UNIVERSITY OF ZAGREB
FACULTY OF ELECTRICAL ENGINEERING AND COMPUTING

SEMINAR

Literature Review of Augmented Reality Piano Learning Solutions

Mirta Moslavac
Mentor: *prof. Lea Skorin-Kapov*

Zagreb, January 2022

CONTENTS

Introduction

The ever-growing area of augmented reality (AR) offers new ways of executing regular assignments with its characteristic of interactability with the overlaid digital information amongst the real-world environment the user is located in. That additional information can amplify or even extend the possibilities available to the user in the physical world and therefore unlock yet unknown techniques for tackling known problems or further advance already familiar approaches to those problems. The opportunities offered by AR can be applied to the area of acquiring a certain skillset, amongst others, like learning how to play a new instrument. Both the academia and the consumers recognized the piano as the ideal instrument for translation of practising within an augmented reality, which resulted in a great amount of AR systems that rely on the piano as the main focus. These available implementations share fundamental features due to the shared nature of required showcasing of piano playing and yet differ in some other specific features, which can be compared and discussed. The objectives of this seminar include:

1. discuss the baselines of music content integration within the AR medium,
2. give an overview of existing theoretical or concrete AR piano learning systems and showcase and compare the proposed systems and features within them, and
3. build on the previously mentioned application features.

The objectives are achieved through the five main chapters of the seminar. Following the introductory chapter, the second chapter discusses how music-related scenarios can be enhanced using AR and how music should be represented in such applications, with an emphasis on piano education due to the nature of this seminar. Multiple existing AR piano teaching solutions are presented and described within the third chapter, while considering the similarities and differences between the given solutions. The fourth chapter highlights specific features within those solutions that could be of larger significance to the end-user. Ultimately, the ideas and comparisons made throughout the seminar are reflected on in the final chapter. Additional chapters provided are a list of references, a list of figures, and a list of abbreviations used in the seminar.

1. Augmented Reality Applications in Music

Augmented reality poses itself as a reasonable option for providing assistance in musical activities due to its nature of unintrusiveness of the physical environment by supplying supplementary virtual data on top of the stimuli incoming from the real world. The musician can then, for example, receive feedback depending on how they are performing while playing a certain instrument, while still being able to perform the intended interactions with the physical instrument undisturbed. Moreover, AR is not the only technology that is fit for those services, but rather the entire umbrella of related technologies called extended reality (XR), to which AR also belongs. Turchet et al. termed the field of music intersecting with such technologies *Musical XR* [25]. Those technologies include virtual reality (VR), augmented reality, augmented virtuality (AV), and mixed reality (MR) - the differences and similarities between these terms are discussed in [19], but most notably, the terms MR and AR can be interchanged in the case of AR implementations supporting interactions between the digital and the physical due to not enough disambiguation between the technologies as of yet. As discussed in [25], no concrete definition of *Musical XR* exists, since different technologies execute certain aspects important for the phenomenon differently, but for an XR technology to be deemed as musical, it must fulfill four minimal prerequisites:

1. *existence of virtual elements* - addition of digital information in varying degrees,
2. *spatial presence* - the user(, physical) and digital elements all coexist in the same space,
3. *interactivity* - inclusion of interaction between the user and the digitally manipulated environment to varying extent, and
4. *sonic organization* - auditory information should serve as irreplaceable fundamental elements of the implementations using the chosen XR technology, influencing the previous three prerequisites.

Any *Musical XR* implementation is encouraged to include the modification of as many

senses as possible to increase the quality of the user's overall experience. In the case of AR, the sensory modalities used in respective implementations include, but are not limited to, auditory, visual and proprioceptive stimuli.
One of the interviewees interviewed by Turchet et al., P. Cook, offers an interesting idea that music itself is also a form of XR since it extends and modifies the environment the person experiencing the music is in, making music and XR technologies a natural pairing. But, according to research done by Turchet et al. for [25], not all XR technologies can be applied to all problems equally - some excel in certain musical applications more than others. AR, for example, is the most appropriate for those musical applications that seek the presence of physical visual stimuli, as well as the additional information, such as when practising an instrument. When looking at general areas of application, VR is mostly used for musical performances, while most of the music-oriented AR implementations focus on the educational aspect of music, as seen later in the seminar with the larger amount of implementations based on an even more specific type of applications focused on piano education that are discussed. Areas of application covered in [25] include: education, entertainment, performance, composition, and sound engineering – the first three applications heavily outweigh the other two in the number of available implementations, either commercially or academically, attributable to the much more narrow targeted audience of composers and sound engineers. The authors also notice how only an insignificant amount of XR systems intended for usage by tutors or musical professionals have been developed, providing a large untapped area for the *Musical XR* expansion.

In terms of music education and pedagogy, and the AR technology, a handful of instruments have been considered as the focal points of AR-aided music applications alongside the piano, such as the acoustic guitar [16] [15] [14] [20], the electric bass guitar [4] and the theremin [13], but with much fewer solutions offered. One reason why that is so is due to the hindrances posed by the nature of playing an instrument and our limited vision if focusing primarily on visual AR elements. The most notable AR/MR devices, and the only ones that can offer true immersion in real time, today are HMDs, worn on top of the user's head and therefore augmenting the user's view of what is visible to them. Many instruments, like the aforementioned guitars, cannot be fully captured by the HMDs due to the barrier of their small FoV, but what is a bigger problem is the position in which the instrument is played in respective to the user's head position, which can be awkward for displaying any additional information on top of or near the instrument in the time of playing or practising.

The piano does not pose such an issue since the user is situated in front of the entire instrument, far enough from it for the user to have an enjoyable experience and not have to struggle with unnatural positioning, even though the AR hardware FoV limitations are still present. The aforementioned lack of user/AR positioning issues and the overwhelming popularity of learning how to play that instrument, alongside the challenged traditional methods of piano teaching, as discussed by Guo [9], new approaches in piano education enhanced by new technologies have been considered, with the domination of AR piano learning systems.

2. Overview of the General Aspects of Certain AR Piano Learning Solutions

A variety of prototyping solutions focused on piano playing using augmented reality have emerged within the last few years, thanks to the popularization of AR and access to the newly released hardware supporting such technology within the academic community, as parts of published papers and researches. Only a few solutions are available to the general public due to the yet unpolished handling of AR technology and existing hardware that is not easily accessible to the average user, thanks to relatively high prices and developer/enterprise market exclusivity [19]. What is also lacking is the definite end product "formula" combining hardware and software, while retaining simplicity and fulfilling the minimum requirements for an enjoyable experience for such a task within AR at the same time. All this amounts to the idea of publishing the developed applications not worthwhile – some of the previously published solutions, like Music Everywhere [7], even ceased to exist, most probably due to the issues previously stated. But even though the prototypes developed aren't market-ready yet, the existing significant amount of proposed solutions focused on improving certain elements of the piano learning aspect using AR can only move in a positive direction towards the end goal of reaching a satisfiable solution that can be easily used by the general public, if the interest in such topics continues to grow, which is significantly influenced by the continuation of hardware advancements.

In this seminar, a variety of papers that propose AR piano learning solutions have been analyzed, resulting in their systematic overview through specific feature and concept emphasis and by highlighting their similarities and differences through various comparisons. A short overview of each of the 17 papers mentioned in the rest of the seminar is given in Table 2.1, all found as a part of the research for this seminar on Semantic Scholar[1] and Google Scholar[2] by searching for AR piano solutions.

[1] https://www.semanticscholar.org/

[2] https://scholar.google.com/

Table 2.1: Overview of Analyzed Papers

Paper	Hardware	Software	Tracking[1]	Players[2]	Focus
[1]	PC, webcam	N/A	1	S	chord and scale education
[2]	Hololens	Unity	N/A	S	game-based learning
[3]	Hololens, LeapMotion	UWP, Vuforia	1	S+P	assistance during piano classes
[5]	ARvision-3D	C#, artoolkitX	1	S	gamified notation literacy development
[6]	Android device, Google Cardboard	Unity, Vuforia	1	S	motivation through a virtual character
[7]	HoloLens	UWP	N/A	S	improvisation and engagement
[8]	non-specific HMD	Unity	N/A	S	fingering technique development
[10]	HoloLens	Unity, Vuforia	1	S	enhanced piano playing/learning
[12]	PC, camera	C++, OpenGL	2	S	enhanced piano playing/learning
[17]	HoloLens	Unity, MRTK	2	S	gamified piano playing/learning
[18]	HTC Vive, ZED Mini	Unity	N/A	S	gamified piano playing/learning
[19]	HoloLens 2	Unity, Vuforia	1	S	enhanced piano playing/learning
[22]	Moverio BT-300	Android	N/A	S+P	enhanced piano playing/learning
[23]	HTC Vive Pro, ZED Mini	Unity	1/2	S	notation literacy improvement
[24]	HoloLens	Unity, Vuforia	1	S	visualized piano playing/learning
[26]	Moverio BT-300	Android, artoolkitX	1	S	improve short-term learning
[27]	Moverio BT-300	Android, artoolkitX	1	P	reduce difficulty of piano learning

[1] Marker-based (1) or markerless (2) approach

[2] Single (S) or paired (P)

2.1. Hardware and Software

Even though today displaying AR, or in more specific cases MR, content is associated with true AR headsets or glasses, also known as AR head-mounted displays (HMDs), contrary to possible initial belief, they do not form an overwhelming majority as the hardware of choice for piano learning solutions. Hackl and Anthes [10], Trujano et al. [24], and Cai et al. [3], to name some, created such AR systems using the Microsoft HoloLens[3] as their primary device for delivering augmented content to the user – owning to the fact that the release date of the successor HoloLens 2[4] was in late 2019, not many solutions were yet able to utilize the new generation of the device, but some do exist such as [19]. On the one hand, Microsoft's devices offer the most advanced features for displaying and merging augmented content with the physical one to date, but on the other hand, it should not be surprising that the research is directed towards some more consumer-ready hardware set-ups since the HoloLens lineup still has crucial limitations that need to be improved upon, such as limited field of view (FoV) [19], and are not attainable for everyday consumption. A popular choice for an MR pair of glasses is Epson Moverio BT-300[5], used by Pan et al. [22], Zeng et al. [26] and Zeng et al. [27], utilizing the optical see-through approach to AR headsets, while Chow et al. [5] used the Trivisio ARvision-3D device[6], which is an example of a video see-through HMD. One type of set-ups that are more readily available to the general user is the combination of a VR headset and a stereo camera that offers MR capabilities creating a video see-through HMD used for AR from a device that is intended for VR use, most often combining HTC Vive[7] and ZED Mini[8] as done by Rigby et al. [23] (pictured in Figure 2.1) and Molloy et al. [18]. Fernandez et al. present the set-up with a mobile phone placed inside a Google Cardboard holder, which resembles the previously described set-up [6], as seen in Figure 2.2. Earlier implementations of piano learning systems, such as the one by Barakonyi and Schmalstieg, relied on a combination of a web camera and a computer monitor to display the augmented content to the user while playing the piano [1].

[3]`https://docs.microsoft.com/en-us/hololens/hololens1-hardware`

[4]`https://www.microsoft.com/en-us/hololens`

[5]`https://www.epson.eu/products/see-through-mobile-viewer/moverio-bt-300`

[6]`https://est-kl.com/manufacturer/trivisio/arvision-3d-hmd.html`

[7]`https://www.vive.com/us/`

[8]`https://www.stereolabs.com/zed-mini/`

Figure 2.1: HTC Vive and ZED Mini Set-Up [21]

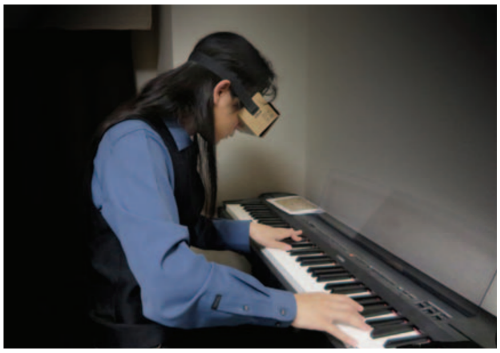

Figure 2.2: Mobile Device and Google Cardboard Set-Up [6]

Most of the solutions used a physical piano or a digital keyboard with integrated Musical Instrument Digital Interface (MIDI) standard[9] support to detect the played keys, while some relied on MR features of interacting with a virtual piano overlay that is pressed at the same time as the user presses the physical key on the piano, e.g. [19]. Other than the MIDI note representation, which mainly models the notes themselves without any additional information about the interaction between notes, MusicXML[10] format can also be used for a truer depiction of the score of the music piece since it models the notation and layout as the one on the music sheet, as described and utilized in [23].

Depending on the device used for displaying augmented content, various software platforms can be used for creating an AR piano learning solution. The majority of applications described in papers that were discussed in this seminar had their entire program logic implemented using Unity3D, followed by Android (with artoolkitX[11], software development kit (SDK)). In case a marker-based approach to tracking was introduced to the system, such as the ones by Fernandez et al. [6] and Cai et al. [3], the most popular tool of choice was the Vuforia Engine[12] or some library inspired by Vuforia like the implementation by Molero et al. [17].

[9] `https://www.midi.org/specifications`
[10] `https://www.musicxml.com/`
[11] `http://www.artoolkitx.org/`
[12] `https://developer.vuforia.com/`

2.2. Combining the Real and the Augmented World

In the case of AR/MR, the most prevalent method of obtaining the final image specifically through an HMD is the see-through method [19], which can be confirmed by the fact that all the HMD-based solutions mentioned within this seminar deliver the final image in this matter. Most implementations opt for the optical see-through approach, which allows the user to blend the true physical perception of the real world with the digital information, while, for example, Rigby et al. [23] choose the other approach, the video see-through method – the method chosen relies on the hardware set-up of the system, which was discussed in the previous section.

If the augmentation does not infer with the user's view, but rather is displayed on an external screen, e.g. monitor or mobile phone screen, then the captured image from the physical world through a camera lens is combined with the virtual information and displayed on the monitor.

In terms of tracking, both the marker-based and markerless approaches seem to be represented relatively equally in the published implementations, not just in the analyzed papers, with the amount of marker-based ones slightly outweighing the markerless, leading to the assumption that no definite choice has been made upon which of the two approaches is more appropriate when developing a system for piano playing. Both approaches have advantages and drawbacks tied to the execution of the systems they're used in, as described in [19]. In short, marker-based tracking relies on visual elements, called fiducial markers, placed in the physical world and, when detected using CV techniques, they are used to appropriately place the virtual elements in the real environment. On the contrary, markerless tracking omits the use of fiducial markers in favour of physical space scanning by only using more advanced CV techniques to determine where to place the aforementioned virtual elements in the analyzed physical environment.

Huang et al. raise a point that the marker-based approach hinders the user's experience of seamless conjunction of the physical and virtual world and therefore Natural Feature Tracking (NFT) methods, i.e. markerless methods, should be used [12]. They propose a system that uses the entire keyboard as a single object that is to be detected and kept track of using computer vision (CV) while displaying the augmented information through a computer monitor, depicted in Figure 2.3 – such an idea could not currently be applied to HMDs due to their limited FoV, nor even in the future, since such wide FoV is probably not obtainable in order to capture an entire set of keys when being situated so close to it. The authors also raise a question about tracking the user's hand, but point out that such approach would lead to diminished accuracy since the tracked hand changes shape during tracking, while static objects such as the keyboard or a regular marker do not, or rather should not.

The authors of the papers that were taken into consideration that opted for the marker-based approach mostly used ready-made toolkits or SDKs like Vuforia and artoolkitX, with image tracking being the most prevalent. The authors of [10] selected the middle C key (the approximately central/starting point on the piano keyboard) as the position where to place a single marker for tracking and the outcomes of the user study conducted by Rigby et al. proposed that the middle C is one of the notes novices learn about and naturally navigate towards it [23], which strengthens the idea of that placement being a good choice since the piece played by the user would most likely be centered around the area of that key. True AR/MR HMD solutions suffer from one of the hardware's biggest drawbacks – limited FoV – which hinders the user experience when trying to navigate through the augmented world. In two of their implementations, Zeng et al. offer a more complex set-up of mixing a VR HMD with an MR stereo camera that offers a much wider FoV, as well as placing multiple markers in the physical world that represent various zones (Figure 2.4), in order to remove the constant need for having the single marker in the user's FoV and that way ensuring reliable tracking and registration of virtual objects [27] [26].

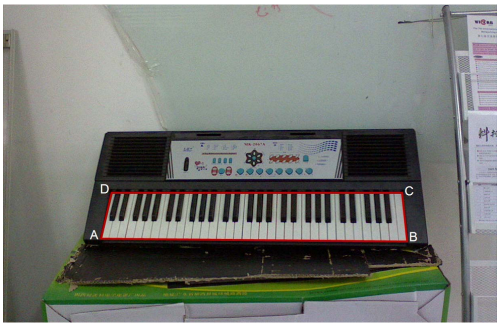

Figure 2.3: Markerless Keyboard Detection [12]

Figure 2.4: Multiple Marker Set-Up [26]

2.3. Differences in Solutions' Foci

As proposed by Hadjakos [11], three main types of instrument pedagogy systems can be identified: *augmented feedback systems*, *demonstration systems* and *exercise generators*, all of which can also deal with some level of motivation strengthening. The piano learning solutions given in papers that are taken into consideration within this seminar can be characterized as being of one or more of these types, depending on their main focus and possible additional features.

2.3.1. Augmented Feedback Systems

Hadjakos describes systems of this type as systems that can gather information, analyze it and provide performance feedback to the piano player. The gathered information is considered to be the player's input, i.e. the played notes by pressing the piano keys. Single captured notes can be viewed as a collective and therefore be compared to each other – chord playing, determining tempo and rhythm of the user's playing, etc. The created context for a group of played notes or just a single note at a time, depending on the analysis model, can be compared to expected notes, with or without additional context, or, ultimately, sheet music. The system can then provide feedback either instantaneously, after a certain note or a group of notes has been played, or at the end of the performance as a cumulative analysis. That feedback should be presented to the user in an augmented form (visual, auditory, olfactory, tactile, ...) so that they are aware of it, i.e. displaying visual information visible within the user's FoV, especially if they are in the middle of a performance. The analyzed systems that provide such features will be further described in section 2.6.

2.3.2. Demonstration Systems

Demonstration systems solely show the player how to accurately play the piano, without the need of knowing how to read sheet music, as stated by Hadjakos. Many of the discussed papers highlight the issue of translating information from the written score directly onto the physical keyboard by locating the proper key, after seeing it for the first time, for novice piano players. This phenomenon is known as sight-reading – the ability to play a musical piece while reading its sheet music at the same, without any prior consultation. Chow et al. [5] and Cai et al. [3] try to address this beginners' issue by offering the feature of a virtual hand model that showcases which notes should be played and how, which will be further described in section 3.2.

2.3.3. Exercise Generators

Defined by Hadjakos, exercise generators are systems that generate exercises for the piano player to practice with. Here, none of the solutions directly generate new exercises without additional human input. A system offered by Cai et al. could be determined as such in a wider sense since another application user is generating exercises specifically for the piano player, in this case, the student's real-life piano teacher [3]. Authors Glickman et al. provide within their solution a complete interface for creating interactive lessons and exercises, but still requiring human interaction to actually craft those levels [7].

2.4. Targeted Audience and Player Count

Since the general focus of the observed solutions is learning how to play the piano, most, if not all, applications share the core demographic of novice piano players, since, according to Chow et al. [5], they are the most probable group to gain benefit from AR-based piano systems. Even though a large percentage of the solutions solely state that they are intended for beginners, those who give out more information start to differ in the age group for which the application is developed. The solution developed by Pan et al. [22] can be used by all age groups, but is mostly focused on the elderly in order to motivate proactive leisure and improve their mental health. Zeng et al. [27] focus on adult novices, due to prevalent challenges in acquiring the initial skills needed to play the piano, such as learning how to read sheet music and mapping it to the physical piano keys, when past the formative years in which they are usually developed.

A few applications name piano students as their targeted demographic and they would use the application during piano lectures. The most notable example stems from a paper by Cai et al. [3], in which the authors discuss the inclusion of the student's teacher into the augmented world controlled by the system. Both the teacher and the student can see the other person's actions through their HMDs in real time, tracking their hands using a dedicated tracker for each player and showcasing them as virtual hand models to the other player, further described in section 3.2. The student can, for example, mimic the teacher's hand movements and fingering methods by overlapping their hands with the teacher's, while the teacher can get instantaneous first-person feedback on the student's performance and therefore intervene much more quickly. The set-up of such a class-oriented system can be seen in Figure 2.5.

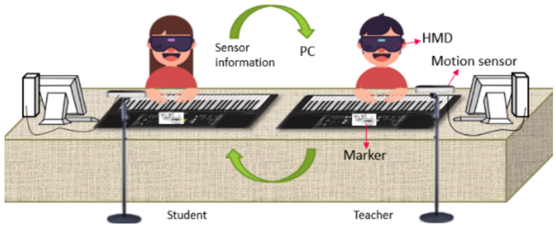

Figure 2.5: Student-Teacher Piano Class Paired AR Playing Set-Up [3]

Learning how to play the piano serves as an ideal problem to be dealt with with AR technology, but features that take advantage of the technology showcased up to now have mostly been appropriate for that and nothing quite more complex. Professional musicians and experienced music students require extreme precision and responsiveness that the currently available hardware and software cannot offer, but the existing limitations do not pose as many problems to beginners who are just getting started with familiarizing themselves with the instrument and therefore executing everything much slower, while not performing overly complicated arrangements, which the AR system should be able to handle and display correctly and on time. What higher-leveled players could take advantage of in terms of AR is displaying metrics and other relevant information during performances or being acquainted with various complex chords and scales during practice, as proposed in [10], but not actual assistance while playing.

Even though the piano is considered to be a single-player instrument, multiple players can play on the same keyboard at once – if so, then generally it is two players that are then divided into the one that plays the main melody, while the other plays the harmonies, or playing the parts that are intended to be played with the left hand of one of the players and the right hand of the other, but no real theoretical constraints are imposed. A couple of papers feature or even entirely focus on paired play, while others provide single-player functionalities.

Zeng et al. divide the players so that each plays the melody intended for one of two hands in order to ease the issues that arise when just starting to play the piano by introducing teamwork [27]. They conclude that, when playing in pairs, the players have a more difficult time trying to sync, but have it easier when playing in a different tempo for each hand, because one player only controls one of their hands when playing, while the single play approach showcases the opposite phenomena. The same principle of division and teamwork is featured in the work of Pan et al., where positive effects of teamwork are apparent, especially when the pair of players know each other well, increasing the playing accuracy even more than when playing solo [22].

2.5. Displaying and Highlighting Notes

Note highlighting refers to presenting augmented information to the piano player which note or a combination of notes they are expected to play next, most often in a visual manner, which will be further discussed, but augmented guidance through other senses, such as auditory, is also possible. That information should not be positioned outside of the player's FoV, which is already possibly limited due to the hardware used for AR content, but preferably placed around their line of sight so that as little time as possible is required from the time the player notices the prompt to their response to it, especially if the player's response to the highlighted next note is expected to be received by the system in a timely manner.

Papers [26] and [10] suggest two different concepts of displaying upcoming notes to the user, but these concepts are recognized in the exact manner or slight variations amongst many other AR piano learning and practising solutions. The two concepts are the *instant* way of note highlighting and the *Beatmania* way, inspired by the concept from a rhythm video game of the same name, both shown in Figure 2.6.

Figure 2.6: Instant (left) and Beatmania-Inspired Note Highlighting (right) Examples [10]

2.5.1. Instant Note Highlighting

The general idea of this concept is to accentuate the to-be-pressed key with some coloured rectangle or border, and after it has been pressed, the augmented highlight is removed and a new one is showcased above the key following the currently pressed one. The currently pressed key can also be highlighted according to the (in)correctness of the performed action in comparison to the expected note, i.e. the key that corresponds to that note. By using such a concept in an AR piano learning solution, users get a non-obtrusive and simple UI that navigates them through a piece of music that they want to play out from note to note (or groups of notes at the same time) – the inclusion of more than one highlight for more following notes, other than the one following the currently played note immediately, can lead to the player becoming extremely confused by not knowing which of the highlighted notes to play next and having to remember which colour represents which state of the key (*to be pressed next*, *to be pressed after the next key is pressed*, and so on), therefore contrasting the primary characteristic of this concept – simplicity. Trying to achieve that same characteristic by enforcing note-to-note playing by highlighting the entire key on top removes necessary information when playing the piano such as the length of the note/key press and the context in which the note is in (knowing the notes that follow it and when should they be played in relation to the currently played note).

This concept is not useful for novice players since they do not possess the experience to know how to predict which note comes next or for demonstration purposes [8]. Some solutions that include this concept: [1], [22], [3] (Figure 2.7), [17] and [8]. Authors Trujano et al. use this concept, in addition to the Beatlemania format, for displaying chord suggestion [24] (Figure 2.8). [26] and [27] both introduce new elements – arrows – that indicate the direction where the player should look in order to locate the next key that is to be played, in case it exceeds the player's current FoV, as seen in Figure 2.9. A solution proposed by Moslavac signals the next key that is to be played by using vertical flags (Figure 2.10), instead of relying on a 2D rectangle [19]. Glickman et al. offer two types of note highlighting within their AR piano learning system [7]. Alongside a variation of the Beatmania concept, the so-called *improvisation overlay indicators* are introduced – they highlight note range that should sound harmonious when played along with chords in a given progression of notes. Unlike previous systems, Rigby et al. do not utilise key highlighting only until after finishing playing and highlight the areas that were shown to be difficult for the player during the performance in the same manner [23].

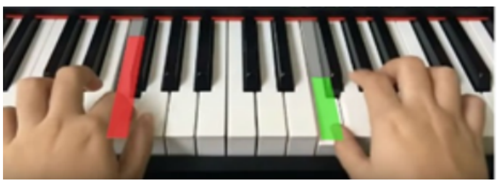

Figure 2.7: Highlighting Correctly and Incorrectly Pressed Keys [3]

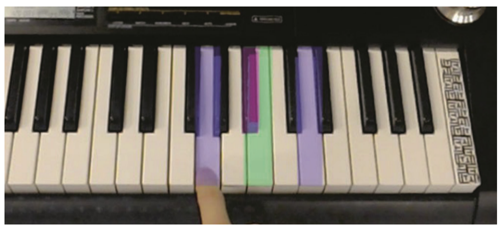

Figure 2.8: Chord Suggestion Supported Through Instant Note Highlighting [24]

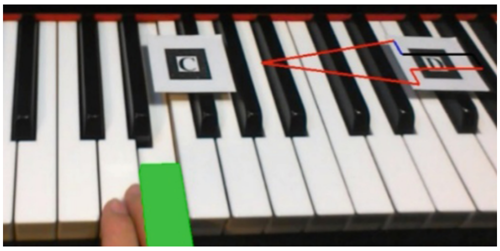

Figure 2.9: Arrow Pointing in the Direction of the Next Note's Location [26]

Figure 2.10: Flags on Top of the Keys Signaling the Next Note [19]

2.5.2. Beatmania-Inspired Note Highlighting

Also referenced in other papers, such as Molloy et al. [18], as the *piano roll* concept, this concept showcases timed upcoming notes as rectangles that approach the top of the corresponding keys directly from the top or in front of the physical keyboard, showcased in Figure 2.11. When the bottom of the rectangle reaches the top of the key, the key should be pressed to play the note and the moment the top of the rectangle reaches the key, the key should be let go. Additional auxiliary lines coming from every key on the keyboard can be added for a better understanding of which key is being approached by a certain rectangle, especially when the note to be played is at a far distance so that the player cannot instantaneously connect the note signal to the key. This concept provides answers to the issues mentioned for the instant note highlighting, but can turn out to be chaotic when a lot of notes have to be played out in a short period of time – a lot of incoming notes that all approach the keyboard at a similar or the same time. Getting an overview of multiple notes ahead can give additional context to the player and give them time to prepare to locate and play the incoming notes. Additional modification of the concept can be made by changing the color of the rectangle as it approaches the physical key, indicating how soon the player has to play the note [5].

Some solutions that include this concept: [6], [5] (Figure 2.13), [10] and [24] (as a part of the song tutorial feature, Figure 2.12). Instead of augmenting the incoming notes directly on top/in front of the real piano on which the player is playing, Birhanu and Rank [2] propose placing a small virtual keyboard in the augmented space above the piano with the same Beatmania concept, as seen in Figure 2.14 – this way, the area around the physical piano is free from any distractions, but the player now must shift their focus from the physical piano to the augmented one, which is not in the player's line of sight, every time they want to see which notes they should be playing next.

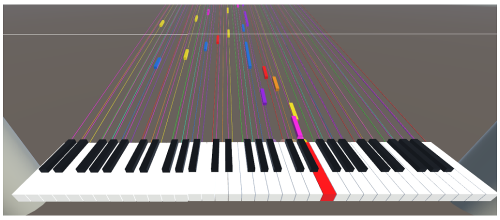

Figure 2.11: Piano Roll Example [18]

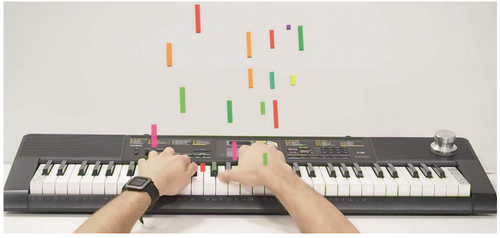

Figure 2.12: Piano Roll Example [24]

Figure 2.13: Piano Roll Example [5]

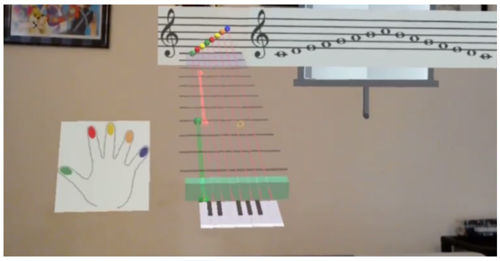

Figure 2.14: Virtual Keyboard With Piano Roll Elements [2]

2.6. Performance Feedback and Accuracy Measurement

The analyzed solutions that are augmented feedback systems, as described in subsection 2.3.1, offer some kind of feedback according to the player's performance in the form of augmented information that is displayed on top of the physical world. Certain feedback can be given instantaneously after pressing a key, depending on whether the correct or incorrect one was pressed and how the performed action ties to previously performed actions, while a complete analysis of the performance can be calculated and presented to the player with additional information that needs to take into account the context of the entire performance.

One type of immediate feedback about pressing the (in)correct key was shortly described in section 2.5, but variations on the used rectangle are present – for example, Molloy et al. use particles of different colors to convey the same information [18]. The same solution also offers combined note and timing accuracy through textual feedback in a gamified manner, explained in more detail in section 3.1, while Molero et al. use another gamification element of a combo multiplier that increases for every correctly played note, and resets after playing an incorrect one, with the information about the current state of the modifier changing after every pressed key [17].

As described in [22] and [27], the simplest accuracy measurement for the entire performance is the ratio of the number of correctly played notes/keys pressed to the total number of notes in the observed performance, with any additional note factors and metrics (such as note duration) taken into consideration. Fernandez et al. [6] combine certain notes that fall into the same sampled time interval and check whether the necessary list of expected notes for a time interval of the song aligns with the user input sample at the same time interval – all notes must match, in order, for the time interval sample to be considered as correct and as such is used in the ratio calculation (now the divisor is the total number of time intervals, instead of notes).

Molloy et al. choose the other route by splitting notes into chunks (by splitting beat intervals), instead of combining notes into larger groups, and calculating the accuracy score for each of those intervals, allowing the player to still get some recognition for pressing the next key, but not on time (either missing the start or releasing too early), instead of completely discarding the partially played note [18]. The authors argue that counting only correctly played notes could lead to cheating since pressing all of the keys at once would also cover the needed note and therefore yield complete accuracy. The same solution also displays and additionally stores the entire piano roll (note highlighting concept) of the performance with the color-coded notes according to the player's actions during the performance so that the player can revisit and analyze the performance, as seen in Figure 2.15. An expanded formula for measuring accuracy is used by Cai et al. [3], taking into account the additional values of associated intensity and duration of the key press, alongside the previously mentioned (in)correctness of the whole note – all extracted from the received MIDI inputs from the used digital piano. Chow et al. present the user with the simplest performance accuracy measurement, but also include information about the average delay, note layout of the entire performance, and the overall grade [5].

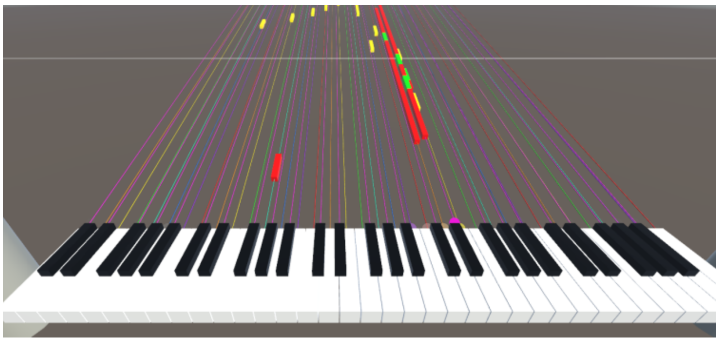

Figure 2.15: Note Feedback After the End of a Session [18]

2.7. Application Modes and Available Modifications

Even though a large number of prototype applications focus on improving and analyzing only one aspect of piano learning through AR in order to get the most isolated results, some applications provide more than one mode, getting one step closer to creating a well-rounded full-fledged piano learning system. Yet, the majority of considered applications in this seminar provide some kind of the ability to make small tweaks to the implemented features in order to personalize the experience while using the application.

From a technical standpoint, solution [10] contains a calibration mode in order to accommodate the need for explicit maker positioning, so that the tracing and registration perform as best as possible. The feature has been added due to the possibility of using a non-standardised set of keys and adjusting the system to the available keyboard.

The solution developed by Moslavac [19] adds the ability to play on a completely virtual augmented piano, without the need of a physical one – instead, only a single image marker is needed and upon recognizing it, the system is ready with a now existing augmented piano in the space, as seen in Figure 2.16. This and some other applications offer the user the ability to toggle notation symbols on top of the keys (Figure 2.16 and Figure 2.17), while also giving the option to toggle the virtual piano sounds on and off, similar to the solution [17].

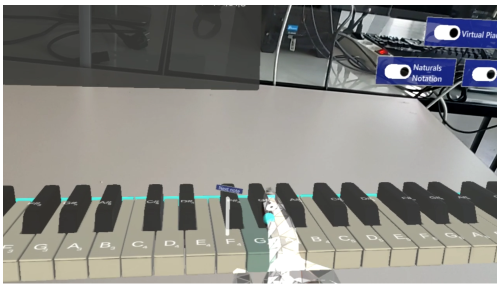

Figure 2.16: Virtual Piano [19]

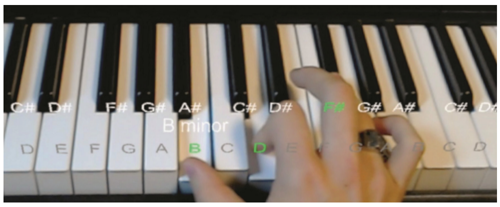

Figure 2.17: Notation Symbols [24]

The implementation by [27] has a dedicated practice mode in which the user can get familiar with the application itself, while [22] use their practice mode as a way to get accustomed to playing the chosen trial song. Glickman et al. [7] also included a similar practice mode in their application, but it additionally presents the user with the choice of choosing different genres covered by the application, such as jazz, and then practicing playing in conjunction with virtual animated musicians in order to strengthen their accompaniment and improvisation skills by synchronising with the musicians playing with the user in real time. Zeng et al. use their practice mode for testing out the results of the classifier model that detects measures that are considered difficult and promptly slows them down while playing in the selected mode [26].

Glickman et al. [7] and Molero et al. [17] provide a lessons mode where the user can learn various concepts and techniques, divided into levels that differ in difficulty and the objectives that are contained within the lesson. Glickman et al. go even further by providing a full-blown user-ready lesson builder so that custom lessons can be built.

Most of the mentioned applications that focus on learning new music pieces and have integrated MIDI elements support the addition of user-generated MIDI files, while those that do not have any MIDI integration most often offer a selection of songs to choose from when in some kind of play mode. The play mode in a handful of applications, such as [27] and [26] also allows players to modify the tempo and the hand with which the user will perform the selected piece. Additionally, both applications have a learning mode for paired play in which every player has their own dedicated hand whose melodies/harmonies they will be playing during the performance.

Hackl and Anthes [10] offer a playback mode, in which the user can see the superimposed virtual keys on top of the physical ones that signal which keys need to be played, with the ability to freely rewind or fast forward to some other part of the selected music piece. Chow et al. propose the addition of a note learning mode in which the application would pause and wait for the user's next note input so that the player can familiarize themselves with the various notes available on the piano [5]. The same system and Molloy et al.'s system [18] include their own game modes, which will be explained in more detail in section 3.1 – Molloy et al. present the game mode as an optional minigame that can be turned on or off in other modes, as well as providing a history mode for accessing all previous playing sessions within the application.

3. AR Piano System Notable Features - Ideas and Executions

While chapter 2 focused on aspects shared by the majority of the solutions presented in the papers that were taken into consideration for this seminar, the following chapter showcases specific features of some of those solutions that seem to be worth discussing further.

3.1. Gamification

Various authors have offered the idea that gamifying the piano playing experience could increase the enjoyment during the playing sessions and therefore increase the motivation to continue coming back to the application and practicing playing the piano. For instance, Trujano et al. [24] suggests that the inclusion of games that require pressing a specific set of keys, which can represent chords, in order to proceed to the next level can influence the user's chord memory and they could be learning all while playing a snappy and entertaining game – especially appropriate for children who struggle with absorbing pure technical knowledge without any special incentives and this way they are exposed to new useful information, without being aware of it directly, in a fun environment. As discussed by Hackl and Anthes, an additional mode for matching note names with the proper keys could also be implemented and result in the aforementioned indirect knowledge acquirement[10].

Molloy et al. include an optional game mode which transforms the regular note matching and song playing scenario into a game similar to the popular game of Space Invaders [18], showcased in Figure 3.1. The goal of this inclusion is to provide additional motivation for the user during the regular song learning process by amplifying it, but keeping it toggleable as an option so that it does not turn out to be distracting if the user ends up in the state of being overwhelmed by all the additional information that needs to be taken in while playing the game. The game turns the used Beatmania note highlighting concept into a space for shooting spacecraft with a gun that can be activated after being charged up whenever the user performs with great accuracy in the given beat interval. The instant feedback is delivered through flashy particle effect, as well as in a textual manner, mimicking the rhythm video games' element of displaying messages such as "Decent", "Impressive", "Godlike" or "Perfection" for every beat interval within a piano learning environment.

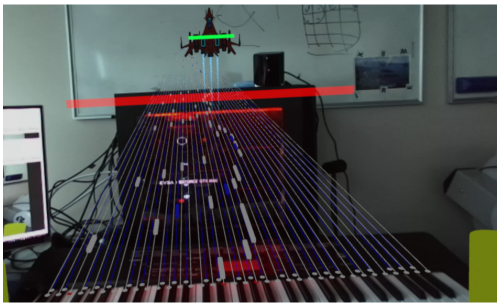

Figure 3.1: Space Invaders-Themed Minigame while Playing a Music Piece [18]

Authors Molero et al. introduce a game mechanism to their system within the score feedback [17]. Firstly, points received when playing the correct note are additionally affected by the combo multiplier – every consecutive correct key press is combined into the current streak and it results in the increase of the value of the multiplier that then increases received points for correctly playing the upcoming notes (Figure 3.2). If the user loses the streak by playing the wrong note, then the multiplier resets and the player should start building the streak again to get even more points with every key press. Secondly, after finishing the performance, the user is given the information about the performance in the form of stars, instead of a numeric score, which indicate how well the player handled the performance (Figure 3.3). In order to proceed to the next song to perform, the player needs to gather a minimum number of stars, implying the minimum threshold of player accuracy deemed to be sufficient when grasping the learning objectives for the selected song that is performed.

Figure 3.2: Streak Multiplier Example [17]

Figure 3.3: Star Grading Example [17]

The reason Chow et al. [5] introduced game elements to their solution is the end goal of removing the possible feeling of needing to fulfill chores when required to practice playing the piano and for the users to keep willingly coming back to brush up on their piano skills by using the application. Their system provides an element of comparing previously achieved scores and creates a competitive environment that strengthens the players' motivation to outdo either themselves or other players using the application. They also suggest the introduction of time limits, obstacles, or random events while trying to achieve a high score. Some other game elements they consider are the increased social interaction through paired or even group play, as well as ranking the available music pieces by difficulty and locking specific groups of songs before completing some prerequisite grouping of songs and achieving the minimum level of accuracy for all of them.

3.2. Virtual Hand - Fingering Assistance

Fingering implies following a set of recommendation guidelines for the best playing experience and accuracy. To novice players, these guidelines can be hard to remember at once, especially if lacking any visual information to refer to.

The nature of AR poses a unique solution of augmenting a virtual hand that plays the music piece in accordance with those fingering guidelines in front of the players, who can then observe it or follow it, matching their own hands to the virtual ones and therefore mimicking the virtual gestures and finger placements, like the implementations of Huang et al. [12], Glickman et al. [7] and Guo et al. [8] (Figure 3.4). In the use case within the classroom, such as the one given by Cai et al. [3] (Figure 3.5), the augmented hand can be the captured teachers hand and the student can then mimic the teachers' showcased playing in real time, while the teacher can give feedback instantaneously, having a first-person view of the student's playing technique.

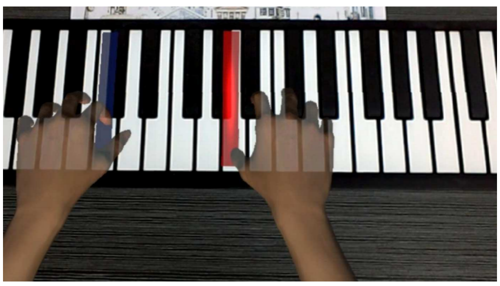

Figure 3.4: Virtual Hand Model Example [8]

Figure 3.5: Virtual Hand Model Example [3]

3.3. Inclusion of Artificial Intelligence Elements

The major everyday advancements in the area of artificial intelligence (AI) result in the increased power of what its elements could be applied to, including music education and therefore piano playing. Machine learning concepts can be introduced to create models that could further enhance the players' experience. Molero et al. [17] suggest extracting behavior patterns from the users within the application in order to determine the difficulties most players are faced with, on a single-user level or the level of the entire application userbase. After classifying the patterns in the observed data, the application could suggest tutorials or hints when such difficulties are observed in the following practices.

Guo et al. [8] developed a system that supports the automatic generation of hand motion animations by using two machine learning concepts for the two stages of generation. As the first concept, a Hidden Markov Model (HMM) is used to translate the given written score, in this case, a raw MIDI file (where one note is represented by one finger tag) to fingering information that is most likely to be adopted by a real pianist – finding the probabilistic model of a piano performance of the given music piece – the model is trained until optimum correspondence between the targeted keys and fingers using an HMM is found. The HMM can be applied to chords as well, i.e. produce multi-finger motion, with the minimal assumption that the fingering state at every note depends only on the previous one, creating a chain of transitions. The second concept utilized is the Viterbi algorithm that is used to generate motions that coordinate with the acquired fingering information from the previously mentioned HMM – the obtained prediction results are combined with existing musical knowledge and as a result, a generation mechanism for coordinating the player's finger motion is created.

In order to improve the short-term learning performance of the beginner piano players, Zeng et al. [26] propose a machine learning model that can determine which measures prove to be more difficult by using a binary k-Nearest Neighbor (kNN) classifier to classify measures within a certain piece of music as easy or difficult. In short, the initial set of information that is required for training the model can have measures classified as difficult if the number of people, who took part in the data collection element, made more mistakes than the predefined threshold of mistakes for deeming the measure as an easy one. After training the model, when practising a new piece of music, the measures the model then classifies as difficult will be slowed down, so that the player can take more time while tackling the more difficult measures.

CONCLUSION

As traditional piano teaching techniques are being challenged more and more, modern approaches to piano pedagogy appear in order to try to revamp the long-established methods – one of such approaches is the incorporation of AR technology into the process of learning how to play the piano. The aforementioned need for change, as well as the rise in popularity of the AR field and the features it offers, resulted in several papers that deal with enhancing the old and adding new piano playing features that were impossible before, which are now achievable thanks to the augmentation of the physical world with virtual objects. A selection of papers themed around this idea was chosen in order to depict the similarities between currently available solutions, as well as highlight the differences that could lead to new areas of exploration and advancement for this specific topic.

While the baseline of most solutions is the same – displaying whether the note played was the correct or incorrect one while playing a selected music piece – a large amount of them chose to additionally focus on highlighting and tackling a specific issue, like the lack of motivation in novice players or improving notation literacy. The most popular system set-up consists of an HMD and an application developed in Unity, with an additional marker-based SDK used for tracking, like Vuforia, while markerless set-ups are gaining more traction. The experience is most often tailored for a single-player experience, but some ideas regarding paired play are explored. Two equally-represented distinct approaches to notifying the user about the incoming notes of the music piece they are playing are recognized – instant and Beatmania-inspired – with a plethora of variations adjusted to the authors' liking. The feedback system of these piano playing solutions is mostly based on showcasing the amount of (in)correctly played notes, while a few experiment with some more comprehensive information processing and displaying the gathered metrics to the user. Alongside countless modification and feature addition possibilities, the most prevalent in the observed papers were the idea of gamifying the entire playing process, incorporating virtual hand models as guides, and the inclusion of AI elements into the system in order to increase feedback accuracy.

References

[1] István Barakonyi Dieter Schmalstieg. Augmented reality agents in the development pipeline of computer entertainment. U *ICEC*, 2005.

[2] Amare Birhanu Stefan Rank. Keynvision: Exploring piano pedagogy in mixed reality. *Extended Abstracts Publication of the Annual Symposium on Computer-Human Interaction in Play*, 2017.

[3] Minya Cai, Muhammad Alfian Amrizal, Toru Abe, Takuo Suganuma. Design and implementation of ar-supported system for piano learning. *2019 IEEE 8th Global Conference on Consumer Electronics (GCCE)*, pages 49–50, 2019.

[4] Ozan Cakmakci, François Bérard, Joëlle Coutaz. An augmented reality based learning assistant for electric bass guitar. U *Proc. of the 10th International Conference on Human-Computer Interaction, Crete, Greece*. Citeseer, 2003.

[5] Jonathan Chow, Haoyang Feng, Robert Amor, Burkhard Claus Wünsche. Music education using augmented reality with a head mounted display. U *AUIC*, 2013.

[6] Carlos A. Torres Fernandez, Pujana Paliyawan, Chun Yin Chu, Ruck Thawonmas. Piano learning application with feedback provided by an ar virtual character. *2016 IEEE 5th Global Conference on Consumer Electronics*, pages 1–2, 2016.

[7] Seth Glickman, Byunghwan Lee, Fu Yen Hsiao, Shantanu Das. Music everywhere - augmented reality piano improvisation learning system. U *NIME*, 2017.

[8] Ruoxi Guo, Jiahao Cui, Wanru Zhao, Shuai Li. Ai and ar based interface for piano training. *2020 International Conference on Virtual Reality and Visualization (ICVRV)*, pages 328–330, 2020.

[9] Yujing Guo. Challenge of multimedia-assisted piano teaching to traditional piano teaching mode based on internet and information technology. U *2021 4th Interna-*

tional Conference on Information Systems and Computer Aided Education, pages 973–977, 2021.

[10] Dominik Hackl Christoph Anthes. Holokeys - an augmented reality application for learning the piano. U *Forum Media Technology*, 2017.

[11] Aristotelis Hadjakos. Sensor-based feedback for piano pedagogy. 2011.

[12] Feng Huang, Yu Zhou, Yao Yu, Ziqiang Wang, Sidan Du. Piano ar: A markerless augmented reality based piano teaching system. *2011 Third International Conference on Intelligent Human-Machine Systems and Cybernetics*, 2:47–52, 2011.

[13] David Johnson, Daniela Damian, George Tzanetakis. Evaluating the effectiveness of mixed reality music instrument learning with the theremin. *Virtual Reality*, 24(2):303–317, 2020.

[14] Joseph R Keebler, Travis J Wiltshire, Dustin C Smith, Stephen M Fiore, Jeffrey S Bedwell. Shifting the paradigm of music instruction: implications of embodiment stemming from an augmented reality guitar learning system. *Frontiers in psychology*, 5:471, 2014.

[15] Fotis Liarokapis. Augmented reality scenarios for guitar learning. U *TPCG*, pages 163–170, 2005.

[16] Jorge Martin-Gutierrez, Marta Sylvia Del Rio Guerra, Vicente Lopez-Chao, René Hale Soto Gastelum, Jose Fernando Valenzuela Bojórquez. Augmented reality to facilitate learning of the acoustic guitar. *Applied Sciences*, 10(7):2425, 2020.

[17] Diana Molero, Santiago Schez-Sobrino, David Vallejo-Fernandez, Carlos González-Morcillo, Javier Albusac. A novel approach to learning music and piano based on mixed reality and gamification. *Multimedia Tools and Applications*, 80:165–186, 2021.

[18] W. Thomas Molloy, Edward Huang, Burkhard Claus Wünsche. Mixed reality piano tutor: A gamified piano practice environment. *2019 International Conference on Electronics, Information, and Communication (ICEIC)*, pages 1–7, 2019.

[19] Mirta Moslavac. Development of an augmented reality-based prototype application for learning how to play the piano using the hololens 2 platform. Faculty of Electrical Engineering and Computing. 2021.

[20] Yoichi Motokawa Hideo Saito. Support system for guitar playing using augmented reality display. U *2006 IEEE/ACM International Symposium on Mixed and Augmented Reality*, pages 243–244. IEEE, 2006.

[21] David Nield. *Augmented Reality, Retail, and IKEA*. New Atlas, 2017. URL `https://newatlas.com/zed-mini-ar-camera/52464/`, last accessed: 18.01.2022.

[22] Honghu Pan, Xingxi He, Hong Zeng, Jia Zhou, Sai Tang. Pilot study of piano learning with ar smart glasses considering both single and paired play. U *HCI*, 2018.

[23] Liam Rigby, Burkhard Claus Wünsche, Alex Shaw. piarno - an augmented reality piano tutor. *32nd Australian Conference on Human-Computer Interaction*, 2020.

[24] Fernando Trujano, Mina Khan, Pattie Maes. Arpiano efficient music learning using augmented reality. U *ICITL*, 2018.

[25] Luca Turchet, Rob Hamilton, Anil Camci. Music in extended realities. *IEEE Access*, 9:15810–15832, 2021.

[26] Hong Zeng, Xingxi He, Honghu Pan. A new practice method based on knn model to improve user experience for an ar piano learning system. U *HCI*, 2019.

[27] Hong Zeng, Xingxi He, Honghu Pan. Funpianoar: A novel ar application for piano learning considering paired play based on multi-marker tracking. *Journal of Physics: Conference Series*, 2019.

List of Figures

Abbreviations

AI	*Artificial Intelligence*
AR	*Augmented Reality*
AV	*Augmented Virtuality*
CV	*Computer Vision*
FoV	*Field of View*
HMD	*Head-Mounted Display*
HMM	*Hidden Markov Model*
kNN	*k-Nearest Neighbor*
MIDI	*Musical Instrument Digital Interface*
NFT	*Natural Feature Tracking*
MR	*Mixed Reality*
VR	*Virtual Reality*
XR	*Extended Reality*
SDK	*Software Development Kit*