The 6th International Conference on Artificial Intelligence in Information and Communication (ICAIIC 2024)

Intuitive, Interactive Training System for the Sense of Sound Using Portable Theremin Device

Tsen-Fang Lin
Dept. of Popular Music Industry
Southern Taiwan University of Science and Technology
Tainan, Taiwan
tflin@stust.edu.tw

Chun-Hung Yang
Dept. of Electronic Engineering,
Southern Taiwan University of Science and Technology
Tainan, Taiwan
eliyang@stust.edu.tw

***Abstract*—The study investigats an interactive Theremin sound pairing assessment system aimed at realizing an intuitive pitch training platform. Utilizing algorithms designed to detect user-controlled pitch and instrument selections, this system facilitates the display of user-friendly information on terminal devices, including portable PCs. Notably, this device achieves a substantial reduction in size and weight compared to the traditional Theremin. It boasts the distinctive, intuitive, air-touch playing characteristics inherent to the Theremin, eliminating the barriers typically associated with learning musical instruments or vocal techniques. Furthermore, it combines educational and entertainment features. Various sound perception feedback methods are explored to create an interactive environment that enhances motivation for practice. The research findings indicate that perceptual skills, such as pitch and timbre discrimination, can be acquired intuitively and visually, obviating the need for a theoretical foundation.**

Keywords—interactive Theremin, sound pairing system, intuitive pitch training, educational and entertainment

I. Introduction

Musical perception, encompassing elements such as pitch discrimination and timbre recognition, plays a pivotal role in shaping our musical experiences. However, for the general populace without prior musical instrument training, musical perception training remains a formidable challenge due to its abstract nature. Since the concept of music itself is an interactive process and technology [1], designing music activities into an interactive form is closer to the characteristics of music itself. Traditional electronic instruments, such as the Theremin, offer an intuitive interface through aerial tactile interaction but are not readily portable.

Therefore, this study introduces an innovative interactive Theremin-based musical perception training system that transcends the confines of conventional training methods. Theremin offers certain possibilities as an IoT device [2]. By employing a specially designed portable Theremin devicintegrated with interactive elements and future plans for artificial intelligence (AI) integration, this system operates at the confluence of music, technology, and artificial intelligence. The interactive teaching music intelligence system provides music learners with better technical support and learning concepts [3]. Utilizing assistive technology can help students minimize stress and allow everyone to better enjoy the music performance experience [4]. Leveraging the Theremin's unique ability to generate sound in response to hand proximity and movement, the system establishes a tangible connection between users and the auditory world, involving tangible hand-eye interactions.

Collaborating with Raspberry Pi and computer interfaces, the system harnesses real-time data to provide users with immediate feedback, thereby refining their pitch perception and instrument recognition skills. Furthermore, a forward-looking enhancement involves the integration of soundscape-style sound effect accompaniments generated by an online artificial intelligence platform, introducing entertainment functionality. The choice of soundscape-style music arises from the Theremin's capacity to produce smooth and seamless electronic tones [5], rendering it an ideal choice for creating immersive musical experiences.

At its current stage, the system has successfully implemented real-time hand region parameter feedback, wireless data communication between devices such as Raspberry Pi and PC, and design features for user feedback and data recording in pitch and timbre selection.

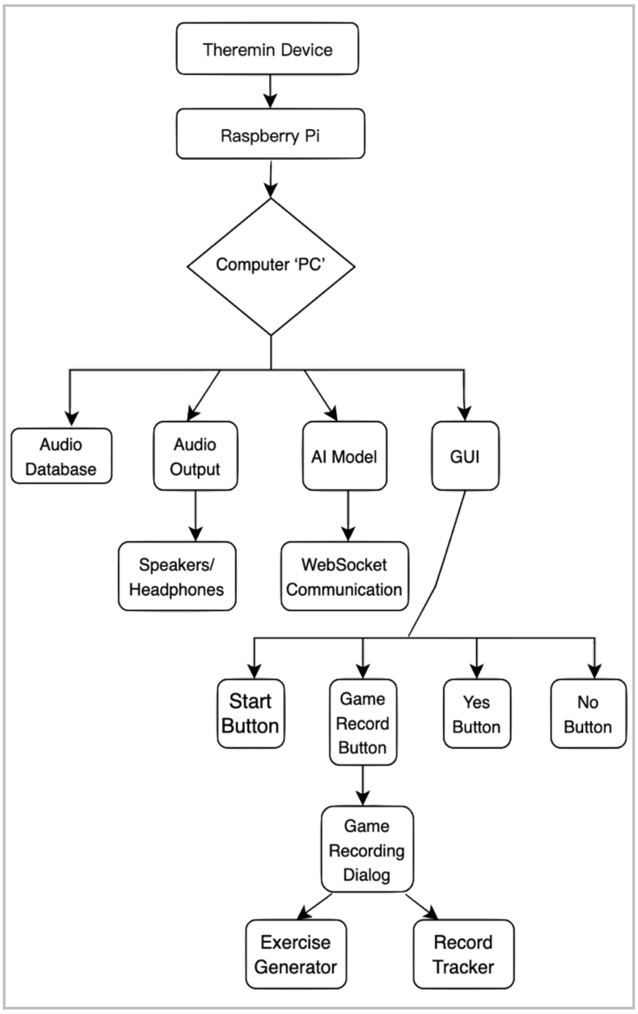


Fig. 1. Current system architecture flow chart

II. Theremin Training System Design Architecture with AI Integration

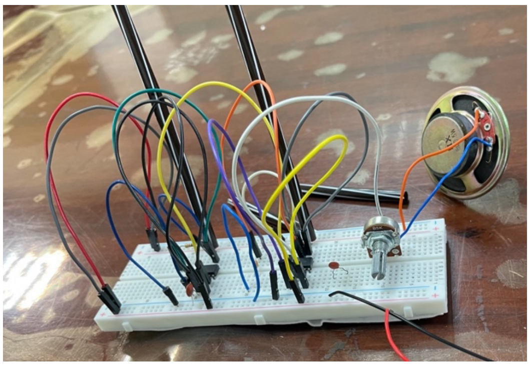
The system aims to enhance users' sound perception, including pitch perception and instrument recognition, using portable Theremin devices, real-time data processing, and plans for artificial intelligence to generate soundscape accompaniments. The following is the structure including future integration. Fig.1 shows the current system architecture flow chart, including parts A-D below.

A. Theremin device: The core of this system is to transform the traditional heavy Theremin into a light and portable device and expand it to have the characteristics of interactive transmission with the computer. Users interact with the Theremin by manipulating the proximity and movement of their hands.

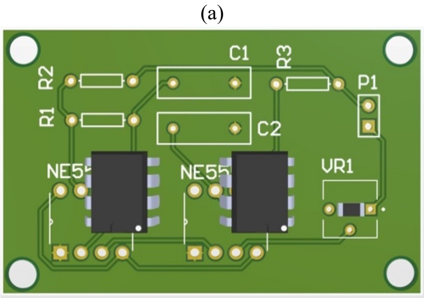
B. Raspberry Pi: The Raspberry Pi acts as a bridge between the Theremin and the computer. It facilitates wireless communication of instantaneous data, including pitch and instrument selections affected by manual movements.

C. Computer (PC): The computer interface hosts the graphical user interface (GUI) and AI model. It receives data from the Raspberry Pi, processes it, and provides instant feedback to the user. The GUI includes buttons for starting training, recording game results, and providing feedback.

D. Primary AI model: The AI model analyzes the data received from Theremin and provides feedback to the user based on their performance, thereby enhancing their pitch perception and instrument identification skills.



(a)



(b)

Fig. 2. (a) A test set and (b)work-in-progress circuit board for the portable Theremin.

Future integration: In the next step, the system will be connected to the artificial intelligence situational music accompaniment platform, so that when the theremin is played, the system can instantly generate soundscape music suitable for the timbre of the theremin to enhance the music experience [6].

To sum up, the Theremin training system with AI integration combine the real-time feedback mechanism of the current architecture with the soundscape accompaniment generated by future AI. This innovative method breaks down the way and efficiency of learning pitch. It not only improves the user's music appreciation, but also stimulates the user's creativity and entertainment experience, making Theremin a multi-functional tool for musicians, educators, and enthusiasts.

III. Portable Theremin Set Design

Fig. 2 is the test set and the circuit board in production of the indicating type Theremin. After repeated testing and adjustments, the timbre, uniformity, and sound quality have been improved to a stable level. This small Theremin electroacoustic model now emits a steady pitch and intensity in response to hand movements. At present, the most stable pitch area is roughly the area where alto girls sing, and the pitch range is about an octave upward from C4. Since there are 12 semitones in an octave, the current pitch test setting range in the interactive sound game is to stabilize 12 consecutive semitones in this area and appear randomly.

Considering that hands-on learning can be an effective way for students who love music but are not familiar with engineering, this study included the assembly process as an option to provide users with fun learning options. All materials of this set of equipment can be assembled without tools. Future product launches will be accompanied by complete assembly instructions. Table 1 is the component specification list.

TABLE I. Specifications and Performance

<table>
<tr><th colspan="2">Component</th><th>Type</th><th>Quantity</th></tr>
<tr><td colspan="2">Chip</td><td>NE555</td><td>2</td></tr>
<tr><td colspan="2">Variable Resistor</td><td>5K</td><td>1</td></tr>
<tr><td colspan="2">Power</td><td>9V Battery</td><td>1</td></tr>
<tr><td colspan="2">Sensor</td><td>Photoresistor</td><td>2</td></tr>
<tr><td colspan="2">Resistance</td><td>1K</td><td>1</td></tr>
<tr><td colspan="2" rowspan="2">Ceramic Capacitors</td><td>0.01u (104)</td><td>1</td></tr>
<tr><td>0.01u (103)</td><td>1</td></tr>
<tr><td colspan="2">Platform</td><td>Breadboard (circuit board in production)</td><td>1</td></tr>
<tr><td colspan="2">Connection</td><td>Dupont line</td><td>1</td></tr>
<tr><td colspan="2">Speaker</td><td>8Ohm 5W (VN4570)</td><td>1</td></tr>
<tr><td rowspan="2">Performance</td><td>Pitch</td><td colspan="2">The current design pitch control is between approximately C3-C4, which is approximately within the range of female vocals.</td></tr>
<tr><td>Sensitivity</td><td colspan="2">The spatial distance between the sensor and the hand from the starting point to one octave upward is about 17cm, which is enough for the hand to have a clear area of differentiation.</td></tr>
</table>

The Theremin settings on the Raspberry Pi side are: (1) Use the MCP3008 ADC to convert the analog signal of the Theremin into a digital value. (2) Program to continuously read the Theremin's pitch and volume values from the ADC. (3) Implement logic to detect changes in pitch and musical instruments. This part is still in testing.

For the part connected to the computer, WebSocket communication is used to instantly send pitch and volume information (which will be converted into instrument selection in the interactive game) to the computer, and the WebSocket library is installed on the Raspberry Pi.

The designed mapping system converts gesture parameters into selected numerical values for pitch and instrument timbre. Move your hand close to the tone antenna to control the tone, move your hand close to the volume antenna to control the volume, and use a horizontal hand position to select one of the sound files for playback. Based on the current spatial sensitivity of horizontal movement, the current default is 8 instrument selections. In other words, there can be 12 different pitches and 8 different instrument timbres each time, for a total of 96 results. The program allows users to switch between these sounds based on hand sensing results.

IV. Interactive Pitch Instrument Interpretation Test Game Design

Figure 3 is a simulation diagram of the usage of the system under test, including the user, the Theremin device, the Raspberry Pi and the computer. At this stage, the program has been initially completed and is under testing. The computer settings and structure of the interactive sound test game are as follows.

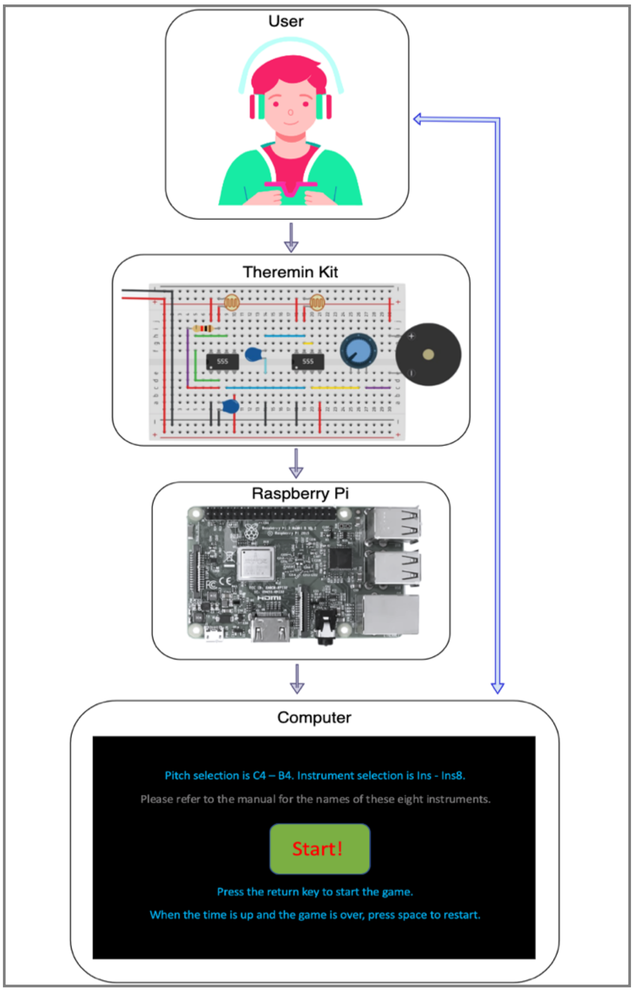


Fig. 3. Schematic diagram of system usage architecture

There are 96 sound files (C4Ins1.wav – B4Ins8.wav) with random pitch and timbre possibilities. The computer side has 96 sound files for pitches and instruments such as C4Ins1.wav – B4Ins8.wav, as well as "Time's up!", "Pitch is right!", "Instrument is right!", "Instrument 1" - "Instrument 8" and other 11 sound files for feedback. A total of 107 audio files of audio. After the user presses the "Start" setting button on the computer, one of them will be played randomly by the computer (uniformly played for 3 seconds) for the Theremin performer to listen to. The theremin player should have 15 seconds to have the theremin produce the correct pitch and select the correct tone that was just played. After more than 15 seconds, the sound file "Time's up!" will be played and this round of testing will end.

During these 15 seconds, if the Theremin player adjusts the pitch to a pitch that is no more than 1/4 of a semitone different from the played sound file, and this lasts for more than 1 second, the computer will immediately issue the sound file "The pitch is correct!" If the Theremin player adjusts to a timbre corresponding to the instrument sound selection of the playing file and continues for more than 1 second, the computer will immediately issue a sound file of "Correct instrument!"

This system requires that when operating Theremin, when the sensing area belongs to Inst1, the computer will emit the sound file of "Instrument 1" at the moment of contact, and then play the sound file corresponding to the block selected by the user (also corresponding to its The selected pitch and timbre), and so on, when the sensing area belongs to Inst8, the computer will emit the sound file of "Instrument 8" at the moment of contact. When the sensing area remains in the same instrument area, sound file notifications will not be issued repeatedly. Once a different instrument area is changed, the sound file notification will be triggered. This way, the user knows exactly which instrument they are triggering.

In addition, the computer side has a window for the user to see, and the window can be resized using the mouse. There are explanatory text and buttons in the window to enter the test, as well as links to view the test records. Test records are used for reviewing learning status and motivating learning. The game logic conditions are listed in order below:

a) Theremin players start the game on their computer by pressing the return key on the keyboard to trigger the "Start" button in the GUI.

b) The computer randomly selects a sound file (pitch and instrument) and plays it once.

c) During 15-second rounds, the theremin player adjusts the pitch and instrument to match the chosen sound.

d) The computer monitors the Theremin's pitch and instrument changes via WebSocket communication.

e) If the Theremin player correctly matches the pitch and instrument, the computer will immediately provide audio feedback such as "Correct pitch!", "Correct instrument!"

f) 15 seconds after the game starts, the computer will provide audio feedback of "Time is up!" based on the performance of the Theremin player. Currently, the text in the button frame in the GUI window changes to "Time is up - press space to restart."

g) After the first round of the game, as long as the user presses space every time the game ends, the text in the button and the background color of the button will return to their initial values.

The traditional way of training pitch perception is too abstract without knowledge of musical instruments, and musical instruments need to be learned before they can be used easily. The interactive Theremin has the characteristics of an intuitive non-playing interface and combines physical and visual aids with perception and hearing. It is an effective way to train the sense of sound, especially for those who have no foundation of music theories.

V. Integration of AI Soundscape-Style Creations

The common time envelope of the instrument sound consists of four parts (time unit "ms", the same below): a. a linear starting attack time, b. followed by a linear decay time of 20% of the maximum amplitude, c. a constant amplitude part sustain time and, d. a linear offset release time. The above four parts are usually referred to as "Attack-Decay-Sustain-Release" (ADSR)[7]. ADSR parameters are crucial to shaping the sound. Except for the human voice and the overtones of wind instruments (which are usually not easily mastered by beginners), almost all musical instruments require fingers to touch another key or endpoint before they can produce another pitch. However, since the theremin relies on gestures in the air to change the sound, the process is continuous and does not require touching the object, so it can produce each sound very smoothly. This quality blends with the nature of pad-like sounds and effects, and the beautiful components of this sound often have a time-varying quality. Theremin is also effective as an interface for producing this type of sound, that is, using two antennas to control the frequency and amplitude of the sound respectively, which has the advantage of being smoother and more intuitive than traditional multimedia instrument digital interface (MIDI). Therefore, Theremin may be the best interface for creating this type of music sound effects. Our next step will be to design the Theremin kit to expand the AI functions of this part to make it more entertaining and practical.

VI. Conclusions and Future Work

In this study, the author proposed an interactive Theremin sound perception test and training system, which allows users without music theory background to learn the perception of pitch and timbre recognition intuitively and concretely. Through the real-time display of the screen, users can get the results of their own sound perception, and through recording, they can observe the changes in the cognitive level of the practice process to improve the motivation and effectiveness of learning. In the future, the part of AI interactive applications will be expanded to make real-time performance more interesting for users.

References

[1] D. Shilton, "Sweet Participation: The Evolution of Music as an Interactive Technology," *Music & Science*, 5, 2022. https://doi.org/10.1177/20592043221084710

[2] D. Čoko, L. D. Rodić, T. Perković and P. Šolić, "Geometry from Thin Air: Theremin as a Playful Learning Device, " 2021 16th International Conference on Telecommunications (ConTEL), pp. 89-96, Zagreb, Croatia, 2021.

[3] F. Liu, "Interactive Music Learning Model Based on RBF Algorithm," *Computational intelligence and neuroscience*, 2022: 5759986, 2022.

[4] A. Acquilino, & G. Scavone, "Current State and Future Directions of Technologies for Music Instrument Pedagogy," *Front. Psychol.*, 13, 2022.

[5] M. Gallagher, "Field recording and the sounding of spaces," *Environment and Planning D: Society and Space*, 33(3), pp.560–576. 2015.

[6] Z. Y. Hong, Y. Alshehri, & N. Alnazzawi, "A decision-support system for assessing the function of machine learning and artificial intelligence in music education for network games," *Soft Comput*, 26, 11063–11075, 2022.

[7] J. Marozeau, & A. Cheveigné, "The effect of fundamental frequency on the brightness dimension of timbre," *The Journal of the Acoustical Society of America*, 121, pp.383-7, 2007.