IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 6 June 2023, accepted 30 June 2023, date of publication 7 July 2023, date of current version 18 July 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3293250

RESEARCH ARTICLE

Exploring Audio Processing in Mixed Reality to Boost Motivation in Piano Learning

INSAN GANANG PUTRANDA, ASIYA MUFIDA YUMNA, YUSEP ROSMANSYAH, AND YUDA SUKMANA
School of Electrical Engineering and Informatics, Bandung Institute of Technology, Bandung 40116, Indonesia

Corresponding author: Insan Ganang Putranda (23521071@std.stei.itb.ac.id)

This work was supported in part by PPMI (Penelitian, Pengabdian Masyarakat, dan Inovasi) Research Program, School of Electrical Engineering and Informatics, Bandung Institute of Technology.

ABSTRACT As technology becomes increasingly important in education, including music, many VR (virtual reality) and AR (augmented reality) applications have been developed to improve skills and knowledge in playing musical instruments, such as piano. However, these applications mainly utilize MIDI input for validating notes. This research explores the potential implementation of mixed reality to enhance piano learning through audio processing capabilities while maintaining users' motivation. The research used the HoloLens 2 device and the FFT (Fast Fourier Transform) method with peak detection and compared various window functions to determine the most accurate one. Blackman-Harris performed the best, with a 97.28% accuracy rate when tested on complex songs. The application was also tested by 31 participants and evaluated using the HMSAM (Hedonic-Motivation System Adoption Model), revealing that curiosity and joy were the most significant factors influencing the use of the application. The effectiveness of the learning was moderate, with an increase of 31.28%. Although there were limitations in the use of audio processing, it could still be utilized and improved further to keep users motivated to learn piano.

INDEX TERMS Mixed reality, digital signal processing, music, games.

I. INTRODUCTION

Technology has played a more important role since the pandemic, not only in education in general, but also in music education, as its sustainability relies on it [1], [2]. The use of VR (Virtual Reality) and AR (Augmented Reality) technology has been considered in the practice of musical skills [3]. AR can present information that is complex and easy to remember for music students to understand a concept [4]. Various designs have been carried out to improve students' skills to learn a popular musical instrument, which is piano [5], [6], [7], [8]. To give real-time feedback, either using the MIDI input or parsing the audio data could be considered. The former is the most popular approach but limits the use of musical instruments to digital, while the latter allows flexibility in musical instruments but is difficult and may be less accurate [8].

Additionally, many students may give up learning before they become good at it, with motivation and feedback being two of the factors in this outcome [9]. Previous VR and AR-based studies provided answers on how to keep users motivated when learning. However, most were performed on digital keyboards using MIDI. This research explores and evaluates the audio processing methods when implemented in a mixed reality application while also giving good feedback and motivation to users using an appropriate evaluation model.

II. LITERATURE REVIEW

A simple systematic literature review method was adopted to answer certain research questions by mapping them with motivation and relevant keywords to be used in database sources [10], [11]. The results were used to help determine the design factor for piano learning using VR and AR approaches, the method and basic knowledge of audio signal processing in validating a melody, and also factors supporting the success of learning musical instruments.

The associate editor coordinating the review of this manuscript and approving it for publication was Tianhua Xu.

A. RELATED WORKS IN VR AND AR

The definition of mixed reality has evolved from the initial concept, which was based on visual display, to include various gestures and locations within the physical and virtual worlds [12], [13]. This research uses the term mixed reality due to the high levels of interactivity between the physical and virtual worlds.

Immersive applications to learn piano have been done by others in different ways. Takegawa [5] used a projector mounted on a full 88-key keyboard to project information on the keys as well as piano roll visualization on a white board behind the keys. Rogers [6] published a similar system using a projector but introduced the concept of social learning, namely attention (by seeing and hearing the song), retention (by practicing), and reproduction (by playing the song in full), as a separate mode of learning.

Chow [7] took a different approach by using an HMD in which a camera in front of the device was used to capture images to be rendered as AR effects. Hackl [8] designed an application named HoloKeys that also uses an HMD device called HoloLens. The device had a translucent screen and added virtual objects directly to it in comparison to the previous one. Hackl also argued that the drawback was the limited field of view, and although the prototype ran well, this limitation cast doubt in its applicability in the AR world.

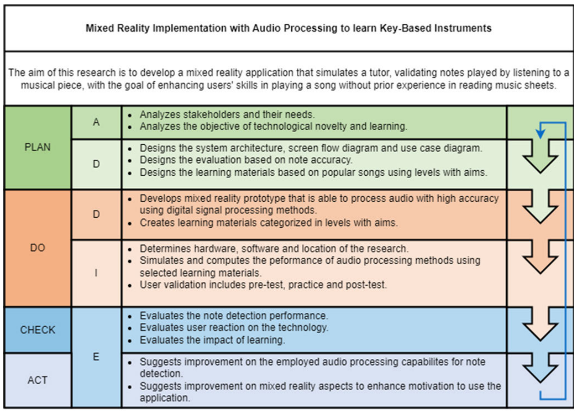
There are also other different approaches, such as using a space invader gamification strategy [14], AR virtual characters to interact with [15], micro projectors to project virtual piano keys and a smartphone for contour and skin color detection [16], simple tripod setup with a smartphone [17], and another HoloLens but with TAM (Technology Acceptance Model) evaluation [18].

Although all the research mentioned shares the common goal of motivating users, the feedback methods to validate note however, except for visual detection, utilized MIDI from digital instruments.

B. AUDIO SIGNAL PROCESSING

A digital representation of sound can be acquired and represented as the sum of sinusoidal functions with various techniques such as FFT (Fast Fourier Transform) [19], [20], [21], [22], [23], Constant-Q Transform [24], and DFT (Discrete Fourier Transform) [25], [26], [27]. FFT has the ability to process signals directly into the spectral domain, opening up possibilities in the field of music [19]. This is in accordance with the requirements of the algorithm for interactive music applications [22], [28].

To get fundamental frequencies, piano audio must first be distinguished from noise, then the adjacent frequencies must be separated, and finally an appropriate window function must be selected [21]. The window function improves the accuracy of peak frequency estimation in measuring piano frequencies due to the phenomenon of spectral leakage, but its performance can vary between applications. Different purposes have shown different performance, which makes window comparison relevant [29], [30], [31].



Mixed Reality Implementation with Audio Processing to learn Key-Based Instruments

The aim of this research is to develop a mixed reality application that simulates a tutor, validating notes played by listening to a musical piece, with the goal of enhancing users' skills in playing a song without prior experience in reading music sheets.

PLAN	A	• Analyzes stakeholders and their needs. • Analyzes the objective of technological novelty and learning.
	D	• Designs the system architecture, screen flow diagram and use case diagram. • Designs the evaluation based on note accuracy. • Designs the learning materials based on popular songs using levels with aims.
DO	D	• Develops mixed reality prototype that is able to process audio with high accuracy using digital signal processing methods. • Creates learning materials categorized in levels with aims.
	I	• Determines hardware, software and location of the research. • Simulates and computes the peformance of audio processing methods using selected learning materials. • User validation includes pre-test, practice and post-test.
CHECK	E	• Evaluates the note detection performance. • Evaluates user reaction on the technology. • Evaluates the impact of learning.
ACT		• Suggests improvement on the employed audio processing capabilites for note detection. • Suggests improvement on mixed reality aspects to enhance motivation to use the application.

FIGURE 1. Establishment process of SLE instance.

C. MOTIVATION IN LEARNING MUSIC

There are factors that contribute to student dropout from piano lessons: two of them are study material and motivation [9]. To address these factors, a variety of songs and gamification techniques, which include feedback mechanisms, can be used to enhance student engagement and learning. Furthermore, despite technological advancements, asynchronous learning in musical instruments may be limited due to sensitivity to signal transmission delays [32]. As a result, real-time validation cannot rely on internet-based applications.

Innovation through audio-visual media using gamification can make learning musical notation more interesting. The three aspects, namely mechanics, dynamics, and aesthetics, are considered the gamification elements in designing the application [33].

To measure the performance of an increase in motivation, TAM and ARCS (Attention, Relevance, Confidence, Satisfaction) are used in a few studies [18], [34]. However, HMSAM (Hedonic-Motivation System Adoption Model) can be used to properly assess the motivation to adopt immersive applications further due to the many cognitive absorption variables considered [35].

III. MATERIALS AND METHODS

The application was developed in C# using Unity and deployed to HoloLens 2. The implementation incorporated the SLEEG (Smart Learning Environment Establishment Guideline) tool, derived from a simplified model of SLE (Smart Learning Environment), to assist in evaluating the effectiveness, efficiency, and engagement of the learning process [36]. The steps can be seen in Fig. 1.

In the planning phase, the intended impact can be depicted using an impact model on Fig. 2 from DRM (Design Research Methodology) [37]. Based on the model, the introduction of a learning application that combines mixed reality and audio processing offers an approach to gather play data for a broader range of musical instruments, bypassing the need for MIDI. The success criterion is an increase in students' ability to play musical instruments, which can be achieved by having

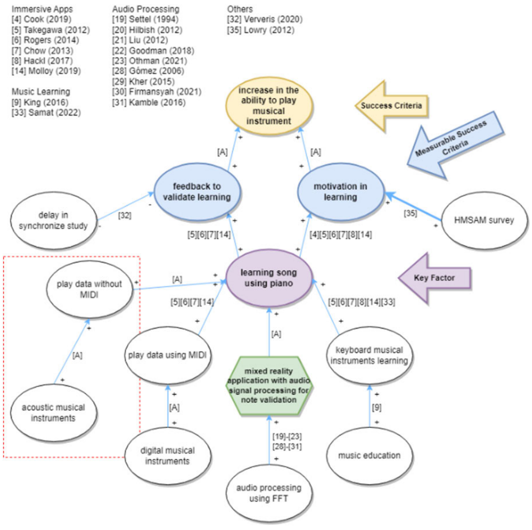


FIGURE 2. The impact model.

TABLE 1. Variables in audio analysis.

Input Variables	Description of Values
MIDI piano songs with various characteristics: texture, tempo and duration.	Textures: monophonic and polyphonic. Tempo: in bpm Duration: in seconds
Window function in FFT	Hanning, Hamming, Blackman or Blackman-Harris
Audio sample rate	44100 Hz, 48000 Hz or 96000 Hz
FFT bin size	2048, 4096, 8192
DSP buffer size	128, 256 or 512
Output Variables	Description of Values
Note accuracy	In percentage (0-100%)
Maximum latency / delay	In ms
Sound artifacts count	Units in frame execution

two measurable success criteria: feedback and motivation. Feedback and motivation can be evaluated by measuring the performance of audio processing to validate notes and by measuring the users' experience when testing the application, respectively.

A. AUDIO ANALYSIS

The research analyzed several samples of piano audio performance that were compared to the MIDI-based file being played in the application. The sound was picked up by the microphone in HoloLens 2 and applied to a window function before being transformed into the frequency domain using FFT. The peaks of the resulting frequencies were corrected using parabolic interpolation and then mapped to the frequency table of a full 88-key piano using a modified binary search. Table 1 shows the variables considered to find the best configuration.

There were three stages to test all the combinations: artifact filtering, the best configuration of a window, and the final frequency analysis. These stages were designed to be efficient, so not all combinations had to be tested in longer songs since each combination had to be individually tested in real time.

The first stage served as a screening process to filter out significant artifacts in a simple and short monophonic song. Sound artifacts lowered accuracy and could be caused by a combination of sample rate and buffer values. They were observable by the buzzing sound they produced during playback. In this stage, one song and one window function were sufficient to evaluate artifacts, resulting in a total of 27 combinations. Combinations that achieved 100% accuracy were selected to proceed to the next stage.

The selected combinations were tested in the second stage using various window functions against a number of short polyphonic songs. The output of this stage was the selection of one combination for each window function with the highest accuracy, totaling four combinations that would proceed to the final stage.

Finally, the remaining four combinations were tested in the last stage with long and complex songs and a greater range to further analyze the frequency response. The result of this analysis was quantitative and based on accuracy and range.

In practice, the audio input in this method was designed to be acquired automatically to simplify the test without the need for manual piano playing in real-time. This was achieved by connecting the HoloLens 2 device to an external speaker using an active converter with a DAC chip and then allowing the device to listen while the song itself was being played by the application. The sound from the speaker was captured by the microphone and analyzed, simulating a piano performance.

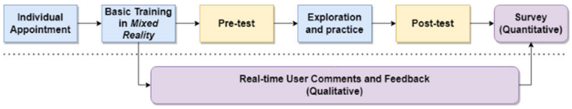


FIGURE 3. The flow of user testing.

B. USER MOTIVATION

The research adopted the Kirkpatrick framework, which had four levels of evaluation, but only the first two levels were executed [38]. Seven variables from HMSAM were used as indicators in a survey given to users in Level 1 (reaction). In Level 2 (learning), a pre-test and post-test stage with a minimum of 15 minutes of exploration and practice in between were conducted. Participants were also encouraged to provide comments and feedback during the test, which added a mixed-methods approach to the evaluation. In addition, participants were given an introduction to the technology and some basic training before they began playing a song. This training included learning the various gestures, such as air tap, hold, and touch, that were used to manipulate the virtual objects. The flow of the whole test is illustrated in Fig. 3.

IV. GAMIFICATION APPROACH

The overview of the game can be described as follows:

- Mechanics: variety of songs, player level, rhythm-based game, scores.
- Dynamics: songs are available based on player level, higher levels are harder, different gameplay modes, interactive features, increase in player's proficiency.

TABLE 2. Description of song levels.

Level	Characteristics	Aim
1	Monophonic texture, slow tempo, under 1 minute	Train one hand in very simple songs.
2	Polyphonic texture (mono for each hand), slow tempo, under 1 minute	Train both hands with two notes at max at a time.
3	Polyphonic texture with chords, slow tempo, under 1 minute	Train left hand with chords while playing melody.
4	Polyphonic texture, medium tempo, under 2 minutes	Train using harder songs with many chord progressions.
5	Polyphonic texture, fast tempo, up to 5 minutes	Train using very hard and fast songs to test the limit.

- Aesthetics: gamification and mixed reality add joy, levels are fun and challenging, audio processing encites curiosity.

The objective of the game is to enhance the user's piano playing proficiency by achieving high scores while playing songs on the instrument. The accuracy of the user's note playing during the game is evaluated, and feedback is provided at the end of each song session. The game provides five levels that are increasingly difficult. The first level is aimed at training one hand, whereas the next levels are for both hands. In order to progress to the next level, a player must achieve a minimum score of 80 in the current level. Table 2 describes the characteristics and aim of each level.

At the start of the program, the user must calibrate the virtual validation line to the back of a real-life piano, which has markers on its keys. After that, user can choose one of the modes adopted from Rogers [6] to practice. These modes are:

- Watch: the user can watch the full performance of the song where the notes approach the validation line and play the sound automatically.
- Train: the user is given a time stop with a 3-second countdown when a note touches the validation line until the user plays the right note.
- Play: the user plays the song normally when the notes touch the validation line. This mode is used to assess the user's proficiency.

The application includes an extra mode called analyze, which is internally used to evaluate the audio processing. This mode is similar to the watch mode in which notes are played, but the system also uses a microphone to validate the sound of the notes being played.

V. RESULTS

A. AUDIO ANALYSIS

For each combination in each stage, a song was played under analyze mode. In the first stage, a Hanning window function and the "Happy Birthday" level 1 song were used. Out of the 27 total combinations tested, 12 were found to have significant artifacts, resulting in reduced accuracy. The other 10 combinations achieved 100% accuracy and were selected to proceed to the next stage of evaluation.

In the second stage, four level 3 songs were used for each combination with each window function, resulting in a total of 160 combinations evaluated. The combination for each window with the highest accuracy was selected to proceed to the last stage, totaling four combinations.

TABLE 3. Window function comparison in final stage test.

Window	FFT Bin	Rate	DSP Buffer	Accuracy
Hanning	8192	48000	512	89.32%
Hamming	8192	44100	512	89.07%
Blackman	8192	44100	512	91.72%
Blackman-Harris	8192	44100	512	97.28%

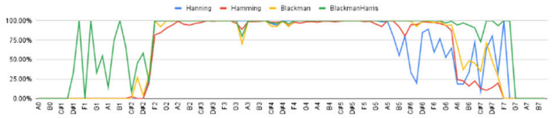


FIGURE 4. Frequency analysis in final stage test.

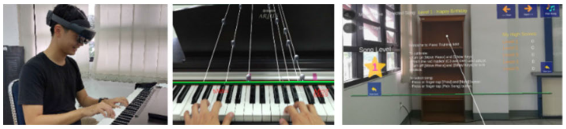


FIGURE 5. User testing the application using HoloLens 2.

In the last stage, each combination was tested with four level 5 songs. Based on the test results, it was found that the Blackman-Harris window function had the highest accuracy at 97.28% for complex songs. The results of this final stage of analysis are summarized in Table 3.

Referring to Fig. 4, it was also observed that all windows experienced fluctuations in accuracy below E2 and above A5. However, the Blackman-Harris window consistently outperformed all others in these regions and was chosen for the user evaluation test because of its highest accuracy.

A limitation to this method was that the presence of harmonics was considered input, which could lead to additional validation by the system. It was also observed that there was a maximum latency of approximately 1000 ms, and this was most likely caused by the hardware and game engine used to capture the audio. These limitations will be discussed in the discussion section.

B. USER MOTIVATION

The user tests were divided into four weekly batches, with a total of 31 participants. Due to limitations in device availability and location at the time, the participants consisted of undergraduate and postgraduate students at the Bandung Institute of Technology who had access to the testing lab. They came from various musical backgrounds and had no prior experience with mixed reality applications for learning piano. Each participant in the test spent an average of 45 minutes in a single session, from the appointment time to the survey. Fig. 5 provides a visual representation of the user testing process using the HoloLens 2 device.

Additionally, a digital piano, the Yamaha Arius YDP-141 with built-in speakers, was used instead of an acoustic piano. Despite this deviation from the original plan, it was still considered acceptable, as capturing sound from a digital piano

TABLE 4. Descriptive analysis of survey questions.

Variable	Possible Range	Actual Range	Mean	Mean per Item	Standard Deviation
JOY	3-15	11-15	14.194	4.731	0.554
CTL	2-10	4-10	8.935	4.468	0.804
IM	2-10	5-10	8.903	4.452	0.843
CUR	2-10	6-10	9.355	4.677	0.621
PEOU	3-15	6-15	11.581	3.860	1.248
PU	3-15	8-15	13.613	4.538	0.700
BIU	2-10	4-10	9.129	4.565	0.738

TABLE 5. HMSAM paths results.

To	From	Tests				Coeff.
		SW	DW	F	t	
PU	PEOU	0.098	-	0.008	0.008*	0.458
CUR	PEOU	0.000	-	0.145	0.145	0.159
JOY	PEOU	0.972	-	0.000	0.000**	0.469
CTL	PEOU	0.528	-	0.002	0.002*	0.356
BIU	PU	0.113	1.611	0.000	0.877	-0.014
	CUR				0.000**	0.610
	JOY				0.005*	0.361
IM	CUR	0.088	1.744	0.030	0.801	-0.052
	JOY				0.005*	0.529
	CTL				0.401	-0.159

*p-value < 0.01, **p-value < 0.001.

with speakers can be treated similarly to capturing sound from an acoustic piano.

The survey used HMSAM indicators and had a total of 17 questions, in which the responses were measured on a Likert scale. All question items were valid using a p-value $<$ 0.05 and reliable using Cronbach's alpha $>$ 0.6. The descriptive analysis can be found in Table 4. Based on the mean per question result, it was found that most participants felt joy, while a few faced difficulties in using the technology (PEOU). The majority of participants expressed positive feelings toward all variables.

When analyzing the paths of HMSAM, several tests were conducted. The results of these tests can be found in Table 5. During the tests, path 2 (PEOU to CUR) was found to be not normal using Shapiro-Wilk (SW) and excluded. No autocorrelation was observed using the Durbin-Watson (DW) test with an alpha value of 0.01. Other tests were the F-test and the partial t-test.

Fig. 6 shows the final result of the accepted model from HMSAM. The results indicated that joy and curiosity had a significant positive influence on users' behavioral intention to use the application, while perceived usefulness and control did not demonstrate a positive impact on the intention to use. To gain further insights, comments and feedback were taken into consideration.

Many participants commented on their surprise while using the technology, particularly when they were able to touch the holograms during navigation and piano calibration in the application. However, there were also several comments about the difficulty of the technology, particularly when it came to calibrating the validation line to the back of the piano keys, even with the help of markers. Some participants needed more time to become more familiar and comfortable with the technology.

Fig. 7 provides a summary of the percentage of user comments regarding the aspects that need improvement in order to enhance their willingness to use the application. Regarding the training aspect of the application, the majority of users found that the training mode and the cue of lines towards the corresponding keys were helpful in guiding their playing. However, some users struggled with hand and finger positioning and suggested the addition of some features to assist them. Some users wanted more time to use the application in order to advance to higher levels.

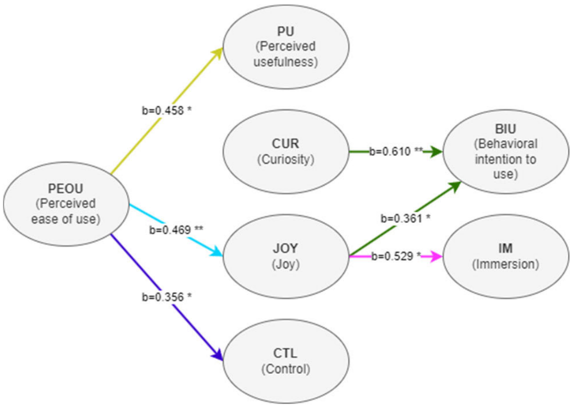


FIGURE 6. HMSAM with coefficient values.

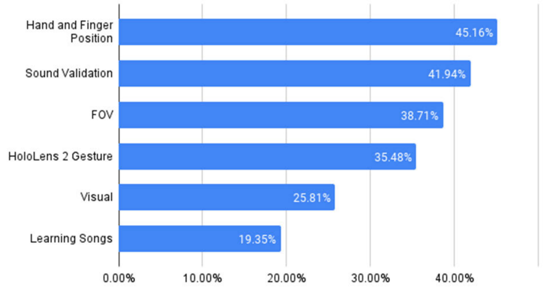


FIGURE 7. Percentage of users commenting on aspects to improve.

Some users also commented on the system's difficulty in detecting frequencies from the lower side. Upon investigation, it was found that the frequency range was smaller than what was tested prior to user evaluation. As a result, some users chose to train and play songs that played at higher frequencies. Others reported that a missed note in a chord was validated occasionally.

Another noteworthy reason was the FOV (field of view) of HoloLens 2. The device has a wider FOV than the first HoloLens used by Hackl [8] and Molero [18]. However, the issue still remains and is most apparent when playing a song with a wider range.

To evaluate the participants' performance, pre-test and post-test scores were calculated, with a possible maximum score of 500 (five levels with a score out of 100 for each level). Several tests were conducted to analyze the collected scores. Scores were normal using Shapiro-Wilk with a p-value $>$ 0.05. The differences between the post-test and pre-test scores from all batches were the same using Levene test statistics (0.793) and a p-value $>$ 0.05. The increase in scores was also significant using the paired t-test. Finally, normalized gain

TABLE 6. Users' score gains.

Pre-tests	Mean	Post-tests	Mean
Song Lv. 1	90.871	Song Lv. 1	98.452
Song Lv. 2	71.226	Song Lv. 2	87.097
Song Lv. 3	40.935	Song Lv. 3	74.323
Song Lv. 4	3.677	Song Lv. 4	35.000
Song Lv. 5	0	Song Lv. 5	3.581
Effectiveness Results			
Total Post - Pre	91.742	Total Max - Pre	293.290
Normalized gain	31.28% (Moderate)		

tests were used to determine the effectiveness of the study, which can be seen in Table 6.

The results indicated that there was an overall increase in scores as participants progressed through the levels. On average, participants were able to pass level 2 after a period of training, that involved both hands. Only a few participants were able to reach level 5, which was expected since the songs were more challenging.

VI. DISCUSSION

The application faced challenges with the latency, range, and accuracy of note detection. Latency could have been caused by hardware limitations and the Unity game engine. In terms of hardware, a delay was also present when using voice commands in HoloLens 2. On the other hand, audio processing in Unity may not be efficient.

The accuracy of note detection in relation to the tested range was found to be slightly less reliable due to variations in sound production among different instruments. For instance, the audio analysis relied on a speaker, while the user testing involved a piano. The accuracy of note detection could also be influenced by harmonics, particularly in crowded chords where multiple notes are present. Minimizing harmonics may be impractical due to the nature of soundwaves and the richness of frequencies in music. However, this characteristic may be leveraged in conjunction with note detection to predict lower frequencies that may have been missed. To improve accuracy, a higher-quality microphone can be used in conjunction with the mixed reality device. Additionally, implementing an AI model can enhance the validation of notes from various sounds based on their peaks. A different game engine can also be explored to address audio processing issues such as latency. Alternatively, a limitation on the playable range can be imposed as another approach.

The results of the user evaluation revealed that users were primarily motivated by curiosity and enjoyment to use the mixed reality-based application, with limited impact on perceived usefulness and control. The research did not find a correlation between ease of use and curiosity (as indicated by its normality), which could be due to individual differences in adaptability and the strong desire to use the application despite its difficulties. Based on these findings, a mixed reality application with audio processing at its current state could only be used as an entertainment tool and as a supplement to traditional learning methods to help maintain motivation. Score gains were also observed to be moderate. To enhance motivation and score, the application can incorporate more intricate level designs that gradually train each hand before playing challenging songs in full and provide additional guidance on hand and finger placement. Moreover, more training sessions can be included to help users familiarize themselves with the technology and master the songs.

The use of audio processing in education can facilitate the learning of various instruments and songs in creative ways. However, inaccuracies exist and can be mitigated through supervision by a tutor or an experienced musician who guides the play. In the case of user testing in this research, an assistant was present.

VII. CONCLUSION

Audio processing in mixed reality applications offers flexibility in learning piano. Based on the results using the three-stage test in audio analysis, FFT with the Blackman-Harris window gives the highest accuracy in note detection. Validating notes through a microphone may be less reliable compared to using MIDI input due to variations in sound production. However, it can still be utilized to keep users motivated and engaged in learning piano in an exciting way. In the future, effectiveness will be improved by addressing these issues and incorporating users' feedback.

REFERENCES

[1] U. H. Salsabila, L. I. Sari, K. H. Lathif, A. P. Lestari, and A. Ayuning, "Peran teknologi dalam pembelajaran di masa pandemi COVID-19," *Al-Mutharahah, Jurnal Penelitian dan Kajian Sosial Keagamaan*, vol. 17, no. 2, pp. 188–198, Nov. 2020, doi: 10.46781/al-mutharahah.v17i2.138.

[2] F. S. H. S. Sinaga, "Sustainabilitas pendidikan musik selama pandemi COVID-19," *Prosiding Seminar Nasional Pascasarjana*, vol. 3, no. 1, pp. 980–988, 2020.

[3] S. Serafin, A. Adjorlu, N. Nilsson, L. Thomsen, and R. Nordahl, "Considerations on the use of virtual and augmented reality technologies in music education," in *Proc. IEEE Virtual Reality Workshop K-12 Embodied Learn. Through Virtual Augmented Reality (KELVAR)*, Los Angeles, CA, USA, Mar. 2017, pp. 1–4, doi: 10.1109/KELVAR.2017.7961562.

[4] M. J. Cook, "Augmented reality: Examining its value in a music technology classroom. Practice and potential," *Waikato J. Educ.*, vol. 24, no. 2, pp. 23–38, Nov. 2019, doi: 10.15663/wje.v24i2.687.

[5] Y. Takegawa, T. Terada, and M. Tsukamoto, "A piano learning support system considering rhythm," in *Proc. ICMC*, 2012, vol. 32, no. 1, pp. 9–25.

[6] K. Rogers, A. Röhlig, M. Weing, J. Gugenheimer, B. Könings, M. Klepsch, F. Schaub, E. Rukzio, T. Seufert, and M. Weber, "P.I.A.N.O.: Faster piano learning with interactive projection," in *Proc. 9th ACM Int. Conf. Interact. Tabletops Surf.*, Nov. 2014, pp. 149–158, doi: 10.1145/2669485.2669514.

[7] J. Chow, H. Feng, R. Amor, and B. C. Wünsche, "Music education using augmented reality with a head mounted display," in *Proc. 14th Australas. User Interface Conf.*, vol. 139, Jan. 2013, pp. 73–79. [Online]. Available: https://dl.acm.org/doi/10.5555/2525493.2525501, doi: 10.5555/2525493.2525501.

[8] D. Hackl and C. Anthes, "HoloKeys—An augmented reality application for learning the piano," *Forum Media Technol.*, pp. 140–144, Nov. 2017.

[9] K. King, "Parting ways with piano lessons: Predictors, invoked reasons, and motivation related to piano Student dropouts," Ph.D. dissertation, Dept. Music, Université d'Ottawa/Univ. Ottawa, Ottawa, ON, Canada, 2016.

[10] G. Lamé, "Systematic literature reviews: An introduction," in *Proc. Design Soc., Int. Conf. Eng. Design*, vol. 1, 2019, pp. 1633–1642, doi: 10.1017/dsi.2019.169.

[11] L. P. Yulianti and K. Surendro, "Implementation of quantum annealing: A systematic review," *IEEE Access*, vol. 10, pp. 73156–73177, 2022, doi: 10.1109/ACCESS.2022.3188117.

[12] P. Milgram and F. Kishino, "A taxonomy of mixed reality visual displays," *IEICE Trans. Inf. Syst.*, vol. E77-D, no. 12, pp. 1321–1329, 1994.

[13] Microsoft Learn. *What is Mixed Reality? Mixed Reality*. Accessed: Apr. 14, 2023. [Online]. Available: https://docs.microsoft.com/en-us/windows/mixed-reality/discover/mixed-reality

[14] W. Molloy, E. Huang, and B. C. Wünsche, "Mixed reality piano tutor: A gamified piano practice environment," in *Proc. Int. Conf. Electron., Inf., Commun. (ICEIC)*, Auckland, New Zealand, Jan. 2019, pp. 1–7, doi: 10.23919/ELINFOCOM.2019.8706474.

[15] C. A. T. Fernandez, P. Paliyawan, C. C. Yin, and R. Thawonmas, "Piano learning application with feedback provided by an AR virtual character," in *Proc. IEEE 5th Global Conf. Consum. Electron.*, Kyoto, Japan, Oct. 2016, pp. 1–2, doi: 10.1109/GCCE.2016.7800380.

[16] C. Sun and P. Chiang, "Mr. Piano: A portable piano tutoring system," in *Proc. IEEE 25th Int. Conf. Electron., Electr. Eng. Comput. (INTERCON)*, Lima, Peru, Aug. 2018, pp. 1–4, doi: 10.1109/INTERCON.2018.8526423.

[17] I. M. Jamal and E. Kiliç, "EasyARPiano: Piano teaching mobile app with augmented reality," in *Proc. Int. Conf. Forthcoming Netw. Sustainability AIoT Era (FoNeS-AIoT)*, Nicosia, Turkey, Dec. 2021, pp. 66–71, doi: 10.1109/FoNeS-AIoT54873.2021.00024.

[18] D. Molero, S. Schez-Sobrino, D. Vallejo, C. Glez-Morcillo, and J. Albusac, "A novel approach to learning music and piano based on mixed reality and gamification," *Multimedia Tools Appl.*, vol. 80, no. 1, pp. 165–186, Jan. 2021, doi: 10.1007/s11042-020-09678-9.

[19] Z. Settel and C. Lippe, "Realtime musical applications using FFT based resynthesis," in *Proc. Int. Comput. Music Conf.* San Francisco, CA, USA: International Computer Music Association, 1994, p. 338.

[20] N. Hilbish, "Multiple fundamental frequency pitch detection for real time MIDI applications," M.S. thesis, Dept. Eng., Virginia Commonwealth Univ., Richmond, VA, USA, 2012, doi: 10.25772/9DZN-A722.

[21] Y. Liu, "Research piano frequency based on DSP and window function," in *Proc. 9th Int. Conf. Fuzzy Syst. Knowl. Discovery*, Chongqing, China, May 2012, pp. 1917–1919, doi: 10.1109/FSKD.2012.6234206.

[22] T. A. Goodman and I. Batten, "Real-time polyphonic pitch detection on acoustic musical signals," in *Proc. IEEE Int. Symp. Signal Process. Inf. Technol. (ISSPIT)*, Louisville, KY, USA, Dec. 2018, pp. 1–6, doi: 10.1109/ISSPIT.2018.8642626.

[23] K. A. Othman and A. A. Z. Abidin, "Signal processing application for musical notes recognition," in *Proc. IEEE 19th Student Conf. Res. Develop. (SCOReD)*, Kota Kinabalu, Malaysia, Nov. 2021, pp. 135–139, doi: 10.1109/SCOReD53546.2021.9652762.

[24] S. W. Foo and E. W. T. Lee, "An innovative approach to transcription of polyphonic signals," in *Proc. Int. Conf. Inf., Commun. Signal Process.*, 2001, pp. 1–5.

[25] A. Klapuri, "Pitch estimation using multiple independent time-frequency windows," in *Proc. IEEE Workshop Appl. Signal Process. Audio Acoust. (WASPAA)*, New Paltz, NY, USA, Oct. 1999, pp. 115–118, doi: 10.1109/ASPAA.1999.810863.

[26] A. P. Klapuri, "Multiple fundamental frequency estimation based on harmonicity and spectral smoothness," *IEEE Trans. Speech Audio Process.*, vol. 11, no. 6, pp. 804–816, Nov. 2003, doi: 10.1109/TSA.2003.815516.

[27] E. Gómez, "Tonal description of music audio signals," Ph.D. dissertation, Dept. Inf. Commun. Technol., Univ. Pompeu Fabra, 2006.

[28] P. de la Cuadra, A. Master, and C. Sapp, "Efficient pitch detection techniques for interactive music," in *Proc. ICMC*, 2001.

[29] V. Kher and T. S. Lamba, "Multi-window comparison of SIR performance in extraction of mono-aural vocal and non-vocal components in REPET," in *Proc. Int. Conf. Signal Process., Comput. Control (ISPCC)*, Waknaghat, India, Sep. 2015, pp. 377–382, doi: 10.1109/ISPCC.2015.7375059.

[30] M. R. Firmansyah, R. Hidayat, and A. Bejo, "Comparison of windowing function on feature extraction using MFCC for speaker identification," in *Proc. Int. Conf. Intell. Cybern. Technol. Appl. (ICICyTA)*, Bandung, Indonesia, Dec. 2021, pp. 1–5, doi: 10.1109/ICICyTA53712.2021.9689160.

[31] H. Kamble and G. S. Phadke, "Frequency response analysis of respiratory sounds and comparative study for windowing techniques," in *Proc. Int. Conf. Signal Process., Commun., Power Embedded Syst. (SCOPES)*, Paralakhemundi, India, Oct. 2016, pp. 210–215, doi: 10.1109/SCOPES.2016.7955809.

[32] A. Ververis and A. Apostolis, "Online music education in the era of COVID-19: Teaching instruments in public music secondary schools of Greece during the 2020 lockdown," in *Proc. Int. Conf. Stud. Educ. Social Sci. (ICSES)*, 2020, pp. 1–9.

[33] J. Samat, A. Baharum, and C. Andin, "Identifying elements of gamification for reading music notation in music education," in *Proc. 13th Int. Conf. Inf. Commun. Technol. Converg. (ICTC)*, Jeju Island, South Korea, Oct. 2022, pp. 563–567, doi: 10.1109/ICTC55196.2022.9952727.

[34] M. Yu-Chun and L. H. Koong, "A study of the affective tutoring system for music appreciation curriculum at the junior high school level," in *Proc. Int. Conf. Educ. Innov. Through Technol. (EITT)*, Tainan, Taiwan, Sep. 2016, pp. 204–207, doi: 10.1109/EITT.2016.47.

[35] P. Lowry, J. Gaskin, N. Twyman, B. Hammer, and T. Roberts, "Taking 'fun and games' seriously: Proposing the hedonic-motivation system adoption model (HMSAM)," *J. Assoc. Inf. Syst.*, vol. 14, no. 11, pp. 617–671, Nov. 2013, doi: 10.17705/1jais.00347.

[36] Y. Rosmansyah, B. L. Putro, A. Putri, N. B. Utomo, and Suhardi, "A simple model of smart learning environment," *Interact. Learn. Environ.*, pp. 1–22, Jan. 2022.

[37] L. Blessing and A. Chakrabarti, *DRM: A Design Research Methodology*. London, U.K.: Springer, 2009, pp. 13–42.

[38] D. L. Kirkpatrick and J. D. Kirkpatrick, *Evaluating Training Programs: The Four Levels*. Oakland, CA, USA: Berrett-Koehler Publishers, 2006.

INSAN GANANG PUTRANDA received the bachelor's degree in software engineering from The University of Melbourne, in 2009. He is currently pursuing the master's degree in informatics with a focus on media and mobile technology with the Bandung Institute of Technology. He is passionate about multimedia applications and game development and actively pursuing his interest in this field.

ASIYA MUFIDA YUMNA received the bachelor's degree in information systems and technology from the Bandung Institute of Technology, in 2022. Her research interests include human-centered computation, interactive computing, and educational technologies. She is currently starting her career plans further into user experience and human–computer interaction.

YUSEP ROSMANSYAH received the bachelor's degree in electrical engineering from the Bandung Institute of Technology, Indonesia, in 1993, and the M.Sc. and Ph.D. degrees from the University of Surrey, U.K., in 1996 and 2003, respectively. He has been a Researcher and a Professor in smart multimedia processing with the School of Electrical Engineering and Informatics, Bandung Institute of Technology. His research interests include multimedia, educational technology, and cyber security.

YUDA SUKMANA received the bachelor's degree in electrical engineering education from the Indonesia University of Education (UPI), and the master's degree in electrical engineering from the Bandung Institute of Technology (ITB), where he is currently pursuing the Ph.D. degree in electrical engineering and informatics. His research interests include educational technology, mobile application, and artificial intelligence.