Research Article

Bela Usabaev*, Anna Eschenbacher*, and Angela Brennecke*

The Virtual Theremin: Designing an Interactive Digital Music Instrument for Film Scene Scoring

https://doi.org/10.1515/icom-2022-0007

Abstract: This paper presents a first prototype of a virtual Theremin instrument for accompanying film scenes with sound. The virtual Theremin is implemented as a hybrid application for the web. Sound control is achieved by capturing user gestures with a webcam and mapping the gestures to the corresponding virtual Theremin parameters pitch and volume. Different sound types can be selected. The application's underlying research is part of the multimodal digital heritage project KOLLISIONEN which targets to open up the private archive of the Russian film maker Sergej Eisenstein to a broader public in digital form. Eisenstein, a film theorist and pioneer of film montage, was particularly intrigued by the Theremin as an instrument for film sound design. The virtual Theremin presented here is therefore linked to a film scene from the 1929 Soviet drama "The General Line" by Sergej Eisenstein which was never set to music originally. In its first implementation state, the application connects music interaction design with digital heritage in a modular, flexible and playful way and uses contemporary web technologies to enable easy operation and the greatest possible accessibility.

Keywords: user-centered computing, sound and music computing, scenario-based design, user interface design, web-based interaction, web applications

1 Introduction and Motivation

KOLLISIONEN is an interdisciplinary research project between the Film University Babelsberg and the Potsdam University of Applied Sciences. In the context of preservation of cultural heritage, it deals with the development of a multi-modal hybrid web and VR application around the Russian filmmaker Sergej Eisenstein. The central goal of the project is to make the achievements and theories of Russian film maker Sergej Eisenstein accessible to a broader public in digital form. Figure 1 illustrates the different research teams that collaborate in the project. The teams focus on data analysis and information visualisation (InfoViz team), virtual reality experiences (VR team), as well as sound integration (3D sound team). Apart from researching how to best link VR and InfoViz with 3D Sound, the 3D sound team is developing an interactive application example, the Theremin app presented here.

Our aim is to bring InfoViz, VR and 3D Sound together as part of one application prototype using multimodal combinations of the three fields and the application programming interfaces (APIs) for the web browser environment.

The virtual Theremin application is designed as a web-based audiovisual digital music instrument (DMI) for film scene scoring. As such, it is intended to be played and explored for sound and music creation by users who are interested in Eisenstein and his work.

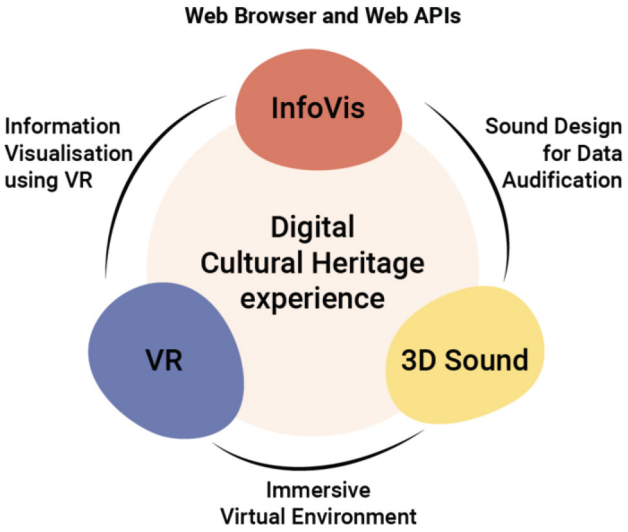

Figure 1: The three modalities of project KOLLISIONEN, and their multimodal combinations for the web-based interactive digital cultural heritage application domain.

***Corresponding authors: Bela Usabaev, Anna Eschenbacher, Angela Brennecke,** Filmuniversität Babelsberg KONRAD WOLF, Potsdam, Germany, e-mails: b.usabaev@filmuniversitaet.de, anna.eschenbacher@filmuniversitaet.de, a.brennecke@filmuniversitaet.de

The original Theremin instrument was developed by one of Eisenstein's contemporaries, Lev Termen in 1920 [15]. The instrument is played by moving both hands around two antennas without touch. The antennas are part of two radio frequency oscillator circuits, where the hands act as capacitive plates. The movement of the hands influences the electromagnetic fields of the two antennas. Movement towards the pitch antenna influences the pitch of the sound, towards the volume antenna the sound volume. The sound itself is produced via a loudspeaker that is connected to an electronic oscillator.[1] We refer to the sound control of pitch and volume simultaneously by hand movement as the Theremin interaction principle. Eisenstein's interest in the instrument adhered to the search for a tool for an unmediated creation of sound for an image. According to his contemporaries, he had planned to use the Theremin for unmediated musical ideation, i. e., to make musical sketches for the film score and soundtrack for his film the General Line.[2] The virtual Theremin application picks up on this.

2 Eisensteins Theory of Vertical Montage

Eisenstein has developed an influential theory of film montage, and has been one of the first formal thinkers of film montage. He has developed the theory of vertical montage and the audiovisual contrapoint [17].

Audiovisual media theory has been researched extensively by Michel Chion [4]. Chion explains the term audiovisual and how sound and image relate to each other in an audiovisual context, where sound and image together lead to more than the sum of its parts. One important aspect of Chions reasoning about the nature of the additional value is that in an artistic work the arrangement of sound and image generates a vacuum that can be filled by the audience, leading to a an experience of more than the components. Artistic researchers have heavily explored and cited Chions work and theory in their own work and research, for example [5].

In his book "Jenseits der Einstellung" [6] Eisenstein gives a detailed example of this vertical montage theory. The sketches remind of an orchestra score, with the image and the score music stacked over each other, while a third stream of information depicts an abstract graphical analysis between the music and the image. Figure 2 shows part of the cited example from Eisensteins book and highlights the described stacked structure.

The composition of the film scene is analysed graphically and analogies in form of graphical shape and form are transferred to the music. Although the units of the musical score differ from the means how composition of a film scene image is described, the graphical representation formed by lines and curves can link both as an abstract device and is reminiscent of a gestural form. The visual and emotional storytelling of the film scene expressed and encoded in terms of a visual composition of perceived lines, curves and line progressions is juxtaposed and transferred to the musical score via the abstract graphical representation. Such are for example the scenes of a prolonged line of force (the army) spanning the whole film shot, as depicted in column VI and VII in Figure 2, or the quietness before the impending battle, as depicted in column XII.

It is clear that the detailed analysis given in the book is post factual, and that at the time of writing there have been no readily available tools for such sketching of melody, and that this is a prototypical analysis example. From todays perspective the visual layout is very reminiscent of video editing software such as Premiere Pro[3] or Avid Media Composer.[4]

During his stay in Paris in 1928, Eisenstein had already started experimenting with graphical music, in terms animation on a film roll. At that time, space for a track of sound had been developed on film material, such that sound could be recorded alongside and synchronous to the image onto the film strip. Although it was only a singular experimental work with drawing on film conducted by Eisenstein, it shows the visual or graphical music approach, that Eisenstein has been exploring.[5]

Leon Theremin wrote in a memoire from 1983 about one encounter between Eisenstein and the Theremin instrument:[6]

> A new version of the Theremin was created for dancers. The sound was controlled by movement. Four dancers giving 4 polyphonic voices. Our laboratories were visited by the cinematographer Eisenstein. After having seen my students dance, he wanted the first demonstration of this new instrument to be held inside the U.S.S.R. I had decided to realise the idea when in 1939 I arrived in Leningrad.

1 https://www.Thereminvox.com/

2 Naum Kleiman Notes on Eisenstein.

3 https://www.adobe.com/de/products/premiere.html

4 https://www.avid.com/de/media-composer

5 Naum Kleiman Notes on Eisenstein.

6 https://www.thereminvox.com/stories/history/les-memoires-de-lev-termen/

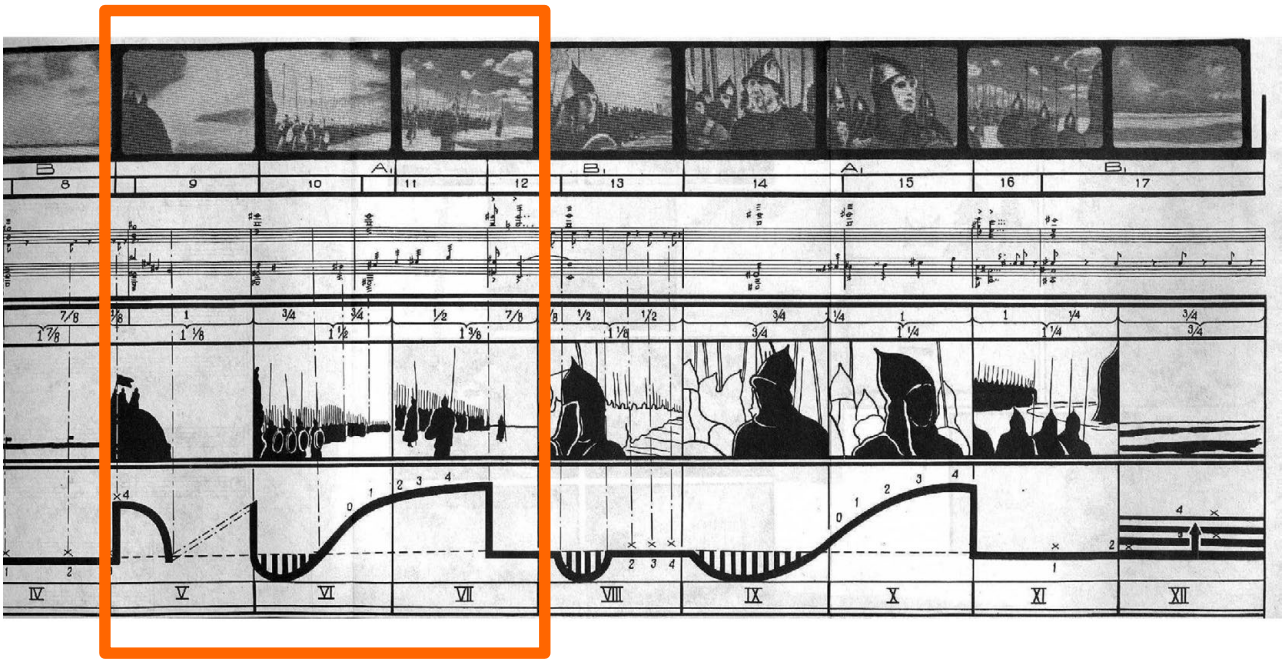

Figure 2: Excerpt from the example of vertical montage analysis described by Eisenstein in [6, 268–269] from his film "Alexander Nevsky". The highlighted area shows the vertically stacked nature of image, sound and an abstract graphic representation as outcome of an analysis of the graphic composition of the image of the film scene and the corresponding musical score.

For his film the General Line Eisenstein had been planning to use the Theremin in order to develop the sound track. The Theremin voice had been aesthetically planned to be used at the end of the film, showing the soviet industrial era, which at that time still was a hazy dream and not unlike a science-fiction future.

In the context of the project we are picking up on the idea of using the Theremin for scoring a famous scene from the film the General Line. In the context of approaching a personal archive for the purpose of digital cultural heritage, we are employing the extended web paradigm, and the current technologies being developed for it, WebXR.

3 Related Work

Although there is a range of sound software and tools for making sound and music, these are usually menu or graphical user interface (GUI) based and are designed for expert users. Interactive tools for sound and music ideation can be found in the field of interactive (web-)applications for sound and music and digital music instrument (DMI) design. Here, exploratory interactive ideation tools for sound are often gesture and movement based [2, 8].

The IVA tool [3] is an interactive audiovisual ideation tool that uses movement and webcam input. Using optical flow and frame by frame feature vector based transformation of image to sound, a correlation between image and sound can be explored and possible sound spaces found. The tool however is not web-based, and the audiovisual mapping technique is not based on Theremin interaction established in the current application. We are using webcam input of users movements in order to apply it within the Theremin interaction principle. A gesture-based ideation application for sound design is presented in [13]. Here, eight different types of target sounds are first crafted by a sound artist as part of an interactive installation. Based on the possible hand-crafted sound categories a feedback loop between the user and the machine learning algorithm based on gesture and speech sound is established, in order for the user to arrive at a target sound iteratively by interaction with the application. The feedback-based workflow of this application is interesting for our sound ideation workflow scenario, however the application is not web-based, and is designed for a range of specific target sounds. In this application gestures are used as a means to illustrate sound, and provide the features from which to iteratively synthesize the target sound. This application is not set in the digital cultural heritage domain and deviates from our design approach of implementing

an analogue instrument virtually. We do not map sounds to movement in order to arrive at pre-defined target sounds but aim for a generative rather than a gesture recognition approach for sound.

A range of web-based interactive tools for musical and sound exploration tied to a visual feedback or visual sound representation on the screen can be found on the platform experiments with Google.[7] The experiments use the webcam to track the user and to playback sounds. The platform is an important source of inspiration with respect to interaction to sound mapping techniques.

The Theremin sound control principle has been used as interaction paradigm by the creative coding community[8][9] and interactive Theremin applications as a web-based app have been implemented.[10]

In the virtual space the Theremin has been implemented [9], for example, for assistance purposes when teaching the instrument, and as web-based VR Theremin instrument[11] requiring a Head Mounted Display (HMD).

In the field of DMI design gesture-based interfaces based on Theremin interaction have been implemented, as described in [20].

Interactive installations are experiences that bridge the field of interactive applications and DMI and can employ Theremin inspired interaction [12]. These are heavily influenced by audience behaviour and thus are tailored to the aspects of interaction as a vital part of the installation. Lastly digital cultural heritage (DCH) applications use archive artefacts to implement sound and audiovisual interaction functionality in gamified terms [16]. DCH applications provide an application scenario and therefore can harness gamification, storytelling or narrative as tools for interaction design. We use the Theremin as archive artefact and a visual representation for the virtual music instrument.

4 The Virtual Theremin Application

As part of the web-based digital cultural heritage (DCH) application prototype the conceptualisation of the virtual Theremin moves between sound interaction interface, digital music instrument (DMI) development and an interactive audiovisual web application. Human computer interaction (HCI) considerations with respect to interactivity, ease of access and intuitive use as well as the enjoyment of the generated sound and the framing visualisation play an important role in the design process of the sound ideation tool. As we are developing the application within the digital heritage context and aiming at an audience interested in Eisenstein and his work, a main design consideration for user interaction interface development is the wide accessibility of the application without the requirement of specialized hardware such as a proprietary HMD.

In the implementation of the virtual Theremin we pursue a hybrid approach that connects the user to a virtual space that includes the film scene. To implement this, we work with tools in the extended web paradigm.[12] It should be possible to easily change the visual web environment the DMI is used in, in order to draw connections to the VR and InfoViz application domains. Thus, the DMI should be used in a 2D or 3D InfoViz as well as in a web-based VR environment. Figure 3 shows the layers of the web application and how the different parts are connected by the extended web paradigm, but decouple visualisation from the DMI.

4.1 Movement-Based Music Interaction and DMI Design

Acoustic instruments are a type of musical instruments that allow the user to perform directly on the sound-producing mechanisms in a tangible way. This interaction type is the main difference when compared to digital musical instruments (DMIs), which are based on control systems that capture the user's intention through input sensors and translate it into parameters to the digital synthesizer [11]. Tangibility can be defined as the capability of being touched.

In the particular case of intangible DMIs, which usually require mid-air gestures to produce sound, building DMIs that are appealing to the user becomes a complex task due to the loss of intimacy between the user and the instrument.

The lines between interface and interaction blur in interesting ways at times, as is stated by „designing interaction, not interfaces“ ([2], [1]). Here also the term embodied listening and situated interaction are introduced, referring to contextualized and sensory-motor feedback-based action-sound loops, that point to questions of learning how to play a DMI.

7 https://experiments.withgoogle.com/

8 https://p5js.org/

9 https://programminginarduino.wordpress.com/2016/03/02/project-06/

10 https://Theremin.app/

11 https://www.makuxr.com/simple-Theremin

12 https://immersive-web.github.io/webxr/

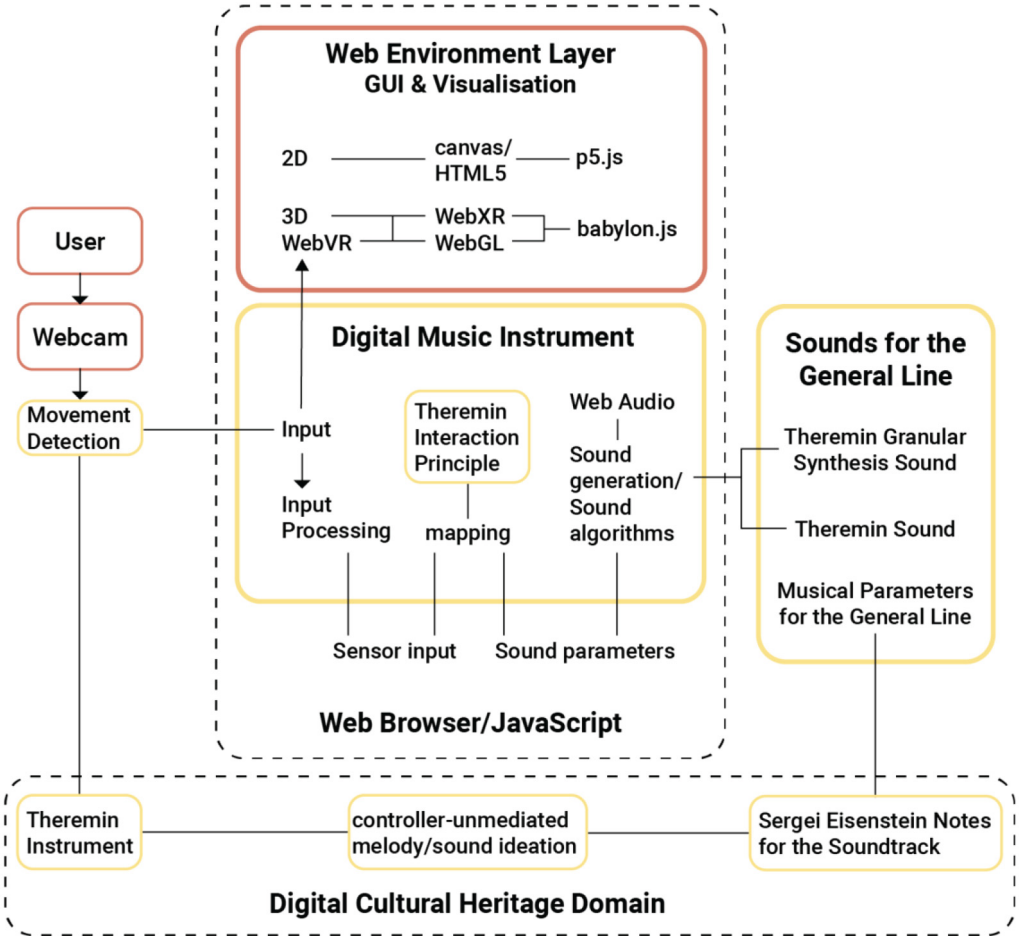

Figure 3: Interactive web application and digital music instrument architecture employing the Theremin interaction principle to link the sound and the visual environment. The DMI can be used in different visual web environments without changing the sound components. We are using p5 JS to develop a 2D web interface and Babylon JS for the 3D web interface respectively. We use the same DMI with both interfaces, and adapt the current mapping function to adhere to the WebGL projection dimensions of the different web environments.

The music instrument that we are developing is assumed to be played, explored, and possibly used for musical creation by users who are interested in Eisenstein and his work.

An important aim of the sound interaction is to encourage the user in a playful use of the tool and to encourage exploratory listening that can provide a positive experience with the sound algorithms [18]. The involvement and user engagement paradigm of this objective roots in the creation of performative installations where a user is engaged without having to be an expert of the tool [14].

DMIs are modular applications for creating and performing music [23], consisting of a sound generation unit, a user interface or control unit and a mapping technique. In this way DMIs are modular applications where the input processing part and the sound synthesis or generation part can be developed independently. The mapping technique of a DMI determines how sensor inputs and interaction are mapped onto sound parameters [2]. In the web-based virtual Theremin we are using the Theremin interaction principle as the basis for the mapping technique. The DMI design paradigm is part of the virtual Theremin application as shown in Figure 3.

As we are implementing the Theremin interaction principle in the virtual space, we are decoupling sound generation from sound control, as defined in [23]. The sound parameters are controlled by the positional data of the hands according to the Theremin interaction principle, however different types of sound and sound generation is produced by and encapsulated as sound algorithms and implemented using the WebAudio API.

In our current setup the film scene and the DMI are juxtaposed in the virtual space and not yet connected on implementation level. The link between the DMI and the film scene are solely the users movements. In this sense

the web application emphasizes sound exploration and focuses on user engagement.

4.2 Web Application Layers of the Virtual Theremin

The web application consists of a web environment layer that determines the visualisation of the virtual space and interaction, as well as a DMI that determines how the application sounds. An interaction layer between these two is implemented by the Theremin interaction principle. As a storytelling element of the web application the Theremin plays a central role, where the choice of the instrument is motivated by the archive domain of Sergej Eisenstein and is an archive artefact employed for storytelling focus and link to the digital cultural heritage web application. The overview of the web layer architecture is shown in Figure 3.

As can be seen in the Mozilla web development specification,[13] all APIs for the browser are based on Javascript, and thus are part of one programming environment. This is reflected in Figure 3, as all three components, sound, visualisation and interaction are implemented using Javascript. The visual part of the application is developed using the WebGL[14] and the WebXR APIs respectively. WebGL is the implementation of computer graphics functions in the browser, WebXR is a framework for the development of virtual augmented and mixed reality web applications, supporting different standardized and custom input devices.[15] Babylon JS is a web-based 3D game engine that implements these APIs, and implements access to the WebAudio API inside the game engine[16] The sound algorithms can be employed in different visual web environments, such that the same sound algorithms can be employed whether the visualisation is implemented in 2D using a p5 JS canvas element or a 3D virtual scene using Babylon JS. As both WebXR and WebAudio use positional data of the hands, the positional data is linking sound and visual information inside the app, without linking both directly. Figure 4 shows how user hand movement is used within the application. It shows how the positional data is mapped to sound parameters and used by the sound algorithms for interactive sound generation. The module to acquire positional data is external to the DMI and the visualisation, making us free to experiment with different acquisition approaches.

4.3 The Virtual Theremin Web-Based DMI

The virtual interactive Theremin is based on the analogue Theremin which is controlled by hand movements. The analog Theremin has two antennas one of which regulates the volume of a sound, the other the pitch. Thus an electronic sound can be played with varying frequency and varying volume. We term this principle of the Theremin to control several sound parameters connected to sound generation by different hand movements as the Theremin interaction principle.

As we are developing the Theremin in a virtual space, different sound parameters can be assigned to the antennas so that a more extensive sound spectrum can be created. The control of sound parameters is organized into sound algorithms, which function as sound modules. A range of different sounds can be created that are not bound by the physical characteristics of the analog Theremin. In this way we can create a digital music instrument that has diverse and parametric sound characteristics.

The sound algorithms are controlled by arm movement using the Theremin interaction principle and movement detection. The sound algorithm currently being controlled can be selected via a GUI. This setup of the application makes it possible to control several sound algorithms with the same movements at the same time. The mapping of user input to sound parameters specifies how the movements control the sound algorithm parameters and therefore the generated sound. Figure 4 shows how the positional data from user interaction is used within the web application and how the different parts of the web application are linked by user interaction.

The web-based virtual Theremin application consists of a visualisation, a sound algorithm implementation and user interaction implementation part. The visualisation parts of the web application are the visualisation of the Theremin, the visualisation of the 3D space where the film scene to be scored is located, the visualisation of the interaction with the virtual Theremin, and the interface for the selection of the sound algorithms.

The sound algorithms are implemented using the WebAudio API.[17] The visual environment and visualisation is implemented using the WebXR API.[18] Both APIs use the

13 https://developer.mozilla.org/en-US/docs/Web/API

14 https://developer.mozilla.org/en-US/docs/Web/API/WebGL_API

15 https://immersive-web.github.io/webxr/

16 https://www.babylonjs.com

17 https://www.w3.org/TR/webaudio/

18 https://immersive-web.github.io/webxr/

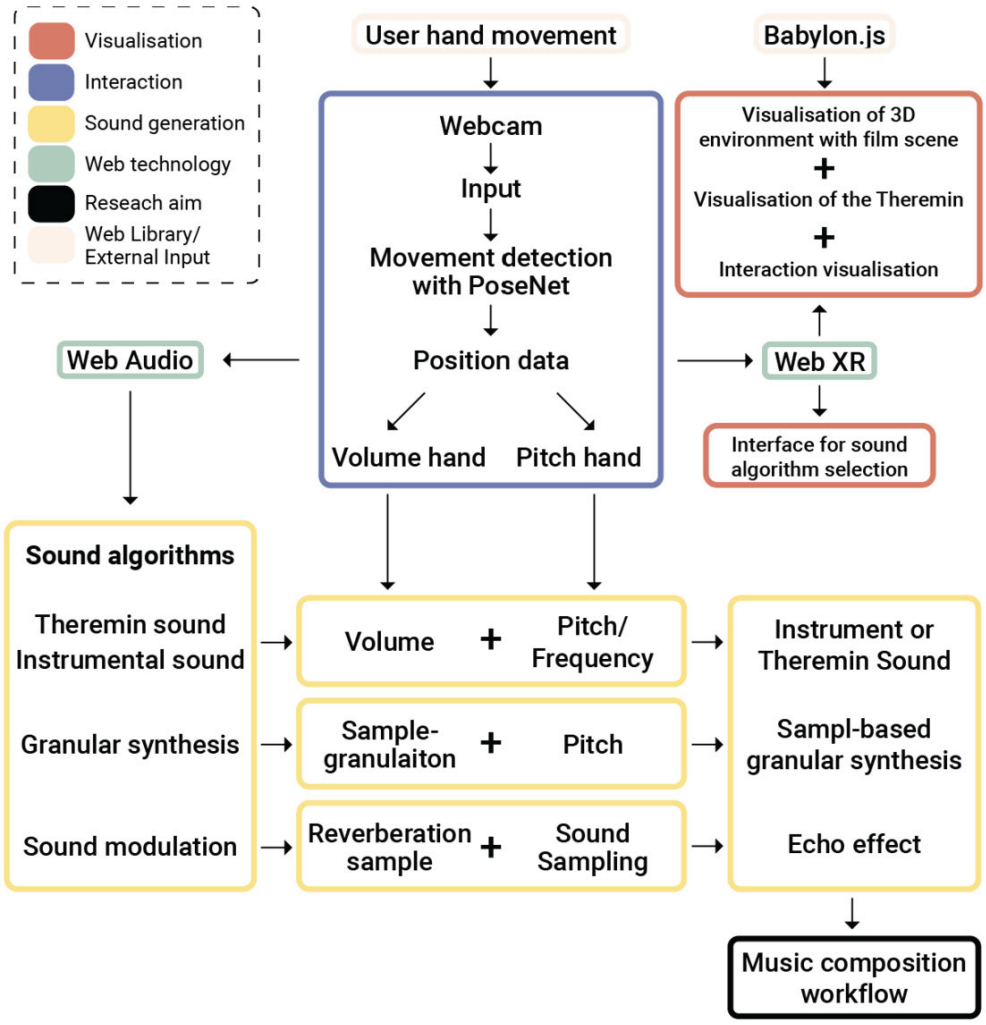

Figure 4: The virtual Theremin interactive web application and the sound algorithms of its digital music instrument (DMI) architecture. The Figure shows how positional data is used by the different parts of the web application. The user interaction links the visualisation and the sound generation part of the web application.

positional data from movement detection. In the current version we have two possible sound algorithms, the classical Theremin sound and a sample-based Theremin granular synthesis algorithm. The Theremin sound provides a very prominent leading voice, also often used in this role in musical compositions. The granular synthesis algorithm is an approach for the creation of atmospheric ambient soundscapes. A longer term goal of the web application is to integrate the sound algorithms as part of an interactive music composition workflow for the film scene. The virtual Theremin prototype web application is available online.[19]

4.3.1 The Visual Web Environment

The virtual space is implemented in the web browser using the Babylon JS[20] web library. The user is presented a virtual room view containing the film scene from the General Line and a 3D rendered virtual Theremin instrument. By means of the webcam image the user can interact with the visualized Theremin instrument. The sound is produced by the DMI. The movement of the hands is visualized by trails of particles following each of the hands, as shown in Figure 5. The visual feedback of the interaction to the user is a realtime animation in the browser, as depicted in the image sequence on the bottom of Figure 5.

19 https://ctechfilmuniversity.github.io/eisenstein-theremin-app/

20 https://www.babylonjs.com

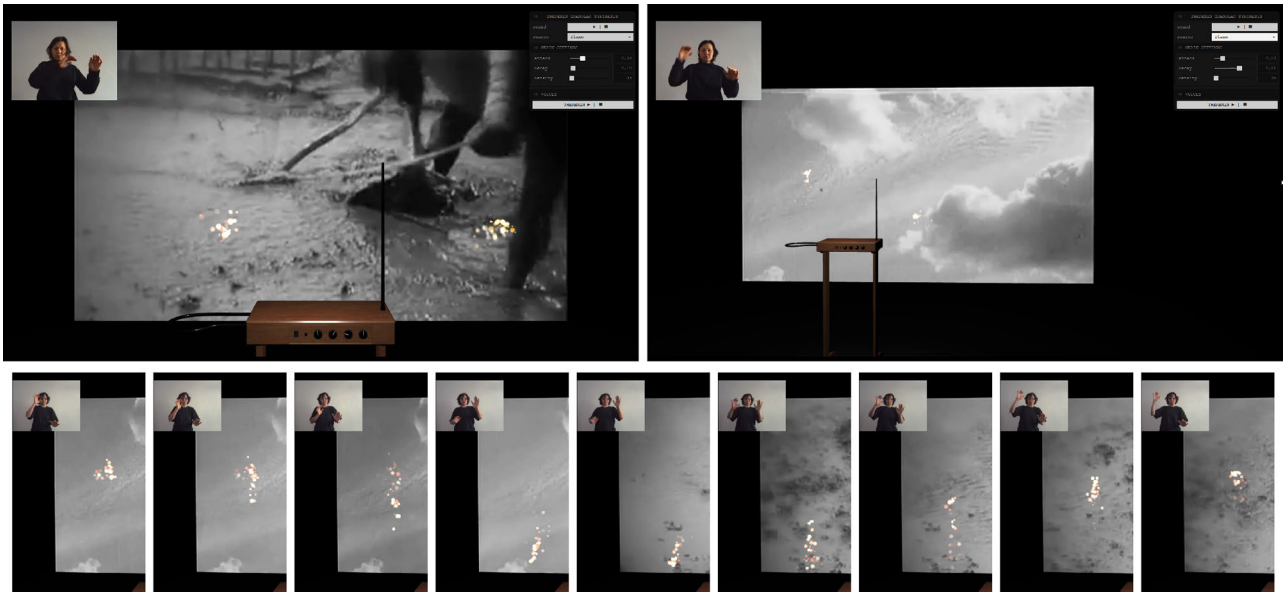

Figure 5: The Figure depicts two views of the virtual Theremin digital music instrument web application. It shows how a user is interacting with a scene from the film "The General Line" by Sergej Eisenstein. The Figure shows the webcam image of the user in the left upper corner, the GUI with the Theremin sound parameters and sound selection in the right upper corner. The film scene is played back in the background with a visualized Theremin instrument graphic in the foreground. The movement of the hands is visualized by a group of particles for each hand. The image sequence on the bottom illustrates the particle visualisation of the hand on the left side of the webcam image.

4.3.2 Connection and Interaction

To interact with the web-based Theremin DMI, the user has to be able to control the two parameters frequency and volume. With an analogue Theremin, these parameters are controlled by hand movement, so that an electronic sound can be played with varying frequency and varying volume. In previous digitized versions, hardware components such as game controllers, fiducial markers, and infrared sensors were used to track the user's motion [7]. Position tracking is becoming increasingly more accessible with advances in computer vision. By using machine learning libraries, no hardware equipment is needed to track a users position, aside from a camera. Thus, the user is free to move around and not constrained by any physical equipment. Being in a web-based environment, our Theremin interaction implementation uses movement detection based on a convolutional neural network with the pre-trained machine learning model PoseNet [10]. It is open-source and part of the ml5.js library based on Tensorflow.js. PoseNet returns an array of 17 key points [10]. These refer to various body parts of the user, such as nose, right and left eye, and both wrists. We use both wrist key points to control the Theremin. Each key point consists of an x and y value, describing the position within the webcam image. We then scale and map the values for the right wrist to a frequency variable. When lifting the right wrist, the frequency value increases and vice versa. The values for the left wrist key point are scaled and mapped to the overall volume of the electronic oscillator. By lowering the left hand, the volume decreases. The maximum volume is reached, when the left hand is positioned at the highest point of the webcam image. This enables the user to play an electronic sound with varying frequency and varying volume according to their hand movement. Aside from the hand interaction, the user is able to control various sound parameters with the help of a GUI on the website.

4.3.3 Sound Generation in the Web Browser

The sound algorithms of the web-based virtual Theremin are implemented using the WebAudio API as the sound synthesis framework in the browser [24].

The WebAudio API is an open source, node-based application programming interface in Javascript [22]. It is similar to flow-graph and node-based sound synthesis systems such as MAX/MSP and Pure Data and node-based sound editors of audio engines in 3D and game development environments such as Unity 3D or the Unreal engine [21]. Thus the WebAudio API is the audio engine counterpart of the web-based game engine Babylon JS and can specifically be used to develop generative sound al-

gorithms and procedural sound behaviour in a web-based gamified 3D environment.

In the browser the WebAudio API can be used to implement sound generation algorithms, but also to implement sound spatialization. Sound does not have to be implemented as part of a virtual 3D space in order to behave spatially in the web browser, as it is directly addressing the loudspeakers connected to the users sound card and thus can emulate 3D sound in a 2D visual web environment.

We tie the sound generation to the position of the virtual Theremin in the virtual environment to adhere to the storytelling purpose of the web experience through the archive artefact, however sound generation is not bound by the virtual scene and different mapping axes can be found, that are not physically motivated in the virtual space.

In the current implementation we are using the node-based flow-graph structure of the WebAudio API for sound generation based on sound parameters generatively controlled by the user.

4.3.4 Sound Algorithms in the Virtual Theremin DMI

In a virtual environment, different sound parameters can be assigned to the antennas of the Theremin to create a more extensive sound spectrum. Using the GUI the user can select or deselect and thus interact with several algorithms at the same time using the same hand movements. The mapping of user input to sound parameters specifies how the movements control the sound algorithms. The sound algorithms can be considered as sound modules, acting upon sound algorithm parameters controlled by user interaction. Figure 6 shows how the hand movements are mapped to the granular synthesis algorithm and how they control the granulation of a sound file.

Granular synthesis is a generative sound synthesis algorithm. A sound sample is cut into tiny sound slices of a length of 100 ms or less, rendering the sound file as a set of sound grains. The larger length of the grains retains the sound characteristic of the original sound sample, while the shorter the grain length, the less the original sound can be recognized. Granular synthesis thus can be used as a means for sound design where different levels of sound timbre can be experimented with.[21] Granular synthesis can be used to change the playback length of a certain sound, by concatenating sound grains sampled from it, retaining the original pitch characteristic of the sampled sound. As we use granular synthesis in a generative synthesis setup, we are constantly sampling the sound file, depending on the users movements. The change in the playback rate of the individual grains also changes the overall characteristic of the synthesized sound.

In the case of the classic Theremin sound, the users hand movements induce similar behaviour as with an analogue Theremin instrument, as described in Section 4.3.2. The movement around the right hand or pitch antenna influences the pitch of the electronic oscillator sound while the movement up and down over the left hand or volume antenna influences the volume of the produced sound.

The position of the hands and the position of the Theremin instrument in the virtual scene are used for the calculation of sound parameters. The data is interpreted according to the Theremin interaction principle as shown in Figure 4.

5 Results and Discussion

In order to assess the web application and the DMI we are conducting a user survey where a group of users is asked to test the web application and to answer a set of 14 questions according to an even valued Lickert-type scale measuring 6 values of agreement.

The questions are targeted towards rating the juxtaposition of sound interaction and video as part of the web interface, the web interface itself and the sound interaction. These are as follows:

- The visual rendering of the web interface was easy to understand
- The menu layout of the web interface was easy to navigate
- It was easy to control sound by hand movements
- The sound control by hand movements led to interesting sounds
- I wished more control over the sound while moving
- The sound control by hand movements and displayed video were interesting together
- The sound control by hand movements was distracting me from the video
- The sound control by hand movements and displayed video together were confusing
- The behaviour of the sound was fitting the video
- The behaviour of the sound was distracting from the video
- I focused more on the sound interaction than on the video
- I focused more on the video than on the sound interaction

21 https://monolake.de/technology/granulator.html

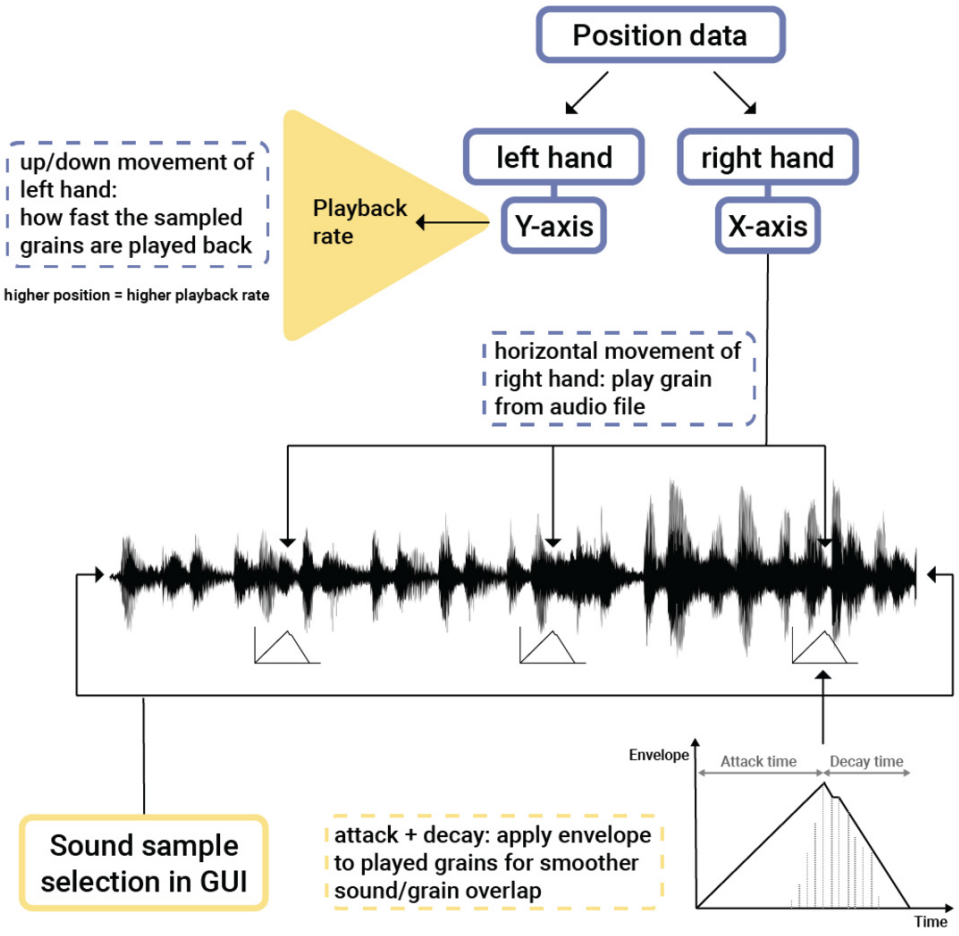

Figure 6: The Figure shows a synchronous granular synthesis algorithm based on granulation of a sound file selected via the GUI of the web application. Horizontal movement of the right hand determines from where in the selected file the sound grains are sampled, while the height of the left hand determines how fast the individual grains are played back, hereby determining the pitch of the overall synthesized sound. The traditional movements for the Theremin are mapped here such that a granular synthesis algorithm can be controlled using the same hand movements.

- I engaged with the video scene more because I played the sound by movement
- I tried the same movements to produce similar sounds

The preliminary evaluation confirmed the first impression of the application. The web interface is simple, but it can take time to get oriented with the hybrid nature of perceptual feedback from the webcam, sound, and particle visualisation. Sound interaction is interesting, but it seems to take time for interest and playfulness to be initiated. Once sound interaction is approached, the users seem to be primarily concentrated on the movement interaction with the sound algorithms.

It is not yet clear how the video and the sound interaction are correlated. The video was not stated to be hindering engagement with sound interaction. The sound interaction itself was perceived as somewhat limited, as users agreed that more control would be desirable, but the interaction was confirmed as easy to use and not distracting.

The juxtaposition of video and sound was agreed as slightly confusing, and the sound not well fitting to the video. Here an inventory of sounds is important in order to keep the users motivated in the task. On the other hand, the sound algorithms might be too abstract and too general for the task at hand and may need to be suggested for use for a certain target task connected to the video. The scope of sound interaction application should be tied to specific parts of the video, allowing for the option of non-applied sound interaction for exploration. All survey participants have agreed to have tried the same movements to produce similar sounds.

Eisenstein has been a pioneer on the audiovisual theory of film montage. His vertical and polyphonic montage theory draws connections between the musical sound-

track and the visual image, coining the term the film and the sound image. The correspondence is described in [6] in form of corresponding geometric shapes. It becomes clear, that such a detailed analysis is post-factual, and at Eisensteins time there were no readily available tools for quick ideation processes.

The design of user interaction of a DMI is based on both human computer interaction research and musical interaction design, with a relevant distinction between a musical gesture as opposed to user movement in the general sense of interaction [8]. Although we follow the original Theremin instrument interaction schema in our implementation, the distinction between a musically meaningful gesture and user movement opens up a fruitful space between an interactive and a musical application, and allows for varied interpretations and use of gesture in our implementation.

The Babylon javascript implementation of the WebGL and WebXR API gives the possibility to add an input device interface and utilize different controllers. Thus the core algorithms of the sound generation unit can be used with different controllers in different web environments that can be realized by employing the Theremin interaction principle. By using VR controllers of an Oculus Quest, we can adapt the application to the web VR environment without changing the sound component. While evaluating the virtual Theremin internally, we have started to develop a 2D interface for the virtual Theremin application. We have found that for the task of sound creation and interaction, a simple 2D web interface may be more suitable than a web VR or web 3D interface, which is more immersive. This finding is still to be confirmed in a larger user evaluation. Our main proposition is that the flat interface may correspond more to the film or moving image viewing habits of users in general, and may be more suitable for a web-based application for film scene scoring.

The web 3D environment however can be populated easier by further storytelling elements and is more suitable to a web experience oriented application, that involves archive artefacts to be revisited or reenacted.

6 Future Work

In order to link the visual and the sound layer together in a more determined way, we aim to develop a sound composition workflow, where the visualisation of the sound interaction and the sound interaction itself can jointly be used as an audiovisual tool to support the interaction process. In order to structure the interactive process a timeline functionality will be implemented. One of the major challenges for the development of the DMI will be the integration of both functionalities into the interactive process developed so far. Currently the tool is aimed at sound exploration and is not tied to the film scene on implementation level.

The current mapping method is still a rough approach to a physical Theremin. One next step would be scaling the positional data algorithmically according to the physics of an actual Theremin to replicate the physical behavior of the variable capacitance as described in [19]. This would ensure a more realistic representation of the analog Theremin.

As a next step towards the communication between the web environment and the DMI, we will evaluate the relation between a 2D and a 3D virtual web environment with respect to the task of film scene scoring.. Here we aim to work more with the virtual projection capabilities of the WebGL library.

With respect to sound generation, similarly to the experiment in [13] we aim to include hand-crafted musical theme and melody compositions to support the interactive sound exploration process.

As we are developing a sound ideation application with the characteristics of a tool, an evaluation procedure in form of a workshop in an artistic research paradigm is a desirable way to evaluate and test the functionality and scope of the virtual Theremin.

The overall development aim is to integrate the DMI in the InfoViz and VR application environments according to the storyline of the digital archive application.

7 Conclusion

We are placing an interactive application for sound generation and manipulation in an extended web environment. Situated in the digital cultural heritage application domain the virtual Theremin is aiming to employ the web experience characteristics of such cultural domain applications, here by implementing the analogue Theremin and its interaction principle in virtual form. As we are developing a digital music instrument in the virtual space using the WebAudio API, a large spectrum of sound implementation possibilities is available. Here a range of possible mapping techniques between sensor input and sound generation opens up, with the possibility of linking the extended virtual web environment with the DMI. Our current application architecture is developed in a modular form, and is applicable to the 2D as well as 3D visual web environment.

The web application though based on the private archive of the filmmaker Sergej Eisenstein is using and combining contemporary web technologies as part of a multi-modal framework. The virtual Theremin application is modular, scalable and adaptable to different kinds of audiovisual contents, with the sound inventory adaptable and extendable to the task at hand, preserving the interaction mechanism of the DMI and the visualisation of the web environment for any kind of audiovisual content. The web environment layer can be changed independently of the DMI, with the mapping being the module to be enhanced if further data is available in case of differing movement sensor input. Movement detection gives the user access to the rendered scene, where the user can play with video and sound.

Acknowledgment: KOLLISIONEN is a research project between Film University Babelsberg and Potsdam University of Applied Sciences funded by the European Regional Development Fund and Land Brandenburg. We would like to thank our project partners at Filmuniversity Babelsberg Tatiana Brandrup and Katrin Springer with the VR team as well as our project partners at Potsdam University of Applied Sciences Prof. Dr. Marian Dörk, Sara Akhlaq and Arran Ridley with the InfoViz team. Last but not least sincere thanks to film historians Naum Kleiman and Vera Kleiman.

Funding: Project KOLLISIONEN is funded by the STaF programme within the European Regional Development Fund (ERDF) of the European Union, grant No. 85037983.

References

[1] Michel Beaudouin-Lafon. Designing interaction, not interfaces. pages 15–22, 01 2004.

[2] Frédéric Bevilacqua and Norbert Schnell. From musical interfaces to musical interactions. In Andrea Gaggioli, Alois Ferscha, Giuseppe Riva, Stephen Dunne, and Isabelle Viaud Delmon, editors, *Human Computer Confluence. Transforming Human Experience Through Symbiotic Technologies*, Lecture Notes in Computer Science, pages 125–140. De Gruyter Open Ltd, Warsaw/Berlin, 2016.

[3] Angela Brennecke, Markus Traber, Simon Stimberg, and Björn Stockleben. Interactive image driven sound. In *Proceedings of the Conference on Mensch Und Computer*, MuC '20, page 85–89, Association for Computing Machinery, New York, NY, USA, 2020.

[4] Michel Chion. *Audio-Vision. Sound on Screen*. Columbia University Press, 1994.

[5] Tadej Droljc. *Composing with isomorphic audiovisual gestalts*. Doctoral thesis, University of Huddersfield, 2018.

[6] Sergej Eisenstein. *Jenseits der Einstellung: Schriften zur Filmtheorie*. Suhrkamp Taschenbuch Wissenschaft, Suhrkamp, 2006.

[7] Christian Geiger, Holger Reckter, David Paschke, and Florian Schulz. Poster: Evolution of a theremin-based 3d-interface for music synthesis. In *2008 IEEE Symposium on 3D User Interfaces*, IEEE, mar 2008.

[8] Simon Holland, Katie Wilkie, Paul Mulholland, and Allan Seago. Music interaction: Understanding music and human-computer interaction. In Simon Holland, Katie Wilkie, Paul Mulholland, and Allan Seago, editors, *Music and Human-Computer Interaction*, Springer Series on Cultural Computing, pages 51–74. Springer, London, 2013.

[9] David Johnson and George Tzanetakis. Vrmin: using mixed reality to augment the theremin for musical tutoring. In Cumhur Erkut, editor, *17th International Conference on New Interfaces for Musical Expression, NIME 2017, Aalborg University, Copenhagen, Denmark, May 15–18, 2017*, pages 151–156. nime.org, 2017.

[10] Alex Kendall, Matthew Grimes, and Roberto Cipolla. Posenet: A convolutional network for real-time 6-dof camera relocalization. In *Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV)*, ICCV '15, page 2938–2946, IEEE Computer Society, USA, 2015.

[11] Mark Marshall, Max Hartshorn, Marcelo Wanderley, and Daniel Levitin. Sensor choice for parameter modulations in digital musical instruments: Empirical evidence from pitch modulation. *Journal of New Music Research*, 38, 12 2009.

[12] Christian Mayer, Patrick Pogscheba, Dionysios Marinos, Björn Wöldecke, and Christian Geiger. An audio-visual music installation with dichotomous user interactions. In *Proceedings of the 11th Conference on Advances in Computer Entertainment Technology*, ACE '14, Association for Computing Machinery, New York, NY, USA, 2014.

[13] Stefano Delle Monache, Davide Rocchesso, Frédéric Bevilacqua, Guillaume Lemaitre, Stefano Baldan, and Andrea Cera. Embodied sound design. *International Journal of Human-Computer Studies*, 118:47–59, 2018.

[14] Hye Yeon Nam and Michael Nitsche. Interactive installations as performance: Inspiration for hci. In *Proceedings of the 8th International Conference on Tangible, Embedded and Embodied Interaction*, TEI '14, page 189–196, Association for Computing Machinery, New York, NY, USA, 2014.

[15] Pavel Nikitin. Leon theremin (lev termen). *IEEE Antennas Propagation Magazine*, 54:252–257, 10 2012.

[16] Cristina Portalés, João M. F. Rodrigues, Alexandra Rodrigues Gonçalves, Ester Alba, and Jorge Sebastián. Digital cultural heritage. *Multimodal Technologies and Interaction*, 2(3), 2018.

[17] Robert Robertson. *Eisenstein on the Audiovisual: The Montage of Music, Image and Sound in Cinema (KINO – The Russian and Soviet Cinema)*. I. B. Tauris, 2011.

[18] Raquel Rodríguez-Carvajal and Oscar Lecuona de la Cruz. Mindfulness and music: A promising subject of an unmapped field. *International Journal of Behavioral Research & Psychology*, 2(3):27–35, 2014.

[19] Kenneth D. Skeldon, Lindsay M. Reid, Viviene McInally, Brendan Dougan, and Craig Fulton. Physics of the theremin. *American Journal of Physics*, 66:945–955, 1998.

[20] Alexei Sourin and Zhong Cai Chock. Playing digital music by waving hands in the air. In *Proceedings of the International*

Workshop on Advanced Image Technology (IWAIT), volume Proceedings Volume 11049, Singapore, 2019.
[21] Richard Stevens and Dave Raybould. *Game Audio Implementation: A Practical Guide Using the Unreal Engine*. Routledge, 2015.
[22] Rex van der Spuy. *Sound with the Web Audio API*, pages 331–368. Apress, Berkeley, CA, 2015.
[23] M. M. Wanderley and P. Depalle. Gestural control of sound synthesis. *Proceedings of the IEEE*, 92(4):632–644, 2004.
[24] Lonce Wyse and Srikumar Subramanian. The Viability of the Web Browser as a Computer Music Platform. *Computer Music Journal*, 37(4):10–23, 12 2013.

Bionotes

Bela Usabaev
Filmuniversität Babelsberg KONRAD WOLF, Potsdam, Germany
b.usabaev@filmuniversitaet.de

Bela Usabaev received her masters degree in computational linguistics from the University of Tübingen with a focus on statistical speech synthesis and machine learning. She is pursuing her diploma in media arts at the Academy of Media Arts in Cologne with a focus on sound environments and is a research associate at the Film University Babelsberg KONRAD WOLF.

Anna Eschenbacher
Filmuniversität Babelsberg KONRAD WOLF, Potsdam, Germany
anna.eschenbacher@filmuniversitaet.de

Anna Eschenbacher is a student and research assistant in the master's program Creative Technologies at Film University Babelsberg KONRAD WOLF in Potsdam. Her research focuses on interactive media for artistic data visualisations and mapping between sound and visual parameters. She works as a visual artist and creative coder based in Berlin.

Prof. Dr.-Ing. Angela Brennecke
Filmuniversität Babelsberg KONRAD WOLF, Potsdam, Germany
a.brennecke@filmuniversitaet.de

Angela Brennecke is professor for "Audio & Interactive Media Technologies/Directing Audio Processing" in the master's program "Creative Technologies" at Film University Babelsberg KONRAD WOLF. Her research interests lie in the field of application development for interactive audiovisual media in artistic contexts.