

Sensing Your Vocals: Exploring the Activity of Vocal Cord Muscles for Pitch Assessment Using Electromyography and Ultrasonography

Kanyu Chen
Graduate School of Media Design
Keio University
Tokyo, Japan
cady.cky@kmd.keio.ac.jp

Rebecca Panskus
Ruhr University Bochum
Bochum, Germany
rebecca.panskus@rub.de

Erwin Wu
School of Computing
Institute of Science Tokyo
Tokyo, Japan
Graduate School of Media Design
Keio University
Tokyo, Japan
wu.e.aa@vogue.cs.titech.ac.jp

Yichen Peng
School of Computing
Institute of Science Tokyo
Tokyo, Japan
peng.y.ag@m.titech.ac.jp

Daichi Saito
School of Computing
Institute of Science Tokyo
Tokyo, Japan
saito.d.ah@m.titech.ac.jp

Emiko Kamiyama
Graduate School of Media Design
Keio University
Tokyo, Japan
emiko.kamiyama.pooh@gmail.com

Ruiteng Li
Fundamental School of Science and
Engineering
Waseda University
Tokyo, Japan
ruiteng.li@toki.waseda.jp

Chen-Chieh Liao
School of Computing
Institute of Science Tokyo
Tokyo, Japan
liao.c.aa@m.titech.ac.jp

Karola Marky
Ruhr University Bochum
Bochum, Germany
karola.marky@rub.de

Kato Akira
Graduate School of Media Design
Keio University
Tokyo, Japan
kato@wide.ad.jp

Hideki Koike
School of Computing
Institute of Science Tokyo
Tokyo, Japan
koike@c.titech.ac.jp

Kai Kunze
Graduate School of Media Design
Keio University
Tokyo, Japan
kai.kunze@gmail.com

Abstract

Vocal training is difficult because the muscles that control pitch, resonance, and phonation are internal and invisible to learners. This paper investigates how Electromyography (EMG) and ultrasonic imaging (UI) can make these muscles observable for training purposes. We report three studies. First, we analyze the EMG and UI data from 16 singers (beginners, experienced & professionals), revealing differences among three vocal groups of the muscle control proficiency. Second, we use the collected data to create a system that visualizes an expert's muscle activity as reference. This system is tested in a user study with 12 novices, showing that EMG highlighted muscle activation nuances, while UI provided insights into vocal cord length and dynamics. Third, to compare our approach to traditional methods (audio analysis and coach instructions), we

conducted a focus group study with 15 experienced singers. Our results suggest that EMG is promising for improving vocal skill development and enhancing feedback systems. We conclude the paper with a detailed comparison of the analyzed modalities (EMG, UI and traditional methods), resulting in recommendations to improve vocal muscle training systems.

CCS Concepts

• **Human-centered computing** → **Sound-based input / output; Visualization design and evaluation methods; Field studies.**

Keywords

Dataset, EMG, Ultrasonography, Microphone, Input Techniques, Usability Study, Vocal Muscles

ACM Reference Format:

Kanyu Chen, Rebecca Panskus, Erwin Wu, Yichen Peng, Daichi Saito, Emiko Kamiyama, Ruiteng Li, Chen-Chieh Liao, Karola Marky, Kato Akira, Hideki Koike, and Kai Kunze. 2026. Sensing Your Vocals: Exploring the Activity of Vocal Cord Muscles for Pitch Assessment Using Electromyography and Ultrasonography. In *Proceedings of the 2026 CHI Conference on Human Factors*



This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.

CHI '26, Barcelona, Spain

© 2026 Copyright held by the owner/author(s).

ACM ISBN 979-8-4007-2278-3/26/04

<https://doi.org/10.1145/3772318.3790965>

in *Computing Systems (CHI '26)*, April 13–17, 2026, Barcelona, Spain. ACM, New York, NY, USA, 20 pages. <https://doi.org/10.1145/3772318.3790965>

1 INTRODUCTION

Vocal training depends on precise control of muscles that most people cannot see or feel directly. Singers must learn to coordinate the thyroarytenoid and cricothyroid muscles to control pitch, yet they typically receive feedback only through sound-based methods such as spectrograms [33] or, in clinical settings, through invasive laryngoscopy [56]. Previous studies have demonstrated the value of metrics like the Singing Power Ratio (SPR) in distinguishing vocal capabilities [67] and have explored respiratory muscle training to enhance vocal function [15]. These approaches remain limited, however, because they provide no intuitive connection between what singers hear and what their muscles are doing. Spectrograms show acoustic output but not the physical mechanisms producing it. Singers are left to discover through trial and error how to engage their vocal muscles to achieve a target sound.

This paper addresses these gaps by investigating electromyography (EMG) and ultrasonographic imaging (UI) as basis for more intuitive and less invasive visualization methods for vocal cords training. EMG and UI are methods to measure muscle movements [6, 26]. Tension of muscles [63], such as the thyroarytenoid and cricothyroid muscles, allows pitch control in voice production. Our goal is leveraging the understanding of muscle activity to improve the visualization of vocal activity. For this, we conducted a series of studies with singers since singing is a specific skill that requires the control of vocal cord muscles.

We conducted three studies to evaluate EMG and UI for vocal training. First, we collected a Vocal Cord Sensing Dataset (VCSD) from 16 singers with varying skill levels, using both EMG and UI to establish baseline patterns of vocal muscle activity. This dataset addresses our first research question:

RQ1: Can EMG and UI be used to capture different levels of proficiency in singing and support singers of various skill levels?

Analysis of the dataset revealed distinct muscle activity patterns between experienced and novice singers, confirming that both modalities can detect proficiency differences. Having established a technical foundation, we next used the data of experienced singers to create reference visualizations for EMG and UI of “correct” singing that novice singers could use for training. In our second study, we tested whether novice singers could use these references to improve their own performance:

RQ2: How do EMG and UI perform in training novice singers?

We recruited 12 novice singers to practice matching their muscle activity to the EMG and UI reference visualizations in real time. Results showed that both modalities support skill development, though with different strengths. EMG feedback enhanced perceived controllability and reduced fatigue, while UI feedback improved vocal cord control for specific pitches but required higher cognitive effort. The UI setup also constrained movement due to its physical size, limiting its use to stationary practice.

Having EMG and UI tested for novices, we created a more mobile setup based on EMG-only and wireless microphones to support performance-oriented scenarios where singer might move. UI was not considered in this study due to its rather bulky setup and the

already established muscle control of experienced singers. Hence, in our final study with 15 experienced singers, we wanted to shed light on the following research question:

RQ3: How does a portable EMG setup perform in supporting experienced singers compared to traditional feedback methods?

Our findings showed that EMG captured nuanced muscle engagement, with tenor singers aligning closer to expert references. In addition, we found a strong correlation with vocal power, or known as Singing Power Ratio (SPR), highlighting superior muscle control in experts. Comparing our sensing setup with traditional methods, we found that EMG can provide technical precision and contextual guidance, showcasing its potential as a portable tool for advanced vocal training. Yet, overall a combination of EMG and traditional methods is likely to yield the best results as both approaches complement each other.

Across all three studies, we found that EMG and UI feedback can be effective for vocal training. While our investigation focused on singing, the underlying techniques have broader applications in speech detection, language learning, and voice capture. We conclude with guidelines for designing effective technology-assisted practice systems.

Contribution. The major contributions of this paper are:

- (1) **Improving vocal pitch assessment:** We propose an accessible approach for collecting and evaluating vocal pitch using EMG and UI sensing based on real-world data.
- (2) **In-depth investigation with singers:** We present three user studies (in total $N=43$) varying untrained (novice) and trained groups to further investigate the effectiveness, usability, skill transferability of vocal sensing techniques and provide a comprehensive design guideline for vocal training systems. We provide an open-access dataset of dual-modality sensing (EMG and ultrasonography) collected from 16 participants with varying levels of vocal training experience, available under an osf link: https://osf.io/gkjsx/?view_only=0e6aee05b63040a5b1f140297775ece9.
- (3) **Broadening EMG and UI for other domains:** We show how our results motivate and inform further research in related domains of vocal training also providing actionable guidelines for effective technology-assisted practice systems.

2 RELATED WORK

This paper builds on four aspects below, using Electromyography (EMG) and ultrasonic imaging (UI) to explore how physiological data can improve vocal skill assessment.

2.1 Vocal Sensing and Feedback in Interactive Music Systems

Interactive music systems have shown potential for musical perception, bodily awareness, and learning outcomes. For example, the integration of digital tools and games [27] to support the voice treatment of parkinson patients demonstrates how gamification can improve vocal therapy by increasing patient participation and vocal loudness. Similarly, Bean Academy [30] simplifies music composition through vocal query transcription, reducing the barriers to music theory comprehension and software proficiency.

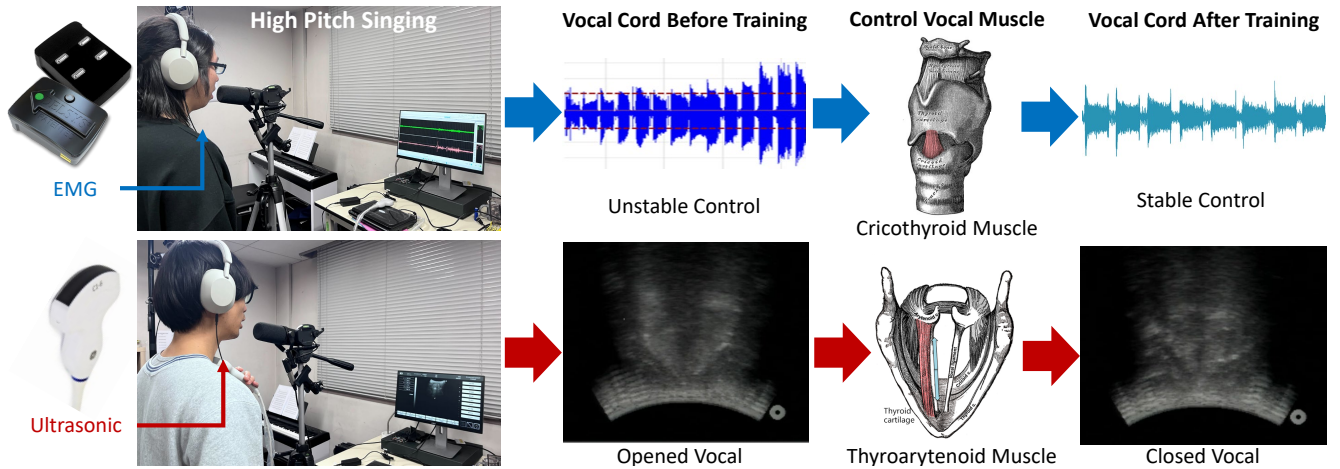


Figure 1: From sensor data collection to visualization: (1) Showing a participant wearing an EMG sensors to detect the activity of the cricothyroid muscle. (2) Showing a participant using a B-type ultrasonography device to visualize the thyroarytenoid muscle.

Research on vocal sensing technologies, such as pitch detection, remains in early stages. Celestia [55] uses pitch detection for real-time interaction in music games, showing the potential of auditory feedback to improve vocal control. Additionally, BrainiBeats [7] uses physiological signals (EMG and EEG) to generate music based on emotional states, merging creativity with biological data. These systems suggest applications for personalized learning environments and therapeutic interventions.

Research has integrated physiological responses into artistic performances [21, 23], and interactive installations have further explored embodied music creation. “The Music Room” [41] is an immersive installation where participants’ movements within the physical space trigger and shape musical elements, demonstrating collaborative, embodied music-making. Similarly, “The Throat III” [64] and “The Vocal Corder” [65, 66] allow opera singers to manipulate their voices dynamically, expanding the expressive possibilities of vocal performance through sensor-based technology. Mixed reality technologies [54] have also shown a significant increase in co-presence for remote collaboration with musical partners, offering opportunities for vocal sensing.

2.2 Physiological Foundations of Vocal Pitch Control

Vocal training combines foundational techniques with physiological understanding to optimize performance. Estill Voice Training (EVT) [57], Speech Level Singing (SLS) [35], and Complete Vocal Technique (CVT) [59] represent three widely-used methods. Traditional approaches to vocal assessment focus on breath control, pitch accuracy, volume modulation, and rhythmic precision [69].

In this paper, our sensing methods to pitch assessment is built on EVT theory [57], which suggests that vocal pitch is determined by the tension in the vocal folds, a factor evident in both speech and song [3, 63]. Previous studies identified the thyroarytenoid and cricothyroid muscles as the primary controllers of vocal pitch [63], with their contractions adjusting the tension of the vocal folds by

rocking the cricoid and arytenoid cartilages. This mechanism alters the pitch of the voice [6, 19, 25].

However, a persistent challenge in vocal pedagogy [44] is the lack of comprehensive feedback mechanisms, particularly in the absence of continuous guidance from professional instructors. Muscle tension is central to discussions of vocal performance pedagogy. Controlled tension is necessary for precise manipulation of the vocal tract, but excessive tension – is often associated with reduced sound quality [40]. Current methods [44] predominantly rely on auditory feedback and visual analysis of spectrograms to evaluate vocal performance, with limited insights into the underlying muscular mechanisms involved in sound production. Emerging trends in vocal training research [8–10, 69] have begun to integrate technological innovations, such as muscle activity visualization and monitoring, to offer more precise and informative feedback to learners.

Therefore, to integrate the sensing technologies in vocal assessment can enhance the efficiency and efficacy of vocal training programs, considering the diverse needs and skill levels of practitioners across various musical genres and educational settings.

2.3 Sensing Technologies for Vocal Analysis

Recent advancements in sensing technologies improve the understanding and control of the human voice. Traditional vocal analysis methods [44] often rely on subjective assessments, which can be difficult to obtain consistently and are prone to interpretation biases. For instance, waveform analysis can detect timing errors and improve note alignment, articulation, and transitions, enhancing recording quality [1]. Recent improvements also enable more detailed processing of complex vocal performances [24]. Singing Power Ratio (SPR) has been used to provide automated feedback in vocal training, assessing resonance and clarity, crucial in classical singing [31, 58]. Despite these advancements, traditional methods still limit a deeper understanding of vocal production mechanisms.

EMG and ultrasonography have enabled researchers to examine vocal physiology in greater detail. Existing research has explored correlations between uttering vowels or sentences and engagement of speech muscles [76], the use of pitch and EMG for omohyoid detection [68], and analysis of face and neck muscle movements during speech [49]. Ultrasonic techniques capture speech dynamics from tongue movement to sound production [26].

Certain wearable designs worn on the throat or chest offer a potential avenue for enhancing bodily signals as input for vocal expression, from a custom wearable collar that monitors vocal performance [50], to a breath-related wearable instrument that enhances somatic awareness (the conscious perception of internal bodily sensations) during singing [13]. Despite significant progress, a fully integrated sensing system specifically tailored for detecting vocal pitch muscles remains absent. Nevertheless, the integration of EMG and ultrasound into practical, wearable systems for real-time vocal analysis is still an area ripe for exploration, with promising applications in both clinical settings and vocal pedagogy.

2.4 Muscle Activity Sensing: From General Applications to Vocal Training

The assessment of muscle engagement has increasingly captivated the Human-Computer Interaction (HCI) community [16], particularly within the realms of motor skill training and rehabilitation [47, 52], owing to its potential to elevate skill acquisition and optimize performance outcomes. Ultrasound stands out as a traditional yet potent modality for visualizing muscle movements. As a medical imaging technique, it furnishes images of internal bodily structures, including muscles [43, 53], thereby offering users a tangible means to visually comprehend muscle movements. In this study, we leverage ultrasonography imaging (UI) to provide real-time visualization of vocal cord movements as a training method.

However, the interpretation of UI can present challenges, and quantifying these images' outcomes often entails complexity [21, 38, 39, 48]. In contrast, Electromyography (EMG), Mechanomyogram (MMG), and Electrical Impedance Tomography (EIT) present alternative methodologies for measuring muscle activity. EMG, by recording the action potentials generated by muscle contractions as a result of neural activity [32, 60], offers a high degree of sensitivity to the electrical activities underlying muscle contractions. This attribute renders EMG particularly adept at capturing the subtle nuances in vocal cord muscle movements, which are characterized by rapid and complex motion. MMG, in contrast, detects mechanical signals from muscle surfaces during contraction, through technologies such as capacitive plane arrays [51] and force-sensitive resistors [17]. However, given its propensity for monitoring mechanical vibrations, MMG is more apt for assessing larger muscle groups rather than the fine movements of the vocal cords.

Electrical Impedance Tomography (EIT) has been explored for capturing hand gesture movements [72, 73] and muscle activity [74, 75]. Unlike EMG and MMG, EIT monitors both contracted and stretched muscles by mapping internal conductivity changes, providing regional and cross-sectional information. However, its precision may not meet the requirements for vocal cord training, where targeted monitoring of specific muscles is essential. EMG sensors

provide localized measurement, making them effective for isolating the activity of small or deeply located muscles such as those controlling the vocal cords.

For vocal training applications requiring precise feedback on small, rapidly moving muscles, EMG offers advantages as a measurement technique. Its ability to provide immediate, specific data on muscle activity makes it valuable for detailed analysis and feedback on vocal performance. Combined with ultrasound visualization, these complementary modalities can address gaps in current vocal training feedback systems.

3 Research Approach & Hardware Setup

Our research program comprised three sequential user studies, each building upon findings from its predecessor. We first describe the iterative investigation structure, then detail the hardware configurations employed across studies.

3.1 Iterative Investigations

We conducted three studies to systematically examine EMG and ultrasound imaging (UI) as modalities for vocal training feedback. Each study addressed distinct research questions while informing subsequent methodological refinements.

- **Study 1 – Data Collection (N=16):** This foundational study investigated whether EMG and UI can effectively visualize vocal cord muscle activity and differentiate between skill levels 4. We recruited 16 singers spanning novice to professional expertise to establish reference datasets. The study addressed two questions: (a) Do measurable differences in vocal cord activity patterns exist between differently skilled singers? (b) Can these differences be rendered as meaningful visualizations? Findings from this study informed sensor selection and feedback design for subsequent investigations.
- **Study 2 – Novice Study (N=12):** In the second study, we specifically focused on novice singers to investigate their temporal stability through EMG and feedback on muscle movements by UI also capturing the task load and usability of the setup. In this study, we used a static setup since the training of basic skills does not require much movement.
- **Study 3 – Experienced Study (N=15):** In our final study, we investigated a more realistic setting for advanced singers requiring a mobile setup that captures the muscle activity of singing during other movements. Because of that, we had to rule out UI sensors, since they are rather bulky and require the singers to be in a static position. Further, we used this study to gather feedback from the experienced singers that compares our setup with traditional coaching.

3.2 Hardware Setup

Based on previous work, we considered the thyroarytenoid and cricothyroid muscles as the primary controllers of vocal pitch [63]. We collected two types of data to directly measure muscle activity across various pitches: EMG and UI. A microphone collected audio.

EMG Setup. To capture EMG, we used Delsys Trigno Wireless EMG sensors to measure vocal muscle activity. Sensors were positioned on the anterior neck between the thyroid cartilage (Adam's

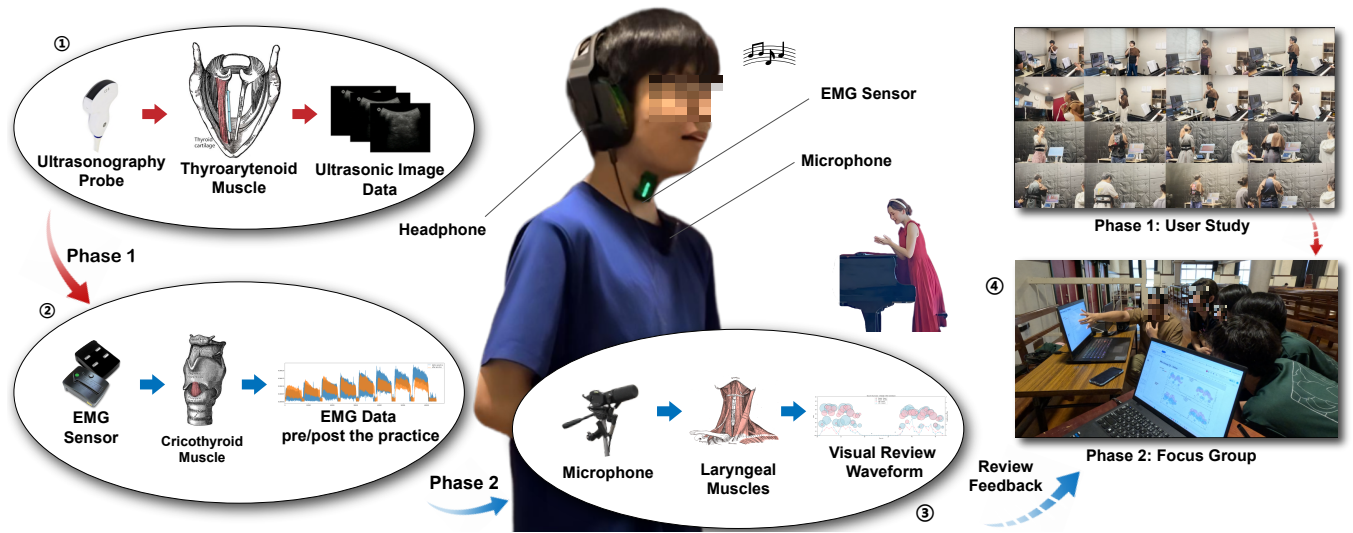


Figure 2: Overview of the Iterative Study Process and the Hardware Setup: (1) Ultrasonography sensors require the manual positioning of the probe on the skin near the vocal cords. (2) EMG sensors need to be placed bilaterally around the user’s larynx to monitor vocal muscle activity. (3) The data recording also includes the use of a noise-canceling microphone.

apple) and the superior border of the cricoid cartilage. The data was captured at 2000Hz in the first study (VCS dataset) and second study (novice study). In the third study (experienced singers study), EMG data was sampled at 4370Hz to segment short-duration pitch events during song passages, and to align with the higher-resolution audio.

Ultrasonography. The CONTEC CMS600P2 Full Digital B-type Ultrasound Diagnostic System was employed. Participants held the probe at the front of the larynx, with ultrasonic operating at a 3.5 MHz transmit frequency and the system recording video at 30 fps. Ultrasonography was used only in the novice study for visual feedback on vocal muscle activity. Due to its bulkiness, this method was not used in the experienced singers study.

Audio Capture. Study 2 employed a Shure SM7B microphone for high-fidelity recording in controlled acoustic conditions. Study 3 substituted a portable wireless microphone to accommodate singer movement during performance tasks.

3.3 Data Analysis Statistical Rationale

All statistical analyses were conducted in Python using numpy, scipy, and statsmodels. Parametric (t -tests) or non-parametric (Wilcoxon signed-rank test) tests were chosen accordingly based on normality assumptions. Multiple comparisons were corrected using the Benjamini–Hochberg FDR procedure [4]. Effect sizes (Cohen’s d or rank-biserial r_{rb}) and 95% confidence intervals were reported following recommended standards [11, 29]. For repeated-measures outcomes (e.g., pitch-wise EMG stability and vocal-fold length), linear mixed-effects models [20] were used with participant as a random intercept and Group and Pitch as fixed effects.

3.4 Ethical Considerations

Our institutional Ethical Review Board (ERB) approved our study design. Throughout our study, we took precautions to treat our participants in an ethically correct manner and adhered to strict privacy laws. Participants of all three studies received a consent form containing detailed information about the captured data, assurances that they could withdraw from participating at any point without negative consequences, and information about their rights in compliance with the General Data Protection Regulation (GDPR) and national data protection laws. Participating in any of the three studies caused no risks beyond everyday life or psychological harm for participants. To protect the privacy of our participants and their environment, we have anonymized the transcripts prior to analysis as well as the photos of our participants used in this paper.

3.5 Positionality Statement

This project was a collaboration between HCI and security and privacy (S&P) researchers from academia and industry. All members of the research team have more than three years of experience in conducting, writing, publishing, and reviewing HCI research, while focusing on end users, children, and singers. We approach this study as researchers located within the HCI and S&P communities. Our interest in this topic stems from our own experience conducting qualitative HCI and S&P research with end users, children, and singers, where we have identified, first-hand, several constraints and challenges and aim to understand the extent to which these constraints and challenges occur on a scale. Nine of the authors grew up in non-Western countries but now collaborate with Western institutions. Three authors grew up in Western countries. The resulting research experiences with end users and children from different cultural backgrounds shape our understanding of these target groups. We acknowledge that our dual roles as conducting this study and as fellow researchers in this space may have influenced

how participants shared their experiences, while also enabling us to engage with them more meaningfully. At the same time, our positions as HCI and S&P researchers limit our ability to interpret findings beyond the perspectives of our own discipline. Our vision for this work is to serve as a starting point for our research communities to work together in addressing the challenges of conducting inclusive research and enacting changes.

4 STUDY 1: INITIAL EMG & UI INVESTIGATION (RQ1)

The primary goal of this study was to conduct an exploratory analysis of vocal muscle activity during singing using EMG and UI, investigating whether these sensing technologies can capture different levels of singing proficiency. To this end, we collected a Vocal Cord Sensing Dataset (VCSD).

4.1 Collected Data

The collected VCSD dataset consists of a total recording of around three hours. The details of the dataset can be found in Table 1. Besides the videos of ultrasonography and the 2-channel EMG data, the dataset also includes the breath sensor values and the reference videos with audio during the collection procedure.

4.2 Study Procedure

1) *Welcome & Familiarization*: Before data collection, the participants were informed about their rights and the implications of their participation. When agreeing to participation and use of their data, they signed a consent form. Then, an initial questionnaire was given to collect the participants' demographics and their vocal training experience. Next, participants were introduced to the sensing devices and the task, followed by a 5-minute practice session to familiarize themselves with the sensing devices and the task.

2) *Singing Task*: Participants were instructed to sing their vocal range as accurately as possible, referenced the scientific pitch notation (SPN). To support pitch accuracy, an 80 bpm piano reference of the scale was played in real time. Participants were asked to follow this reference closely and maintain each pitch for two seconds. Each participant performed the task with both sensing modalities (EMG and UI). Each session was repeated four times, yielding eight recording rounds per participant. Session order was counterbalanced across participants to mitigate learning effects.

3) *Exit Interview*: Finally, the participants were interviewed about vocal training and their impressions of the two sensing technologies.

4.3 Participants & Recruitment

We recruited 16 participants (6 female, 10 male, aged 21-33 years, mean=25.7) from two local institutes. Ten participants were beginners with little vocal training experience, and three participants were intermediate amateurs who had basic vocal knowledge. The remaining three participants were experts who experienced professional vocal training for more than 10 years. Novice and experienced participants were provided a 1000 yen gift card per hour as reimbursement for their participation in the study. Expert participants

were reimbursed with 10,000 yen as compensation for their time and effort. The collection process was approved by the IRB of the local institutes. All data collected within the scope of this study are anonymous and are only released with the approval of the participants.

4.4 Results

Since the raw data contains noise and redundant information, we first post-processed the data and analyzed the results from different perspectives, to obtain insights from the collected data.

4.4.1 Data Processing.

EMG. Since our focus was on the stability and controllability of the cricothyroid muscle, we tried to extract stability information from the data. The raw data was first denoised through a moving average filter (window size = 10ms), then a Hilbert transform was performed to calculate envelopes of the signal [12, 45]. The stability of muscle activity [18] s was then calculated as follows:

$$s = \frac{1}{N-1} \sum_{t=1}^{N-1} \left\| 20 \log \frac{A_{t+1}}{A_t} \right\| \quad (1)$$

A_t denotes the previous envelope value of the filtered EMG at timestep t . This equation is designed with reference to shimmer measurement in voice, which is frequently used in the field of acoustic analysis [62]. Dividing among envelopes allows calculating the stability of each pitch and comparing the differences in scales between each participant.

UI. The UI video had to be quantified for analysis. Therefore, we developed a landmark detector for tracking the vocal cord muscle from the ultrasonic images. Here, we focused on five important key points in the video: start points (connection) of two vocal cords, ends of the inner side of vocal cords, and the end of the outer side of vocal cords, as shown in Figure 3. These five points discern changes in the true vocal cord structure and cartilage position based on previous research [28]. Since the shape of the vocal cord differs for each participant, we manually annotated the key points on an initial frame for each session. With the positional data from the five key points, we computed the length of the true vocal cords (depicted in red in Figure 3) as follows:

$$L = \frac{1}{2} * \left(\text{Dist} \left(P_{VS}, \frac{P_{VL1} + P_{VL2}}{2} \right) + \text{Dist} \left(P_{VS}, \frac{P_{VR1} + P_{VR2}}{2} \right) \right) \quad (2)$$

4.4.2 *Data Analysis*. To understand the effect of the two sensing data as well as to answer RQ1, we further analyzed the collected data among the participants.

Following recommendations for statistical reporting [11, 29], we report effect sizes (partial η^2) alongside p -values. First, the processed data was statistically analyzed using linear mixed-effects models (LME) [20] among the three groups of different level participants (beginners, intermediate amateurs, experts). *Pitch* was treated as a repeated factor, *Group* of different levels as a between-subject factor, and *Gender* as a covariate. Participant number was

Dataset	Sampling	Novices (1-10)										Experienced (11-13)			Professionals (14-16)			Total Size
Subject	-	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	16
Pitch Range (the number of pitch)	G2 - E6 (27)	F3 - F4 (14)	C3 - C4 (8)	F3 - G4 (9)	C3 - B4 (14)	G3 - D5 (12)	G3 - D5 (12)	D3 - C5 (14)	G3 - C5 (11)	F3 - D5 (13)	F3 - D5 (13)	F3 - E5 (14)	F2 - C5 (19)	G2 - E5 (20)	E3 - E6 (22)	D3 - E6 (22)	G2 - C6 (25)	G2 - E6
2-channel EMG Data	2000 hz/s	316s	253s	179s	383s	348s	277s	342s	130s	245s	329s	298s	249s	421s	139s	495s	524s	4928s
Ultrasonography Data	30 fps	273s	333s	264s	287s	217s	213s	280s	336s	232s	288s	333s	288s	368s	483s	583s	360s	5138s

Table 1: Statistics of the VCSD datasets: EMG, UI, and pitch range for 16 users across skill levels (Novice, Intermediate, Expert)

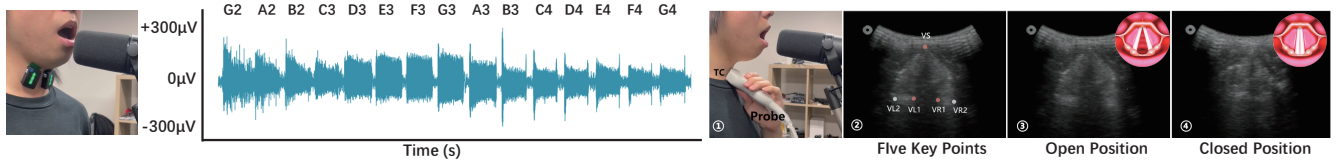


Figure 3: Left: EMG sensor placement and visualization of a sample of raw EMG data accompanied by muscle/cartilage position annotations. Right: Positioning of the ultrasonography probe and sample of raw ultrasound imaging data accompanied by muscle/cartilage position annotations.

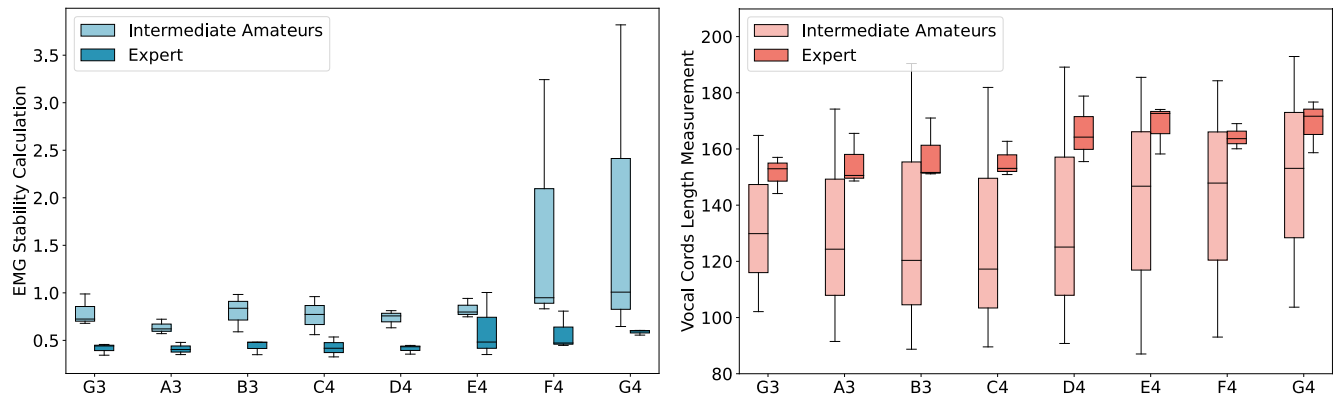


Figure 4: Comparison of the stability calculation derived from EMG data and the measured vocal cord length via ultrasonography across a pitch range from G3 to G4. Left a set of a participant's raw EMG imaging data, right a set of a participant's ultrasound imaging data, each spanning over one octave.

included as a random intercept to account for within-subject dependency across pitches. Since most beginners cannot “correctly” sing the notes, we focused on the data between the intermediate and expert-level participants in the following analysis.

EMG. For the stability score, we picked up the common range (G3-G4) of the intermediate and expert groups to perform a deeper investigation. The mixed-effects analysis revealed a significant main effect of Group, with experts showing markedly lower EMG instability than intermediate singers ($\beta = -0.48$, 95% CI $[-0.83, -0.12]$, $p = .013$, partial $\eta^2 = .18$). Because higher EMG stability scores indicate poorer temporal muscle control, this result reflects substantially more stable cricothyroid activity among expert singers across the octave. The main effect of Pitch was not significant ($p = .28$), although EMG instability tended to increase at higher notes (F4-G4). The Group $\times$ Pitch interaction was not significant ($p = .62$),

suggesting a consistent stability advantage for experts across the octave range. Gender had no significant influence on EMG ($p = .26$).

The results per pitch are shown in the left figure of Figure 4. Despite an overall better temporal stability of the expert group, we found an obvious difference in the higher pitches (F4 and G4), which indicates the ability to control vocal muscles in high-pitch sounds of the experts.

UI. For the vocal length data estimated from the UI, there was no significant difference between intermediate and expert singers ($\beta = -1.82$, 95% CI $[-6.97, 3.34]$, $p = .49$). Pitch did not reach statistical significance ($p = .29$), although an expected increase in length was observed toward higher notes. Gender showed a significant main effect ($\beta = 12.43$, 95% CI $[4.85, 20.01]$, $p = .003$, partial $\eta^2 = .25$), indicating systematic differences in UI measurements between male

and female singers. No significant Group $\times$ Pitch interaction was observed ($p = .99$).

The results suggest a similar trend as the EMG data that the range of vocal cords is more stable for the experts (see right part of Figure 4). Although the UI does not provide any temporal information, it shows that expert singers can more precisely manage their vocal cords.

Interpretation Across Pitches. As visualized in Figure 4, expert singers demonstrate lower EMG instability (better temporal muscle control), particularly at higher pitches. They also show more consistent vocal cord positioning in the UI data, indicated by narrower distributions of measured lengths. Intermediate singers, in contrast, exhibit greater variability in both EMG and UI, highlighting less stable laryngeal control strategies.

4.4.3 Result Summary: RQ1. In sum, EMG and UI can both effectively capture the vocal cord muscle activity and be used to differentiate the level of singers in the scale task. The EMG feedback was easy to understand and provided a better representation of the temporal stability of vocal cords. UI offers a direct visual cue of vocal cords, providing intuitive feedback of muscle activities.

5 STUDY 2: NOVICE STUDY (RQ2)

Study 1 demonstrated that EMG and UI can differentiate vocal cord muscle activity across skill levels, with EMG capturing temporal stability and UI providing direct visual feedback of vocal cord positioning. Building on these findings, we conducted a study with novice singers to evaluate how each feedback modality supports vocal training. Participants practiced matching their muscle activity to expert reference visualizations derived from the Vocal Cord Sensing Dataset (VCSD) collected in Study 1.

5.1 Study Procedure

The study employed a within-subjects design in which each participant completed training under three conditions: audio-only baseline, EMG feedback, and UI feedback. Condition order was counterbalanced across participants. The procedure (see Figure 5) consisted of five phases:

- 1) *Welcome & Familiarization:* First, we welcomed the participants and obtained their consent.
- 2) *Warm-Up & Pre-Training:* We did a warm-up for 5-10 minutes to prepare the participants for the vocal exercise. Next, we measured a pre-training baseline for EMG and UI.
- 3) *Practice Sessions:* The participants practiced for 5 minutes under one of the three conditions with a vocal guide which is brief demonstration on vocal muscle structure and the vocalis mechanism. The participants then learned from experts' reference recording (that we obtained from the VCSD collected in Study 1) for each practice session.
- 4) *Post-Training:* Next, we measured the post-training performance without a vocal guide and asked the participants to fill in a questionnaire about the condition. To capture the perceived task load,

we used the NASA-TLX [22]. We further captured a subset of the Sense of Agency Scale (SoAS) [61]. Here, we extracted five pertinent questions from the SoAS, focusing on evaluating aspects of controllability. The steps (2) to (5) were repeated for three conditions. The order of EMG and UI was counterbalanced.

5) *Exit Interview:* The participants were interviewed about their preferences, evaluation, and suggestions for the system. Further, we captured the participants' feedback on the advantageous and less beneficial aspects of each training methodology.

5.2 Participants & Recruitment

To investigate the training effects of visual cues from both above vocal cords sensing, 12 novice participants (6 male, 6 female; aged 24-34 years, mean=27.4) were recruited. Regarding the singing (vocal training) experience, none of the participants had experienced professional training before, while all participants are singing enthusiasts who visit Karaoke (or similar activities) at least once a month. Participants received a 1000 yen gift card per hour as reimbursement for their participation in the study. The study was approved by the local Institutional Review Board.

5.3 Results

5.3.1 Sensing: EMG Stability Decreased, While Mean Vocal Cords Lengths Increased After the Practice.

Stability Controlling (EMG). We evaluated muscle activity during vocal training, calculating the stability $s \downarrow$ of EMG signals (see Section 4.4.2) before and after practicing with the EMG training method. For each pitch (G3–G4), we conducted paired Wilcoxon signed-rank tests to compare pre-post stability within participants and applied Benjamini-Hochberg false discovery rate (FDR) correction [4]. Before correction, three pitches (B3, C4, E4) showed significant increases in instability after training (B3: $W = 6$, $p = .048$, $r_{tb} = .62$, 95% CI [.08, .88]; C4: $W = 8$, $p = .065$, $r_{tb} = .47$, 95% CI [−.15, .82]; E4: $W = 14$, $p = .009$, $r_{tb} = .74$, 95% CI [.22, .93]). After applying FDR correction, only E4 remained statistically significant ($p_{FDR} = .036$), while the remaining pitches did not survive correction ($p_{FDR} > .10$). The post-training stability was generally lower than the pre-training levels (see Figure 6 (left)). This finding stands in contrast to the self-reported perceptions of increased controllability as indicated in the responses to the adapted SoAS questionnaire. Notably, this decrease in stability was statistically significant, especially for pitches B3, C3, and E4, with $W = 6$, $p = 0.048$, $W = 3$, $p = 0.049$, and $W = 14$, $p = 0.009$ respectively, as determined by the Wilcoxon's signed-rank test.

To address potential confounding factors, we fitted a linear mixed-effects model [20] with fixed factors Pre/Post, Pitch, Order, and Gender, and random intercepts for participants. The model revealed a significant main effect of Pre/Post, indicating increased instability after training ($\beta = 0.41$, 95% CI [0.12, 0.69], $p = .018$, partial $R^2 = .21$). Order ($\beta = 0.08$, 95% CI [−0.11, 0.27], $p = .41$) and Gender ($\beta = 0.10$, 95% CI [−0.12, 0.32], $p = .33$) were not significant. No significant interactions were observed.

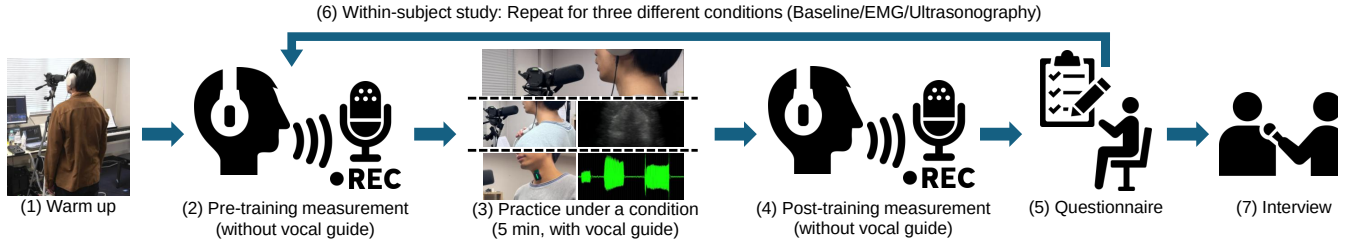


Figure 5: The procedure of the novice study.

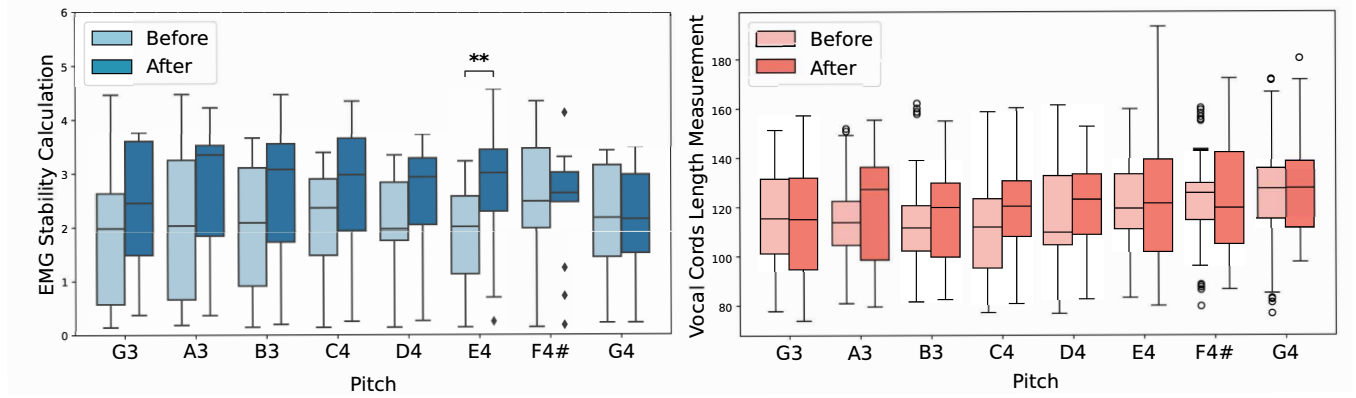


Figure 6: Left: Results of the stability $s \downarrow$ test on the vocal muscle activity following the Equation 3, pre- ("before" in light blue) and post-training ("after" in dark blue). Right: Results of the manually annotating on the vocal length form ultrasonography according to instruction in Section 4.4.2, pre- ("before" in light red) and post-training ("after" in dark red). In both graphs the Wilcoxon's test results are indicated by the bridges between boxes (**: $p < 0.01$).

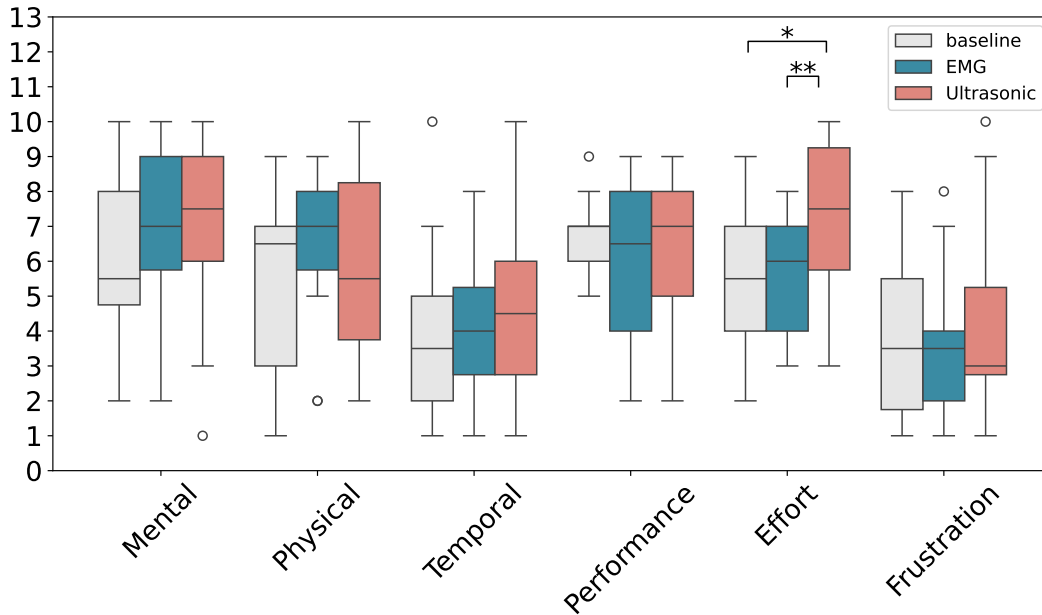


Figure 7: NASA-TLX Scores Comparison Across Baseline, EMG, and UI Methods with the Wilcoxon signed-rank test (*: $p < 0.05$, **: $p < 0.01$).

The decrease in EMG stability could be attributed to distraction, as participants primarily focused on maintaining stable visual feedback while managing the cognitive demands of interpreting raw EMG data. We further address this in the discussion section.

Vocal Lengths Controlling (UI). We analyzed UI recordings by annotating five frames per pitch (compared to one frame in Study 1) to increase measurement reliability. Mean vocal cord length increased for each pitch after UI-feedback training (Figure 6, right panel). Before multiple-comparison correction, we observed medium-to-large effects and trends toward significance at pitches B3 ($W = 3$, $p = .089$, $r = .49$, 95% CI $[-.12, .83]$) and C4 ($W = 7$, $p = .058$, $r = .52$, 95% CI $[-.08, .84]$), and a significant increase at D4 ($W = 4$, $p = .049$, $r = .57$, 95% CI $[-.01, .86]$). However, after controlling the false discovery rate across all eight pitches using the Benjamini-Hochberg procedure [4], none of these effects remained statistically significant (B3: $p_{FDR} = .24$, C4 and D4: $p_{FDR} = .23$). Despite the lack of FDR-corrected significance, the consistent positive effect sizes suggest a systematic tendency toward longer vocal cord configurations after training, which aligns with the pattern observed in the VCSD, where expert singers exhibited longer and more stable vocal cord lengths than intermediate participants.

5.3.2 Perception: EMG Enhanced Perceived Controllability, While UI Increased Mental Workload.

NASA-TLX. We compared the workload between three methods: (1) training with audio-only feedback (baseline), (2) training with audio and EMG visual feedback, and (3) training with UI feedback (see Figure 7). Our results show that both the training with EMG and UI feedback have a slightly higher workload than the baseline. Further, training with UI has a slightly higher workload than training with EMG.

When comparing UI against the Baseline condition, we observed medium-to-large effects for Performance ($W = 6$, $p = .071$, $r_{tb} = .56$, 95% CI $[-.09, .86]$, $p_{FDR} = .14$) and Effort ($W = 14$, $p = .044$, $r_{tb} = .63$, 95% CI $[-.03, .88]$, $p_{FDR} = .10$), with the latter approaching statistical significance after correction. Other dimensions (Mental, Physical, Temporal, Frustration) showed smaller effects and did not reach significance ($p_{FDR} > .30$). Comparing UI to EMG revealed a similar pattern: Effort showed a large and statistically significant increase for UI ($W = 1.5$, $p = .0094$, $r_{tb} = .71$, 95% CI $[-.18, .92]$, $p_{FDR} = .028$), while Mental and Physical workloads showed moderate but non-significant trends ($p_{FDR} = .18$ and $.13$). No significant differences were observed for Temporal, Performance, or Frustration ($p_{FDR} > .30$). It is expected that both of our sensing methods offer participants a direction of active practice to use their vocal muscles to sing, instead of adjusting the way to pronounce passively based on the pitch from audio feedback. Participants perceived a greater improvement in their vocal *Performance* following training with the proposed systems, as compared to solely relying on auditory feedback from hearing their own audio. This subjective assessment underscores the enhanced efficacy of the training methodologies implemented in our study.

Controllability. The analysis of the adapted SoAS questionnaire (see Figure 8), indicates an increased perception of controllability among participants using our proposed vocal muscle sensing system, as compared to three alternative training methodologies.

Before correction [4], we compared EMG-based feedback against UI feedback using Wilcoxon signed-rank tests that participants showed significantly higher controllability ratings for EMG in Q1 ($W = 3$, $p = .048$, $r_{tb} = .67$, 95% CI $[-.06, .91]$), Q2 ($W = 3$, $p = .045$, $r_{tb} = .68$, 95% CI $[-.07, .91]$), and Q5 ($W = 7$, $p = .013$, $r_{tb} = .74$, 95% CI $[-.22, .93]$). After FDR correction, the effect for Q5 remained significant ($p_{FDR} = .032$), while Q1 and Q2 showed marginal trends ($p_{FDR} = .072$). Q3 and Q4 did not show significant differences ($p_{FDR} > .40$), consistent with their smaller effect sizes ($r_{tb} < .30$). Notably, the EMG approach demonstrated superior controllability, possibly because of the more intuitive representation of EMG signals. Furthermore, the EMG methodology notably outperformed in Q5, which pertains to self-presentation confidence [61].

5.3.3 Audience Involvement: EMG Enhances Control and Reduces Fatigue; UI Visualizes Muscles but Lacks Precision. Audience involvement integrates expert evaluations and performer self-assessments to provide a comprehensive evaluation of vocal performance and the usability of sensors. Therefore, we first invited an expert group for scoring participants recorded audio across our three practice methods. Second, participants engaged in a playback session of their own recordings during an exit interview, incorporating their reflections into the evaluation.

Expert Group. Here is a brief description about three experts' background and their scoring criteria.

- Exp1: 5 years experience in music production and piano practice since childhood, this expert focused on evaluating both bass and tenor recordings, noting fluctuations in performance due to vocal fatigue.
- Exp2: An acoustic engineer with 15 years of piano practice and vocal knowledge, this expert prioritized vocal stability, accuracy, and loudness, emphasizing the impact of gender and vocal range on performance.
- Exp3: A former professional musical stage actor with over 10 years of experience, this expert assessed recordings based on vocal stability and pitch control, noting difficulties in low notes, particularly for female participants.

Rating of Experts. To ensure reliability, all audio evaluations were conducted on both pre- and post-training recordings using consistent recording equipment and environmental settings.

In our investigation, the expert evaluations focused on scoring participants recorded audio to appraise participants vocal performances separating in low&high pitch conditions of participants before&after they underwent training based on three critical aspects: Stability, Asthenia, and Harmonic. According to the data provided in Figure 9, the experts observed that vocal performances enhanced across all three aspects following training with the UI method. In contrast, training with the EMG method yielded somewhat diminished performance levels, as per expert assessments. Notably, the scores reflecting improvement in high-pitched vocalizations were more pronounced compared to those in lower pitches, exhibiting average enhancements of approximately 0.157 and 0.113 respectively, on a 5-point likert scale (see Appendix B). This scale was developed through expert discussions and references to previous research [14, 71].

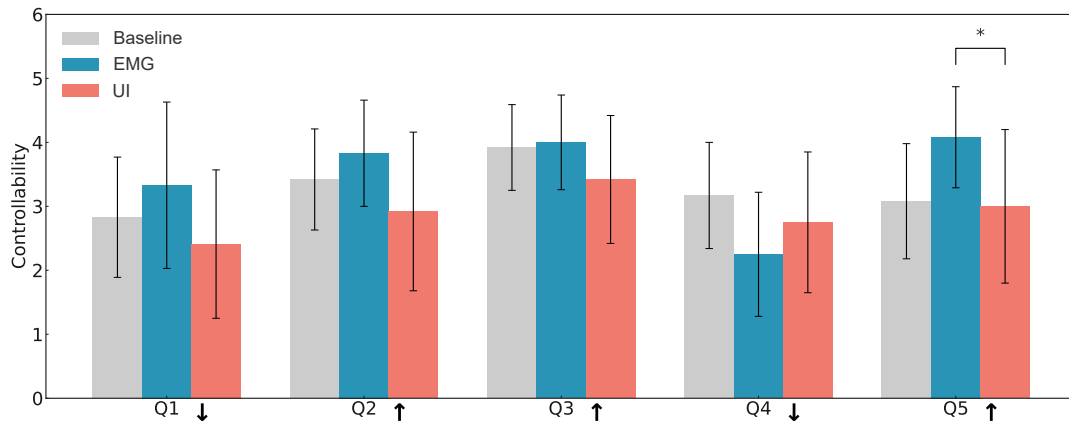


Figure 8: The results of the pick-up SoAS questionnaire with the Wilcoxon signed-rank test. (*: $p < 0.05$. Error bars represent standard deviations. Above questions (Q1-Q5) lists as following: (Q1) My actions just happen without my intention. (Q2) My behavior is planned by me from the very beginning to the very end. (Q3) I am completely responsible for everything that results from my actions. (Q4) My movements are automatic - my body simply makes them. (Q5) I am in full control of what I do.

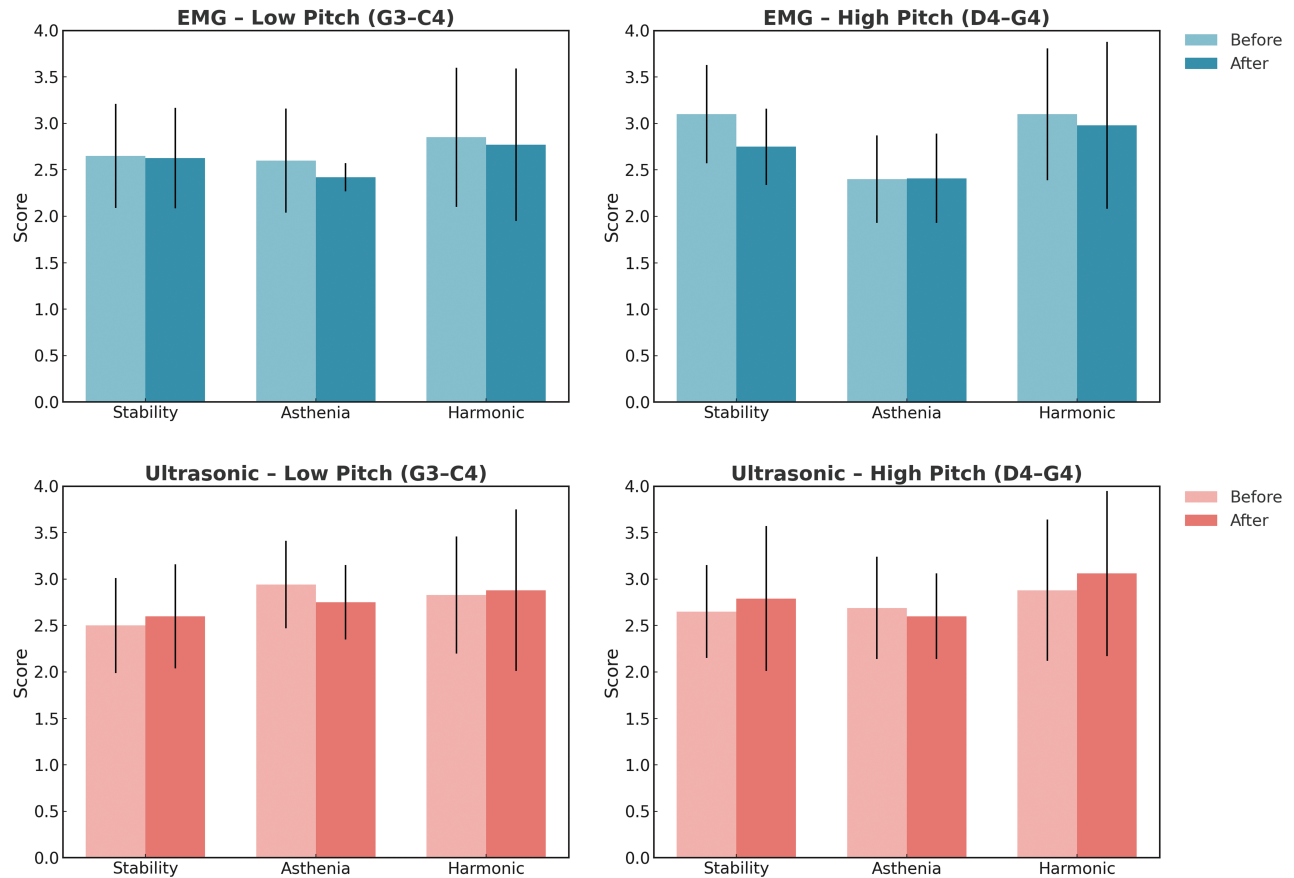


Figure 9: The outcomes of the expert evaluations regarding participants' singing performances are quantified in three distinct aspects: Stability↑, Asthenia↓, and Harmonic↑, shown in the bar chart above.

Participants and Experts Interviews. *Practice Feedback Suggestion from Experts:* Experts observed notable differences in the performance of vocal stability, harmonics, and asthenia across the training methods. Exp2 rated audio (condition 1 with expert's reference practice) and UI (condition 3 with ultrasonograph practice) feedback higher, stating, "1, 3 sound clearer and louder than 2 [with EMG practice]; there are fewer noises, less hoarseness and bubbling sounds." (Exp2). Exp3, however, noted the consistent difficulties participants had with low notes: "Females generally have difficulty in stably pronouncing the low notes... it may be better to change the key for female practice." (Exp3).

Participants' Reported Feedback on Sensing Methods. From the participant feedback, EMG proved more effective in enhancing control and reducing fatigue. As P6 noted, "Following the chart helps improve vocal position in real-time, and reduces vocal fatigue." (P6). Participants also highlighted how EMG offered real-time insights, with P7 commenting on the teacher's vocal stability: "I noticed that the teacher's vocal EMG changes were stable during vocalization." (P7).

Conversely, UI provided valuable visualizations but was less intuitive for control. P4 mentioned, "Ultrasound clearly shows the muscle movement process." but P8 found it challenging to control: "I find it hard to control my vocals to the exact position I want." Exp2 confirmed this, stating, "Ultrasonography sounds it takes slightly more efforts than EMG." (Exp2).

Participants found the baseline method to rely heavily on understanding the expert's reference. As P3 said, "I try to adjust pitch based on reference sound feedback.", indicating reliance on auditory feedback rather than active vocal control. EMG, on the other hand, facilitated real-time correction, while UI offered clear visualizations but lacked precision for improving vocal positioning.

5.3.4 Result Summary: RQ2. In sum, EMG feedback significantly enhanced the participants' perceived sense of control and self-presentation confidence, outperforming both the baseline and UI feedback. Despite this increased perception, actual muscle stability decreased post-practice, possibly due to cognitive overload from visual feedback interpretation. In contrast, UI feedback led to measurable improvements in vocal control, particularly in lengthening vocal cords for specific pitches, although some participants struggled with precise control due to usability challenges with the UI probe. Overall, both methods showed benefits, but with different strengths and challenges in practice.

6 STUDY 3: EXPERIENCED SINGERS STUDY (RQ3)

In our final study, we investigated the perceptions of experienced singers to investigate RQ3: How does a portable EMG setup perform in supporting experienced singers compared to traditional feedback methods?

6.1 Study Procedure

The study was approved by the local IRB, following a systematic procedure to ensure consistency across data collection:

1) *Welcome & Warm-Up:* The study started with a 10 minutes warm-up, guided by a vocal expert with over 10 years of professional musical stage experience.

2) *MVC Calibration:* Before the recording trials, participants performed a maximum voluntary contraction (MVC) task by sustaining a vowel (/a:/) at their highest comfortable intensity for about three seconds. This task was repeated multiple times within a 35-second calibration window, with a five-second rest period between each sustained vocalization. The EMG signal was monitored in real time using the Avanti Sensor interface (see the third picture in Figure 10), which automatically computed the maximum amplitude and baseline noise. This MVC value was used to normalize subsequent EMG data.

3) *Recording Trials:* Each participant practiced and mimicked *Christine's audition segment* from the musical "You Are Music"¹. Headphones were used to ensure accurate feedback and prevent interference from ambient noise.

4) *Feedback Sessions:* Next, participants were divided into focus groups, where they first received traditional feedback from the vocal expert. Second, they were presented with the EMG and audio analysis of their singing, and finally discussed with their peers about questions stated by the group moderating researcher.

6.2 Participants & Recruitment

We recruited 15 male experienced singers with the average age of 15 years ($Min=13$; $Max=17$; $SD=1.146$), all of whom had received continuous formal vocal training experience for more than one year. Consent of the parents was given before the study. The participants were either bass or tenor singers.

For the focus group discussion, the participants were randomly split into four groups, leading to one tenor-only, one bass-only and two mixed groups. For an overview of the allocation of participants to the focus groups see Table 2.

6.3 Results

Before analyzing our data, we had to process it to extract key features, such as EMG Normalized Root Mean Square (EMG RMS) and Singing Power Ratio (SPR). The EMG RMS reflects vocal cord muscles activation [14, 34], indicating muscle coordination and strength, while SPR [67] measures energy distribution in the voice, enabling accurate tracking of vocal power and resonance across different frequencies. The Raw EMG data (sampled at 4370Hz) was normalized using the Maximum Voluntary Contraction (MVC) method [2] to allow for consistent muscle activity comparisons. Data were segmented into 200ms windows to calculate Root Mean Square (RMS) values. For the SPR, microphone array data (sampled at 48kHz) were first processed with a band-pass filter (500–4000 Hz) to suppress low-frequency rumble and high-frequency noise. Singing Power Ratio (SPR) was then calculated using the Librosa package [37], following standard acoustic definitions of SPR [67]. Librosa

¹The musical notation in Fig 11 is an original illustration created by the author based on "You Are Music" by Maury Yeston, © 1991, purchased from Musicnotes.com. This usage falls under the Fair Use doctrine for academic purposes.



Figure 10: Photos of participants during the experienced singers study: (1) Stage rehearsal for the Phantom musical. (2) Warm-up sessions ensuring participants' proficiency in the selected solfeggio before recording. (3) Participant wearing the sensing setup for the solfeggio recording. (4) The focus group discussion were conducted together with experts, amateurs and researchers.

Focus Group	Participants	Expertise	Year	PerAb ↑	SinAb ↑	RMS Mean (SD) ↑	SPR Mean (SD)↑
F1	P1.1	Tenor	2	41	25	-57.30 (165.18)	-0.248 (0.367)
	P1.2	Tenor	1	39	23	-75.33 (154.78)	-0.247 (0.367)
	P1.3	Tenor	2	43	34	-33.77 (111.40)	-0.247 (0.367)
	P1.4	Tenor	> 10	43	37	-33.74 (131.66)	-0.246 (0.366)
F2	P2.1	Bass	2	28	32	-125.87 (157.67)	-0.245 (0.366)
	P2.2	Bass	4-5	38	12	-27.00 (162.21)	-0.247 (0.367)
	P2.3	Bass	4-5	44	33	-	-0.246 (0.366)
	P2.4	Bass	< 1	28	24	-88.35 (150.10)	-0.244 (0.365)
F3	P3.1	Tenor	4-5	40	30	-95.24 (153.60)	-0.247 (0.369)
	P3.2	Bass	1	44	30	-90.01 (159.65)	-0.247 (0.367)
	P3.3	Tenor	> 10	39	26	-76.93 (146.11)	-0.246 (0.367)
	P3.4	Bass	4-5	40	32	-90.12 (153.37)	-0.244 (0.365)
F4	P4.1	Tenor	3	46	24	-58.34 (175.46)	-0.248 (0.367)
	P4.2	Bass	1	38	26	-58.63 (129.48)	-0.238 (0.358)
	P4.3	Bass	2	51	33	-93.59 (155.78)	-0.244 (0.366)

Table 2: Participant to Focus Group Allocation of our experienced singers study. Demographics are collected by the Goldsmiths Musical Sophistication Index (Gold-MSI) [42]: Year (Experience), PerAb (Perceptual Ability), SinAb (Singing Ability). Descriptive Statistics of EMG and SPR are calculated separately for students and expert by the difference of normalized EMG Root Mean Square (RMS) and SPR.

was used for signal loading, Short-Time Fourier Transform (STFT) computation, and spectral energy extraction. Specifically, we applied `librosa.stft()` to obtain the magnitude spectrogram, from which SPR was derived as the ratio of energy in the high-frequency band (2–4 kHz) to that in the low-frequency band (0.5–1 kHz). The band energies were computed by summing the squared magnitude values across their respective frequency bins over time as follows:

$$SPR = 10 \cdot \log_{10} \left(\frac{\int_{2 \text{ kHz}}^{4 \text{ kHz}} P(f) df}{\int_{0.5 \text{ kHz}}^{1 \text{ kHz}} P(f) df} \right) \quad (3)$$

6.3.1 Sensing: Tenor Group Shows Higher Muscle Engagement; Expert Displays Strong EMG-SPR Correlation.

Tenor versus Bass. The Tenor Group had RMS Mean differences (RMS and SPR differences were calculated by the student's *Mean Value* minus the expert's) ranging from -33.73 to -95.23, with moderate variability (see Table 2). In contrast, the Bass Group exhibited larger differences (-125.87 to -26.9972) and greater variability, indicating less consistent muscle engagement. Both groups demonstrated stable SPR differences around -0.245, with minimal variability.

The Tenor Group displayed higher and more consistent EMG RMS values, indicating greater muscle engagement, while the Bass Group exhibited wider variability and lower muscle activation. These differences are further illustrated in the group comparison Figure 11, where the tenor group appears to have more controlled muscle activation, closer to expert performance.

Expert EMG demonstrated significantly higher and more flexible RMS values than the trained group, suggesting more controlled and efficient muscle activation. These results align with the dataset findings, reinforcing the expert's superior performance.

EMG versus SPR. We downsampled EMG RMS and SPR, aligned them with the notes and calculated Pearson correlations across the expert and student participants. Pearson correlation coefficients between EMG and SPR revealed a strong positive correlation for the expert ($r = .75, p < .01$), indicating a significant relationship between vocal muscle activation and singing power.

Specifically, from the observation of Pearson values in Figure 12, we see that the expert and experienced student show similar acoustic patterns, with more positive correlations, suggesting that their acoustic characteristics align despite differences in physiological responses, indicating similar techniques.

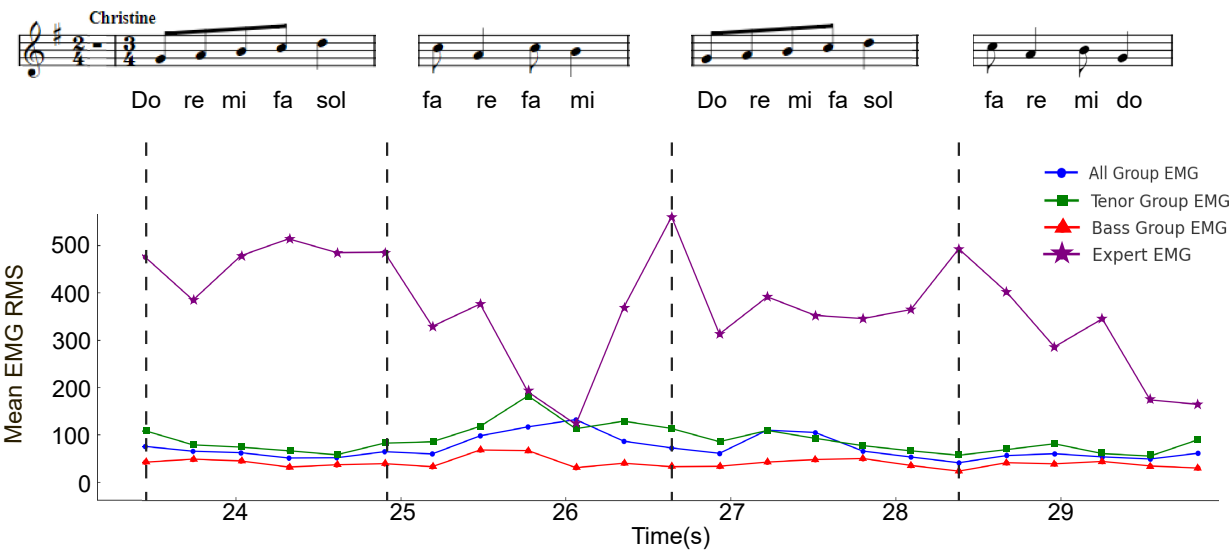


Figure 11: Upper: The song 'You Are Music' has a tempo of 102 BPM, with 3 beats per measure and a beat duration of 0.588 seconds, equivalent to a quarter note or two eighth notes (0.292 seconds). Bottom: The line chart compares EMG RMS values for different participant groups and an expert over the time period from 23s to 30s. It displays the average EMG RMS values for the All Group (blue), Tenor Group (green), Bass Group (red), and the Expert (purple), illustrating differences in muscle activity and performance between the expert and the student groups.

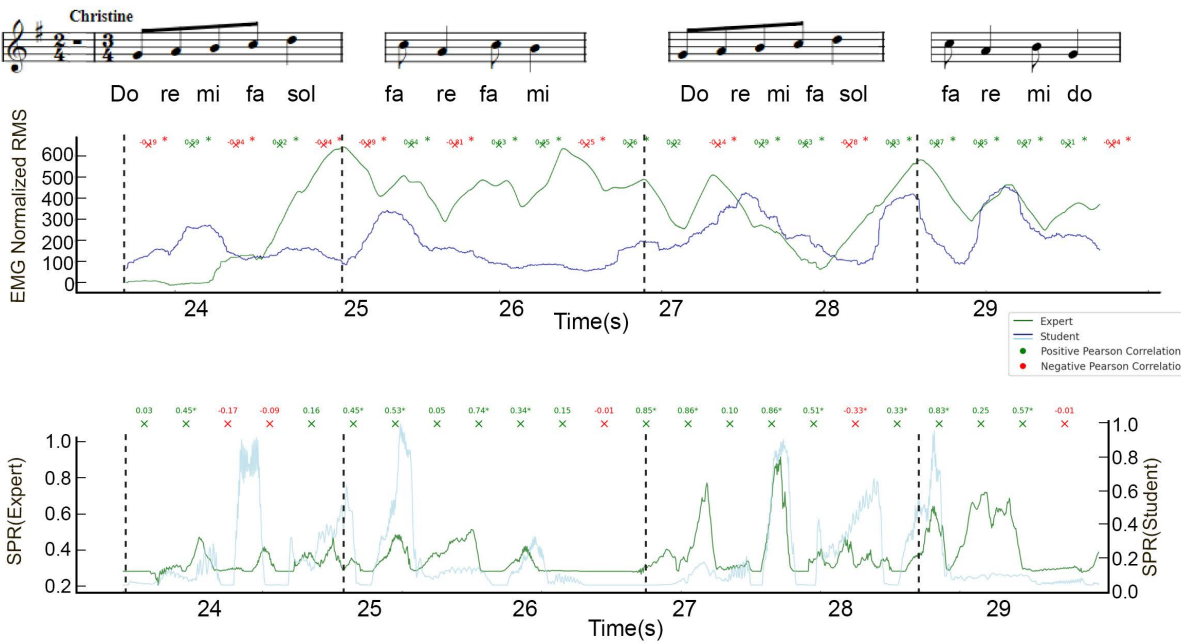


Figure 12: In the period (21s-28s) of the singing segment (upper graph), the trends in EMG and SPR of an expert, and a student who leads the Christine role, are shown. Significance markers (asterisks) are placed above certain correlation values to denote statistical significance.

Professional participants' correlations varied from approximately $r = .09$ to $r = .53$, showing different levels of engagement with the sensing feedback. This finding highlights a closer relationship between SPR and EMG signals in experts, potentially reflecting higher coordination between vocal performance and muscle control.

Principal Component Analysis (PCA). We further applied PCA and statistical tests using Python to explore differences between professional groups (Tenor and Bass) based on three Factors of PerAb (Perceptual Ability), SinAb (Singing Abilities), and RMS Mean. One participant (P2.3) was excluded as the outlier since EMG sensor could not be securely attached due to their moisture skin.

PCA reduced the dataset's dimensionality by extracting two principal components that explained 86.29 percent of the total variance. Principal Component 1 (PC1), which explained 44.26 percent of the variance, was influenced by PerAb and RMS Mean Difference. Principal Component 2 (PC2), explaining 42.03 percent, was driven by SinAb and RMS Mean Difference. These components effectively capture the variation in perceptual and singing abilities and muscle signals.

In the PCA space, the Tenor group appeared more clustered, indicating consistency among participants, while the Bass group was more spread out, particularly along PC1, suggesting greater individual variability. To test for group differences, we conducted t -tests and a Wilcoxon signed-rank test. The t -tests showed no significant differences in PerAb ($t(8.64) = 0.91, p = .388$, Cohen's $d = 0.44$, 95% CI $[-4.05, 9.44]$), SinAb ($t(12.74) = 0.21, p = .838$, $d = 0.11$, 95% CI $[-6.37, 7.72]$), or EMG Mean Difference ($t(11.02) = 1.40, p = .189$, $d = 0.75$, 95% CI $[-11.68, 52.51]$). The Wilcoxon signed-rank test also did not reveal significant differences.

6.3.2 Focus Group: Participants appreciated the precision of technology-assisted training but recommended combining it with traditional coaching for clearer feedback. To analyze the interview data from the focus groups, we first transcribed the interviews into written form. Next, we did an inductive coding. Therefore, first one researcher familiarized themselves with the data and proposed a codebook. Second, two coder applied the codebook to half of the transcripts. Finally, both coder validated each others' assessments, and disagreements were resolved in a review meeting. We followed suggestions by literature, *not* to do multiple independent codings and calculate Inter Coder Reliability (ICR) to prove reliability [5, 36, 46], acknowledging the influence of the researcher on the process. After this, both researchers grouped the codes into our main themes.

Theme 1: Impressions of Technology-Assisted Training. When we first explained the visualization to the participants, some struggled to understand it immediately. After answering participants' questions, all participants were confident to gain an even better understanding the more frequent they engage with the visualization. Aspects of the visualization participants were most interested in were (1) the TIME INTERVAL, (2) their applied MUSCLE FORCE represented in the visualization, and (3) VOICE DIFFERENCES between the "user", i.e. male participants, and the female expert-baseline, their singing was compared against.

Participants from F1 and F2 recognized that the visualization showed timing differences between when they started to sing compared to the expert-baseline. P1.1 suggested to use a 0.1 seconds TIME INTERVAL (instead of 0.2 seconds) for visualizing the data even more accurately.

Further, participants showed an interest in how their applied MUSCLE FORCE is represented to see how they have to adjust their own muscle, as P1.3 asked "Well, this vertical axis is the amount of muscle power applied, right?" (P1.3). At the same time, the visualization of the MUSCLE FORCE made some participants realize that they could improve their singing by increasing the muscle control.

Finally, participants wondered if the gender of the expert that sang in the training-baseline should match the gender of the person training with the baseline. P1.4 and P1.2 concluded that "There are differences between male and female voices." (P1.4), so "It's better to do it with the same gender." (P1.2). Participants from the bass group F2 considered it as helpful for their singing to see the different amplitudes in the graph of female and male singers: "The second graph shows the difference between the male and female voices, so the amplitude of their voices is completely different." (P2.1).

Theme 2: Limitations of Traditional Vocal Training. The participants' discussion about their first impression of the visualization transitioned into discussing their singing voice types (bass or tenor), and how traditional training is limited to address specific aspects of it.

Participants from the tenor group especially experienced challenges with reaching high pitches. They discussed that they often reached a threshold which they cannot overcome, and that additional difficulties can occur for persons singing in *falsetto*². Because of this, some participants mentioned to sing "based on instinct" (P1.4) to achieve best performance, or trying to imitate the original singer of a song. Participants referred to this as "intuitive training", meaning that they do not follow a dedicated training schedules or routine, but rather sing intuitively during their day, as P3.1 said: "I sing in the bath" (P3.1).

In contrast, other participants mentioned to have dedicated training routines including listening to others singing a song before learning it themselves, singing karaoke, doing certain vocal exercises, or recording themselves singing to listen to it. Additionally, P3.4 mentioned to sing during his workouts to include various body muscles in his singing: "I was told that if I push a desk while singing, I can learn how to use my stomach, so I'm doing it." (P3.4). However, solely relying on auditory feedback was perceived as less effective to improve the learners' vocal *Performance* compared to training with a technology-assisted method as also found in our novice study (see Section 5).

Within their traditional training, participants experienced challenges in controlling their natural and singing voice, finding a balance between their normal voice and falsetto, and memorizing their training without visual cues that they can re-check at a later time.

²Falsetto [70] is the vocal register occupying the frequency range just above the modal voice register and overlapping with it by approximately one octave. It is produced by the vibration of the ligamentous edges of the vocal cords, in whole or in part.

Theme 3: Advantages of Technology-Assisted Training. Our participants identified several advantages of training with our visualization compared to traditional training. First, they considered the visualization to be more precise, as it directly points out which parts of the singing need improvement. This makes it easier for participants to “see” the volume and the pitch of their voice at one glance. Further, participants reported that this helps them to easier understand on which parts of their muscle they need to focus: “*So, when I use such a graph, it is easier to understand the specifics from my muscles.*” (P1.4). Participants also felt that the visualization was beneficial when it came to language barriers between them and their traditional coaches. Further, participants liked that they were able to confirm their expectation of singing with the visualization. Participants from the bass group compared the amplitude of their singing with the tenor group: The bass groups’ visualization had less fluctuation compared to the tenor group, which is in line with expectations when comparing bass and tenor singers. Participants therefore expressed a feeling of relief when identifying this behavior with the visualization: “*I’m in the bass group, so I’m high stability that’s how it is. That’s good. I’m relieved.*” (P2.1). Similar observations were made in the mixed group F4. There, participants discussed the difference between tenor and bass voices, comparing the different types to the expert-baseline.

Theme 4: Concerns towards Technology-Assisted Training. Our participants not only considered advantages of technology-assisted training but also expressed some concerns. Participants considered that singing coaches need to set some expert singing to be the standard-baseline for singing when they employ the visualization feedback in their current training. This led to the concern that students might not be able to fulfill these criteria anymore. Further, they wondered whether the perception of “good” singing might change, once there is a technical solution to evaluate someone’s singing: “*When the machine thinks it’s good, everyone thinks it’s good.*” (P4.3), only leaving little room for subjective interpretation. Participants further missed the human interaction with their coach, and P1.4 argued that interacting only with a numerical representation in the visualization is not as natural, since “*A human being is not a machine. So, it is not something that can be moved by inputting numerical values. In the end, we use our nerves, our senses, and our bodies, so I think it’s a little difficult to understand even if you look at it numerically.*” (P1.4).

Theme 5: Future of Technology-Assisted Training. Our participants came up with ideas how they would like to incorporate the technology-assisted training in their current routines. The participants’ ideas to **include the visualization as it is** are (1) to get an overview of their progress, (2) as a substitute for finding a learning group, (3) as a substitute for a coach when training at home, or (4) as an addition when training karaoke.

However, participants also had several suggestions on how to combine the visualization with traditional training which we detail below. When comparing traditional feedback from their coach to the feedback provided by the visualization, participants considered the form of feedback: While their coach can provide them feedback combined with suggestions how to improve their singing, with the visualization participants had to find out by trying themselves what they need to do in order to improve. Hence, participants expressed

a need to **supplement the visualization by advice**, similar to the advice a coach would give.

In contrast to supplementing the visualization by advice, P1.4 suggested to **use the visualization as a supplement to traditional feedback**. Participants argued that it is hard to understand the correlation between the visualization and muscle movement solely from the visualization. Making participants suggest that a connection between what they see and feel is necessary, and could be achieved by discussing the visualization with their coach during traditional training.

Overall, the participants expressed a preference to **use a mixture of traditional feedback and the new visualization** method, as P1.2 stated “*I don’t think it is right to choose only one of them.*” (P1.2).

6.3.3 Result Summary: RQ3. Traditional feedback currently encounters several challenges in terms of availability when no coach is present, language-barriers between coach and learner, and precision, e.g., when learners have to review recordings of their singing, leading to delayed inaccurate feedback. Despite some concerns how technology-assisted training could lead to higher expectations and singing standards, all our focus groups found positive aspects and opportunities on how to employ it. Participants therefore came up with ideas on how to combine traditional and technology-assisted training in their current routines.

7 DISCUSSION

Our three studies demonstrate that EMG and ultrasonography provide complementary feedback modalities for vocal training, with effectiveness varying by skill level and voice type. Study 1 established that these sensing technologies can differentiate muscle activation patterns between novice and experienced singers. Study 2 showed that novices improved vocal performance with both technologies, though each imposed distinct cognitive demands. Study 3 revealed that experienced singers exhibited strong coordination between muscle activation and acoustic output, with notable differences between tenor and bass voice types. Below, we synthesize these findings to address how physiological sensing can augment vocal pedagogy.

7.1 Summary of Key Study Results

Both EMG and UI demonstrated promising capabilities in detecting and differentiating vocal muscle activity across different skill levels.

In our second user study with novice singers, EMG enhanced the participants’ perceived control, even though the stability decreased post-training, while UI improved vocal cord length control but required more effort to interpret. The decrease of the stability likely is a temporary effect that can have two possible explanations. First, this temporary decline should not be interpreted as performance loss, but rather as a transitional adaptation phase. Novices are required to recruit and coordinate muscles in a way that differs from intuitive singing, and such re-coordination is often accompanied by short-term instability. Prior vocal pedagogy literature similarly notes that increased or misdirected tension can momentarily degrade sound quality when new techniques are introduced [40]. Our findings align with this view: the sensing feedback drew attention

to specific muscle groups, prompting novices to consciously engage muscles they previously controlled implicitly, resulting in short-term variability. Second, novices need to “re-learn” their ability to control their vocal cords in a new way compared to singing by intuition. More long-term training lets individuals regain this stability. As shown in our first study, novices typically have less stability in their EMG signals compared to more experienced singers.

In our third study, (the experienced singers study) we revealed a strong correlation between EMG and singing power in experts, with Tenor participants showing more consistent muscle engagement than Bass participants. While participants appreciated the precision of the sensing feedback, they also expressed concerns about over-reliance on numerical data and emphasized the importance of integrating technology with traditional vocal training for a more balanced approach.

Our studies collectively showed that both EMG and UI provide valuable insights into vocal muscle control, but their impact on perception differs. EMG offers intuitive, real-time feedback, improving muscle engagement and control, while UI provides detailed visual cues for vocal cord movement, helpful in early-stage training or refining specific techniques. The need for such visual feedback was also evident in our final study with the experienced singers where we investigated EMG exclusively. Our first study revealed significant differences between beginners and experts, with EMG excelling in temporal stability and UI in visualizing muscle activity. The novice study highlighted a trade-off between perceived control and cognitive overload with EMG, while UI required more effort to interpret but improved vocal control. The experienced singers study further demonstrated that tenor singers exhibited better muscle engagement, and a strong correlation between EMG and SPR was observed in expert singers, indicating higher coordination between muscle activity and vocal performance.

7.2 Enhancing Muscle Control and Vocal Performance

7.2.1 Traditional versus Technical. One current challenge of singing training revealed by our experienced singers study is that frequent training is necessary, but often no learning groups for extracurricular learning (without coach) are available. This leads to singers currently rely solely on auditory feedback from recordings of their own voice. But results of our novice study showed that learners can reach a greater improvement of their vocal *Performance* when utilizing UI or EMG as training methods.

7.2.2 Tenor versus Bass. In comparing EMG and UI for skill transferability, our findings showed that EMG stability decreased post-training for novices, especially in lower pitches, while UI improved vocal cord length control, particularly in higher pitches. In the professional group, tenors exhibited more consistent muscle control than bass singers, who showed greater variability and difficulty in skill transfer. Pearson correlations between EMG and SPR highlighted stronger coordination in the expert, with students, especially bass singers, showing lower correlations. PCA analysis confirmed that tenors had more consistent performance, while bass singers displayed higher variability, indicating the need for more

targeted feedback to improve muscle stability. Overall, EMG benefits tenors more, while UI aids novices in muscle coordination, with bass singers requiring additional support.

7.3 Implicating the Design Guideline for Musical Training and Performance

In designing effective vocal training systems, the integration of EMG and UI offers significant advantages by providing detailed feedback on muscle activity and vocal cord movements. EMG excels in capturing fine motor control in vocal cords, offering precise data for vocal performance, while UI provides direct visual cues, helping users understand complex vocal mechanisms.

Building on advances in Human-Computer Interaction for music performance, such as BrainiBeats [7], these technologies can enhance not only vocal pedagogy but also applications in speech therapy and personalized music education. However, usability challenges with UI, especially as a wearable, highlight the need for refining its design for broader accessibility. Combining these sensing technologies into user-friendly, real-time systems could support continuous feedback, improving training for users at all skill levels.

7.4 Limitations and Future Works

Longer-term Training Periods. One limitation we identified is that evaluating pitch immediately after training may not fully reflect long-term effects. To explore this, we recalled 6 novice participants to assess their performance on “Little Star.” The EMG method showed average improvements of 8.16 Hz in pitch accuracy and 7.53 Hz in consistency [37], while the UI method had smaller gains. These results suggest both methods enhance pitch control, with EMG showing stronger effects. Future work should focus on how sensory feedback can support long-term improvements in muscle control and vocal performance.

Interface Design. We recognized the cognitive load in interpreting raw EMG and UI data in the novice study and the need for improved visual feedback in the experienced singers study. Future systems should offer more intuitive feedback aligned with expert models and extended training routines. Additionally, current sensing is limited by Bluetooth interference between the EMG sensor and ultrasound probe, which we aim to resolve with a dual-modal interface and automated UI using neural networks.

Group Variability and Sample Size Limitations. Although PCA analysis revealed different distributions between the Tenor and Bass groups in the principal component space, the statistical tests failed to show significant group differences. This may be attributed to the small sample size or high within-group variability, indicating that further research with a larger sample size is needed to explore these potential differences in more depth.

Gender and Developmental Confounds. In Study 3, our male adolescent participants (ages 13–17) were compared against a female expert baseline. We chose this baseline because the musical piece (“You Are Music”) is written for female soprano, but this creates a confound: male and female singers have different laryngeal anatomy, so differences in EMG and SPR (Figure 11) may reflect sex rather than skill level. Our adolescent participants were also undergoing vocal development, adding further variability. Future

work should use sex- and age-matched baselines to better isolate training effects.

8 CONCLUSION

In conclusion, this research explored the vocal cord muscles activities using novel sensing (EMG and UI) for vocal performance assessment. Through our user studies with novice, experienced and expert singers (44 in total), we demonstrated that these sensing technologies can effectively capture vocal muscle activity and provide valuable feedback for improving vocal control. While both methods showed promise in enhancing pitch accuracy and stability, future work is needed to optimize user experience and investigate long-term training effects. Our findings highlight the potential of integrating advanced sensing technologies into vocal training, offering more precise, intuitive feedback to support diverse user needs in vocal education and performance.

Acknowledgments

This work is conducted under the IoT Accessibility Toolkit Project supported by JST Presto Grant Number JPMJPR2132 and funded by JST SPRING H09GQ24152, collaborating with Koike lab at the Institute of Science Tokyo under the fundings by JST CRONOS JPMJCS24N8.

References

- [1] John B Allen. 2019. Advanced waveform analysis for improving the quality of vocal recordings. *Journal of the Acoustical Society of America* 146, 4 (2019), 2457–2468.
- [2] Sridhar P Arjunan, Dinesh K Kumar, and Ganesh Naik. 2014. Computation and evaluation of features of surface electromyogram to identify the force of muscle contraction and muscle fatigue. *BioMed research international* 2014, 1 (2014), 197960.
- [3] Michel Belyk, Yune S Lee, and Steven Brown. 2018. How does human motor cortex regulate vocal pitch in singers? *Royal Society Open Science* 5, 8 (2018), 172208.
- [4] Yoav Benjamini and Yoel Hochberg. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal statistical society: series B (Methodological)* 57, 1 (1995), 289–300.
- [5] Virginia Braun and Victoria Clarke. 2013. *Successful Qualitative Research: A Practical guide for Beginners*. SAGE Publications, London.
- [6] Fritz Buchthal. 1959. Electromyography of intrinsic laryngeal muscles. *Quarterly Journal of Experimental Physiology and Cognate Medical Sciences: Translation and Integration* 44, 2 (1959), 137–148.
- [7] Caterina Ceccato, Ethel Pruss, Anita Vrins, Jos Prinsen, and Maryam Alimardani. 2023. BrainBeats: A dual brain-computer interface for musical composition using inter-brain synchrony and emotional valence. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) (CHI EA '23). Association for Computing Machinery, New York, NY, USA, Article 56, 7 pages. doi:10.1145/3544549.3585910
- [8] Kanyu Chen, Zhuang Chang, Qianyan Zou, and Kai Kunze. 2025. Exploring Singing Breath: Physiological Insights and Directions for Breath-Aware Augmentation in Mixed Reality Design. In *Companion of the 2025 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. 702–706.
- [9] Kanyu Chen, Akira Kato, and Kai Kunze. 2025. A Multimodal Wearable Sensing System for Vocal Muscle Biofeedback in Singing Pitch Training. In *Companion of the 2025 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. 449–454.
- [10] Kanyu Chen, Erwin Wu, Daichi Saito, Yichen Peng, Chen-Chieh Liao, Akira Kato, Hideki Koike, and Kai Kunze. 2024. Novel Sensing Methods for Vocal Technique Analysis: Evaluation on Electromyography and Ultrasonography. In *2024 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*. IEEE, 121–125.
- [11] Jacob Cohen. 2013. *Statistical power analysis for the behavioral sciences*. routledge.
- [12] Leon Cohen. 1995. *Time-frequency analysis*. Vol. 778. Prentice hall New Jersey.
- [13] Kelsey Cotton, Pedro Sanches, Vasiliki Tsaknaki, and Pavel Karpashevich. 2021. The Body Electric: A NIME designed through and with the somatic experience of singing. In *NIME 2021*. <https://nime.pubpub.org/pub/ntm5kbux>.
- [14] Marc S De Bodt, Floris L Wuyts, Paul H Van de Heyning, and Christophe Croux. 1997. Test-retest study of the GRBAS scale: influence of experience and professional background on perceptual rating of voice quality. *Journal of voice* 11, 1 (1997), 74–80.
- [15] Maude Desjardins, Lucinda Halstead, Annie Simpson, Patrick Flume, and Heather Shaw Bonilha. 2022. Respiratory muscle strength training to improve vocal function in patients with presbyphonia. *Journal of Voice* 36, 3 (2022), 344–360.
- [16] Ethan Eddy, Erik J Scheme, and Scott Bateman. 2023. A Framework and Call to Action for the Future Development of EMG-Based Input in HCI. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) (CHI '23). Association for Computing Machinery, New York, NY, USA, Article 145, 23 pages. doi:10.1145/3544548.3580962
- [17] Daniele Esposito, Emilio Andreozzi, Antonio Fratini, Gaetano D Gargiulo, Sergio Savino, Vincenzo Niola, and Paolo Bifulco. 2018. A piezoresistive sensor to measure muscle contraction and mechanomyography. *Sensors (Basel)* 18, 8 (Aug. 2018).
- [18] Mireia Farrús, Javier Hernando, and Pascual Ejarque. 2007. Jitter and shimmer measurements for speaker recognition. In *8th Annual Conference of the International Speech Communication Association; 2007 Aug. 27-31; Antwerp (Belgium)*. [place unknown]: ISCA; 2007. p. 778-81. International Speech Communication Association (ISCA).
- [19] Thomas Gay, Marshall Strome, Hajime Hirose, and Masayuki Sawashima. 1972. Electromyography of the intrinsic laryngeal muscles during phonation. *Annals of Otolaryngology, Rhinology & Laryngology* 81, 3 (1972), 401–409.
- [20] Andrew Gelman and Jennifer Hill. 2007. *Data analysis using regression and multilevel/hierarchical models*. Cambridge university press.
- [21] Jiawen Han, George Chernyshov, Moe Sugawa, Dingding Zheng, Danny Hynds, Taichi Furukawa, Marcelo Padovani, Kouta Minamizawa, Karola Marky, Jamie A Ward, et al. 2022. Linking Audience Physiology to Choreography. *ACM Transactions on Computer-Human Interaction* (2022).
- [22] Sandra G Hart and Lowell E Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Human mental workload* 1, 3 (1988), 139–183.
- [23] Yan He, George Chernyshov, Jiawen Han, Dingding Zheng, Ragnar Thomsen, Danny Hynds, Muyu Liu, Yuehui Yang, Yulan Ju, Yun Suen Pai, et al. 2022. Frisson waves: exploring automatic detection, triggering and sharing of aesthetic chills in music performances. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 6, 3 (2022), 1–23.
- [24] Peng Huang, Yuan Li, and Zheng Wang. 2021. Enhanced waveform analysis for detecting vibrato and pitch modulation in vocal performances. *Journal of the Acoustical Society of America* 150, 3 (2021), 1753–1762.
- [25] Gail B Kempster, Charles R Larson, and Michael K Kistler. 1988. Effects of electrical stimulation of cricothyroid and thyroarytenoid muscles on voice fundamental frequency. *Journal of Voice* 2, 3 (1988), 221–229.
- [26] Naoki Kimura, Michinari Kono, and Jun Rekimoto. 2019. SottoVoce: An ultrasound imaging-based silent speech interaction using deep neural networks. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–11.
- [27] Markus Krause, Jan Smeddinck, and Ronald Meyer. 2013. A digital game to support voice treatment for parkinson's disease. In *CHI '13 Extended Abstracts on Human Factors in Computing Systems* (Paris, France) (CHI EA '13). Association for Computing Machinery, New York, NY, USA, 445–450. doi:10.1145/2468356.2468435
- [28] Amarjeet Kumar, Chandni Sinha, Akhilesh Kumar Singh, and Umesh Kumar Bhadani. 2017. Vocal cord dysfunction: Ultrasonography-aided diagnosis during routine airway examination. *Saudi Journal of Anaesthesia* 11, 3 (2017), 370–371.
- [29] Daniël Lakens. 2013. Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. *Frontiers in psychology* 4 (2013), 863.
- [30] Jaeeun Lee, Hyeyoon Cho, and Yonghyun Kim. 2023. Bean Academy: A Music Composition Game for Beginners with Vocal Query Transcription. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) (CHI EA '23). Association for Computing Machinery, New York, NY, USA, Article 591, 6 pages. doi:10.1145/3544549.3583824
- [31] Jihun Lee, Hyun Kim, and Soyeon Park. 2022. Real-time feedback system for classical singing training using singing power ratio. *Journal of Voice* 36, 1 (2022), 129–136.
- [32] Pedro Lopes, Alexandra Ion, and Patrick Baudisch. 2015. Impacto: Simulating Physical Impact by Combining Tactile Stimulation with Electrical Muscle Stimulation. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology* (Charlotte, NC, USA) (UIST '15). Association for Computing Machinery, New York, NY, USA, 11–19. doi:10.1145/2807442.2807443
- [33] Kally M Luck, Dorothea C Lerman, Wai-Ling Wu, Danielle L Dupuis, and Louisa A Hussein. 2018. A comparison of written, vocal, and video feedback when training teachers. *Journal of Behavioral Education* 27 (2018), 124–144.
- [34] Hiroyuki Manabe, Akira Hiraui, and Toshiaki Sugimura. 2003. Unvoiced speech recognition using EMG-mime speech recognition. In *CHI '03 extended abstracts*

- on Human factors in computing systems. 794–795.
- [35] Josef William McClellan. 2011. A comparative analysis of speech level singing and traditional vocal training in the United States. (2011).
 - [36] Nora McDonald, Sarita Schoenebeck, and Andrea Forte. 2019. Reliability and Inter-rater Reliability in Qualitative Research: Norms and Guidelines for CSCW and HCI Practice. *Proc. ACM Hum.-Comput. Interact.* 3, CSCW, Article 72 (nov 2019), 23 pages. doi:10.1145/3359174
 - [37] Brian McFee, Colin Raffel, Dawen Liang, Daniel PW Ellis, Matt McVicar, Eric Battenberg, and Oriol Nieto. 2015. librosa: Audio and music signal analysis in python.. In *SciPy*. 18–24.
 - [38] Xiaru Meng, Yan He, and Kai Kunze. 2023. Towards enhancing a recorded concert experience in virtual reality by visualizing the physiological data of the audience. In *Proceedings of the Augmented Humans International Conference 2023*. 330–333.
 - [39] Xiaru Meng, Yulan Ju, Christopher Changmok Kim, Yan He, Giulia Barbareschi, Kouta Minamizawa, Kai Kunze, and Matthias Hoppe. 2025. A Placebo Concert: The Placebo Effect for Visualization of Physiological Audience Data during Experience Recreation in Virtual Reality. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. 1–16.
 - [40] Richard Miller. 1996. The structure of singing: System and art in vocal technique. (*No Title*) (1996).
 - [41] Fabio Morreale, Raul Masu, Antonella De Angeli, and Paolo Rota. 2013. The music room. In *CHI '13 Extended Abstracts on Human Factors in Computing Systems* (Paris, France) (*CHI EA '13*). Association for Computing Machinery, New York, NY, USA, 3099–3102. doi:10.1145/2468356.2479620
 - [42] Daniel Müllensiefen, Bruno Gingras, Jason Musil, and Lauren Stewart. 2014. The musicality of non-musicians: An index for assessing musical sophistication in the general population. *PLoS one* 9, 2 (2014), e89642.
 - [43] Masaaki Nagae, Hiroyuki Umegaki, Akito Yoshiko, and Kosuke Fujita. 2023. Muscle ultrasound and its application to point-of-care ultrasonography: a narrative review. *Ann. Med.* 55, 1 (Dec. 2023), 190–197.
 - [44] Andrew Naseth. 2012. Constructing the voice: Present and future considerations of vocal pedagogy. *The Choral Journal* 53, 2 (2012), 39.
 - [45] Alan V Oppenheim. 1999. *Discrete-time signal processing*. Pearson Education India.
 - [46] Anna-Marie Orloff, Matthias Fassl, Alexander Ponticello, Florin Martius, Anne Mertens, Katharina Krombholz, and Matthew Smith. 2023. Different Researchers, Different Results? Analyzing the Influence of Researcher Experience and Data Type During Qualitative Analysis of an Interview and Survey Study on Security Advice. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) (*CHI '23*). Association for Computing Machinery, New York, NY, USA, Article 864, 21 pages. doi:10.1145/3544548.3580766
 - [47] Michalis Papakostas, Varun Kanak, Maher Abujelala, Konstantinos Tsiakas, and Filia Makedon. 2019. Physical fatigue detection through EMG wearables and subjective user reports: a machine learning approach towards adaptive rehabilitation. In *Proceedings of the 12th ACM International Conference on Pervasive Technologies Related to Assistive Environments* (Rhodes, Greece) (*PETRA '19*). Association for Computing Machinery, New York, NY, USA, 475–481. doi:10.1145/3316782.3322772
 - [48] Michael T Paris and Marina Mourtzakis. 2021. Muscle composition analysis of ultrasound images: A narrative review of texture analysis. *Ultrasound Med. Biol.* 47, 4 (April 2021), 880–895.
 - [49] Courtney Reed, Andrew McPherson, et al. 2020. Surface electromyography for direct vocal control. International Conference on New Interfaces for Musical Expression (NIME).
 - [50] Courtney N Reed, Sophie Skach, Paul Strohmeier, and Andrew P McPherson. 2022. Singing knit: soft knit biosensing for augmenting vocal performances. In *Proceedings of the Augmented Humans International Conference 2022*. 170–183.
 - [51] Julius Cosmo Romeo Rudolph, David Holman, Bruno De Araujo, Ricardo Jota, Daniel Wigdor, and Valkyrie Savage. 2022. Sensing Hand Interactions with Everyday Objects by Profiling Wrist Topography. In *Proceedings of the Sixteenth International Conference on Tangible, Embedded, and Embodied Interaction* (Daejeon, Republic of Korea) (*TEI '22*). Association for Computing Machinery, New York, NY, USA, Article 14, 14 pages. doi:10.1145/3490149.3501320
 - [52] Osamu Saisho, Shingo Tsukada, Hiroshi Nakashima, Hiroshi Imamura, and Kyoichi Takaori. 2019. Enhancing support for optimal muscle usage in sports: coaching and skill-improvement tracking with sEMG. In *Proceedings of the 2019 ACM International Symposium on Wearable Computers* (London, United Kingdom) (*ISWC '19*). Association for Computing Machinery, New York, NY, USA, 206–210. doi:10.1145/3341163.3347722
 - [53] Fabio Sarto, Jörg Spörri, Daniel P Fitze, Jonathan I Quinlan, Marco V Narici, and Martino V Franchi. 2021. Implementing ultrasound imaging for the assessment of muscle and tendon properties in elite sports: Practical aspects, methodological considerations and future directions. *Sports Med.* 51, 6 (June 2021), 1151–1170.
 - [54] Ruben Schlagowski, Daria Nazarenko, Yekta Can, Kunal Gupta, Silvan Mertes, Mark Billinghurst, and Elisabeth André. 2023. Wish You Were Here: Mental and Physiological Effects of Remote Music Collaboration in Mixed Reality. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) (*CHI '23*). Association for Computing Machinery, New York, NY, USA, Article 102, 16 pages. doi:10.1145/3544548.3581162
 - [55] Yang Shi and Cheng Yang. 2013. Celestia: a vocal interaction music game. In *CHI '13 Extended Abstracts on Human Factors in Computing Systems* (Paris, France) (*CHI EA '13*). Association for Computing Machinery, New York, NY, USA, 2647–2650. doi:10.1145/2468356.2479485
 - [56] Phillip Song. 2013. Assessment of vocal cord function and voice disorders. *Principles and practice of interventional pulmonology* (2013), 137–149.
 - [57] Kimberly M Steinhauer and Mary McDonald Klimek. 2019. Vocal Traditions: Estill Voice Training®. *Voice and Speech Review* 13 (2019), 354–359. Issue 3. doi:10.1080/23268263.2019.1605707
 - [58] Johan Sundberg. 2020. Singing power ratio and vocal health in professional classical singers. *Journal of Voice* 34, 5 (2020), 690–696.
 - [59] Johan Sundberg, Maddalena Bitelli, Annika Holmberg, and Ville Laaksonen. 2017. The “Overdrive” Mode in the “Complete Vocal Technique”: A Preliminary Study. *Journal of Voice* 31 (2017), 528–535. Issue 5. doi:10.1016/j.jvoice.2017.02.009
 - [60] Akifumi Takahashi, Jas Brooks, Hiroyuki Kajimoto, and Pedro Lopes. 2021. Increasing Electrical Muscle Stimulation’s Dexterity by means of Back of the Hand Actuation. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) (*CHI '21*). Association for Computing Machinery, New York, NY, USA, Article 216, 12 pages. doi:10.1145/3411764.3445761
 - [61] Adam Tapal, Ela Oren, Reuven Dar, and Baruch Eitam. 2017. The Sense of Agency Scale: A Measure of Consciously Perceived Control over One’s Mind, Body, and the Immediate Environment. *Frontiers in Psychology* 8 (09 2017), 1552. doi:10.3389/fpsyg.2017.01552
 - [62] João Paulo Teixeira and Paula Odete Fernandes. 2015. Acoustic analysis of vocal dysphonia. *Procedia Computer Science* 64 (2015), 466–473.
 - [63] Ingo R Titze and Brad H Story. 2002. Rules for controlling low-dimensional vocal fold models with muscle activation. *The Journal of the Acoustical Society of America* 112, 3 (2002), 1064–1076.
 - [64] Carl Unander-Scharin, Kristina Höök, and Ludvig Elblaus. 2013. The throat III: disforming operatic voices through a novel interactive instrument. In *CHI '13 Extended Abstracts on Human Factors in Computing Systems* (Paris, France) (*CHI EA '13*). Association for Computing Machinery, New York, NY, USA, 3007–3010. doi:10.1145/2468356.2479596
 - [65] Carl Unander-Scharin, Åsa Unander-Scharin, and Kristina Höök. 2014. The vocal chorder: empowering opera singers with a large interactive instrument. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Toronto, Ontario, Canada) (*CHI '14*). Association for Computing Machinery, New York, NY, USA, 1001–1010. doi:10.1145/2556288.2557050
 - [66] Carl Unander-Scharin, Åsa Unander-Scharin, Kristina Höök, and Ludvig Elblaus. 2014. Interacting with the vocal chorder: re-empowering the opera diva. In *CHI '14 Extended Abstracts on Human Factors in Computing Systems* (Toronto, Ontario, Canada) (*CHI EA '14*). Association for Computing Machinery, New York, NY, USA, 603–606. doi:10.1145/2559206.2574798
 - [67] M Usha, YV Geetha, and YS Darshan. 2017. Objective identification of prepubertal female singers and non-singers by singing power ratio using matlab. *Journal of Voice* 31, 2 (2017), 157–160.
 - [68] Jennifer M Vojtech, Michael D Chan, Bhawna Shiwani, Serge H Roy, James T Heaton, Geoffrey S Meltzner, Paola Contessa, Gianluca De Luca, Rupal Patel, and Joshua C Kline. 2021. Surface electromyography-based recognition, synthesis, and perception of prosodic subvocal speech. *Journal of Speech, Language, and Hearing Research* 64, 6S (2021), 2134–2153.
 - [69] Graham F Welch, David M Howard, and John Nix. 2019. *The Oxford handbook of singing*. Oxford University Press.
 - [70] Wikipedia contributors. 2024. Falsetto — Wikipedia, The Free Encyclopedia. <https://en.wikipedia.org/w/index.php?title=Falsetto&oldid=1240557845> [Online; accessed 12-September-2024].
 - [71] Emi Juliana Yamauchi, Satoshi Imaizumi, Hagino Maruyama, and Tomoyuki Haji. 2010. Perceptual evaluation of pathological voice quality: A comparative analysis between the RASATI and GRBASI scales. *Logopedics Phoniatrics Vocology* 35, 3 (2010), 121–128.
 - [72] Yang Zhang, Robert Xiao, and Chris Harrison. 2016. Advancing Hand Gesture Recognition with High Resolution Electrical Impedance Tomography. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology* (Tokyo, Japan) (*UIST '16*). Association for Computing Machinery, New York, NY, USA, 843–850. doi:10.1145/2984511.2984574
 - [73] Enhao Zheng, Jiacheng Wan, Lin Yang, Qining Wang, and Hong Qiao. 2021. Wrist Angle Estimation With a Musculoskeletal Model Driven by Electrical Impedance Tomography Signals. *IEEE Robotics and Automation Letters* 6, 2 (2021), 2186–2193. doi:10.1109/LRA.2021.3060400
 - [74] Junyi Zhu, Yuxuan Lei, Aashini Shah, Gila Schein, Hamid Ghaednia, Joseph Schwab, Casper Harteveld, and Stefanie Mueller. 2022. MuscleRehab: Improving Unsupervised Physical Rehabilitation by Monitoring and Visualizing Muscle Engagement. In *Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology* (Bend, OR, USA) (*UIST '22*). Association for Computing Machinery, New York, NY, USA, Article 33, 14 pages. doi:10.1145/3526113.3545705
 - [75] Junyi Zhu, Jackson C Snowden, Joshua Verdejo, Emily Chen, Paul Zhang, Hamid Ghaednia, Joseph H Schwab, and Stefanie Mueller. 2021. EIT-kit: An Electrical Impedance Tomography Toolkit for Health and Motion Sensing. In *The 34th*

Annual ACM Symposium on User Interface Software and Technology (Virtual Event, USA) (*UIST '21*). Association for Computing Machinery, New York, NY, USA, 400–413. doi:10.1145/3472749.3474758

- [76] Mingxing Zhu, Xin Wang, Hanjie Deng, Yuchao He, Haoshi Zhang, Zhenzhen Liu, Shixiong Chen, Mingjiang Wang, and Guanglin Li. 2022. Towards evaluating pitch-related phonation function in speech communication using high-density surface electromyography. *Frontiers in Neuroscience* 16 (2022), 941594.