Listen by Looking: a framework to support the development of serious games for live music

Giulio Pitteri, Edoardo Micheloni, Carlo Fantozzi, Nicola Orio

University of Padova,
Padova, Italy

Abstract

This paper presents an ongoing project focused on the framework *Listen by Looking*, that includes several tools to support the implementation of digital games, allowing the interactive display of music sheets, piano rolls and dialog windows with contextual information. Using this framework, a serious game for mobile devices was designed to entertain and teach music to users by listening to songs and answering questions about the melodies and musical structures learned during the game. The usefulness of game was assessed in a controlled setup, by comparing the answers of the app users with those of people who listened the the same music without the app. Furthermore, an external synchronization tool through a data-over-sound channel has been tested in order to simultaneously manage a wide network of devices. The results are encouraging for future developments of the framework, with the possibility of using it during a public performance and with different games.

Keywords: Gamification, Graphics Framework, Musical Notation, Score Following, Content Delivery, Data over Sound

Email addresses: pitteri.giulio@studenti.unipd.it (Giulio Pitteri), micheloni@dei.unipd.it (Edoardo Micheloni), carlo.fantozzi@unipd.it (Carlo Fantozzi), nicola.orio@unipd.it (Nicola Orio)

1. Introduction

Music enjoyment depends on a number of individual factors. Even while attending the same concert, people in the audience can experience different levels of involvement depending on their previous knowledge of the pieces, on their ability to pay attention to the dimensions of the music language – melody, harmony, rhythm, orchestration, timbre – and in general on their background in music education. The same experience can thus be absorbing for some, boring for the ones who do not have the cultural tools to decode the message, and even anxiety-provoking for those who do not feel competent enough.

Technology offers a variety of ways to bridge the gap between users and content. Studies on human-computer interaction have always striven to simplify and make communication with analogue and digital contents more intuitive through the design and evaluation of new interfaces. Natural User Interaction [1], Tangible Interfaces [2, 3] and Physically-based Interaction [4, 5] offer many examples in this sense. Among all the products for end users, probably video games provide their players with the higher variety of opportunities for interactive engagement. The user interface involves specific hardware control devices, in order to achieve a goal or perform a task in a more or less realistic virtual scenario. Over the past few decades, video game controllers have undergone continuous changes, starting from classical input devices such as the joystick, the mouse and the keyboard up to more contemporary and unusual trends. In these cases, the user interface is the means that should ease the access or allow to carry out the task.

But what if the interface were just the means to learn and enrich the experience of the user in a real scenario? In this case, the objective would not be to design an intuitive and easy-to-use interface, but to create content and an interaction system tailored to a context of learning and augmented reality. In this paper, a framework is presented and discussed that allows the implementation of serious game interfaces that enrich the experience of music lovers while listening to music. Furthermore, it increases their engagement by implementing a question-and-answer game about music content that can be played while listening to the music. The goal is to provide the user with an environment that is visually engaging and entertaining.

This framework is implemented in *Listen by Looking*, a mobile application through which the user can learn and interact with pieces of music i) by playing with music notation, ii) piano roll and iii) answering to questions that address the characteristic of the music content using contextual information. An innovative feature of the framework consists in the adoption of the app for live performances, using a *score following* algorithm to recognize the different sections of the music and a *data over sound* system to communicate the online alignment to mobile devices. *Listen by Looking* is the first game implemented as an app that focuses on learning about structures, rhythm, notations and melodies of tonal Western music [6]. In this paper we describe the complete architecture of *Listen by Looking* with a special focus on the technological solutions that have been recently introduced for its application to live music performances. The possibility to visualize in real time a representation of the score during a live concert, be engaged by the serious game and – at will – come back to the visualization without loosing synchronization, is an important contribution of our work. The effectiveness of the approach has been assessed with a number of subjects, with encouraging results.

The paper is structured as follows. Section 2 introduces several concepts and references about serious games: after a general introduction, Subsection 2.1 delves into the theme of music serious games, then Subsection 2.2 elaborates on how to measure the effectiveness of serious games, and, lastly, Subsection 2.3 adds some remarks on music-based interaction in commercial software. Section 3 is devoted to Listen by Looking. Listen by Looking was born as a mobile application in 2014, after some conversations with pianist Giovanni Umberto Battel on how to take advantage of the then-innovative technologies in smartphones to improve the experience with live concerts. Over the years, Listen by Looking has become a laboratory for different multi-disciplinary research questions, including the problem of synchronizing a score with live music and the problem of broadcasting data to devices in a concert room without specialized hardware. The structure of Section 3 is reminiscent of this journey: Subsection 3.1 introduces the base structure of the app, Subsection 3.2 summarizes the work on score following, and Subsection 3.3 presents our previously unpublished efforts on data transmission over Wi-Fi and ultrasonic audio frequencies. Section 4 is devoted to the recent shift from Listen by Looking as an app to Listen by Looking as a framework to develop educational activities. A specific serious game we implemented is described in Subsection 4.1. The results of the assessment of the game are presented in Section 5. Finally, conclusions are given in Section 6.

2. Background

The term *serious game* has been applied to define a number of different approaches, from adding a purpose to a traditional gameplay – with challenge, levelling up and fun as essential parts – to creating an environment that provides experiences and emotions with minimal gaming characteristics. The work [7] proposes to consider the whole gamut of games as serious games along a continuum, depending on the relative importance of traditional gaming elements. One common feature, as stated in [8], is that serious games address the challenge of combining the actions of learning and playing by: generating intrinsic motivation for the player to engage in the game; providing a responsive game environment that gives the player immediate feedback; delivering complex content that allows for ample learning opportunities.

Serious games are applied to many contexts such as teaching (e.g., learning a foreign language), training (e.g., becoming a skilled operator) or raising social awareness (e.g., understanding different points of view). All these activities require an internal motivation that is often paired with an external commitment, and in principle do not need to be *fun* to play. Thus, a possible definition of serious games is: *"games that are designed for a primary purpose other than pure entertainment"* [9]. Yet, it may be argued that serious games do not adhere to the common definitions of "game", for instance "the voluntary attempt to overcome unnecessary obstacles" [10] or "a free activity standing quite consciously outside ordinary life as being not serious, but at the same time absorbing the player intensely and utterly" [11]. As a matter of fact, users can be bound to play a serious game because it is part of their mandatory training. On the contrary, some authors argue that, even in a serious game, entertainment is more important than pedagogy, which is a secondary goal [12].

We believe that the presence of a *voluntary effort* is of prime importance when the game is aimed at bridging the gap between users and cultural content. In many contexts, such as museums and expositions, concerts halls and theatres, people may feel incompetent and afraid of the cognitive effort needed to fully enjoy the cultural experience, but they can be still motivated to learn and understand. Serious games might be a tool to provide additional motivation, but they need to be entertaining because the public is obviously free to choose whether to participate or not.

2.1. Music Serious Games

Even if music is considered an entertaining activity, learning music theory or listening to complex pieces can become boring and dull. To this end, a number of researchers have focused on serious games about music, addressing different aspects – defined as *layers* in the overview on music serious games [13] – of this art form.

A number of games have addressed music education for children. For instance, the research work presented in [14] targets preschoolers, hence adopts a simple game mechanics that relies on sounds of cartoon animals. Another example is the work presented in [15], aimed at teaching the position of the notes on the score to music students in primary school; in this case the fun element is the presence of a camera-based gesture interface and the game mechanics is based on collecting points through correct answers. Music reading, in particular the ability of following a score while listening to music, has been addressed in [13] using complex polyphonic scores. Even in this case the gaming element is reduced to collecting points proportional to the number of correct alignments. More abstract concepts related to body movements and music, such as fluidity or energy, are the subject of the two-player game described in [16] where a player earns points when her movements match the characteristics of the audio content.

Social aspects of music have been taken into account in the development of serious games. For instance, in [17] two players cooperate by creating a rhythm using two MIDI pads and they score points depending both on the quality of individual drumming and on the level of interplay. Interaction and movement are prevalent over game elements in [18]: in this installation, the four channels of a mixer are controlled by the body movements of four participants inside a room, hence the goal is straightforward and there is little game structure. A more complex interactive environment is described in [19], where the position of a single user inside given areas in a large room communicates her intentions in a game to teach simple harmony. The same environment is applied also to explore the different layers of a complex composition.

Music has also been used for healthcare applications. An interesting experiment is presented in [20], where music is used as the informative cue in a simplified *Memory* game (i.e. the game where players have to match identical cards) aimed at the early diagnosis of dyslexia. Personalized music has been proposed in [21] as a way to give positive feedback on tasks developed to help people with dementia. Although music is not part of the game mechanics itself, its potential for reminiscence has been judged particularly relevant for these patients.

2.2. The Effectiveness of Serious Games

An important issue in the development of serious games is measuring their effectiveness. As stated in [22], which carried out a meta-analysis on more than 100 papers, the most common approach is based on the use of questionnaires assessing usability and learning outcomes. Although a comprehensive study [23] on hundreds of digital and board games showed empirical evidence of the positive effect of playing both in terms of increased engagement and improved learning, this effect cannot be taken for granted. In fact, a report on 40 selected studies on learning effectiveness [24] showed that serious games are not always superior to other types of learning material: seven out of 40 did not lead to improvements, and two additional games even exhibited a negative effect. Since the evaluation has been carried out by the developers of the game in more than half of the cases, it may be that actual results are even less in favour of serious games.

In the large majority of serious games described in scientific studies, users can be grouped in cohorts that share similar characteristics (i.e., age, interests, background knowledge). This can not be the case for cultural heritage applications, where the public is much more heterogeneous. Yet, the importance of evaluation for cultural heritage applications has been stressed in [25] where a framework is proposed for designing, developing and evaluating serious games with cultural content. The effectiveness of a serious game with cultural content for mobile devices has been studied in [26], where factors related to age emerged as significant in the ability to acquire factual knowledge. In the evaluation presented in our paper, we limited the age range of participants and we normalized the variable related to background knowledge.

2.3. Music-based Interaction in Commercial Products

Commercial software for mobile devices aimed at music teaching is already available. Some gaming elements are present in many of them, even if, according to [27], "there remains significant scope for improvement in their learning management" because most of the games exploit simple mechanics and provide little feedback to the player. Nowadays, there is a growing trend to bring more physical movement and social interaction into commercial games while the benefits of computing and graphical systems are spreading in the gaming industry [28]. Well-known examples are the Xbox Kinect and the Nintendo Wii. Alongside this tendency, a push is witnessed towards new ways of interfacing the user with musical games. Successful examples from the past decade are *Guitar Hero*, *Singstar* and *Rockband*, all of whom proposed controllers pretty similar to real instruments. Other projects focused on representing some real aspects of the music, like the shape of the instruments and the music sheet. Two examples are *Magic Piano* by Smule and *Piano Notes!* by Visions Encoded.

Typically, music and sound have been used by interface designers to convey information to users. The principles behind this choice were analyzed by William W. Gaver [29] and can be summarized under the concept of *auditory icons*. The opposite process, i.e. controlling computers via musical messages, has been somewhat neglected instead. Some valuable motivations to continue accurate research in this field are described by James L. Alty [30]: music is pervasive in daily life, is memorable and durable, contains a large quantity of highly structured information, and involves the simultaneous transmission of a set of complex ideas related over time within an established semantic framework. Video games that can convey this complex set of ideas must work at both a high and low level of abstraction. For example, some projects have put the "real instrument" as the core interface to interact with the game: *Yousician*, *Flowkey* in cooperation with Yamaha, or closed projects such as *Guitarbots*, *WildChords*, *PlayMusa*[31]. All these software products are designed to teach playing instruments using graphics and a gamification approach [32, 33]: exercises to learn playing an instrument are integrated in basic principles of the games (challenges and scores).

3. The Listen by Looking App

The *Listen by Looking* app was designed to make live concerts of classical music more accessible to listeners who lack basic knowledge of the rules in this form of art. Indeed, as for any art form, some background is required to fully appreciate music. In order to address a heterogeneous audience, Listen by looking presents music in a graphic form which aims to be more easily comprehensible; the graphic form moves dynamically in sync with live music and is accompanied by useful information for understanding piece context as i) the scoring period, ii) the music sections and iii) the themes presented. Listen by Looking supports two graphic forms: the actual staves with notes ("Score"), and an immediate representation of duration and height of the notes with absolute reference ('Piano Roll"). We refer to the second form as Piano Roll because it recalls the rolls of perforated paper that controlled the mechanics of late nineteenth-century automated pianos. This representation was introduced in the app for users who can not understand sheet music. The app can be used in "Live" mode while actually listening to music in a concert hall, or in "Archive" mode using recordings stored inside the mobile device. The information needed to access a piece can be easily downloaded from within the app. The package associated with a piece contains in a single file:

- the score;
- the piano roll;
- the recording of the piece in MP3 format and a pre-computed synchronization table for Archive mode;
- the information necessary for live synchronization in Live mode;
- additional multimedia information, such as texts and images on the structure of the music piece or its composer.

Once the package is downloaded and stored locally, Listen by Looking can work without an Internet connection.

3.1. Structure of the App

The structure of Listen by Looking includes three blocks, highlighted in Figure 1, that correspond to the main screens shown to the user. Each of the screens allow the user to open a fourth, Preferences screen to change the application parameters.

The first block (Play mode selection screen) corresponds to the opening screen whom the user sees when starting the app. The screen is shown in Figure 2. The user is presented with two choice buttons for Archive and Live

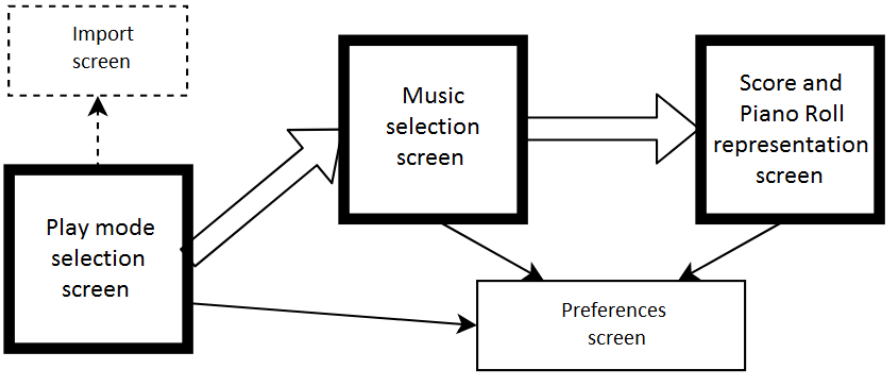

Figure 1: Generic flow structure of the Listen by Looking app. An arrow represents the user moving to a different screen by pressing a button on the user interface.

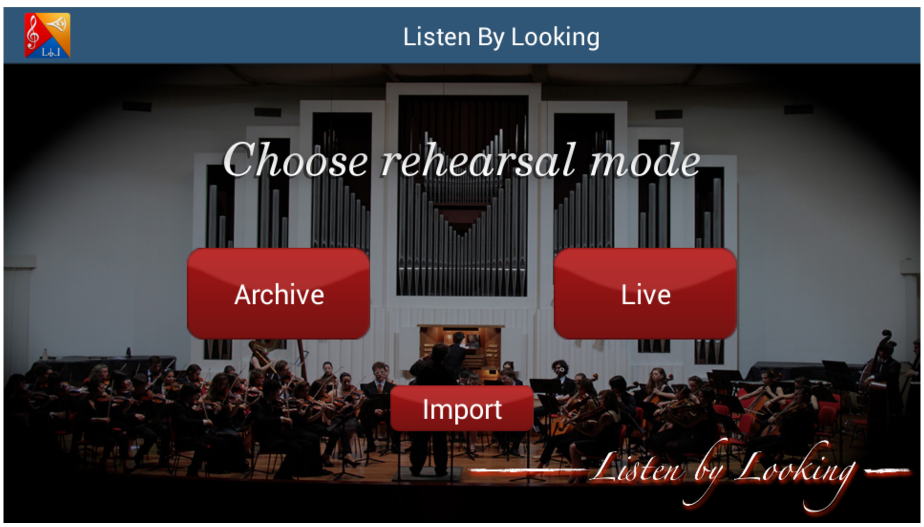

Figure 2: First screen of the app.

modes, and one "Import" button located at the bottom of the screen. This button allows the user to scan for pieces available in an external server, and to download them locally. It is worth reiterating that this is the only function in the app that requires an Internet connection.

The second block (Music selection screen) is associated with the list screen, showed in Figure 3. In this screen, the user is presented with a list of pieces

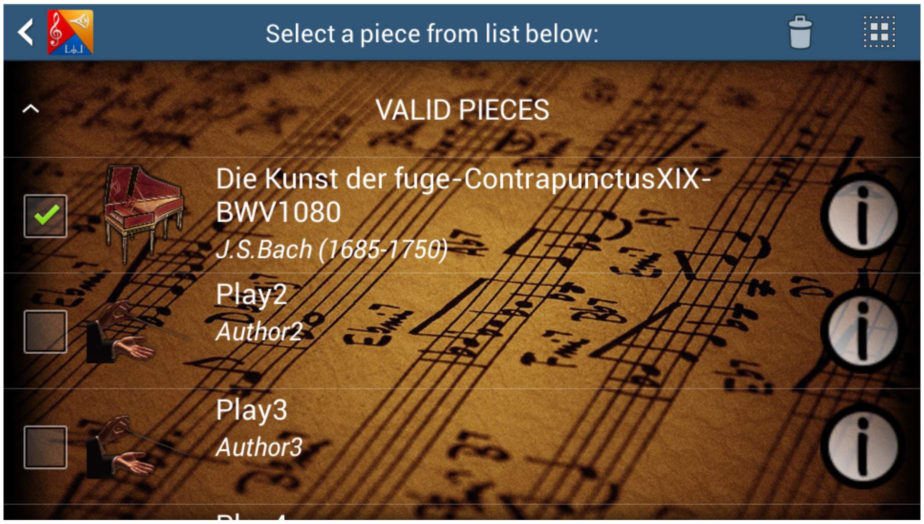

Figure 3: Second screen of the app.

stored in the local memory and available for selection. This screen also allows the user to delete pieces from memory. Each row of the list represents a different piece and displays its title and its author, along with an information button that pops up a dialog with more details about the piece.

The final block (Score and Piano Roll representation screen) is associated with the main screen of the application. In this screen, the selected score representation is presented to the user, along with control buttons on the left and two bars on top and on bottom of the screen, as showed in Figure 4. The bottom bar displays the progress of the audio while graphically partitioning its entire duration into music sections. The top bar controls the way in which the different score voices are displayed on the screen. The control buttons on the left enable the user to horizontally zoom the score (+,0,-), to enable score annotations (coloring certain notes, for example different motifs in a fugue), to switch from the Score to the Piano Roll representation, to re-enable score synchronization with audio in case the user decides to manually scroll across the piece, and to start/stop the audio playback. Additionally, the score global zoom system is controlled via some user gestures: a single tap zooms in, a double tap zooms

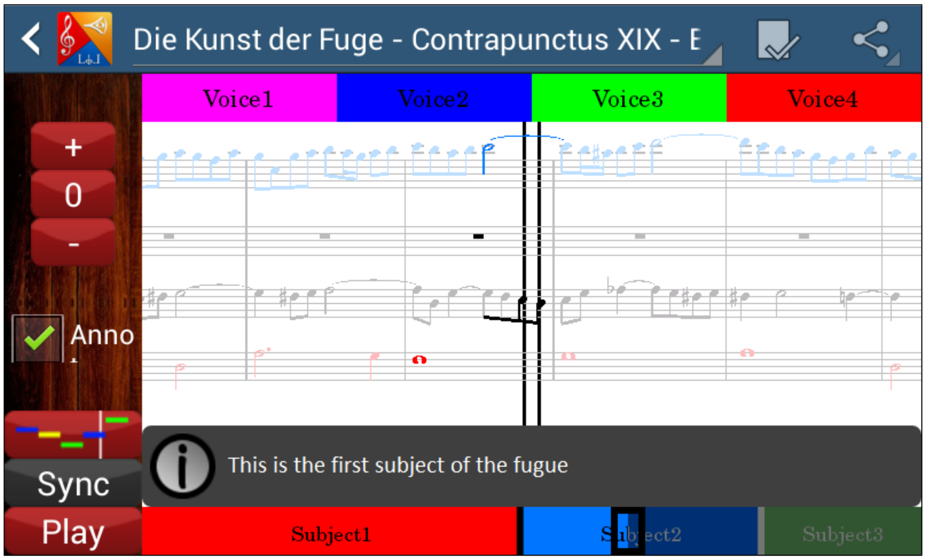

Figure 4: Main screen of the app with score representation.

out, and a long press resets zoom. The Piano Roll displays notes as rectangles whose length represents the duration of the note, while the vertical position of the rectangle represents its pitch without an absolute reference, meaning that the lowest-pitched note is at the bottom of the screen and the highest-pitched is at the top. An example of Piano Roll is shown in Figure 5.

It is worth noting that every major music notation software company develops and uses its own proprietary format, along with proprietary fonts, to represent scores. As a consequence, for the purpose of this application there was a need to develop an ad-hoc format to represent the various score objects in a hierarchical way. Our format is based on the paradigm of markup languages, while the fonts are derived from the font set included in the open source software LilyPond. Both the Score and the Piano Roll representations are described in two separate text files by human-readable graphics directives that are parsed at loading time[1].

[1] There is indeed a music interchange format call MusicXML, but it still needs an interpreter to position the notes exactly on the staves.

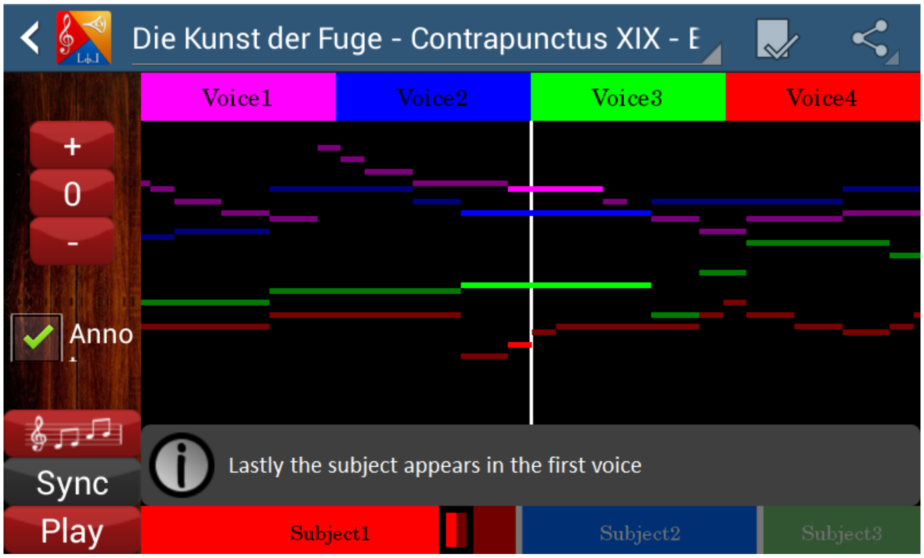

Figure 5: Main screen of the app with Piano Roll representation.

3.2. Score Following

Listen by Looking must display events on the screen perfectly synchronized with music events. The real-time alignment of the live music performance with the Score and Piano Roll representations is obtained through a software module called a score follower. The general task of music-to-score alignment is to identify, for each time frame of a recorded or live performance, the corresponding position on a symbolic representation of the score that is being played. In the case of live applications, a local alignment is computed but the system should be able to adjust it, even substantially, when new information is available.

Early approaches to score following focused on the alignment of a monophonic solo instrument, and were based on external pitch trackers. However, current approaches directly deal with the audio signal, usually exploiting statistical models [34]. In particular, hidden Markov models have proved to be a useful tool to align complex polyphonic scores in real time [35, 36], and hierarchical hidden Markov models [37] have been used to anticipate the musical gestures. An overview of the different approaches to score following is presented in [38]. The work in this paper builds upon previous research on score following, applied to an identification task [39].

The idea behind the proposed approach is that a performance can be considered as the realization of a process that converts the representation a performer has about a music piece into a sequence of acoustic features. Even if the representation is usually read from a music score, the process is non-deterministic because different performances correspond to the same music piece, depending on a number of parameters which are only partially known. For instance, these parameters depend on the timbre of the singing voice, on the choice of the instrumentation and on the playing techniques, but also on background noise, room reverberation, and recording techniques. Moreover, high-level features – e.g., notes – depend on unknown parameters such as variations due to ornamentation or simply to performing errors.

For the current application, the real-time alignment of a performance with a score is obtained through the use of a hidden Markov model. The model works on two levels. The lower level compares the features of the incoming signal with the ones that are expected, given the information in the score. In this level the main features are the signal energy – to detect note attacks and sustain phases as well as possible rests – and the presence of peaks in the signal that are related to the expected notes in the score. Given the considerations about the possible differences between different performances of the same score, peaks are modeled in a very general way by a set of filter banks where each filter is centered around the first four harmonics of the active notes. Groups of states of the lower level are embedded in states at the higher level. The model is strictly left-to-right: self-transition probabilities of states at the lower level model the expected duration of each event in the score; forward-transitions between states at the higher level model the linear relation between subsequent events in the score.

The position of the performer on the score is computed through a decoding technique alternative to classic Viterbi decoding. Basically, the model computes the probability of being in a given high-level state – which corresponds to a given event in the score – after observing a sequence of audio features extracted from the performance in real-time.

In Listen by Looking, alignment is computed either off-line and included in the package for Archive mode, or in real-time for Live mode. In the second case, assuming that there could be many listeners at the same concert, we propose that the alignment is computed centrally by a dedicated server and sent as a synchronization signal to the devices of the listeners. In this way it is possible to capture the audio signal in an optimal condition, and to minimize the energy consumed by the devices.

3.3. Data Over Sound

Our initial approach to send a synchronization signal from a central server to the mobile devices was based on a non-canonical exploitation of Wi-Fi beacon frames [40]. We successfully tested this flavour of Wi-Fi transmission on devices running up to version 4.4.2 of Android. However, from version 5 Android interprets this use of Wi-Fi as a malicious activity and thus stops it. Moreover, different Android devices have varying low-level implementations of the Wi-Fi software stack: this fact affects the latency of beacon packets, which ranges from less than a second to three or four seconds. Moreover, the low-level implementation is not accessible to app developers: as a consequence, there is no way to set the exact packet rate.

This problem could be addressed by connecting all the devices to a centralized server through canonical Wi-Fi connections, but in a typical concert hall, seating about 200 people, this would require the setup of an ad-hoc Wi-Fi infrastructure because consumer equipment can not manage so many devices at the same time. Another solution would be transmitting the data in broadcast mode, so that the devices do not have to maintain a two-way communication channel with the server. Unfortunately, most Android devices do not support broadcast transmission at the hardware level. This analysis led us to abandon Wi-Fi technology altogether and start distributing the synchronization signal with a technique known as *data over sound* (DoS). Research on DoS stems from the need to transmit data in environments (e.g., ATEX environments) where transmission with electromagnetic waves is not allowed and cable transmission is not feasible, or to transmit data in an environment (e.g., under water) filled with a medium that hampers long-range electromagnetic propagation.

Low-latency broadcast communication from a server to a relatively high number of devices, as in our scenario, can be obtained using an audio channel, in particular using a high-frequency carrier wave that is almost inaudible (ultrasonic) for the public but is captured by the microphones of most smartphones and mobile devices [41]. Ultrasonic waves present the added benefit of not interfering with electronic devices and are basically filtered by the walls of a room, making them a good option also for security applications [42] and for mobile payment systems [43]. In this paper, we base our design on the model to communicate through high-frequency sounds proposed in [44].

A DoS transmission channel is no different from a radio channel, except for the fact that it relies on air pressure variations instead of photons. In our case we adopt a simplex channel: the transmitter sends data and the receiver acquires them without the need for the latter to inform the former that data have been received correctly. This greatly simplifies the design of the transmission system and let the programmer avoid the need to calculate back and forth transmission timing between devices.

The design of a DoS channel needs to comply with some additional constraints with respect to a conventional radio channel. The first constraint arises from the fact that sound waves propagate in air at a much lower speed than electromagnetic waves, therefore there is a significant delay between the transmission of a packet and its acquisition by the receiver. A second, and more severe, constraint arises from the fading effect caused by multi-path propagation. In other words, a transmitted signal can reach the receiver twice along two different paths (for instance, bouncing off an obstacle) and at different times, possibly causing destructive interference.

These two constraints are the main limiting factors for the symbol transmission rate, which we set to 10 sym/s considering a scenario of a typical concert hall approximately 30 meters both wide and long, and assuming sound wave velocity to be around 340 m/s. With the states symbol rate, the physical separation between symbols is around 34 meters. If we assume that the worst multi-path propagation length (which still possesses enough signal strength to effectively disrupt the transmission) is about 1.1/1.2 times the normal propagation distance, then in the worst case there is below 25% of overlap between the echoed symbol and the next one, allowing for correct symbol decoding at the receiver. It is worth noting, however, that these are all theoretical calculations based on assumptions that may not be verified in all concert halls, hence the symbol rate of 10 sym/s should be treated as a default value, whereas in a particularly echoing room with a much more severe fading effect the only practical mitigation would need be for the symbol rate to be lowered.

A third constraint arises from the fact that the audio signal must not interfere with the fruition of live music, thus restricting the usable band of the channel to ultrasonic frequencies. Moreover, commercial microphones tend to introduce an amplitude attenuation that increases with the frequency of the signal and is usually non-negligible over 17 kHz. Taking these two factors into consideration, the transmission band for our DoS system was limited to only 200 Hz between 18.0 and 18.2 kHz.

Having satisfied the aforementioned constraint, the next design step was to decide the transmission scheme. The choice was made to adopt an OFDM modulation scheme [45]. OFDM employs a simple frequency shaping technique based on the fact that a rectangular function in the time domain becomes a sinc function in the frequency domain, and vice versa. By applying an amplitude modulation to rectangular functions in the time domain (using a set of appropriately spaced frequencies), the result is a series of sinc functions in the frequency domain, spaced in such a way that the peak of one function lies exactly in the minima of the others and its amplitude is not altered. In the light of this property, the functions are more easily implemented in the frequency domain with an FFT library. Given our low bit rate, we found it adequate to use the FFT implementation in pure Java provided by the jTransforms library[2].

The final design step is defining the number of bits per symbol, that is, the number of sinc function to superimpose in the frequency domain. The number of functions is proportional to the number of bits that can be transmitted. However, as the number of frequency lines (sinc functions) increases, the signal-to-noise ratio of a single line is degraded and more power is required to successfully transmit with a given amount of background noise. In our encoding scheme, one bit per frequency line amplitude (higher or lower than a given reference) is transmitted, and the phases of the lines are set in such a way that each group of three lines carries the phase of an amplitude-modulated sinusoid quantized into eight different phases, thus carrying 3 bits of information. The overall bit rate is set to 240 bits per second.

To test our DoS channel, a simple transmitter was implemented in a desktop programming environment (Wolfram Mathematica), while the receiver is being incorporated into Listen by Looking. To perform the experiments described in this article, we implemented the receiver as a stand-alone Android app that exposes debug information. Furthermore, a simulator of the Android receiver has been programmed in Mathematica as a fully automated procedure that reads audio data from a WAV file instead of a microphone, thus allowing an offline analysis of the audio data emitted from different speakers and recorded by different devices in various environmental setups. We tested our channel in three different setups: the first inside an acoustic anechoic chamber with a single Genelec 8030C speaker as transmitter and a Behringer Ecm8000 omnidirectional microphone as receiver, aligned with each other in line of sight. We considered this environmental setup as a reference because the conditions for transmission are ideal, there are no reverberations and echoes in the chamber and the frequency response of both the speaker and the microphone in the transmission band is flat. The second and third setups we considered are real-world scenarios: two domestic environments without sound-absorbing walls and

[2] https://github.com/wendykierp/JTransforms

potentially non-flat frequency response of the transmitter and the receiver. In both setups the receiver was a real Android device: a Sony Xperia X smartphone in one case, a Google Pixel 3 XL in the other. The emitters were a Tangent Spectrum X5 speaker in the former case, and an Insignia Computer 8A18A speaker in the latter.

In all three setups we performed the experiments using four different distances between the transmitter and the receiver (0.5, 1, 1.5 and 2 meters), and four different values of the signal to noise and distortion ratio (SNDR), specifically 0dB, 10dB, 20dB and infinite (that is, without any background noise added to the signal). The added noise was white and Gaussian. Each of these signal stimuli carried two distinct types of data, one being a 100-character human-readable string and the other being a 129-chunk long sequence of 8 bits chunks generated randomly in a 32 bit long array and then right rotated in a round robin fashion to obtain the successive 4 chunks. Each transmission was quantitatively evaluated by measuring the number of correct bits received and the mean SNDR level seen by the receiver, along with collecting graphics of the distribution of the amplitudes and the phases of received symbols. In the first setup the SNDR level was measured with a phonometer at 1 m distance from the speaker. In the other setups the SNDR level was set directly at the signal generation level, hence the SNDR level later measured at the receiver can be significantly lower. The results with the first setup (controlled environment) are not reported because the measured bit error rates did not provide any actual insight: all bits were received correctly. It is worth noting that the minimum SNDR measured by the receiver was 1.5 dB. Results with the second and third setups (domestic environments) are reported in Table 1 for the string, and in Table 2 for the sequence of bits.

It can be observed in the tables that for distances up to 2 meters the bit error rate remains under 15% when the SNDR is greater than zero. Assuming that the bit errors are grouped in time, meaning that they are caused by some kind of noise burst, then using a standard Reed-Solomon error correcting code (ECC), with 8 ECC symbols every 12 data symbols, the receiver can recover up to 15%

	Bit error rate (%)		Measured SNDR (dB)	
	Setup 2	Setup 3	Setup 2	Setup 3
0.5 mt distance				
No noise	0.0	0.0	37.4	33.5
20 dB SNDR	0.0	0.0	21.3	25.0
10 dB SNDR	0.0	0.0	11.5	15.6
0 dB SNDR	0.0	0.0	2.3	6.3
1 mt distance				
No noise	0.0	5.6	31.5	27.9
20 dB SNDR	0.0	5.4	21.1	22.1
10 dB SNDR	0.0	6.9	11.8	13.4
0 dB SNDR	0.0	6.9	1.5	4.2
1.5 mt distance				
No noise	30.9	23.1	20.6	19.7
20 dB SNDR	27.1	25	15.6	10.1
10 dB SNDR	22.5	22.6	7	7.8
0 dB SNDR	26.4	24.1	-2.6	-1.5
2 mt distance				
No noise	10.9	7.8	24.7	20.3
20 dB SNDR	12	7.5	17.9	16.6
10 dB SNDR	12.2	8.1	9	8.3
0 dB SNDR	7.2	11.8	-0.9	-0.6

Table 1: Table reporting DoS transmission results measured in the second and third test environments (domestic environments) for the string type test signal.

	Bit error rate (%)		Measured SNDR (dB)	
	Setup 2	Setup 3	Setup 2	Setup 3
0.5 mt distance				
No noise	0.0	0.4	38.5	33.4
20 dB SNDR	0.0	0.2	10	14
10 dB SNDR	0.0	0.0	0.8	4.1
0 dB SNDR	57.7	0.1	-12.8	-5.6
1 mt distance				
No noise	0.0	11.4	33.6	27.8
20 dB SNDR	0.0	8	10.4	12.1
10 dB SNDR	0.0	6.5	-0.3	2.2
0 dB SNDR	49.1	13	-10.8	-7.8
1.5 mt distance				
No noise	35.6	26.9	21.5	21.3
20 dB SNDR	31.3	24.1	5.2	6.7
10 dB SNDR	30.3	48	-4.1	-3.4
0 dB SNDR	46.9	48	-15	-13
2 mt distance				
No noise	10.7	14	24.9	21.7
20 dB SNDR	11.8	7.2	6.6	6.8
10 dB SNDR	11.8	12.2	-3.3	-2.8
0 dB SNDR	48.3	N/A	-13.3	N/A

Table 2: Table reporting DoS transmission results measured in the second and third test environments (domestic environments) for the bits type test signal.

of bit errors with a time delay at most of two seconds. However, in the case that the bit errors are uniformly spread in time Reed-Solomon is quite ineffective and a different correction approach should be considered, like Low-Density Parity-Check codes, for which the overall maximum bit error rate acceptable must be re-evaluated. In any case, this second phenomenon is unlikely to happen in a real concert performance.

In the offline analysis of the sound waves recorded by the Android devices, we noted that many of the wrong bits were due to a misaligned decision level in the amplitude constellation. In this case, considering that for each bit the receiver chooses one of three possible answers provided by three different adaptive algorithms, many of those wrong bits could be recovered even without applying the ECC correction by simply trying the two discarded answers. Another minor cause of errors was found to be a misaligned phase constellation, which can be mitigated using a one-bit successive coding for the phases, such as a Gray coding.

In both tables it can be seen that at a distance of 1.5 meters bit error rates increase dramatically, and after an analysis of the raw audio samples the cause was found to be an unfortunate situation where the received symbol is nearly superimposed to a strong echo of the previous symbol. From an additional frequency analysis it also emerged that a small part of the errors are due to both a mismatch between the nominal and the transmitted audio sampling frequency, and to the non-linearity of the transmitter-receiver system, mainly introduced by the speaker.

4. Educational Use of the App

As stated in Section 3, the Listen by looking app is built on three main software components, which constitute a framework that can be adopted to develop different educational or edutainment activities. In summary, the components are: a) a rendering engine to visualize a music score in different forms, together with additional graphical and textual hints; b) a music-to-score alignment system; c) a communication tool, based on a DoS approach, which can be used to broadcast a synchronization signal as well as other data, such as educational contents and/or questions. The markup format used for defining the Score and the Piano Roll representations has, among other characteristics, a peculiar advantage: it allows to separate the definition of the score from its drawing logic. In other words, it allows to feed the drawing engine of the application with high-level music directives rather than graphical primitives, and this makes the drawing engine suitable as a framework (that is, a base of code) for building different kinds of games and learning experiences.

To explain how the three components can interact, we consider the scenario of a live concert. Before the show, every spectator installs the Listen by Looking app on her smartphone or tablet and downloads a package containing a) the music scores (codified in the markup language) and eventual lyrics of the pieces that are being played, b) audio recordings of the pieces, and c) additional information about various historical or structural aspects of the pieces. The audio recordings allow to design at-home listening activities both before the concert, to prepare and motivate the audience, and after the concert, to reinforce the knowledge acquired during the live session. During the concert, the score follower allows real-time synchronization between the music played on the stage and the graphical and/or textual information presented on the mobile devices. Moreover, the DoS module allows to broadcast asynchronous messages, for example questions that direct the attention of the audience towards specific musical aspects, such as the presence of different musical themes or particular rhythmic patterns. In this way, the motivation of the audience can be improved by means of gamification techniques, such as rewarding points and badges for answering the questions.

4.1. An Example of Serious Game

In this section, we present a serious game (detailed with educational purposes) developed for Android mobile devices using the Listen by Looking framework. This game is intended to be played at home before the concert, to give the listeners some information and hints that help him to understand and appreciate peculiar aspects of the musical language used by the composer. The game was used to test the Listen by Looking framework, following the experimental set-up detailed in Section 5.

The game is structured as follows. After having installed the app, the user can load a package containing three music pieces, complete with audio recordings, and a set of hints and questions. After the user has chosen a piece, the audio starts playing and the score begins scrolling in sync with the music, letting the user choose whether she prefers the Score or the Piano Roll representation. During playback, when scheduled, the user is presented with messages containing useful information about the piece: historical background, piece instrumentation, information about the composer, and so on. Figures 4 and 5 show two examples with the two different graphic representations. Then, at various points during playback, and chiefly at the end of a section or during a pause, the user is presented with questions (see Figure 6 for an example) she must answer before continuing to listen to the piece. The questions, listed in Table 5 and num-

Figure 6: Main screen of the app when music playback is interrupted by a question.

bered from B1 to B6, were chosen to address two specific pedagogical aims: a) to stimulate the listening attention towards the different melodic lines of the counterpoint (B1-B3), and b) to guide the auditory memory in recognizing new and recurring melodic motifs and musical themes (B4-B6). Such basic competences are fundamental to understand and recognize more complex structures and musical *macroforms*, such as fugue or sonata structures.

5. Assessment

5.1. Subjects

The app was assessed in laboratory with the collaboration of 20 subjects recruited on a volunteer basis. 10 subjects were men and 10 women, with an average age of 33.95 years (standard deviation: 16.76 years); 6 of them play, or played, a musical instrument. No subject reported hearing defects. The selection was made without any criterion based on age, gender, education or profession. The subjects were randomly divided into two groups of 10 people each:

1. an experimental group that used the app on a device provided by the tester;
2. a control group that listened to music by means of a conventional audio player.

The 6 players of a musical instrument were uniformly distributed between the two groups.

5.2. Materials

The music pieces selected for the game and used in this assessment were:

- J. S. Bach - *Contrapunctus XVIII*, Die Kunst der Fuge (The art of the fugue),
- M. Ravel - *Les Entretiens de la Belle et la Bête (Conversation of Beauty and the Beast)*, Ma Mere l'Oye,

- M. Mussorgsky - *Bydlo*, Pictures at an Exibition.

Music was reproduced through a pair of Sennheiser HD-280 PRO headphones. The app was installed on a Samsung Galaxy Core Plus smartphone with Android version 4.2.2; we remark that, even if the device employed in the assessment is not very recent, the Listen by Looking app is compatible with most devices running the latest versions of Android. The device with the application installed was always the same, and provided by the tester, in order to obtain more uniform results and to avoid differences in the user interface due to customization from the vendors of the devices.

5.3. Procedure

The assessment procedure was organized into three different phases: information collection, exploration (experimental group only), active interaction.

In the first phase the users in both groups were asked to anonymously provide, using Google Forms, some general information about themselves such as age, education, whether she/he plays a musical instrument and can read a music score.

In the second phase, the users in the experimental group were first allowed to explore the app freely for five minutes. Immediately after, they were asked whether they agreed with a number of statements on the usability of the app; the statements are listed in Table 3. The answers were collected with Google Forms and were provided as a value between 1 and 5 (Likert scale) for each statement, where 1 corresponds to “totally disagree” and 5 corresponds to “totally agree”. The last statement, about piano roll, differs from the others by the fact that allows only a Yes or No answer. The users in the control group were not asked to answer these questions.

In the third and last phase, the users in the experimental group played the game described in Section 4.1 with the three pieces listed in Section 5.2. The app repeatedly stopped during playback and challenged the user to answer a question, then playback resumed. This procedure was done for each of the three pieces. Instead, the users in the control group were allowed to listen to the

Usability Assessment
A1 - Understanding the interaction with the touch-screen was simple.
A2 - Touchscreen was a comfortable way of interaction.
A3 - Interacting with the app was simple.
A4 - The interactive experience was positively different from other fruitions I previously had.
A5 - The interactive experience was easy to understand.
A6 - The interactive experience was enjoyable.
A7 - The interactive experience was entertaining.
A8 - Information were presented in a clear way.
A9 - The navigation structure was easy to use.
A10 - The on-screen information was useful/interesting.
A11 - Texts, icons and representations were clear and unambiguous.
A12 - The colors used didn't create difficulties for the navigation actions.
A13 - Visual effects were consistent with the action performed.
A14 - The interface was visually appealing.
A15 - After using the app I understood what a piano roll is.

Table 3: Questions about usability.

Samples Assessment
B1 - How many melodic lines can you identify?
B2 - How many musical instruments can you identify?
B3 - Do you sense a melodic line more important than the others?
B4 - How many motifs have you found in the piece until this point?
B5 - In this point what motif (if any) did you recognize?
B6 - In this other point what motif (if any) did you recognize?

Table 4: Questions about music.

audio of the three pieces and during reproduction were prompted by the tester to answer the same questions faced by the experimental group. The questions are the same for each of the pieces and are reported in Table 4.

5.4. Results

Table 5 summarizes the mean scores for for the first four questions about the pieces. A score of 1 was assigned each time an answer was correct, otherwise 0 points were attributed. Table 6 reports the average score over the four questions for each piece.

The last two questions, that is, B5 and B6, must be treated differently because the user was requested to indicate what motifs could be identified, so a score of 1 was assigned for each motif recognized and a score of -0.5 for each motif incorrectly recognized. Table 7 contains the scores for the two questions for each piece. Furthermore, in Table 8 the total scores obtained by each user in both groups are reported; the maximum score is 22 points. The scores of the experimental group are significantly higher than the control one; according

Question	**Exp. group**	**Control group**
	Mean / St.Dev.	Mean / St.Dev.
B1	0.7 / 0.47	0.47 / 0.51
B2	0.63 / 0.49	0.6 / 0.5
B3	0.9 / 0.31	0.83 / 0.38
B4	0.7 / 0.47	0.37 / 0.49

Table 5: Mean scores for questions B1-B4 over all the pieces.

Piece	**Exp. group**	**Control group**
	Mean / St.Dev.	Mean / St.Dev.
Piece 1	0.65 / 0.48	0.5 / 0.51
Piece 2	0.75 / 0.44	0.52 / 0.51
Piece 3	0.8 / 0.41	0.67 / 0.47

Table 6: Mean scores for single piece.

to the Mann-Whitney Test, the difference is statistically significant ($U = 4.5$, $z - score = 3.4$, $p < .01$).

Finally, Table 9 contains the interface usability scores coming from the users of the experimental group.

5.5. Discussion

The tables in the previous section show a significant increment in the results obtained by the experimental group, supporting the hypothesis that Listen by Looking is a useful tool for learning structural aspects of music. Indeed, the tables containing the mean score for each question highlight a higher score for the experimental group. However, this difference is less evident for the third piece. The last row of Table 7 shows that the difference between the means of the groups for piece 3 is significantly smaller than the same difference for pieces 1 and 2. This can be explained by observing that the motifs on this piece are easier to recognize and the piece itself is scored for fewer instruments.

Regarding the usability statement, almost all of them earned a positive rat-

Question	**Exp. group**	**Control group**
	Mean / St.Dev.	Mean / St.Dev.
Piece 1		
B5 (2)	1.8 / 0.42	0.8 / 0.63
B6 (2)	1.55 / 0.69	0.5 / 0.53
Piece 2		
B5 (2)	1.9 / 0.32	0.8 / 0.42
B6 (2)	1.7 / 0.67	0.8 / 0.63
Piece 3		
B5 (1)	0.8 / 0.26	0.65 / 0.47
B6 (1)	0.85 / 0.34	0.7 / 0.48

Table 7: Mean scores for questions B5 and B6. The numbers in brackets are the maximum scores assigned for each question.

User #	**Exp. group score**	**Control group score**
1	16	15
2	17	14
3	21	11
4	20	9
5	22	10
6	17.5	11
7	18.5	13
8	11.5	12
9	15	8
10	15.5	7.5
Mean	17.4	11.05
St. Dev	3.13	2.5

Table 8: Scores of all users under test.

Question	Min	Max	Mean	Standard deviation
A.1	2	5	4.1	1.2
A.2	4	5	4.7	0.48
A.3	3	5	3.6	0.84
A.4	3	5	4.3	0.82
A.5	2	5	3.9	0.99
A.6	3	5	4.4	0.7
A.7	3	5	4.2	0.79
A.8	2	5	3.9	1.2
A.9	3	5	3.9	0.88
A.10	3	5	4.5	0.71
A.11	2	5	3.7	1.16
A.12	2	5	4.3	0.95
A.13	3	5	4.4	0.84
A.14	4	5	4.5	0.53
A.15	7 Yes		3 No	

Table 9: Interface usability scores from the experimental group.

ing with values between 3 and 5. In particular, statement A.2 received scores between 4 and 5 with a mean value of 4.7, and statement A.14 received values between 4 and 5 with a mean value of 4.5. These data are very encouraging. However, the assessment highlights the need for further improvements. Indeed, questions A.1, A.5, A.8, A.11 and A.12 earned a minimum rating of 2 while receiving a mean rating of 4.1, 3.9, 3.9, 3.7 and 4.3, respectively, meaning that the interface may need a tuning regarding its graphical aspects to please a greater number of users.

6. Conclusions

In this paper we presented a framework, named *Listen by Looking*, and a mobile application, bearing the same name, designed to entertain and teach basic aspects of classical music during public performances. The user interface provides access to the score – either the usual music score or a simplified piano roll representation of it – which is synchronized with the performance in real time. Online synchronization, also known as score following, is carried out by a central server The synchronization signal is sent to the mobile devices via a data-over-sound technique: the signal is modulated in near-ultrasound frequencies that are inaudible by many listeners and masked by the sound of the ongoing performance. The use of data-over-sound allows for the simultaneous synchronization of a large number of devices. An important technical detail of the app is that the drawing engine allows to abstract from the primitive graphics directives, so the programmer can feed it with high-level music drawing instructions. Hence, Listen by Looking can be considered a framework for building serious games and interactive music learning tasks.

In this paper we also presented a serious game, developed and tested with a number of participants. The game aims at raising awareness about the structural elements of the music language that can be highlighted using a synchronous visual representation: motives, instrumentation, polyphony, and superposition of different melodic lines. In its present state of development, the game is intended to be played before attending a concert, in order to to familiarize with the concert program in a playful way. Basically, the user listens to a piece of music using the app, which provides visual tips and additional information about what is happening as the piece unfolds; then, at various points during playback, the user is prompted to answer specific questions on what she/he just listened. To the best of our knowledge, Listen by Looking is the first app that aims at teaching structures, rhythm, notations and melodies of tonal Western music while actually listening to a performance, both recorded or live. The presence of a score following module, and the possibility of efficiently distributing a synchronization signal through inaudible frequencies, allows the user to decide the amount of time to devote to the gaming activity without loosing synchronization with the ongoing performance. Moreover, by answering the questions, the user becomes increasingly aware of the different elements of the music language while still enjoying the listening experience. The assessment on the game was performed with 20 people divided into two groups (experimental and control). Results show that the scores obtained by the users in the group utilizing the app is generally higher and that Listen by Looking can be effective.

In light of these results, we are driven to further develop the application by consolidating the framework built around the drawing engine, so as to present the programmer with a general way of building serious games based on music. Other aspects for future development reside in expanding the possibilities for music score representation to introduce advanced graphical effects like animations and fading.

References

[1] D. Wigdor, D. Wixon, Brave NUI world: designing natural user interfaces for touch and gesture, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2011.

[2] H. Ishii, Tangible bits: beyond pixels, in: Proceedings of the 2nd interna-

tional conference on Tangible and embedded interaction, ACM, New York, NY, USA, 2008, pp. xv–xxv.

[3] L. Turchet, D. Zanotto, S. Minto, A. Rodà, S. K. Agrawal, Emotion rendering in plantar vibro-tactile simulations of imagined walking styles, IEEE Transactions on Affective Computing 8 (3) (2017) 340–354.

[4] R. J. Jacob, A. Girouard, L. M. Hirshfield, M. S. Horn, O. Shaer, E. T. Solovey, J. Zigelbaum, Reality-based interaction: a framework for post-wimp interfaces, in: Proceedings of the SIGCHI conference on Human factors in computing systems, ACM, New York, NY, USA, 2008, pp. 201–210.

[5] L. Turchet, A. Rodà, Emotion rendering in auditory simulations of imagined walking styles, IEEE Transactions on Affective Computing 8 (2) (2017) 241–253.

[6] G. Pitteri, E. Micheloni, A. Rodà, C. Fantozzi, N. Orio, Listen by looking: a mobile application for augmented fruition of live music and interactive learning, in: Proceedings of the 5th EAI International Conference on Smart Objects and Technologies for Social Good, 2019, pp. 136–141.

[7] T. Marsh, Serious games continuum: Between games for purpose and experiential environments for purpose, Entertainment Computing 2 (2011) 61–68. doi:10.1016/j.entcom.2010.12.004.

[8] U. Ritterfeld, M. Cody, P. Vorderer, Serious Games: Mechanisms and Effects, Taylor & Francis, 2009.

[9] D. R. Michael, S. L. Chen, Serious Games: Games That Educate, Train, and Inform, Muska & Lipman/Premier-Trade, 2005.

[10] B. Suits, The Grasshopper: Games, Life and Utopia, Broadview Press, 2005.

[11] J. Huizinga, Homo Ludens: A Study of the Play-Element in Cult, Beacon Press, 1949.

[12] M. Zyda, From visual simulation to virtual reality to games, Computer 38 (2005) 25 – 32. doi:10.1109/MC.2005.297.

[13] A. Baratè, M. G. Bergomi, L. A. Ludovico, Development of serious games for music education, Journal of e-Learning and Knowledge Society 9 (2) (2013) 93–108.

[14] P. Raziunaite, A. Miliunaite, R. Maskeliunas, R. Damasevicius, T. Sidekerskiene, B. Narkeviciene, Designing an educational music game for digital game based learning: A lithuanian case study, 2018, pp. 0800–0805. doi:10.23919/MIPRO.2018.8400148.

[15] M. Renzi, S. Vassos, T. Catarci, S. Kimani, Touching notes: A gesture-based game for teaching music to children, in: Proceedings of the Ninth International Conference on Tangible, Embedded, and Embodied Interaction, TEI '15, Association for Computing Machinery, New York, NY, USA, 2015, p. 603–606.

[16] K. Kolykhalova, P. Alborno, A. Camurri, G. Volpe, A serious games platform for validating sonification of human full-body movement qualities, in: Proceedings of the 3rd International Symposium on Movement and Computing, MOCO '16, Association for Computing Machinery, New York, NY, USA, 2016. doi:10.1145/2948910.2948962.
URL https://doi.org/10.1145/2948910.2948962

[17] F. Carnovalini, A. Rodà, P. Caneva, A musical serious game for social interaction through augmented rhythmic improvisation, in: Proceedings of the 5th EAI International Conference on Smart Objects and Technologies for Social Good, GoodTechs '19, Association for Computing Machinery, New York, NY, USA, 2019, p. 130–135. doi:10.1145/3342428.3342683.
URL https://doi.org/10.1145/3342428.3342683

[18] E. T. Khoo, T. Merritt, V. L. Fei, W. Liu, H. Rahaman, J. Prasad, T. Marsh, Body music: Physical exploration of music theory, in: Proceedings of the 2008 ACM SIGGRAPH Symposium on Video Games, Sandbox '08, Association for Computing Machinery, New York, NY, USA, 2008, p. 35–42. doi:10.1145/1401843.1401850.
URL https://doi.org/10.1145/1401843.1401850

[19] M. Mandanici, F. Altieri, A. Rodà, S. Canazza, Inclusive sound and music serious games in a large-scale responsive environment, British Journal of Educational Technology 49 (4) (2018) 620–635.

[20] M. Rauschenberger, L. Rello, R. Baeza-Yates, E. Gomez, J. P. Bigham, Towards the prediction of dyslexia by a web-based game with musical elements, in: Proceedings of the 14th Web for All Conference on The Future of Accessible Work, W4A '17, Association for Computing Machinery, New York, NY, USA, 2017. doi:10.1145/3058555.3058565.
URL https://doi.org/10.1145/3058555.3058565

[21] L. Ferreira, S. Cavaco, S. B. i. Badia, A usability study with healthcare professionals of a customizable framework for reminiscence and music based cognitive activities for people with dementia, in: Proceedings of the 23rd Pan-Hellenic Conference on Informatics, PCI '19, Association for Computing Machinery, New York, NY, USA, 2019, p. 16–23. doi:10.1145/3368640.3368654.
URL https://doi.org/10.1145/3368640.3368654

[22] A. Calderón, M. Ruiz, A systematic literature review on serious games evaluation: An application to software project management, Computers & Education 87 (2015) 396 – 422.

[23] E. A. Boyle, T. Hainey, T. M. Connolly, G. Gray, J. Earp, M. Ott, T. Lim, M. Ninaus, C. Ribeiro, J. a. Pereira, An update to the systematic literature review of empirical evidence of the impacts and outcomes of computer games and serious games, Compututer & Education 94 (C) (2016) 178–192.

[24] P. Backlund, M. Hendrix, Educational games - are they worth the effort? a literature survey of the effectiveness of serious games, in: Games and Virtual Worlds for Serious Applications (VS-GAMES 2013 5th International Conference on, 2013, pp. 1–8.

[25] R. Andreoli, A. Corolla, A. Faggiano, D. Malandrino, D. Pirozzi, M. Ranaldi, G. Santangelo, V. Scarano, A framework to design, develop, and evaluate immersive and collaborative serious games in cultural heritage, J. Comput. Cult. Herit. 11 (1) (Dec. 2017). doi:10.1145/3064644. URL https://doi.org/10.1145/3064644

[26] I. Rubino, C. Barberis, J. Xhembulla, G. Malnati, Integrating a location-based mobile game in the museum visit: Evaluating visitors' behaviour and learning, J. Comput. Cult. Herit. 8 (3) (May 2015). doi:10.1145/2724723. URL https://doi.org/10.1145/2724723

[27] C. Pierce, C. J Woodward, A. Bartel, Learning management models in serious mobile music games, in: Proceedings of the Australasian Computer Science Week Multiconference, ACSW '20, Association for Computing Machinery, New York, NY, USA, 2020. doi:10.1145/3373017.3373069. URL https://doi.org/10.1145/3373017.3373069

[28] C. Magerkurth, A. D. Cheok, R. L. Mandryk, T. Nilsen, Pervasive games: bringing computer entertainment back to the real world, Computers in Entertainment (CIE) 3 (3) (2005) 4–4.

[29] W. W. Gaver, Auditory icons: Using sound in computer interfaces, Human-computer interaction 2 (2) (1986) 167–177.

[30] J. L. Alty, Can we use music in computer-human communication?, in: BCS HCI, Cambridge University Press, New York, NY, USA, 1995, pp. 409–423.

[31] E. Micheloni, M. Tramarin, A. Rodà, F. Chiaravalli, Playing to play: a piano-based user interface for music education video-games, Multimedia Tools and Applications 78 (10) (2019) 13713–13730.

[32] S. Deterding, D. Dixon, R. Khaled, L. Nacke, From game design elements to gamefulness: defining gamification, in: Proceedings of the 15th international academic MindTrek conference: Envisioning future media environments, ACM, New York, NY, USA, 2011, pp. 9–15.

[33] D. Johnson, S. Deterding, K.-A. Kuhn, A. Staneva, S. Stoyanov, L. Hides, Gamification for health and wellbeing: A systematic review of the literature, Internet Interventions 6 (2016) 89 – 106.

[34] L. Grubb, R. Dannenberg, A stochastic method of tracking a vocal performer, in: Proceedings of the International Computer Music Conference, Michigan Publishing, Thessaloniki, Greece, 1997, pp. 301–308.

[35] C. Raphael, Automatic segmentation of acoustic musical signals using hidden markov models, IEEE Transactions on Pattern Analysis and Machine Intelligence 21 (4) (1999) 360–370.

[36] N. Montecchio, N. Orio, Automatic alignment of music performances with scores aimed at educational applications, in: Proceedings of the International Conference on Automated Solutions for Cross Media Content and Multi-channel Distribution, IEEE Computer Society, Washington, DC, USA, 2008, pp. 17–24.

[37] A. Cont, Realtime audio to score alignment for polyphonic music instruments using sparse non-negative constraints and hierarchical hmms, in: Proceedings of the IEEE International Conference in Acoustics and Speech Signal Processing, IEEE, Toulouse, France, 2006, pp. V245–V248.

[38] N. Montecchio, E. D. Buccio, N. Orio, An efficient identification methodology for improved access to music heritage collections, Journal of Multimedia 7 (2) (2012) 145–158.

[39] N. Orio, Automatic identification of audio recordings based on statistical modeling, Signal Processing 90 (4) (2010) 1064–1076.

[40] G. Pitteri, E. Micheloni, A. Rodà, C. Fantozzi, N. Orio, Listen by looking: A mobile application for augmented fruition of live music and interactive learning, in: Proceedings of the 5th EAI International Conference on Smart Objects and Technologies for Social Good, GoodTechs '19, Association for Computing Machinery, New York, NY, USA, 2019, p. 136–141. doi:10.1145/3342428.3342684.
URL https://doi.org/10.1145/3342428.3342684

[41] S. Kim, H. Mun, Y. Lee, A data-over-sound application: Attendance book, in: 2019 20th Asia-Pacific Network Operations and Management Symposium (APNOMS), 2019, pp. 1–4.

[42] W. Jiang, W. M. D. Wright, Progress in airborne ultrasonic data communications for indoor applications, in: 2016 IEEE 14th International Conference on Industrial Informatics (INDIN), 2016, pp. 322–327.

[43] A. Vizzarri, F. Vatalaro, m-payment systems: Technologies and business models, in: 2014 Euro Med Telco Conference (EMTC), 2014, pp. 1–6.

[44] P. Getreuer, C. Gnegy, R. F. Lyon, R. A. Saurous, Ultrasonic communication using consumer hardware, IEEE Transactions on Multimedia 20 (6) (2018) 1277–1290.

[45] S. B. Weinstein, The history of orthogonal frequency-division multiplexing, IEEE Communications Magazine 47 (11) (2009) 26–35.