Interspeech 2025
17-21 August 2025, Rotterdam, The Netherlands

Web-Based Application for Real-Time Biofeedback of Vocal Resonance in Gender-Affirming Voice Training: Design and Usability Evaluation

Tara McAllister[1], *Collin Eagen*[1], *Yi Shan*[2], *Peter Traver*[2], *Daphna Harel*[3], *Tae Hong Park*[4], *Vesna Novak*[5]

[1]Communicative Sciences and Disorders, New York University, USA; [2]Music and Performing Arts Professions, New York University, USA; [3]Applied Statistics, Social Science, and Humanities, New York University, USA; [4]Department of Music, Purdue University, USA; [5]Electrical and Computer Engineering, University of Cincinnati, USA

tkm214@nyu.edu, cme9578@nyu.edu, ys4945@nyu.edu, pdt9929@nyu.edu, daphna.harel@nyu.edu, thp@purdue.edu, novakdn@ucmail.uc.edu

Abstract

Gender-affirming voice training (GAVT) can reduce voice-gender incongruence for transgender and gender-diverse individuals, but access is often limited by high costs and a lack of qualified providers. Interactive software could expand access, but existing GAVT apps are typically limited in functionality. This paper describes the development and testing of software to address one of the most challenging aspects of GAVT: modifying vocal tract resonance. We introduce a biofeedback system that uses a real-time linear predictive coding (LPC) spectrum to visualize changes in resonance as learners adjust their vocal tract shape. Visual targets for brighter (feminine) and darker (masculine) resonances help guide users toward their desired voice characteristics. In-lab user testing with 10 trans women yielded an average System Usability Scale (SUS) score of 75.25, supporting the acceptability of the tool as an adjunct to resonance training in the GAVT context.

Index Terms: biofeedback, human-computer interaction, speech visualization, transgender, gender

1. Introduction

1.1. Gender-affirming voice training

Transgender and gender-diverse individuals (henceforth, "trans people"), defined as those whose gender identity does not align with the one assumed at birth, comprise roughly 1.6% of the U.S. population and may make up as much as 5% of young adults [1]. This group often faces a significantly reduced quality of life compared to the general population [2], a disparity largely attributed to societal and systemic discrimination, as well as poor mental health associated with gender dysphoria [3]. One specific challenge is voice dysphoria, a form of distress stemming from the sensation that one's voice does not align with one's gender identity [4]. While some trans people pursue surgical interventions for voice-gender incongruence, surgery is an expensive and invasive option [5]. Alternatively, gender-affirming voice training (GAVT), often guided by speech-language pathologists, offers a non-invasive way to reduce voice-gender incongruence [6]. However, access to GAVT is constrained by factors such as a shortage of qualified providers [7] and the high cost of care, which requires numerous sessions [8] and is rarely covered by insurance [9].

One potential way to improve access to GAVT is by developing interactive software that guides users through voice training exercises. Such software could complement in-person sessions or support learners who are unable to access a suitable clinician. While a number of smartphone apps have been created for GAVT, their functionality is limited, and they are not widely embraced by trans individuals [10, 11, 12]. Surveys of user needs have identified a particular demand for software that offers visual-acoustic biofeedback [13]. Biofeedback involves using instrumentation to deliver real-time information about a physiological function, providing insights into movement or behavior that may not be easily perceived under ordinary circumstances [14]. The software provides a visual target or model; the learner then modifies their output in an effort to match the model and monitors the visual feedback to evaluate the success of their attempts. Existing GAVT apps that include biofeedback primarily focus on voice pitch, which corresponds to the rate of vocal fold vibration. However, we are aware of no existing GAVT app that provides biofeedback representing the resonant frequencies of the vocal tract, or formants. This paper focuses on a novel piece of software to provide biofeedback targeting resonance in the GAVT context. The resonance module described here is part of a larger project to develop a multi-component web-based software for GAVT called TruVox.

1.2. Resonance in GAVT

Vocal resonance makes a substantive contribution to the perception of speaker gender and is thus an important target for GAVT [15]. While it is partly anatomically determined, it is also behaviorally modifiable [16]. However, most individuals have relatively poor insight into what resonance represents and how it can be controlled. In the context of voice feminization, which represents the focus of this study, the goal is generally to shift resonance toward a higher frequency. Various terms can be used to describe differences in resonance associated with gender: lower/higher, darker/brighter, larger/smaller, etc. [10]. Here we use the term "brighter" to refer to higher resonant frequencies generally associated with more feminine-perceived voices and "darker" for lower resonant frequencies generally associated with more masculine-perceived voices.

One previous lab-based study [17] evaluated the feasibility of using visual-acoustic biofeedback to train trans women to alter vocal resonance during speech. The study focused on the second resonant frequency (henceforth, R2), which correlates strongly with perceived voice femininity [18]. After one biofeedback training session, participants were observed to exhibit statistically significant increases in both R2 frequency and perceived speech femininity at the single word level. That study used PC-based software that is too costly and insufficiently user-friendly to be widely adopted. The present study reports the development and initial testing of a web application to visu-

10.21437/Interspeech.2025-1050

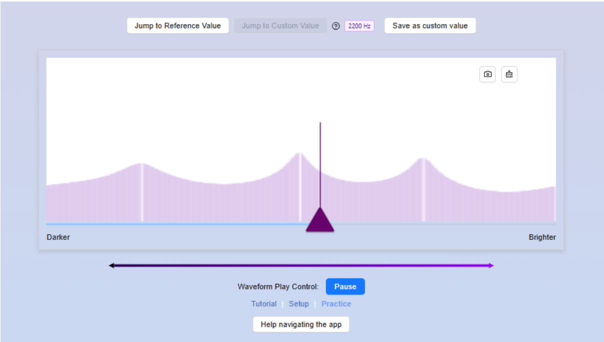

Figure 1: *Screenshot of the TruVox resonance module.*

alize and modify resonance in the GAVT context. Here we provide technical description of the resonance module and its clinical exercises, as well as preliminary evaluation of the acceptability of this software through user testing with trans women.

2. Software prototype

2.1. Real-time LPC display

To target resonance in the GAVT context, learners view a dynamic representation of formants during speech. In the software reported here, this display takes the form of a real-time LPC spectrum [19] implemented in JavaScript. At the present time, it is supported only in the Chrome browser.

LPC is a source-filter model of the voice, where the source is meant to mimic the glottis (typically modeled as noise or pulse train) and the filter mimics the frequency response of the vocal tract. In the frequency domain, LPC can be summarized by the equation below, where $X(n)$ is the speech signal, $H(z)$ is the filter and $E(z)$ is the noise source [20][21]:

$$X(z) = H(z)E(z)$$

$H(z)$ can be modeled as an all-pole infinite impulse response (IIR) filter:

$$H(z) = \frac{1}{A(z)} = \frac{1}{1 - \sum_{k=1}^{p} a_k z^{-k}}$$

$H(z)$ estimates the spectral structure of the vocal tract via calculation of the LPC coefficients a_k, derived using Levinson-Durbin recursion on a given buffer of audio [22].

Fig. 1 shows the LPC spectrum with frequency on the x-axis and amplitude on the y-axis; the peaks in the spectrum are the resonant frequencies of the vocal tract. When a user speaks into the device microphone while the resonance software is running, the visual rendering of the spectral envelope is dynamically updated. White lines superimposed over the spectrum are the output of a peak-picking algorithm that finds local maxima exceeding a preset threshold for magnitude. An LPC spectrum is an appropriate visualization for sounds with strong formant structure, including vowels and vowel-like consonants such as "r" and "l" sounds [23]. The resonance module reported here focuses exclusively on vowels as targets.

2.2. Target slider

An adjustable-frequency slider superimposing a vertical line over the spectrum serves as the target for training. Practice in our software focuses on the second resonant frequency, which we call R2 (avoiding the term "formant," which is unfamiliar to most users). The target for R2 is always defined in relation to a target vowel, since R2 differs across vowels, sometimes dramatically. For the testing reported here, R2 targets were defined from a reference sample of adult speakers of American English [24]. Based on pilot testing, R2 targets were set to 1.5 standard deviations below the mean R2 frequency reported for cisgender (cis) women. 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 app usability.

By default, when a vowel is selected, the slider is navigated to its target position for R2 based on the reference sample of cis women talkers. However, the user can drag the slider to different frequencies, and they can click "save custom target" to cache the current position of the slider for later use in the same browser (Fig. 1). Buttons allow the user to toggle between the reference value and the custom value.

2.3. Navigation and help dialogue

From all views in the app, the user can access three pages, described below: "Tutorial", "Setup", and "Practice". In addition, there is a "Help navigating the app" button. For the Setup page, the help dialogue provides additional text to contextualize the available options. For the Practice page, the help dialogue provides some contextualizing text (e.g., "Here you will see the wave and the adjustable slider") and some tips and reminders (e.g., "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").

2.4. Tutorial

Even trained professionals like speech pathologists often have a limited understanding of what vocal resonance represents and how it can be controlled, and learners without previous vocal or phonetic training are expected to have minimal familiarity with the concept. Therefore, we created a series of brief video tutorials (roughly ten minutes total watch time) in which C. Eagen, a trained vocalist and speech pathologist, explains concepts related to resonance in a voiceover while recording a video of the real-time LPC display. The tutorial covers the following topics:

1. *What is resonance?* This module explains that resonance is determined by the size and shape of the vocal tract above the level of the vocal folds. It shows the LPC wave as the narrator sustains different vowel sounds. The narrator introduces R2 as the focus of resonance practice and introduces the adjustable slider used to set a target for R2.
2. *Bright and dark resonance*: This module provides still images of LPC spectra to make the following points:
   - R2 is generally higher (further to the right) in brighter/more feminine-perceived voices and lower (further to the left) in darker/more masculine-perceived voices.
   - However, the location of R2 depends on the vowel, and differences in R2 between different vowels are often larger than differences between brighter and darker voices.
3. *Modifying resonance*: The narrator sustains a vowel and demonstrates shifting R2 to both brighter and darker targets.
4. *Strategies for resonance*: The narrator describes and demonstrates articulatory strategies that can move R2 toward a higher frequency [25, 26], such as retracting the lips, carrying the tongue further forward in the oral cavity, and narrowing

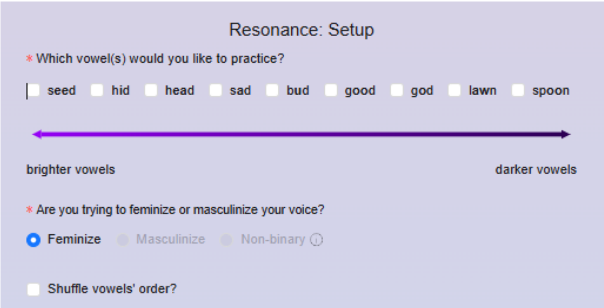

Figure 2: *Screenshot of the setup page.*

the space in the back of the throat between the tongue and posterior pharyngeal wall.

5. *Resonance chains*: The narrator describes and demonstrates the resonance chaining exercise, detailed below.
6. *Pitch versus resonance*: This module presents a dual display with real-time pitch tracking on one side and the real-time LPC for resonance on the other. The narrator demonstrates the ability to change pitch without resonance, and vice versa.

2.5. Setup page

On this page (Fig. 2), the user can check one or more vowel sounds to target in practice. The target vowels are represented with the key words *seed, hid, head, sad, bud, good, god, lawn, spoon*. The vowels are ordered from highest to lowest R2 frequency based on the reference data from American English women in [24]. The setup page also asks for the user's voice goal (more feminine, more masculine, or nonbinary), but only the voice feminization option is active at this time. The user can also select if the vowels will be shuffled in order.

2.6. Practice page

As shown in Fig. 1, here the user sees the real-time LPC spectrum and the adjustable slider, which is automatically navigated to a vowel-specific frequency as described above. A text banner encourages the user to plug in an external microphone in order to improve the signal-to-noise ratio and reduce the risk of vocal strain. The target vowel is represented in text form and in a sample word (e.g., "eh"; "text"). In this view, the user is expected to sustain the target vowel sound and visually evaluate whether R2 is coming close to the target slider. If not, the user can try the strategies described in the tutorial video to attempt to better approximate the target. If the user is satisfied with their practice for a given vowel in isolation, they can click a button to "try a different vowel." Alternatively, they can practice the same vowel in a more complex context, as described below.

2.7. Resonance chains

Practicing a vowel in isolation is of limited functional relevance, since real-world communication involves rapid transitions between different speech sounds. However, it is difficult to use real-time biofeedback in a connected speech context, because the visual target will change as different sounds are articulated. To bypass this problem, the TruVox app features "phrase chains" that range in length from one to five syllables and feature the same vowel in all syllables. For example, a phrase chain for the vowel "uh" could be *"uh"-"some"-"someone"-"someone was"-"someone was up."* As modeled in the tutorial, the learner is encouraged to step through the chain, drawing out the target sound in each syllable and visually comparing it to the visual target. This results in a somewhat unnatural, chanted production. The tutorial suggests that once the learner approximates their target frequency at the highest level of the chain, they can repeat the chain with a more natural cadence and intonation contour. The software contains 27 chains (3 per vowel).

3. User testing

3.1. Protocol

Ten trans women who self-described as native speakers of American English completed a single-session evaluation of usability in the laboratory setting. Participants provided written informed consent and basic demographic characteristics, plus information about their previous history of voice training (either in a singing/acting or GAVT context). They also completed a pure-tone hearing screening consisting of tones at 500, 1000, 2000, and 4000 Hz at 20 dB HL. Three participants showed signs of high-frequency hearing loss (failure to detect 4000 Hz tones in one or both ears). However, these participants were retained because the frequencies targeted in GAVT are generally at 3000 Hz and below. Participants had an average age of 45.4 years (SD = 17 years, range = 27 to 72 years).

In the user testing session, participants were first engaged in one minute of vocal warmup targeting resonance (e.g., humming). They then watched the first two resonance tutorial videos, followed by a brief comprehension check. Next, participants watched two more tutorial videos, followed by one minute of unstructured interaction with the biofeedback display. Participants then watched the "Strategies for resonance" tutorial, followed by a more extended period of practice using the feedback display for three target vowels. The participant and experimenter discussed an appropriate target for each vowel. If the participant's R2 peak fell at a lower frequency than their target, the experimenter guided them in using strategies (e.g., tongue fronting) to try to bring their peak closer to the target. This practice continued for roughly ten minutes. Participants were then cued to watch tutorial video 5 ("Resonance chains"), followed by guided practice with the resonance chains functionality. If participants were not ready to practice above the word level, they were encouraged to try a chain just for comprehension, then returned to practice at the vowel or single word level. At the end of the session, participants completed the System Usability Scale (SUS) [27] and a questionnaire asking for comments on strengths and weaknesses of each part of the app. All data capture used the REDCap system [28, 29] hosted by NYU.

3.2. Results

3.2.1. SUS scores

Across the ten participants, the average overall SUS score was 75.25 out of 100 (standard deviation = 10.7, range = 57.5-87.5). 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 not unexpected, since resonance is a challenging concept and users were required to review an extended tutorial before embarking on resonance practice. SUS scores are compared to benchmarks of 70 ("acceptable") and 80 ("good") [30]. While the scores provided by testers indicated room for improvement, the average score cleared the bar to proceed to preliminary clinical testing.

3.2.2. *Qualitative feedback*

- Feedback on the **tutorial videos** was overwhelmingly positive. Testers noted that they were "clear and concise" and praised the narrator for "great delivery" and "vocal skill to illustrate the points." One participant suggested that the tutorials could be made more interactive, e.g. by encouraging the user to pause and try out the strategies modeled in each video. Another participant suggested that the videos could be more repetitive or more explicitly directive ("do this, now do this") for users who prefer a structured learning experience.
- Users had divided opinions about the **help dialogue**. Five users described the text as "clear and effective" and "really helpful." However, four users indicated some difficulty related to the "density of the text block" and suggested reorganizing the content into more subsections or adding visual elements to make it easier to parse.
- The majority of users expressed satisfaction with the **setup component**, describing it as simple and easy to use. One user dissented, noting that she had some difficulty understanding the logic and sequence of the setup. Another user reported an issue that arose during her testing session: the slider had been set to a previous tester's custom value, but this was not apparent from the on-screen display. Based on a suggestion from the user, we adjusted the buttons for custom and reference targets so that the one in current use is grayed out.
- Users had a number of comments pertaining to the **Practice component** while producing **sustained vowels**. Multiple users expressed appreciation for the core visualization of resonance, which is not to our knowledge included in any other GAVT tools. One user wrote "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." Several users proposed ideas for improving this component, such as an option to record and play back one's own voice and the associated movement of the wave. Another user suggested that practice with the wave could be interleaved with pre-recorded model utterances illustrating the target resonance for different vowels. A different user pointed out that it could be helpful if some of the strategies for modifying resonance from Tutorial 4 were reiterated during practice.
- For the **Practice component** while producing **resonance chains**, five users gave mainly positive comments (e.g., "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"). Four users made suggestions related to the challenge of visually tracking R2 for the target vowels when they were being produced in alternation with other sounds. Suggestions were to include more video models of target chains (potentially interleaved with practice), additional visuals and more verbal explanation, or a different visualization with a "fading trailing edge" that could potentially smooth across multiple sounds. Lastly, one user noted that some of the chains had a mix of vowel sounds and suggested that the chains be further reviewed for errors.
- A final open-ended text field asked for any other feedback about the software. Two users asked if the app could encompass higher levels of practice, making similar requests 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." The second user is correct to point out that the complexity of such a display could have drawbacks for usability, but it is a possibility worth exploring for advanced users. Lastly, two testers pointed out that the user interface needs to be refined for general audiences. One noted, "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." This is an important point because many of our testers work in technology and engineering (a consequence of outreach through a Discord channel that reached multiple testers with a similar occupational background).

4. Limitations and future directions

Our next steps will focus on implementing suggestions provided by testers, such as making the help dialogue more user-friendly and providing expanded opportunities for users to go back and forth between video demos and practice with the wave. Refining the interface for users who are less tech-savvy is an important goal, but improvements to the core technology will likely be needed to address this objective in full. For instance, a user suggested that the display would be easier to understand if a change in color occurred as R2 gets close to its target. We have not pursued such a feature at this time because automatic detection of R2 is insufficiently stable with the naive peak-picking algorithm described above. While the resonant peaks are often detected accurately, the algorithm suffers from a known spurious peak phenomenon [31, 32], and occasionally also misses a peak. This results in flickering behavior that could confuse users rather than help them. State-of-the-art technology has been shown to achieve more reliable detection of resonant peaks across talkers with different vocal tract sizes, though not in a real-time context. Members of our research team are currently exploring adapting contemporary formant detection algorithms, such as KARMA [33] and DeepFormants [34], for enhanced accuracy of formant tracking in real time.

Once reliable resonance tracking is achieved, we will also be able to achieve our goal of setting R2 targets in a more customized manner, instead of relying on reference values from cis talkers. In particular, we can measure the user's R2 for different vowels at baseline, then allow them to set a custom target representing a degree of shift (e.g., "slightly brighter") relative to that target. A related goal is to build out practice routines for users who identify as transmasculine or nonbinary. In principle, the basic mechanism of practice with the app can work just as well for darkening resonance as for brightening it. However, additional testing is needed to identify suggested targets for trans men, since transmasculine voices are impacted by gender-affirming hormone replacement therapy [35] and therefore present different considerations than transfeminine voices.

Because the average SUS score indicated that this prototype was acceptable, we will proceed to a clinical feasibility trial that integrates the resonance module described here with a suite of biofeedback tools focused on vocal pitch developed as part of the TruVox software. In the immediate term, we will focus on evaluating whether time spent in home practice of speech skills is greater for trans women randomized to use our software suite, compared to trans women assigned to work with preexisting acoustic visualization software that is not customized for GAVT. Follow-up research will compare the magnitude of perceptually and acoustically measured change in trans women's voices after practice with our tools, as well as changes in user satisfaction with their vocal presentation.

5. References

[1] A. Brown, "About 5% of young adults in the u.s. say their gender is different from their sex assigned at birth," Pew Research Institute, Report, 2022. [Online]. Available: https://www.pewresearch.org/fact-tank/2022/06/07/

[2] A. Nobili, C. Glazebrook, and J. Arcelus, "Quality of life of treatment-seeking transgender adults: A systematic review and meta-analysis," *Reviews in Endocrine and Metabolic Disorders*, vol. 19, no. 3, pp. 199–220, 2018.

[3] C. Dhejne, R. Van Vlerken, G. Heylens, and J. Arcelus, "Mental health and gender dysphoria: A review of the literature," *Gender Dysphoria and Gender Incongruence*, pp. 56–69, 2018.

[4] A. B. Hancock, J. Krissinger, and K. Owen, "Voice perceptions and quality of life of transgender people," *Journal of Voice*, vol. 25, no. 5, pp. 553–558, 2011.

[5] T. E. Song and N. Jiang, "Transgender phonosurgery: a systematic review and meta-analysis," *Otolaryngology–Head and Neck Surgery*, vol. 156, no. 5, pp. 803–808, 2017.

[6] K. A. Chadwick, R. Coleman, K. Andreadis, M. Pitti, and A. Rameau, "Outcomes of gender-affirming voice and communication modification for transgender individuals," *The Laryngoscope*, vol. 132, no. 8, pp. 1615–1621, 2022.

[7] J.-J. Matthews, A. Olszewski, and J. Petereit, "Knowledge, training, and attitudes of students and speech-language pathologists about providing communication services to individuals who are transgender," *American Journal of Speech-Language Pathology*, vol. 29, no. 2, pp. 597–610, 2020.

[8] M. P. Gelfer, "Voice treatment for the male-to-female transgendered client," *American Journal of Speech-Language Pathology*, vol. 8, no. 3, pp. 201–208, 1999.

[9] E. K. DeVore, S. K. Gadkaree, K. Richburg, E. M. Banaszak, T. V. Wang, M. R. Naunheim, and D. A. Shaye, "Coverage for gender-affirming voice surgery and therapy for transgender individuals," *The Laryngoscope*, vol. 131, no. 3, pp. E896–E902, 2021.

[10] T. Kowalchuk, "What 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.

[11] A. A. Ahmed, "Trans competent interaction design: A qualitative study on voice, identity, and technology," *Interacting with Computers*, vol. 30, no. 1, pp. 53–71, 2018.

[12] A. A. Ahmed, L. Kim, and A. L. Hoffmann, "'This app can help you change your voice': Authenticity and authority in mobile applications for transgender voice training," *Convergence*, vol. 28, no. 5, pp. 1283–1302, 2022.

[13] E. J. Bush, B. I. Krueger, M. Cody, J. D. Clapp, and V. D. Novak, "Considerations for voice and communication training software for transgender and nonbinary people," *Journal of Voice*, vol. 38, no. 5, pp. 1251–e1, 2024.

[14] R. A. Volin, "A relationship between stimulability and the efficacy of visual biofeedback in the training of a respiratory control task," *American Journal of Speech-Language Pathology*, vol. 7, no. 1, pp. 81–90, 1998.

[15] Y. Leung, J. Oates, and S. P. Chan, "Voice, articulation, and prosody contribute to listener perceptions of speaker gender: A systematic review and meta-analysis," *Journal of Speech, Language, and Hearing Research*, vol. 61, no. 2, pp. 266–297, 2018.

[16] C. Leyns, T. Papeleu, P. Tomassen, G. T'Sjoen, and E. D'haeseleer, "Effects of speech therapy for transgender women: A systematic review," *International Journal of Transgender Health*, vol. 22, no. 4, pp. 360–380, 2021.

[17] D. Kawitzky and T. McAllister, "The effect of formant biofeedback on the feminization of voice in transgender women," *Journal of Voice*, vol. 34, no. 1, pp. 53–67, 2020.

[18] D. G. Childers and K. Wu, "Gender recognition from speech. Part II: Fine analysis," *The Journal of the Acoustical society of America*, vol. 90, no. 4, pp. 1841–1856, 1991.

[19] T. McAllister Byun and E. R. Hitchcock, "Investigating the use of traditional and spectral biofeedback approaches to intervention for /r/ misarticulation," *American Journal of Speech-Language Pathology*, vol. 21, no. 3, pp. 207–221, 2012.

[20] T. H. Park, *Introduction to digital signal processing: Computer musically speaking*. World Scientific, 2009.

[21] H.-S. Kim, "Linear predictive coding is all-pole resonance modeling," *Center for Computer Research in Music and Acoustics, Stanford University*, 2023.

[22] J. Franke, "A Levinson-Durbin recursion for autoregressive-moving average processes," *Biometrika*, vol. 72, no. 3, pp. 573–581, 1985.

[23] E. R. Hitchcock, L. C. Ochs, M. T. Swartz, M. C. Leece, J. L. Preston, and T. McAllister, "Tutorial: Using visual–acoustic biofeedback for speech sound training," *American Journal of Speech-Language Pathology*, vol. 32, no. 1, pp. 18–36, 2023.

[24] C. G. Clopper, D. B. Pisoni, and K. De Jong, "Acoustic characteristics of the vowel systems of six regional varieties of american english," *The Journal of the Acoustical society of America*, vol. 118, no. 3, pp. 1661–1676, 2005.

[25] L. Carew, G. Dacakis, and J. Oates, "The effectiveness of oral resonance therapy on the perception of femininity of voice in male-to-female transsexuals," *Journal of Voice*, vol. 21, no. 5, pp. 591–603, 2007.

[26] S. Hirsch, "Combining voice, speech science and art approaches to resonant challenges in transgender voice and communication training," *Perspectives of the ASHA Special Interest Groups*, vol. 2, no. 10, pp. 74–82, 2017.

[27] J. Brooke *et al.*, "Sus-a quick and dirty usability scale," *Usability evaluation in industry*, vol. 189, no. 194, pp. 4–7, 1996.

[28] P. H. RT, B. Minor, V. Elliott, M. Fernandez, L. O'Neal, L. McLeod, G. Delacqua, F. Delacqua, J. Kirby, and S. Duda, "The REDCap consortium: Building an international community of software partners," *J Biomed Inform*, vol. 95, p. 103208, 2019.

[29] P. A. Harris, R. Taylor, R. Thielke, J. Payne, N. Gonzalez, J. G. Conde *et al.*, "A metadata-driven methodology and workflow process for providing translational research informatics support," *J Biomed Inform*, vol. 42, no. 2, pp. 377–81, 2009.

[30] J. R. Lewis and J. Sauro, "Item benchmarks for the system usability scale." *Journal of Usability Studies*, vol. 13, no. 3, 2018.

[31] S. M. Kay and S. L. Marple, "Spectrum analysis—a modern perspective," *Proceedings of the IEEE*, vol. 69, no. 11, pp. 1380–1419, 1981.

[32] P. Alku, S. R. Kadiri, and D. Gowda, "Refining a deep learning-based formant tracker using linear prediction methods," *Computer Speech & Language*, vol. 81, p. 101515, 2023.

[33] D. D. Mehta, D. Rudoy, and P. J. Wolfe, "Kalman-based autoregressive moving average modeling and inference for formant and antiformant tracking," *The Journal of the Acoustical Society of America*, vol. 132, no. 3, pp. 1732–1746, 2012.

[34] Y. Dissen, J. Goldberger, and J. Keshet, "Formant estimation and tracking: A deep learning approach," *The Journal of the Acoustical Society of America*, vol. 145, no. 2, pp. 642–653, 2019.

[35] A. Ziegler, T. Henke, J. Wiedrick, and L. B. Helou, "Effectiveness of testosterone therapy for masculinizing voice in transgender patients: A meta-analytic review," *International Journal of Transgenderism*, vol. 19, no. 1, pp. 25–45, 2018.