feminization on reddit*. Master's thesis. Dalhousie University, Halifax, Nova Scotia; 2020. Available at ⟨http://hdl.handle.net/10222/79490⟩.
19. Adler RK, Hirsch S, Pickering J. *Voice and communication therapy for the transgender/gender diverse client: a comprehensive clinical guide.* Plural Publishing; 2018.
20. Leyns C, Papeleu T, Tomassen P, T'Sjoen G, D'haeseleer E. Effects of speech therapy for transgender women: a systematic review. *Int J Transgender Health.* 2021;22:360–380.
21. Kawitzky D, McAllister T. The effect of formant biofeedback on the feminization of voice in transgender women. *J Voice.* 2020;34:53–67.
22. Bultynck C, Pas C, Defreyne J, Cosyns M, Heijer M, T'Sjoen G. Self-perception of voice in transgender persons during cross-sex hormone therapy. *Laryngoscope.* 2017;127:2796–2804.
23. McKenna VS, Gustin RL, Hobek AL, et al. Factors related to treatment attendance for patients seeking gender-affirming voice therapy. *Am J Speech Lang Pathol.* 2023;32:216–233.
24. Shuster LI, Ruscello DM, Toth AR. The use of visual feedback to elicit correct r. *Am J Speech Lang Pathol.* 1995;4:37–44.
25. McAllisterByun T, Hitchcock ER. Investigating the use of traditional and spectral biofeedback approaches to intervention for /r/ misarticulation. *Am J Speech Lang Pathol.* 2012;21:207–221.
26. Hitchcock ER, Ochs LC, Swartz MT, Leece MC, Preston JL, McAllister T. Tutorial: using visual–acoustic biofeedback for speech sound training. *Am J Speech Lang Pathol.* 2023;32:18–36.
27. Hirsch S. Combining voice, speech science and art approaches to resonant challenges in transgender voice and communication training. *Perspect ASHA Spec Int Gr.* 2017;2:74–82.
28. Childers DG, Wu K. Gender recognition from speech. Part II: fine analysis. *J Acoust Soc Am.* 1991;90:1841–1856.
29. Coleman RO. Acoustic correlates of speaker sex identification: Implications for the transsexual voice. *J Sex Res.* 1983;19:293–306.
30. Gelfer MP, Mikos VA. The relative contributions of speaking fundamental frequency and formant frequencies to gender identification based on isolated vowels. *J Voice.* 2005;19:544–554.
31. Clopper CG, Pisoni DB, De Jong K. Acoustic characteristics of the vowel systems of six regional varieties of American English. *J Acoust Soc Am.* 2005;118:1661–1676.
32. Hijleh K, Pinto C. Realizing the benefits of SOVTEs: a reflection on the research. *J Singing.* 2021;77:333–344.
33. Harris P, Taylor R, Minor B, et al. The REDCap consortium: building an international community of software partners. *J Biomed Inform.* 2019;95:103208.
34. Harris PA, Taylor R, Thielke R, et al. A metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform.* 2009;42:377–381.
35. Brooke J. SUS: A 'quick and dirty' usability scale. In: Jordan PW, Thomas B, Weerdmeester BA, McClelland IL, eds. *Usability Evaluation in Industry.* London: Taylor & Francis; 1996:189–194.
36. American Speech-Language-Hearing Association. *Guidelines for Audiologic Screening [Guidelines]*; 1997. Available from the American Speech-Language-Hearing Association ⟨https://www.asha.org/policy/GL1997-00002/⟩.
37. Boersma P, Weenink D. *Praat: doing phonetics by computer*. Version 6.2.10; 2022.
38. Ryan RM. Control and information in the intrapersonal sphere: an extension of cognitive evaluation theory. *J Person Soc Psychol.* 1982;43:450.
39. Lewis JR, Sauro J. Item benchmarks for the system usability scale. *J Usability Stud.* 2018;13:158–167.
40. Yiu EM-L, Lo MC, Barrett EA. A systematic review of resonant voice therapy. *Int J Speech Lang Pathol.* 2017;19:17–29.
41. McAuliffe M, Socolof M, Mihuc S, Wagner M, Sonderegger M. Montreal forced aligner: trainable text-speech alignment using Kaldi. *Interspeech.* 2017;2017:498–502.
42. Barreda S. Fast track: fast (nearly) automatic formant-tracking using Praat. *Ling Vanguard.* 2021;7:20200051.
43. Munson B. The acoustic correlates of perceived masculinity, perceived femininity, and perceived sexual orientation. *Lang Speech.* 2007;50:125–142.
44. Oates J, Södersten M, Quinn S, et al. Gender-affirming voice training for trans women: effectiveness of training on patient-reported outcomes and listener perceptions of voice. *J Speech Lang Hearing Res.* 2023;66:4206–4235.
45. Ziegler A, Henke T, Wiedrick J, Helou LB. Effectiveness of testosterone therapy for masculinizing voice in transgender patients: a meta-analytic review. *Int J Transgenderism.* 2018;19:25–45.
46. Kay SM, Marple SL. Spectrum analysis—a modern perspective. *Proc IEEE.* 1981;69:1380–1419.
47. Alku P, Kadiri SR, Gowda D. Refining a deep learning-based formant tracker using linear prediction methods. *Comp Speech Lang.* 2023;81:101515.

48. Mehta DD, Rudoy D, Wolfe PJ. Kalman-based autoregressive moving average modeling and inference for formant and antiformant tracking. *J Acoust Soc Am*. 2012;132:1732–1746.
49. Dissen Y, Goldberger J, Keshet J. Formant estimation and tracking: a deep learning approach. *J Acoust Soc Am*. 2019;145:642–653.
50. McKenna VS, David K, Wilkens ME, et al. *Multi-session pitch and resonance training with visual-acoustic feedback and structured exercise for vocal feminization*. In: US Professional Association for Transgender Health, Santa Ana Pueblo, NM; 2025.
51. Dacakis G, Erasmus J, Nygren U, Oates J, Quinn S, Södersten M. Development and initial psychometric evaluation of the self-efficacy scale for voice modification in trans women. *J Voice*. 2024;38:1251–e21.
52. Gibbon FE, Paterson L. A survey of speech and language therapists' views on electropalatography therapy outcomes in Scotland. *Child Lang Teaching Ther*. 2006;22:275–292.
53. Preston JL, McCabe P, Rivera-Campos A, Whittle JL, Landry E, Maas E. Ultrasound visual feedback treatment and practice variability for residual speech sound errors. *J Speech Lang Hearing Res*. 2014;57:2102–2115.
54. McAllister T, Hitchcock ER, Ortiz JA. Computer-assisted challenge point intervention for residual speech errors. *Perspect ASHA Spec Int Gr*. 2021;6:214–229.