ViVo: Piano Learning Through Visualizing Vocalizations on a Lighted Keyboard

Maya Caren
Palo Alto, California
maya.c.caren@gmail.com

Abstract

Vocalization and visualization are recognized as two powerful methods for internalizing music that are effective with beginner and skilled musicians alike. Despite the well-researched benefits of each practice, integrated visualization of vocalizations for instrument learning has seen little attention in the music technology community. This paper introduces the design and implementation of ViVo, a piano learning tool that connects the embodied sense of pitch offered by vocalization with the spatial intuition provided by in situ visualization. ViVo offers two modes: a real-time mode that hears live user vocalizations to concurrently illuminate the corresponding piano keys, and a practice mode that visualizes recorded vocalizations for repeated practice. By providing an integrated system to foster and visualize vocalizations, ViVo aims to leverage the noted benefits of both practices to make learning piano more effective, intuitive, and engaging.

Keywords

Vocalization, Visualization, Learning, Piano, HCI

Figure 1: Picture of System

1 Introduction

Vocalization, the practice of using the voice to render pitches, melodies, and rhythms out loud, has been found to be a powerful mechanism for internalizing music that holds great benefit for learning musical instruments. Prominent music pedagogies such as Dalcroze, Gordon, Kodály, Orff, and Suzuki [7, 8, 10, 11, 15], as well as many skilled musicians, employ vocalization as a crucial step in developing the intuitive sense of music that is foundational for instrument mastery, and additionally as an effective technique for building improvisation and composition skills. Evidence suggests that vocalization does improve many aspects of musicianship: it has been shown to improve beginning musicians' sense of pitch and melody [5]; vocal melodies are better remembered than instrumental melodies by both musicians and non-musicians [16]; and vocalization has been found to be an important tool in facilitating self-guided music learning, which is crucial for improvement outside of structured classroom and lesson contexts [2].

However, despite extensive evidence that vocalization is highly effective for learning music, it is not regularly utilized in practice. Many high school and college band directors, while understanding the value of vocalization, do not use it in their instruction [1]; the same tendency has also been found among elementary school teachers [13].

Similarly, vocalization as a tool specifically for learning musical instruments has seen little attention in the music technology community—while vocalization has a well-established presence at NIME, a taxonomy of research involving voice at NIME [12] points to its use primarily as a control mechanism for a range of interactions or as the focus of explorations into voice itself, rather than as a mechanism for understanding another instrument.

Learning tools based on *visualization*, in contrast, have been well-explored in music technology. Existing piano systems include ones which provide real-time instruction and feedback on improvisation, composition, and general music understanding. They use a variety of approaches including light strips mounted above a piano keyboard [3], projected animations [17], rendered virtual models [14], and augmented reality (AR) overlays [4, 9].

But despite the prevalence of visualization as an instrument learning tool, visualizing *vocalization* expressly for learning a musical instrument has seen little attention.

ViVo is designed to enable the real-time visualization of vocalization dedicated to learning piano. It integrates these two methods of intuitive musical understanding by associating the embodied sense of pitch offered by vocalization with the spatial intuition provided by in situ visualization. By providing an integrated system to visualize vocalizations, ViVo aims to leverage the noted benefits of both practices to make learning piano more effective, intuitive, and engaging.

2 Design

ViVo is envisioned to be a tool that can be easily incorporated into music instrument practices. It consists of three main components: a microphone input and a central processor housed together in an enclosure, and an electronic piano keyboard with embedded key lights (Figure 2). The interface is intentionally simple to allow for easy activation of the mic during instruction in order to encourage vocalization, and for quick capture of impromptu arrangements to support improvisation.

NIME '25, June 24–27, 2025, Canberra, Australia

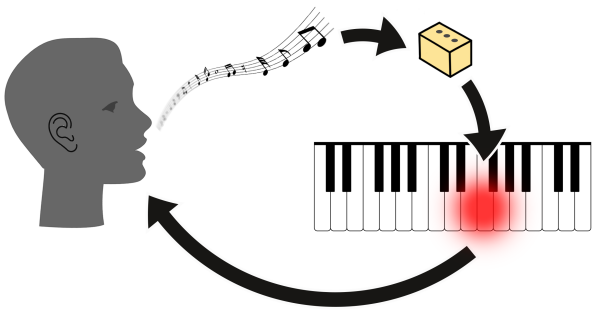

Figure 2: Schematic Diagram

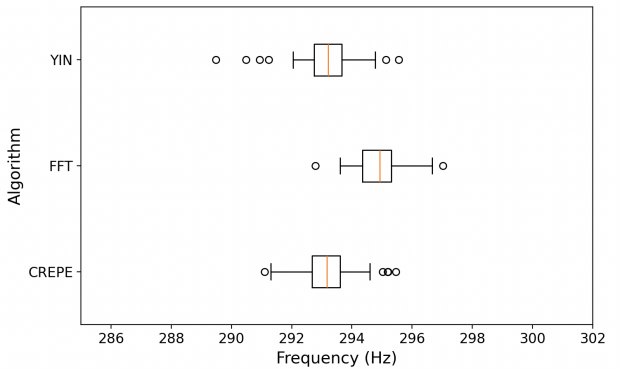

Figure 3: Pitch Detection - Voice Singing D4 (293.7 Hz); Frequency Threshold Range For D4 is 285 Hz to 302 Hz

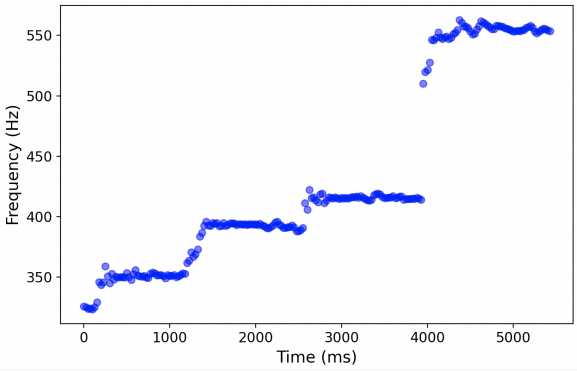

Figure 4: First 4 Notes of Autumn Leaves (YIN) - F4, G4, G#4 and C#5

2.1 Modes of Operation

The system has two modes of operation: Real-time mode and Practice mode.

The default state is Real-time mode, where user vocalizations are processed from the live microphone input. Note onsets and pitches (i.e. fundamental frequencies f_0 of notes) are extracted from the input audio, and corresponding pitches on the keyboard are illuminated in real-time. The pitches extracted from the input can optionally also be simultaneously played by sending modified MIDI messages to the keyboard.

Practice mode is activated either by pressing the Record button to record a vocalization, or by pressing the Playback button to play back the recording on-demand with associated pitches concurrently illuminated and sounded using extracted note onset and pitch information. The system reverts back to Real-time mode when not recording or when a playback recording concludes.

2.2 Goals

ViVo's main use case as a device employed during instrument practice imposes several system-level constraints: portability, a reasonably low materials cost, and sufficiently real-time operation for a seamless interactive experience.

For performance metrics, three key characteristics were identified as important goals for high-quality visualization: pitch accuracy, pitch range, and note onset/offset fidelity.

2.3 Hardware

ViVo's system is built around the Teensy 4.1, a low-cost microcontroller that offers real-time audio processing capabilities in a small form factor. Including the microphone, custom PCB, enclosure, and all other components (excluding the MIDI keyboard), the entire system can be built for under $100 USD. All code and design files are open-source and publicly available.[1]

2.4 Algorithms

ViVo is constructed with two candidate pitch detection algorithms to extract pitch from vocalizations: YIN [5] and a heuristics-based approach using an 1024 point FFT with parabola fitting.

Although using parabolic and Gaussian interpolation to more accurately identify the most prominent pitch in an FFT frequency spectrum has been well-explored [6], ViVo introduces several modifications specialized for singing f_0 detection: ViVo first identifies the bins that correspond to the pitch range of most vocalizations by novice musicians (approximately 100–1000 Hz) and computes the average energy μ_E across that range; then, the algorithm traverses the bins from low to high and identifies the first three consecutive bins with energy greater than μ_E, performs a three-point parabola fit, and returns the vertex of the parabola as the detected f_0, optionally smoothed by an IIR lowpass filter. This method makes the system more robust to environmental noise, and avoids the common error mode of being off by an octave that results from the first harmonic sometimes having more energy than the fundamental.

[1] https://github.com/mayacaren/ViVo.git

3 Evaluation

ViVo was evaluated against the three stated goals of pitch accuracy, range and note onset/offset detection. For consistency and repeatability, ViVo was tested using the Spitfire Audio Folk Voices library played through a speaker placed 70 cm away at normal talking volumes, to simulate common usage conditions.

3.1 Pitch Detection

Both pitch detection approaches had approximately equal latencies (less than 30ms) and memory footprints. The YIN and FFT results both compared well with the CREPE algorithm (Figure 3) which was run on the test file directly using a MacBook. The YIN algorithm was more accurate than the 1024 point FFT by 1-2 Hz. Similar to previous findings on gains from parabola fitting

[7], ViVo's FFT parabola fitting algorithm experienced a 10-20x improvement in resolution compared with the raw bin size of 43Hz. While the YIN algorithms performed slightly better than the FFT-based approach, the FFT may be attractive in other applications, especially in compute-restricted environments where it is already being computed for visualization or other feature extraction purposes.

Although both tested pitch detection methods perform well for monophonic vocalization, variation in extracted pitch, resulting from users themselves singing inexact or inconsistent pitches, is unavoidable. Many singers "bend into" notes by starting at an arbitrarily lower pitch and transitioning up into their intended pitch (Figure 4), which creates a period of pitch uncertainty at the onset of each note that may span several hundred milliseconds. This can be reduced by instructing users to begin notes with a consonant (e.g. "doo" or "bah"), but is still significant, especially with novice singers.

3.2 Range

The YIN, FFT and CREPE algorithms were tested from E3 (165Hz) to C5 (523Hz), and all three algorithms were able to span the full range with similar results to Figure 3 for each note. Of the three algorithms, the FFT was slightly less robust as it occasionally could not detect the specific combination of low pitches (E3 to G3) at low volume. The root cause appeared to be issues with the parabola fitting only receiving 2 accurate data points from the FFT, instead of the 3 points needed.

3.3 Note Onset/ Offset

Note onsets and offsets were detected by recognizing when the RMS signal power exceeds a relative threshold after first applying a bandpass filter of 100 to 1000 Hz. Using this approach, ViVo had a signal to noise ranging from 30 to 40 dB during testing.

4 Future Work

In future work, ViVo's pitch detection system will be extended to support simple polyphonic input, which would enable visualization of inputs beyond user vocalization to include content such as recorded music or online music lesson videos.

Further development would also include implementing ViVo in a larger music education or classroom context to evaluate how it could help build musical intuition and support self-guided practice at scale. Another potential user study would investigate ViVo's effect on the development of improvisation skills in jazz pianists by assessing the degree to which the repeated practice of vocalizing improvised musical phrases integrated with ViVo's spatial visualizations of corresponding pitches improves their learning process.

More broadly, ViVo's integrated visualization of vocalization could be employed to build musical intuition in students learning a variety of new instruments. While the current prototype uses a piano keyboard, future iterations of ViVo could be implemented on other target instruments like guitar (e.g. through a lighted fretboard), or even offer an intuitive route to learn novel musical interfaces that have uncommon pitch layouts (such as for first-time users of the Linnstrument, which has both MIDI-controlled lighted keys and a layout unfamiliar to most musicians).

While developed primarily as an instrument learning tool, ViVo also opens novel live performance opportunities by presenting a new mode of musical interaction that utilizes voice to engage with and actuate a range of digital systems. This real-time vocal agency afforded by ViVo offers possibilities for applications in live stage productions, collaborative and interactive creative works, and improvisational music performances.

5 Ethical Standards

This research was conducted using low-cost hardware and open source software with funding provided by the researcher. The software developed for this project is completely open source. All participation was voluntary and there are no conflicts of interest.

References

[1] H Christian Bernhard. 2003. Singing in instrumental music education: Research and implications. *Update: Applications of Research in Music Education* 22, 1 (2003), 28–35.

[2] Amalia Casas-Mas, Guadalupe López-Íñiguez, Juan Ignacio Pozo, and Ignacio Montero. 2019. Function of private singing in instrumental music learning: A multiple case study of self-regulation and embodiment. *Musicae Scientiae* 23, 4 (2019), 442–464.

[3] Isabela Corintha, Giordano Cabral, and Gilberto Bernardes. 2019. AMIGO: an assistive musical instrument to engage, learn and create music. In *Proceedings of the international conference on New Interfaces for Musical Expression.*

[4] Shantanu Das, Seth Glickman, Fu Yen Hsiao, and Byunghwan Lee. 2017. Music everywhere–augmented reality piano improvisation learning system. In *Proc. International Conference on New Interfaces for Musical Expression (NIME).* 511–512.

[5] Charles A Elliott. 1974. Effect of vocalization on the sense of pitch of beginning band class students. *Journal of Research in Music Education* 22, 2 (1974), 120–128.

[6] M Gasior and JL Gonzalez. 2004. Improving FFT frequency measurement resolution by parabolic and Gaussian spectrum interpolation. In *AIP Conference Proceedings*, Vol. 732. American Institute of Physics, 276–285.

[7] Edwin Gordon. 2007. *Learning sequences in music: A contemporary music learning theory.* Gia Publications.

[8] Mícheál Houlahan and Philip Tacka. 2008. *Kodály Today: A Cognitive Approach to Elementary Music Education.* Oxford University Press.

[9] Feng Huang, Yu Zhou, Yao Yu, Ziqiang Wang, and Sidan Du. 2011. Piano AR: A Markerless Augmented Reality Based Piano Teaching System. In *2011 Third International Conference on Intelligent Human-Machine Systems and Cybernetics*, Vol. 2. 47–52.

[10] Émile Jacques-Dalcroze. 1967. *Rhythm, Music amd Education.* Barclay Press.

[11] Gunild Keetman. 1974. *Elementaria: First Acquaintance with Orff-Schulwerk.* Schott.

[12] Rébecca Kleinberger, Nikhil Singh, Xiao Xiao, and Akito van Troyer. 2022. Voice at NIME: a Taxonomy of New Interfaces for Vocal Musical Expression. In *NIME 2022.* PubPub.

[13] S.R. Lee. 1996. The effects of vocalization on achievement levels of selected performance areas found in elementary instrumental bands. In *ERIC Document Reproduction Services No 418030.* Salem-Teikyo University.

[14] Chih-Chun Lin and Damon Shing-Min Liu. 2006. An intelligent virtual piano tutor. In *Proceedings of the 2006 ACM international conference on Virtual reality continuum and its applications.* 353–356.

[15] Shin'ichi Suzuki. 2013. *Nurtured by Love: The Classic Approach to Talent Education.* Alfred Music.

[16] Michael W. Weiss, Sandra E. Trehub, and E. Glenn Schellenberg. 2012. Something in the Way She Sings: Enhanced Memory for Vocal Melodies. In *Psychological Science.* Association for Psychological Science, 1074–1078.

[17] Xiao Xiao, Basheer Tome, and Hiroshi Ishii. 2014. Andante: Walking Figures on the Piano Keyboard to Visualize Musical Motion.. In *NIME.* Cambridge, MA, 629–632.