A Web-based Real-Time Kinect Application for Gestural Interaction with Virtual Musical Instruments

Athanasia Zlatintsi
School of Electr. & Comp. Enginr.,
National Technical Univ. of Athens
Athens, Greece
nzlat@cs.ntua.gr

Panagiotis P. Filntisis
School of Electr. & Comp. Enginr.,
National Technical Univ. of Athens
Athens, Greece
filby@central.ntua.gr

Christos Garoufis
School of Electr. & Comp. Enginr.,
National Technical Univ. of Athens
Athens, Greece
el11125@central.ntua.gr

Antigoni Tsiami
School of Electr. & Comp. Enginr.,
National Technical Univ. of Athens
Athens, Greece
antsiami@cs.ntua.gr

Kosmas Kritsis
Inst. for Language and Speech Proc.,
Athena RIC
Athens, Greece
kosmas.kritsis@ilsp.gr

M.A. Kaliakatsos-Papakostas
Inst. for Language and Speech Proc.,
Athena RIC
Athens, Greece
maximos@ilsp.gr

Aggelos Gkiokas
Inst. for Language and Speech Proc.,
Athena RIC
Athens, Greece
agkiokas@ilsp.gr

Vassilis Katsouros
Inst. for Language and Speech Proc.,
Athena RIC
Athens, Greece
vsk@ilsp.gr

Petros Maragos
School of Electr. & Comp. Enginr.,
National Technical Univ. of Athens
Athens, Greece
maragos@cs.ntua.gr

ABSTRACT

We present a web-based real-time application that enables gestural interaction with virtual instruments for musical expression. Skeletons of the users are tracked by a Kinect sensor, while the performance of the virtual instruments is accomplished using gestures inspired from their corresponding physical counterparts. The application supports the virtual performance of an air guitar and an upright bass, as well as a more abstract conductor-like performance with two instruments, while collaborative playing of two or more players is also allowed. The multimodal virtual interface of our application, which includes 3D avatars, allows users, even if not musically educated, to engage in innovative interactive musical activities, while its web-based architecture improves its accessibility and performance. The application was qualitatively evaluated by 13 users, in terms of its usability and enjoyability, among others, accomplishing high ratings and positive feedback.

CCS CONCEPTS

• **Human-centered computing → Web-based interaction**; **Gestural input**; *Auditory feedback*; *Virtual Reality*;

KEYWORDS

gestural interaction, virtual musical instruments, web technologies, real-time applications, avateering

AM'18, September 12–14, 2018, Wrexham, United Kingdom

ACM ISBN 978-1-4503-6609-0/18/09…$15.00
https://doi.org/10.1145/3243274.3243297

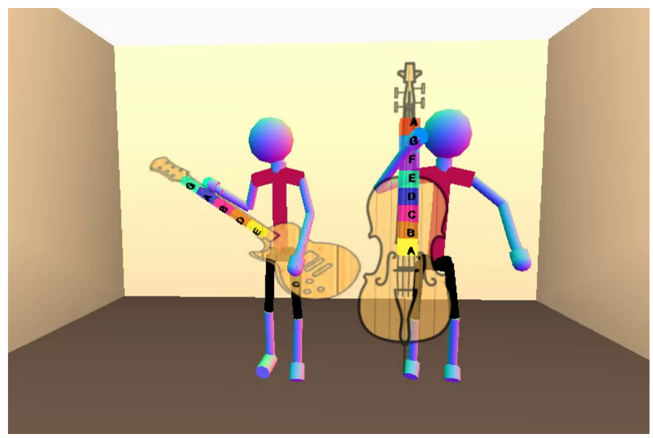


Figure 1: Snapshot of the web-based application for real-time gesture-based collaborative interaction with virtual musical instruments.

ACM Reference Format:
Athanasia Zlatintsi, Panagiotis P. Filntisis, Christos Garoufis, Antigoni Tsiami, Kosmas Kritsis, M.A. Kaliakatsos-Papakostas, Aggelos Gkiokas, Vassilis Katsouros, and Petros Maragos. 2018. A Web-based Real-Time Kinect Application for Gestural Interaction with Virtual Musical Instruments. In *Audio Mostly 2018: Sound in Immersion and Emotion (AM'18), September 12–14, 2018, Wrexham, United Kingdom.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3243274.3243297

1 INTRODUCTION

During the last years, we have seen many advancements in computer technology and interfacing in both music and digital media, and a plethora of attempts have been made in order to create virtual

In Proc. Audio Mostly 2018: Sound in Immersion and Emotion (AM'18), Wrexham, United Kingdom, Sep. 2018

User Interfaces (UI) for sound generation and the control of musical expression. In addition, the rapid development and availability of low cost sensing technologies has also created a wide interest in developing virtual reality (VR) applications. Such advances in sensing technology, such as the Microsoft Kinect sensor and others, combined with the improvement of gesture recognition algorithms, have enabled the development of gesturally controlled musical instruments, which increase the users' degree of freedom to move and to express their emotions, or musicality, without actually contacting a physical object. As also defined in [9], a virtual music instrument is something analogous to a physical musical instrument, a gestural interface, that could however provide for much greater freedom in the mapping of movement to sound.

In this work, we implement a web-based real-time application for gesture-based interaction with virtual instruments using a Microsoft Kinect v2 sensor. The user is able to musically "air control" virtual instruments, such as a guitar and an upright bass, without any physical contact. For the web-based visualization, see Fig. 1, we employ virtual 3D avatars for increased control and better manipulation of the virtual instruments as well as facilitation during the generation of music by the user.

Camera-based motion tracking has been widely used in many applications including music human-computer interaction (HCI). In this case, cameras or infrared sensors are coupled with computer vision and machine learning algorithms to estimate a subject's pose, position and specific motion. Such approaches are advantageous since not only they do not constrain the user's movements but they also support full body motion tracking. Even though some drawbacks can be found in the literature, as for instance the placement of the cameras that could limit the interaction workspace to the camera's field of view, the tracking errors when subjects are occluded or the limited frame rates, they are still considered as affordable solutions that are unobtrusive [4], constituting a good solution to implement high precision motion tracking.

Many works have used gesture-based interaction, using Kinect cameras for motion sensing, for similar applications. For instance, in [12] an augmented piano performance system was created, using gestures and a small projected area above a keyboard to create new extended techniques for manipulating the sounds of the piano. The "Third Room" [6] is also a Kinect-based hybrid 3D space for music composition and performance, where users are placed in a virtual environment to interact with new instruments for musical expression. In [11] gesture recognition and the Kinect's skeleton tracking functionality are used to track the position of the mallets for the performance of percussion instruments, while in [7] the same technology is used to experiment with three virtual musical instruments, a drum, a guitar and the spider king. All three instruments virtually set the relevant sensing input areas, i.e., strings of the guitar or cymbals of the drum, and the user can control the instrument through those virtual inputs. A gesture-based virtual interaction for a drumkit is also explored in [10]. This real-time system detects user-generated drum-hitting gestures incorporating different machine learning solutions, based on the analysis of velocity and acceleration or Wiener filters, in order to compensate for various latency effects. In [5], head tracking from depth images was performed for motion analysis of musical ensembles, while in [3] a real-time gestural interface was presented for musical expression, combining images and sounds specific to percussionists. In addition to the Kinect sensor, hand glove controllers were also used to sense accelerations, kicks, inclinations, orientation movements and relative angles of the hands.

Other approaches, using other sensing technologies than cameras, include wrist mounts that couple the Leap Motion optical sensor with inertial measurement units to combine the benefits of wearable and camera-based motion tracking [1], while in [8] user input from data gloves was used for the gestural control of four virtual instruments: a xylophone, a gestural FM Synthesizer, a membrane and an air guitar. Motion capture devices can be another alternative for sensing motion; for instance, as used in [2] to track, interpret and sonify the performer's movements and gestures in relation to a virtual object.

This work aims at developing an application, where the user will engage in innovative interactive music activities with advanced multimodal interfaces, using only a Kinect sensor, which actually gives the ability to the user to move freely in the physical space, unconstrained and without any other sensors attached to his body. Additionally, the proposed application is web-based, making it widely accessible to everyone; that means that anyone with a Kinect sensor could virtually play music even at home. The gestures are intuitive, including "motion templates" that have some referential similarity to the gestures that a musician does when performing the specific instruments, thus no extensive training or effort to memorize them is required.

The remainder of this paper is organized as follows: Section 2 presents a general overview of our application and elaborates on the gesture-based interaction regarding the different modes as well as the specifications of each virtual instrument. Section 3 outlines the methods and technologies used for the implementation of the web-based real-time application and the visualization of the 3D virtual environment, using only raw skeleton data from the Kinect sensor. In Sec. 4, we analyze the results of the qualitative human evaluation, showing that the application gained really positive feedback, and finally in Sec. 5 we present our conclusions and thoughts for future work.

2 WEB-BASED APPLICATION FOR INTERACTION WITH VIRTUAL INSTRUMENTS

A snapshot of the web-based application can be seen in Fig. 1. The application consists of three different modes of interaction in which the user gesturally controls virtual instruments. Next, we provide a general overview of our application and description of the three modes.

2.1 Gesture and Virtual Reality Interaction for Performing Virtual Instruments

In order to build the gesture interaction module, we have employed one Kinect v2 sensor for Xbox One, an RGB-D sensor by Microsoft. This decision was based on the fact that it is an inexpensive solution that minimizes intrusiveness [4] and thus constitutes a good solution to implement high precision motion tracking. The Kinect sensor can provide the required visual information for the task, since it can capture Full HD RGB video at 30fps (frames per second), depth information that is recorded by the infrared camera embedded in the sensor, and, via the Kinect SDK[1] interface, more streams, such as the skeletons (including 25 joint coordinates, thus the major parts of the user's body) of up to 6 people concurrently.

The skeletons provided by the Kinect are inferred using depth data and their coordinates are provided both on the image (x, y-axis) and on the 3D world (x, y, z-axis). Our application uses all 25 joint positions that are provided by the Kinect v2, to generate a full body 3D virtual avatar that is used in order to improve the interaction and the user interface. In addition, specific joints (such as the position of the hands) are used for recognition of specific gestures that, depending on the selected mode of interaction, generate music.

2.2 Modes of Interaction

The gesture-based interaction with the virtual instruments supports three different modes. The aim of a "simple" and more intuitive gesture interaction in this case is to provide the users (especially those that are not musically educated) the ability to perform various virtual instruments without constraints. Specifically, in the first two modes, our application enables the users to play the virtual instruments by performing specific "motion templates" (thus predefined gestures) that have some referential similarity to the gestures that a musician does when performing the specific instruments. The virtual instruments accessible from these modes include: i) the air-guitar, from the first mode and ii) the upright bass, from the second mode (played using a virtual bow). The third mode incorporates movements that could be abstractly likened with the movements of a conductor, while each hand is assigned with one of the two previously named instruments. In addition to that, the application enables collaborative playing, hence two, or more, users can collaborate and create music together, either by using the same virtual instrument (from the ones mentioned above) or different ones.

Finally, in order to facilitate and guide the interaction of the user, as well as improve the visualization of music generation, each mode provides specific visual aids, apart from the virtual instruments, namely, colored bars accompanied with letters, which show to the user which note is going to be generated if he performs a sound activation gesture. According to the mode, the associated hands of the user are colored likewise. The different modes and the gesture templates depend on the selected instrument type and are described in more detail next.

Air-Guitar Interaction: The user, by selecting the air-guitar mode, will be able to perform gestures similar to ones that a guitar-player does.

Gesture 1: In order to enable and activate the "sounding" event the user brings the right hand around the waist height and moves it vertically (downwards or upwards), simulating this way the moving hand of a guitar player, using as for instance a plectrum. As long as the right hand is performing the specific up and down movement, a pitch is generated.

Gesture 2: In order to be able to change the "sounding" pitch of the string, the user has to move the left hand diagonally from the height of the head to below the waist, as if she/he stops the string on the fretboard determining this way the pitch of the fingered note. This particular gesture is enabled only when *Gesture 1* is also active.

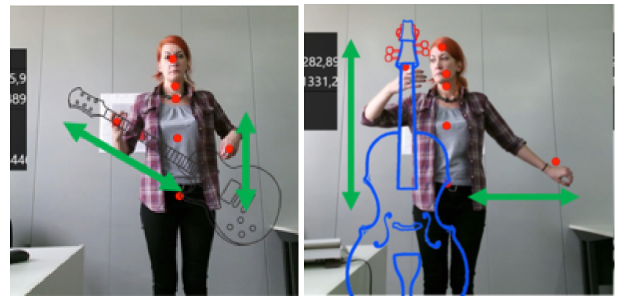


Figure 2: Visualization of the air-guitar (left) and the upright bass (right) gestures for triggering sounding events of various notes. (Left figure) Right hand, showing the vertical movement, that triggers the sound event and Left hand, showing the diagonal movement across the virtual guitar fretboard, that triggers different pitches. (Right figure) Right hand, showing the horizontal movement, that triggers the sound event and Left hand, showing the vertical movement, that triggers different pitches.

For the air guitar mode, two different mappings are predefined, hence the different positions of the left hand are mapped to: i) a pentatonic scale including the notes: G3, A3, B3, D4, and E4 or b) to predefined chords, which are D4, F4, G4, G#4, which when played in the correct order (provided to the user) can simulate the sound of a well-known guitar riff. Figure 2 (left) shows a visualization of the gestures that has to be performed, in order to trigger the various notes.

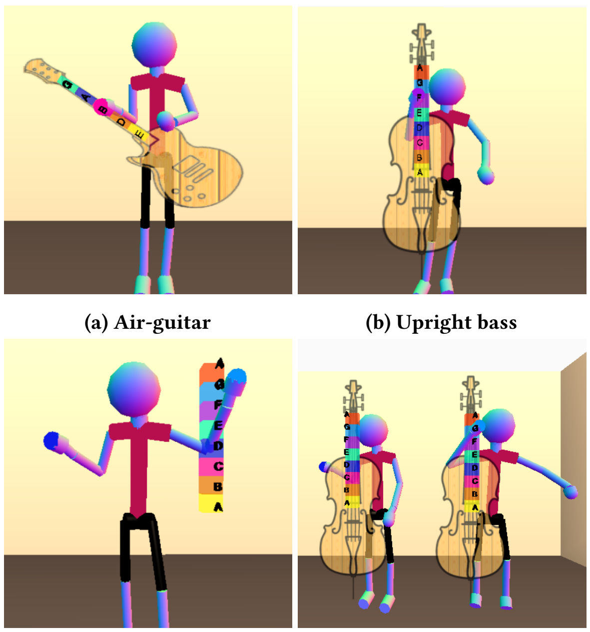
In addition, in this mode, a semi-transparent guitar is rendered on the 3D world that follows the user. Moreover, along the guitar fretboard, a colored bar is overlaid, that consists of different colors and letters that show which note/chord is generated at each different position. We also note that the hand of the user that moves along the fretboard is also colored correspondingly, in order to visualize to the user the note that will be generated if he performs Gesture 1 (sound activation). A snapshot of the application that shows this mode can be seen in Fig. 3a.

Upright Bass - Bowing Interaction: The user by selecting the Upright Bass and the bowing will be able to perform gestures similar to the ones that an upright bass (or a cello) player does.

Gesture 1: In order to enable and activate the "sound", the user should bring the right hand around the waist height and move it horizontally (from right to left and the other way around), simulating this way the bowing movements, as if the bow is dragged horizontally across the string, see Fig. 2 (right). As long as the right hand is performing the specific continuous movement/gesture, we assume that the bow is in contact with the string and a pitch should be generated. When no such movement of the right hand is performed, a sound event is not triggered.

Gesture 2: In order to be able to change the pitch of the string of the upright bass the user has to move the left hand vertically (i.e., downwards or upwards) from the head to the waist height, as if it

[1] https://www.microsoft.com/en-us/download/details.aspx?id=44561



(a) Air-guitar (b) Upright bass

(c) Conductor (d) Multiplayer

Figure 3: The available interaction modes of the web application and their visualization.

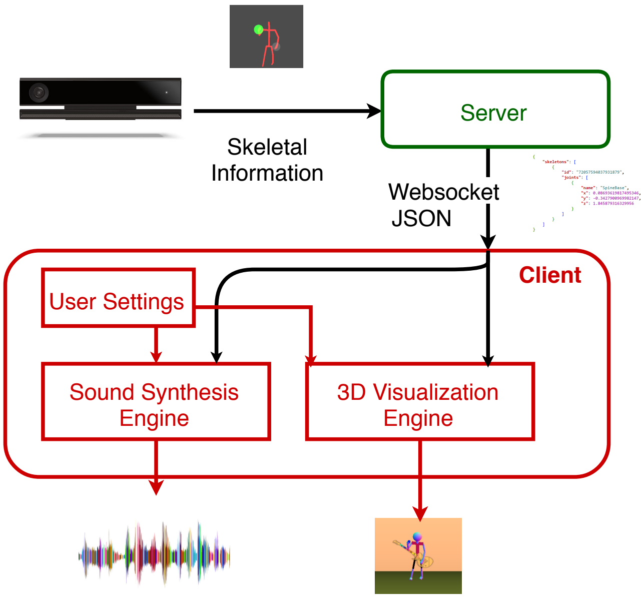


Figure 4: Architecture of the web-based application.

stops the string on the fretboard, determining this way the pitch of the fingered note. This particular gesture is enabled only when Gesture 1 is active, generating different pitches, in order to trigger the various notes, refer to Fig. 2 (right figure) for a visualization of the gestures that has to be performed.

The notes that are generated depending on the position of the user's left hand, from top to bottom of the fretboard, are the eight notes of a scale; hence A2, B2, C3, D3, E3, F3, G3, and A3. As in the guitar mode, in the upright bass mode, the visualization engine renders an upright bass that follows the user. Along the fretboard of the bass there is also a colored bar with letters that denote the played note, in order to facilitate the interaction as before. A snapshot of this mode can be seen in Fig. 3b.

The Conductor (Two Hands) Interaction: In the conductor mode, where the user's movements could be abstractly likened with the movements of a conductor, each of the user's hands is assigned with one of the two instruments. The movements for triggering a sound event are the vertical movements of the two hands, which correspond to various pitches, namely, A3, B3, C4, D4, E4, F4, G4 and A4. When these vertical movements are combined with horizontal movements of the hands, then the volume of each instrument is altered, obtaining a higher volume when the two hands are further apart, while silencing the instruments when the hands are positioned close to the user's spine. In this mode, the user can actually "air-draw" with the hands, listen to consonant and dissonant musical intervals and generally experiment with the virtual music performance in a more engaging and creative way.

As in the previous modes, a colored bar with note names is shown vertically denoting which notes are played at each different height level. This mode can be seen in Fig. 3c.

Multiplayer Interaction: In order to enable collaboration between various players, the gesture interaction for music performance can allow the collaboration of two or more players. So in this mode, the users can either play virtually the same instrument, i.e., guitar, upright bass (or even the conductor mode), or choose to play the two different instruments simultaneously, as shown in Fig. 1.

The applications allows for more than two players to stand in front of the Kinect sensor, get assigned with an instrument and be instantly able to interact. The gestures for the performance are the ones previously named, depending on the instrument. A snapshot that shows two players playing the upright bass can be seen in Fig. 3d.

3 SYSTEM ARCHITECTURE

In this section, we describe the system architecture and the various methods that we employed for building the web application.

3.1 Server and Client Modules

The architecture of the application is shown in Fig. 4. As it can be seen, the application consists of two concrete modules: i) the server, which handles the data received from the Kinect v2 sensor, which are sent in an appropriate JSON format via a Websocket, and ii) the client, which runs in the user's browser and handles the visualization, the sound synthesis and the User Interface as a whole (e.g., user settings).

The server part is implemented in the C# language and leverages the Kinect v2 API, in order to receive skeletal information from

In Proc. Audio Mostly 2018: Sound in Immersion and Emotion (AM'18), Wrexham, United Kingdom, Sep. 2018

the Kinect at 30fps. The skeletal information is then converted to an intermediate JSON format, appropriate for transfer via a Websocket. Since we only transfer skeletal information and not the other Kinect streams, the created JSON has a negligible memory footprint and there is no bottleneck regarding the bandwidth of the user's connection. Consequently, there is no delay in transferring the data to the client, even in the case of multiple users playing simultaneously.

On the client side, the 3D world that depicts the user and the instruments is built using the three.js library[2], which provides, among others, a WebGL renderer for lightweight 3D drawing. Sound is generated via WebAudioFont[3] - a set of resources and associated technologies that uses sample-based synthesis to play musical instruments in browsers.

The client application, upon receiving the skeletal information, passes the information to both the sound synthesis engine and the visualization engine, which process the movements of the skeleton and output their result to the user's speakers and browser, respectively.

Through this architecture, a user can use the web application at home, by plugging a Kinect sensor in his computer, and running an executable (the server part) which connects to the browser and transfers the skeletal information. The application supports all major browsers that are capable of WebGL rendering.

3.2 3D Visualization Engine

The visualization engine maps the world coordinates (x, y, z) that are received for each skeletal joint directly to the joints of the 3D world avatar in the same metric units. Upon receiving new skeletal information, the engine checks if the skeleton is already in the 3D world and, depending on that, updates his/her joints or creates a new skeleton. If a user stops getting tracked, then the engine removes his skeleton from the world. As stated in Subsection 2.2, depending on the mode that is selected, the 3D visualization engine also renders semi-transparent virtual instruments, and overlaid colored bars with letters that denote the notes that are generated.

3.3 Audio Engine for the Virtual Instruments

In the sound synthesis engine, music generation is accomplished via the WebAudioFont library. The library employs the Web Audio API for playing music. The library includes an extensive catalog of instruments and soundfonts, two of which have been chosen for the web application: a Distortion Guitar and a Contrabass. Apart from playing only notes, the library allows playing full chords (several notes simultaneously), as well.

4 QUALITY EVALUATION AND USABILITY TESTING

All modes of the application were subjectively evaluated by 13 users, in terms of the usability, the gestural interaction, the intuitiveness of the performance, the visualization of the User Interface, the sound quality, and the enjoyability of the different modes. The majority of the users (10 out of 13) had at least second-hand experience with using a Kinect as a motion tracking tool, while about half of them

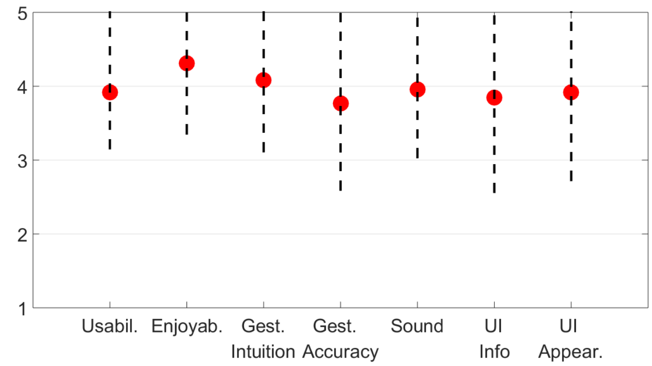


Figure 5: A mean – standard deviation plot of the users' scores for the overall usability and enjoyability of the application, the intuition and performance accuracy of the gestures, the sound quality, and the information provided by the User Interface, as well as its appearance. Results were provided in a five-point Likert scale.

were adequately familiar with playing a musical instrument, more specifically with playing the guitar.

First, a brief explanation about the interaction, the various modes and the predefined gestures was given and then the users were asked to perform in each mode for as long as they needed; with a total duration of 10-15 min. to be enough for each user to explore all modes. Thus, the users were able to interact with the different virtual musical instruments and evaluate the various modes, as well as play in collaboration with a second user. Afterwards, they filled a questionnaire, where they could express their opinion about various aspects of their interaction with the virtual instruments in five-point Likert scales (where 1 denoted that something was bad or difficult, not informative nor enjoyable and 5 denoted the maximum value for as for instance intuitiveness, informativeness, enjoyability etc.). In addition, they could also write down their suggestions for potential feature improvements. The results of these scores, obtained by the Likert scales, are presented in Fig. 5, in the form of a mean – standard deviation plot.

In general, the overall usability of the web-based application was highly rated (as can be also seen in Fig. 5), and the interaction with the virtual instrument was considered very enjoyable. The gestures, while being regarded as intuitive enough by the users, were deemed as being quite hard to accurately perform, especially in the cases of the bowing gesture of the upright bass, while some difficulties were also mentioned while trying to play neighboring guitar notes. We assume, as also stated by some users, that this challenge that was faced in terms of performance could be actually minimized with more practice. Even in physical musical instruments skilled players mature after several years of practice, so virtual musical instruments could also demand some practice, especially given that no actual physical contact with an "object" exists. Additionally, the users were satisfied with the sound quality of the instruments as well as with the visualization of the displayed information, since, as they commented, it provided great assistance towards a more enjoyable performance. Finally, the overall user interface received

[2]https://threejs.org/
[3]https://github.com/surikov/webaudiofont

really positive feedback, due to the real-time responsiveness and the minimal latency of the application.

Regarding the preferred interaction mode and the comparison between the three modes, the air guitar was overall the most popular, with a user noting that "it felt like I could play an actual song". On the other hand, the – more abstract – conductor mode was enjoyed by users who considered easier the gestures used for the sound generation. Additionally, the simultaneous support of two instruments, played in various volumes, was found to be less creatively restricting. The ability to accommodate collaborative playing with multiple players was also well-liked.

Finally, a popular suggestion among several users, was to increase the variability of musical output. More specific proposals included the addition of more instruments, like drums or keyboards, and the extension of the available pitch range of the existing instruments. However, we may note here that due to tracking limitations of the Kinect sensor not all instruments are suitable to be implemented for virtual interaction. Similarly, the extension of the range of the available notes would require more fine-grained movements and thus could lead to tracking inaccuracies. Concluding, we could point out that the gesture-based interaction with the virtual instruments was met with enthusiasm by all the users.

5 CONCLUSIONS

In this work, we present a web-based real-time application, that employs skeleton tracking, in order to facilitate gestural interaction between players and virtual musical instruments, via motions that resemble their physical counterparts. A subjective evaluation of the application with human users gave positive results, and verified that it can provide a creative and enjoyable experience in terms of music composition and performance for musical expression. In addition, since the application uses a web-based architecture, it is easily accessible. We also emphasize the fact that through the incorporation of visual aids in the 3D virtual world, even users that did not have any musical education were able to enjoy the interaction.

For future work, we intend to increase the number of instruments, developing for instance models for virtual interaction with i.e., drums or xylophone, enabling this way full virtual ensembles. Additionally, by extending the recognized motion templates/gestures for the various modes, we wish to improve the creative control for the users. In the same way, we wish to add more refined/detailed movements that give the user more control over the musical output, both by improving the user tracking, as well as employing more recent sensors. Finally, by further improving the visual and audio aids, we aim at achieving an interaction that can provide a step by step guidance and increase the educational aspect of the application, achieving heightened user experience and enjoyability.

ACKNOWLEDGMENTS

This work is supported by the iMuSciCA project that has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 731861.

We would like to thank all the students and colleagues at National Technical University of Athens, that participated in the subjective evaluation and for their valuable comments and discussions regarding the system's functionality for its improvement.

REFERENCES

[1] D. Brown, N. Renney, A. Stark, C. Nash, and T. Mitchell. 2016. Leimu: Gloveless music interaction using a wrist mounted leap motion. In *Proc. Int'l Conf. on New Interfaces for Musical Expression (NIME-16)*. Brisbane, Australia.

[2] G. Burloiu, S. Damian, B. Golumbeanu, and V. Mihai. 2017. Structured interaction in the SoundThimble real-time gesture sonification framework. In *Proc. Audio Mostly Conference (AM'17)*. London, UK.

[3] J.-M. Fernandez, T. Koppel, N. Verstraete, G. Lorieux, A. Vert, and P. Spiesser. 2017. GeKiPe, a gesture-based interface for audiovaudio performance. In *Proc. Int'l Conf. on New Interfaces for Musical Expression (NIME-17)*. Copenhagen, Denmark.

[4] M. Gleicher and N. Ferrier. 2002. Evaluating video-based motion capture. In *Proc. Computer Animation Conf. (CA-02)*. Switzerland.

[5] A. Hadjakos, T. Grobhauser, and W. Goeb. 2013. Motion Analysis of Music Ensembles with the Kinect. In *Proc. Int'l Conf. on New Interfaces for Musical Expression (NIME-13)*. Daejeon, South Korea.

[6] C. Honigman, A. Walton, and A. Kapur. 2013. The Third Room: A 3D Virtual Music Paradigm. In *Proc. Int'l Conf. on New Interfaces for Musical Expression (NIME-13)*. Daejeon, South Korea.

[7] M. H. Hsu, W. G. Kumara, T. K. Shih, and Z. Cheng. 2013. Spider King: Virtual musical instruments based on microsoft Kinect.. In *Proc. Int'l Joint Conf. on Awareness Science and Technology & Ubi-Media Computing (iCAST-UMEDIA-13)*. Aizuwakamatsu, Japan.

[8] T. Maki-Patola, J. Laitinen, A. Kanerva, and T. Takala. 2005. Experiments with Virtual Reality Instruments. In *Proc. Int'l Conf. on New Interfaces for Musical Expression (NIME-05)*. Vancouver, Canada.

[9] A. Mulder. 1994. Virtual Musical Instruments: Accessing the sound synthesis universe as a performer. In *Proc. Brazilian Symposium on Computer Music*.

[10] A. Rosa-Pujazon, I. Barbanco, L.J. Tardon, and A. M. Barbancho. 2015. A Virtual Reality Drumkit Simulator System with a Kinect Device. *Int'l. Journal Creative Interfaces Computer Graphics (IJCICG)* 6, 1 (Jan. 2015), 72–86.

[11] S. Trail, M. Dean, T. F. Tavares, G. Odowichuk, P. Driessen, W. A. Schloss, and G. Tzanetakis. 2012. Non-invasive sensing and gesture control for pitched percussion hyper-instruments using the Kinect. In *Proc. Int'l Conf. on New Interfaces for Musical Expression (NIME-12)*. Michigan, USA.

[12] Q. Yang and G. Essl. 2012. Augmented Piano Performance using a Depth Camera. In *Proc. Int'l Conf. on New Interfaces for Musical Expression (NIME-12)*. Michigan, USA.