2025 IEEE World Haptics Conference (WHC)
Suwon Convention Center, Suwon, Korea
July 8 ~ 11, 2025

Vocal Guide System with Force–Vibrotactile Haptic Feedback using Real-Time Pitch and Rhythm Detection in VR*

Malka Rania Iqbal[1], Minha Jeon[2], Sang-Woo Seo[1,3], and Seokhee Jeon[1, †]

***Abstract*— This study proposes a VR-based vocal training system integrating real-time force, vibrotactile, and visual feedback to enhance singing accuracy and expressiveness. Utilizing gesture-based haptic kinesthetic guidance, discrete vibrotactile cues, and immersive visual feedback in virtual reality, the system addresses limitations of traditional training methods, offering interactive and multimodal guidance. A pitch and rhythm detection module identifies errors promptly, triggering immediate corrective feedback.**

I. Introduction

Traditional vocal training primarily relies on auditory or limited modality feedback, often leading to passive learner engagement and fatigue. Recent research underscores the significance of gesture and movement integration into musical instruction for improved pitch, rhythm, and expressiveness. Motivated by these insights, we propose a novel VR-based vocal training system that combines kinesthetic force feedback, vibrotactile alerts, and intuitive visual cues. This multimodal approach aims to facilitate active learning, immediate error correction, and enhanced musical immersion, thereby optimizing vocal performance training.

II. Related Works

To enhance the musical experience and vocal performance of learners, researchers have explored various feedback modalities. Hoppe et al. proposed a system that provides real-time visual feedback to help singers improve pitch accuracy and vocal control [1]. Similarly, Lee et al. developed a training system that uses real-time vibrotactile feedback to intuitively guide drum performance [2]. However, these systems primarily target technical accuracy (e.g., pitch, rhythm, vocalization) with unidirectional feedback, which limits interactive engagement and reduces the learner's active role in the training process [1], [2]. Moreover, prolonged reliance on a single feedback modality (visual or vibrotactile) can lead to user fatigue. Recent research has highlighted the importance of incorporating gestures and body movements into vocal training. Nafisi et al. found that integrating gestures into vocal instruction significantly improved learners' rhythmic precision, pitch accuracy, performance energy, and emotional expressiveness [3]. Meissl et al. similarly analyzed conductors' spatial gestures and demonstrated that these movements effectively convey musical dynamics and emotional intent [4]. In addition, Allcoat et al. showed that immersive virtual reality (VR) environments can foster higher engagement, more positive emotions, and better learning outcomes than traditional methods [5]. Motivated by these findings, we propose a VR-based vocal training system that employs real-time haptic feedback to guide and respond to users' gestures and body movements. By providing interactive, real-time feedback in an immersive virtual environment, the system aims to overcome the limitations of traditional visual and auditory feedback modalities, thereby enhancing learners' immersion, expressiveness, and overall performance.

*This work was supported by the Institute of Information & Communications Technology Planning & Evaluation(IITP) grant funded by the Korea government(MSIT) (IITP-RS-2025-02214780, Generative Haptics and Fine Response Inference for Flexible Tactile Interfaces, 50%), and the MSIT(Ministry of Science and ICT) under the metaverse support program to nurture the best talents (IITP-2024-RS-2024–00425383, 50%).

[1]Department of Computer Engineering, Kyung Hee University, Yongin, Gyeonggi-do 17104, Republic of Korea (e-mail: malkarania15@gmail.com)

[2]Department of Software Convergence, Kyung Hee University, Yongin, Gyeonggi-do 17104, Republic of Korea (e-mail: harororo721@khu.ac.kr)

[3]Creative Content Research Division, Electronics and Telecommunication Research Institute, Daejeon 34129, Republic of Korea (e-mail: swseo@etri.re.kr).

†Corresponding author : Seokhee Jeon (e-mail : jeon@khu.ac.kr)

III. System Design and Implementation

The Vocal Guide System integrates real-time audio analysis with multimodal feedback to help users improve singing accuracy. The system is composed of a pitch and rhythm error detection module and three feedback modalities (force feedback, vibrotactile cues, and VR visual displays), along with indicators of system status as shown in Figure 1. The following sections describe each component and their implementation in the prototype.

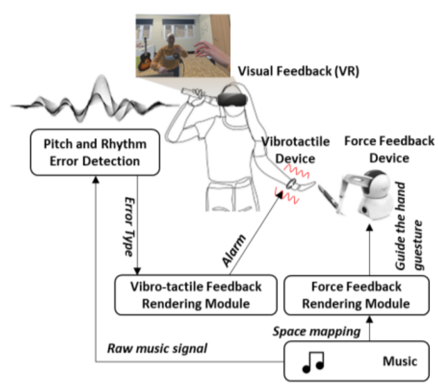

Figure 1: Overview of the proposed vocal guide system.

A. Pitch and Rhythm Error Detection

The system continuously monitors the user's voice to identify pitch and timing deviations from a target melody. A microphone captures the audio, and a signal processing module extracts the fundamental frequency (F0) trajectory and note onset times in real time. We use a robust note recognition approach similar to recent deep learning-based methods [6] to track the sung notes with high accuracy. This module detects when the user's pitch deviates beyond a permitted tolerance (intonation error) or when a note is started too early/late or missed entirely (rhythm error). Each performed note is automatically compared to the reference musical score: if the pitch is off by more than a set threshold (e.g., 50 cents) or the timing of a note onset differs significantly from the expected beat, an error is registered. By analyzing the sequence of detected notes and their timing, the system can pinpoint pitch inaccuracies and rhythm mistakes during the performance. These detected errors are then used to trigger corrective feedback in the haptic and visual modules. The error detection runs with low latency so that feedback can be given almost immediately as the user sings, allowing real-time guidance.

B. Force Feedback

The Vocal Guide System uses a 3D Systems Touch haptic device to deliver kinesthetic guidance. The user's wrist is attached to the device via a fixed rod, enabling direct actuation of wrist motion. The haptic device continuously produces smooth force cues along predefined gesture trajectories, guiding the user's movements. Pre-calculated gesture trajectories were parameterized following the spatial–dynamics correspondences identified by Meissl et al. [4], whereby forte events are rendered as vertical ascents or lateral arm abductions, whereas piano events are encoded as downward contractions toward the torso. These trajectory templates provide a semantically grounded reference frame for the haptic guidance controller when steering the user's wrist along expressive conducting motions. Configured for low-impedance operation, the Touch device provides gentle, compliant force feedback that softly pulls the user toward the target trajectory. This guidance is designed to be minimally resistive: it exerts only enough force to steer the movement, yielding readily to user-initiated motions and avoiding significant opposition. When not actively guiding, the device remains nearly transparent, allowing the wrist to move freely.

C. Vibrotactile Feedback

The system also uses vibrotactile cues to convey discrete feedback events and enhance the user's awareness of errors. Small vibrating actuators are embedded in the lightweight wearable device. These vibrators deliver brief vibration patterns whenever a significant pitch or rhythm error occurs, serving as immediate alerts. For instance, if the user entirely misses a note or deviates greatly from the correct rhythm, a short vibration pulse is triggered as a warning. Vibrotactile pulses are also used as a metronome: the device can pulse at the song's tempo or on each downbeat, giving the user a clear sense of timing. Unlike the continuous force feedback, the vibrations are discrete and quick, which effectively capture the user's attention without overwhelming them. Different vibration patterns encode different information – a double pulse might indicate a pitch error while a single longer buzz could indicate a rhythm mistake – allowing the performer to distinguish the type of feedback.

D. VR Visual Feedback

To provide additional guidance, the Vocal Guide System includes a VR interface. The user wears a VR headset in which a custom training environment displays real-time visual feedback about their performance. The VR scene presents the target melody and the user's singing in an intuitive visual form. As the user sings, a marker moves according to the detected pitch and rhythm on a piano roll in front of users. Target notes are shown at their correct pitch heights and timing positions in the VR space. If the user sings the note correctly, their marker aligns with the target note. If they sing off-key, the marker will appear above or below the target note, immediately indicating the error. Similarly, if the timing is off, the marker lags behind or jumps ahead of the target note's position on the timeline. The VR display also uses color cues (for example, turning a note red if it was sung incorrectly) to highlight errors after they occur. In addition to note-by-note feedback, the virtual environment can present summary information, such as an error count or score, at the end of a practice session. By immersing the user in a visual representation of their singing, the system leverages visual learning modalities alongside haptic feedback.

IV. Conclusion

The proposed VR-based vocal training system effectively integrates real-time error detection with multimodal haptic, vibrotactile, and visual feedback. Kinesthetic guidance via force feedback gently directs expressive gestures, while vibrotactile cues quickly alert users to pitch and rhythm errors. Visual feedback within an immersive VR environment reinforces understanding of musical accuracy. Collectively, these interactive components offer a comprehensive and engaging training experience, addressing the limitations of conventional single-modality feedback systems. Future work will involve extensive user evaluations to quantify improvements in musical skill acquisition and learner engagement.

References

[1] D. Hoppe, M. Sadakata, and P. Desain, "Development of real-time visual feedback assistance in singing training: a review," *Computer Assisted Learning*, vol. 22, no. 4, pp. 308–316, Aug. 2006, doi: 10.1111/j.1365-2729.2006.00178.x.

[2] I. Lee and S. Choi, "Vibrotactile guidance for drumming learning: Method and perceptual assessment," in *2014 IEEE Haptics Symposium (HAPTICS)*, Houston, TX, USA: IEEE, Feb. 2014, pp. 147–152. doi: 10.1109/HAPTICS.2014.6775447.

[3] J. S. A. Nafisi, "Gestures and body-movements in the teaching of singing: a survey into current practice in Australia and Germany".

[4] K. Meissl, P. Sambre, and K. Feyaerts, "Mapping musical dynamics in space. A qualitative analysis of conductors' movements in orchestra rehearsals," *Front. Commun.*, vol. 7, p. 986733, Nov. 2022, doi: 10.3389/fcomm.2022.986733.

[5] D. Allcoat and A. Von Mühlenen, "Learning in virtual reality: Effects on performance, emotion and engagement," *Research in Learning Technology*, vol. 26, no. 0, Nov. 2018, doi: 10.25304/rlt.v26.2140.

[6] D. Chang, "Vocal performance evaluation of the intelligent note recognition method based on deep learning," *Sci Rep*, vol. 15, no. 1, p. 13927, Apr. 2025, doi: 10.1038/s41598-025-99357-2.