Sonic Gesture Challenge: A Music Game for Active Listening

Amanda Andrén, Tove Grimstad Bang, Carlo Barone,
Gabriella Dalman, Johannes Loor, Karl Simu, Markus Wesslén
KTH Royal Institute of Technology
{amandr, tgbang, cbarone, gdalman, loor, ksimu, mwesslen}@kth.se

1. INTRODUCTION

Active music listening has shown to be beneficial for improving listening enjoyment in both hearing persons and persons with hearing loss [1,2]. With this paper we present the process behind the development and pilot testing of an audio game where the user, through goal oriented music listening tasks, is required to listen actively in order to succeed with the tasks.

The game, *Sonic Gesture Challenge* was created for web deployment, intended for use on handheld devices, using JS (JavaScript) and WebAudioXML, a recently released JS library developed by Hans Lindetorp, in which XML syntax is used for WebAudio applications [3]. Each of the seven authors created a sound design to go with the game. All sound designs were tested in a pilot study, with the goal of assessing the feasibility of the implementation method, and gaining some understanding of what goes into an enjoyable, yet challenging 'enough' sound design, aiming to support the experience of active listening.

In the game interface, the user is met with a goal oriented music listening tasks built for active listening. The user is first asked to listen to a prerecorded sound and then, by moving their finger within a touch area, recreate the sound. Different movements on the touch area create different sounds, depending on the characteristics of the different sound designs. Through this gestural controlling and interaction with the sound, the user is required to actively and mindfully notice any small changes in the sound and adjust their gesture accordingly in order to obtain a match and succeed the task.

2. BACKGROUND

Previous research has shown that active, attentive or focused music listening is likely to improve listening enjoyment, and games and other goal oriented tasks have been used to facilitate active listening on numerous occasions [1, 2,4,5]. Hansen and Hiraga developed an audio-based game for focused listening aiming to promote hearing training for hearing impaired users [1, 4]. In this game, for Android devices, the user first listens to a sound file, which is then chopped into smaller pieces and spread out randomly across a graphical user interface. Now the player's task is to drag these snippets of sound into the right order, much like in a jigsaw puzzle, in order to solve the puzzle and return to the sound played at the beginning of the game. This game included sound files with both music, speech and a combination of both, and following user tests with both hearing and hearing impaired uses, music was found to be most enjoyable and also most challenging.

As with digital music instruments, an audio game "must strike the right balance between challenge, frustration and boredom" as pointed out by Jordà [6] in his article exploring efficiency and apprenticeship in the relation between an instrument and a player. Digital instruments or devices, depending on their area of application and use context, might serve as tools for interaction with already existing music, functioning more as toys for musical explorations, rather than new music instruments. In such cases, one could argue that the *learning curve* of the device should be quite steep, such that the player can interact with the device without any prior knowledge or guidance, as well as finding the interaction rewarding without spending too much time exploring the actions or gestures involved.

Musical tasks, such as performing scales and arpeggios or musical phrases, have been proposed by Orio et al. as a means to evaluate different musical input devices [7]. The factors pointed out as important to take into consideration when deciding on a musical task for a device, are *learnability*, as discussed by Jordà above, *controllability*, the precision of the timing and musical features, as well as the device's capability for *exploration* and different gestures and gesture nuances available, which is of particular importance when asking a user to perform musical tasks where they are to replicate a given sound.

3. METHOD

3.1 Game Implementation

The game was developed with the HTML, CSS, JS webstack and the WebAudioXML JS library, to be hosted online and later accessed through a browser on several devices. When accessed the user is met with an interface featuring three elements: (1) a play button, which when clicked plays a sound which is to be imitated, (2) an interactive area, where touch events trigger sound and (3) a compare button, which plays the sound made by the player and compares it to the stored sound (see Figure 1).

The core of the game is in an XML file. Using the WebAudioXML library, a number of audio objects are used

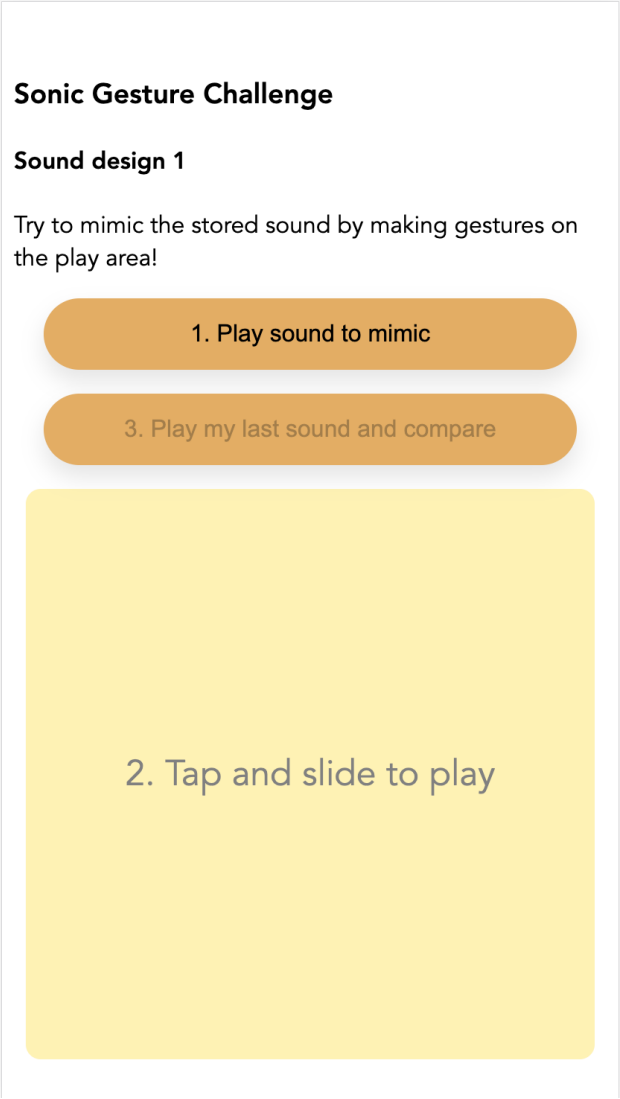

Figure 1. The interface of the game, as seen on mobile.

to create sound (for further detail see the WebAudioXML documentation [3]). This is also where the seven authors implemented their own sound designs. The XML file is linked to the interface of the game mainly through the interactive area, via an HTML *element* tagged with an "interactionArea" attribute. When a player puts their pointer anywhere inside the interactive area, WebAudioXML calculates the relative horizontal X and vertical Y positions of the pointer, referring to the interactive area *element* [8]. Through JS PointerEvents, the player's interaction is then be mapped to the different audio objects. Moving the pointer around inside the interactive area will then change the characteristics of the sound. The interactive area therefore gives players and designers three sound changing dimensions to play and experiment with connected to the gesture. These are; the horizontal axis and the vertical axis within the touch area, as well as the speed/time of the movement.

Central to the game is also the WebAudioXML concept of "gestures"; a way to record, store and output data from JS Pointer Events on the interactive area. Between a Pointer Down and Pointer Up Event pair, every Pointer Events relative X and Y value as well as time passed since the initial Pointer Event can be stored in a gesture [9]. These gestures are then what is used to play and control sound within the game. What makes up the sound which is to be imitated, is a previously recorded gesture stored on file. When a player clicks the play button, WebAudioXML plays the sequence from the data of the stored gesture. Similarly, when the player has made an attempt and clicks the compare button, a comparison function is used to compare the data from the stored gesture with the data from the last gesture, added by the player. The comparison function itself works by comparing the relative X and Y values from the player's entered gesture, to the X and Y values of the stored gesture within a certain position margin and a certain time margin. These two margins were set independently by each designer to fit their design. Additionally, a third margin is used to control the percentage of data points which are allowed to be wrong.

3.2 Pilot Study

All the sound designs took part in a pilot study and were evaluated within the group, all hearing persons. The evaluation aimed to gain insight into the feasibility of the implementation method, using WebAudioXML, the game's potential effect on active listening, and any potential correlation across enjoyment, difficulty, mapping and the nature of the sounds in such a game. This evaluation was done using an online questionnaire, where each of us, on mobile phone wearing headphones, interacted with a sound design (including our own), and then rated questions on:

1. our own attentiveness during the task (1 through 7),
2. number of trials before getting it right (see range in Section 5),
3. how much we enjoyed the task (1 through 7),
4. how we found the gesture-sound mapping, i.e. the connection between sound and gesture (1 through 7).

And after interacting with all seven sound designs, we rated the following:

1. which sound design we enjoyed the most,
2. which sound design we found the most challenging,
3. which sound design we found the least challenging.

4. SOUND DESIGNS

Each of the seven authors developed their individual sound design to go with the game, all aiming to enhance a user's engagement with the game.

4.1 Freq Blender by Amanda Andrén

This sound design uses the standard waveforms included in the OscillatorNode. I used sine-, sawtooth-, square-, and triangle waveforms. The one that it did not use was the custom waveform, that is because I thought that four different waveforms were enough. All the waveforms were included in the Mixer. Two sine waves were in the Mixer, this was because I wanted the second one to be the first overtone to the first one with the fundamental frequency.

All of the waveforms were mapped to the x-axis with different start- and end frequency with different spans. All of the waveforms except the triangle waveform were mapped from left to right in increasing frequency. The reason for mapping them in that direction is because it is the intuitive way of doing it, what the user would expect. For example, on a piano, you have the lower frequencies to the left and the higher frequencies to the right. The overall sound is following that pattern, but when it comes to the triangle waveform I went for the unexpected and mapped it the opposite way. The frequencies for the sound design are between 50 Hz and 900 Hz. All intervals are multiples of the fundamental frequency except the triangular oscillator which spans from 500 Hz to 200 Hz along the x-axis.

As for the mapping of the low-pass filter, up generally means more, an increase. The filter in this case is designed to limit the frequencies the further down you go on the play area.

An envelope is used to control the characteristics of the sound. The ADSR envelope was used for this purpose. No theories were applied for this, instead I changed the parameters and listened how the changes affected the sound and used that to fit my preferences. I found that I preferred to have the decay for a long time in proportion to the rest.

In the last step a delay was added. I chose to do this so that you would be more aware of the gestures that you were making. Without the delay I felt that the user would hurry through the gesture and not stop to think about the correlation. At the same time, I did not want it to be noticeable and I still wanted the game to be enjoyable. Having a long delay will make it substantially harder and the connection between the gesture and the outcome will be lost.

I changed the comparison function to make the game easier. I noticed that I never got the gesture right no matter how many times I tried. I changed both the position margin and the ratio margin to be more generous with the margin of error. I changed it around a few times before settling on values that had a good success rate from my side. I chose not to change the time margin. That is because I chose to add the delay but I also because I felt that it was easier to get the tempo right even if the rest of the gesture was incorrect.

4.2 Polyphonic Loop by Tove Grimstad Bang

This sound design consists of a looping sample of strings, from freesound.org [10], and a synthesised sound with a continuous pitch change mapped to the gesture across the horizontal axis in the touch area, from low pitch on the left to high pitch on the right. The synthesised sound is made up of two triangle wave oscillators, with an offset in between, and spans over two octaves from D2-77.78 Hz to D4-311.13 Hz and A2-116.54 Hz to A4-466.16 Hz [11]. The octave was set to start at D2, and thus repeat D three times across the horizontal axis due to its recurring harmonising with the sample.

Both the synthesised sound and the sample have a delay, with the intention of creating a reverberation effect. They are also passed through each their low pass filter, with cutoff frequencies between 150 Hz and 2000 Hz mapped to the gesture along the vertical axis. The cutoff frequencies for the synthesised sound and the sample are set in opposite directions, such that, at the top you can almost not hear the sample, and at the bottom, you can almost not hear the synthesised sound. This way of almost isolating the two sounds was done in order to provide a way for the user to identify the two and their mapping more easily.

Wanting to work with gesture control in multiple dimensions in this sound design, this string sample was chosen with the aim of making the time dimension, through tempo and melody, a central element in the design. The sound synthesis was added as a way to open up for exploration of harmonies and gestures within the touch area, and aims to incite the user to search for harmonies between the dynamic sample and the synthesised sound [7]. The sound design is intended to pull the user's attention such that, even with the visuals of the interface available, the auditory feedback from the gesture would be enough to explore the sound and succeed with the task. Solving the task with ones eyes closed should be just as feasible as with the eyes open.

The string sample was chosen because of its rich melody, and potential to engage and pull in the user. Other samples with more pronounced tempo and rhythm were also tested, but the strings were found to be a bigger listening challenge, were the user is forced to really pay attention to the changing melody.

The stored gesture, the one that the user is set to imitate, was rather long, and included a good portion of the sample. This was done with the intention of introducing the user to the melody right away, from the first click, with the goal of intriguing the user to listen and explore the sound further.

After informally testing the sound design in the context of the game with five different people, all the margins in the comparison function were adjusted to make it easier, as all of them were having trouble succeeding with the task.

4.3 Air Whistle by Carlo Barone

This sound design was created keeping in account the results obtained by Godøy et al. [12], who analyzed the relationship between sound and gestures performed on a 2D surface, and the nature of such gestures, depending also on the musical ability of the participants. In this experiment certain correlations between gestures and heard sounds, which had been empirically described previously, nevertheless lacking a scientific proof before the aforementioned study, were more methodically analyzed and established; for instance, it was found out how to an ascending pitch it was associated an ascending curve and vice versa, or how the graphical idea associated to a percussion roll followed by a decay was a vibration pattern followed by a descending curve.

The sound design consists of four waves, opportunely enveloped and mixed:

1. a *sine wave*, varying along the X axis, whose pitch decreases moving the finger rightwards on the designed 2D surface;
2. a *sawtooth wave*, varying along the Y axis, whose pitch decreases moving the finger downwards;
3. a *sine wave*, varying along the Y axis, whose pitch decreases moving the finger downwards. On this sine wave it was applied a low-pass filter, whose cutting action follows in turn the movement along the Y axis, increasing as long as it goes downwards.
4. a *square wave*, varying along the Y axis, whose pitch decreases moving the finger downwards;

,

The designed gesture is a descending curve, approximately a parabola, which should recall the descending motion of a body, with the sound aiming to describe the whistling sound it makes while going through air friction. The error margin was enlarged, in order to make more feasible for the participants to get the design right.

Hence, the first idea was to produce a sound whose pitch lowered going downwards, giving that part of the synthesized sound most importance. For recalling the ”whistle” effect, the chosen waveforms would be pretty sharp in the timbre; besides, the low-pass filter applied on the sine allows to obtain a more sound fullness at the end of the gesture in a lower point on the graph, without affecting it in the beginning, since the sound is supposed to be high pitched and without overtones.

Thus, a sine wave was added on the horizontal axis, for giving importance to such component as well and enhancing the perception of bidimensionality, although without making this component more important than the vertical one, being it most representative of a descending motion.

The direction of the horizontal variation of the sound is towards right, perhaps influenced by the designer's and the testers' environment, especially as far as it concerns general graphical representations - as a matter of fact, many of the graphical elements in the European society are created on a left-to-right basis, essentially caused by the writing system, which operates in that direction. In this sense, changes are operable with little effort, for better adapting the content to different ”directional” backgrounds.

4.4 Spring Harmony by Gabriella Dalman

There were quite a lot of possibilities in this sound design because the gesture were supposed to be of drawing a finger across a two dimensional space. This meant that there were two parameters that the user could control, except from the time parameter in which the movement was done. One mapped to the movement across the x-axis and the other one mapped to the movement across the y-axis. In order for it to be a musical instrument one axis was mapped to frequency right from the start. My initial thought was to create a frequency modulation synthesis that sounded like a saxophone on one axis and letting the other axis control the tremolo or vibrato of the synthesis. After some experimentation I decided that I wanted to challenge the ear in a different way. By mapping frequency to both axes the user could create harmonies with the two axes and train their ear on hearing different intervals.

For the synthesis that was mapped to the x-axis I designed a synthesis that had a harder sound to it than the synthesis on the y-axis. It was important that the two synths sounded different to one another so that the user could hear which tone was made by which synthesis. I had three oscillators mapped with different frequency bands to the x-axis: a sine wave with a bandpass filter from 300 Hz to 500 Hz, a triangle wave with a bandpass filter from 800 Hz to 1000 Hz and a square wave with a bandpass filter from 1500 Hz to 1700 Hz. I also added a 200 milliseconds delay for some reverb. The frequency on the x-axis was quantified to a C-major scale using MIDI values from 60 to 72 from left to right.

The synthesis on the y-axis was mapped to the same C-major scale and also consisted of three oscillators: a sawtooth wave with a bandpass filter from 300 Hz to 500 Hz, a sine wave with a bandpass filter from 800 Hz to 1000 Hz and an other sine wave with a bandpass filter from 1500 Hz to 1700 Hz. A 400 milliseconds delay were added to the oscillators. The synthesis on the y-axis sounded softer so there would be easier to differentiate between the axes.

For the gesture that is stored in the game I chose to create a simple suspension that resolves in a major third. The x-value stays at the tonic, in this case a C, while the y-value goes from the fourth to the third and then to the second and resolves back to the major third. In chords it would be sus4 to major third to sus2 and back to major third, but without the fifth. After the y-value has landed on the major third the x-value slides up an octave to finish on the high C while the y-value holds the major third note. This way the user was able to train the ear on the intervals of a forth, major third, and a second.

There are possibilities to create different harmonies with this synthesis and for musicians and non musicians it is important and fun to train the ear on hearing intervals. An extension to the synthesis could be to map the axes to other musical scales or even scales that are microtonal. Microtonality is intervals that are smaller that semitones and requires more training of the ear. By changing the scale of the axes this synthesis can provide different difficulty levels that can be adjusted for musical novices or a virtuoso.

4.5 Cinematic Chaos by Johannes Loor

When designing my sound, I started with the idea of going from chaos to order. This notion is very broad and can mean many different things, so to narrow down the scope of my idea I started exploring the web in search for inspiration. This led me to rediscover a very recognisable cinematic experience, the classic THX intro [13]. This intro embodies my idea of chaos to order by going from a very complex sound landscape with tones at several, seemingly random, frequencies to finishing in one note at a few different octaves. This slowly changing movement gives a great sensation of satisfaction at the end, probably due to the end result being a pattern we finally can recognize after trying to make sense of all the different tones sliding upwards and downwards in frequency. Given the perfect match between my idea and the THX-intro (not to mention the nostalgic feelings kicking in), I decided to make it my main inspiration and design a sound that, when played correctly, would resemble the classic piece of cinematic history.

The sound I created consists of several different sine and sawtooth waves, created with the OscillatorNode in WebAudioXML. These waves slide upwards and downwards in pitch, most of them starting somewhere between 100-200 Hz, and are mapped to the pointer position on the x-axis if they increase in frequency and to the y-axis if they decrease. This made the tones mapped to the x-axis produce their lowest note along the left side of the playable area and their highest along the right side, while the y-axis tones produced the highest note along the top and lowest at the bottom.

Because of how the mapping was designed, the gesture needed to produce the desired sound was a diagonal slide, starting at the top left corner and ending in the bottom right. When reaching the bottom right, most nodes would play the same note (D) in five different octaves: D3-D7 with the frequencies of 36.71, 73.42, 146.83, 293.66 and 587.33 Hz, respectively. To add some vibrato, two nodes had a tiny offset of 1Hz making their endpoint instead land at 292.66 and 586.33 Hz. A final OcillatorNode was mapped to the y-axis moving between 686.33 Hz and 0, to add a faster sweeping notion. Reverb was also added by giving some nodes a delay of 300-500ms.

All mappings of OscillatorNodes were set between 0-95% of the playable area along each axis, making the end chord easier to achieve as it did not change over the last 5%. To control the sound levels, the OscillatorNodes were grouped using the Mixer element which made it possible to control the gain of each group separately. Finally a low-pass filter, with its cutoff frequency of 300-6000 mapped to the x-axis, and an Envelope element controlling adsr (attack, decay, sustain, release) was added.

The choice of using a mix of sine and sawtooth waves as the building blocks for the sound was a result of trial and error. I tried adding other types of the OscillatorNode in different combinations, such as square and triangle, but they often covered the sine wave too much and did not match how I pictured the sound to be. To add some complexity to the sound, I also tried making a simple FM-synth (by following the example found here [14]) but again the result did not fit the desired sound landscape and was removed.

4.6 Galaxy Blues by Karl Simu

The design follows a basic synth structure where sound is generated by oscillators whose output is passed and shaped through filters. Signals are then passed through an envelope generator controlling the attack, decay, sustain and release (ADSR) parameters before being outputted. Additionally, signals are also routed to a delay effect.

The core of this sound design is frequency modulation (FM). Developed by John Chowning in the mid 1960s, FM synthesis is the idea of using a modulator to modulate the frequency of a waveform for sound synthesis. [15]. In this design, one sine wave oscillator is used as a root to drive a modulating signal to two separate sawtooth oscillators, one of which is detuned by 10 cents. The modulating sine signal's frequency is mapped to the pointer position on the x-axis, in MIDI notes 91-103 (1567.98 Hz - 3135.96 Hz). Similarly, the frequencies of the sawtooth carrier signals are also mapped to the pointer position on the x-axis, in MIDI notes 67-79 (392.00 Hz - 783.99 Hz). As the frequencies of all oscillators are controlled by the same parameter (position on the x-axis), the modulating frequency will always be the 4th harmonic of the carrier frequency, synthesizing a harmonic sound [16]. Further variation was added by also mapping the modulation index, i.e gain of the modulation signal to the x-axis. The modulation index was mapped seemingly confusingly to MIDI notes 55-67, here representing a gain of 196 dB through 392 dB. Exiting the oscillator stage, the mixed signal is routed to four separate filters. One of the filters is a lowpass filter with a cutoff frequency set to 392 Hz. The other three filters are all bandpass filters which centre frequencies are also mapped to the pointers relative x position. The filters are mapped to MIDI notes 67 - 79, 79 - 91 and 91 - 103 respectively. See figure 2. The roll-off of all filters is -12dB/octave. The Q value of the bandpass filters, representing the filter bandwidth is exponentially mapped to the pointer position on the y-axis, and range from 0.1 to 10.

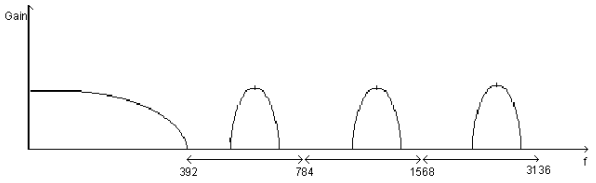

Figure 2. The four filters, axes not in linear scale

Some panning was also added to each of the three resulting signals which passed the bandpass filters. Before exiting as output, signals are passed through an envelope generator. The ADSR values are static and set to 50, 100, 100, 10 (ms) respectively. Upon exit, the final signal is also routed to a delay effect. After a 300 ms delay, the signal is passed through a lowpass filter with a cutoff frequency of 500 Hz and a gain decrease of -6dB/iteration. A second gain control was added to control the wetness of the delay effect, mapped to the y-axis between gain values 0 and 1 (0% to 100% wetness).

In summary the relative x value of the user's finger on the interactive area controls the modulation frequency, the modulation index, the carrier frequency and the centre frequency of the three bandpass filters. The relative y value controls the Q value of the bandpass filters as well as the wetness of the delay. However as the game is played, the player needs not to note all of these mappings. As the user moves their finger in the x-direction, what is mainly heard is the change in pitch, from one tone to another. Moving in the x-direction then follows the notation e.g. of a piano where pitch increases from left to right. As the user moves in the y-direction the filters opening/closing and the wetness of the delay effect can both be heard. The aim was to have the sound appear somewhat "discrete" and "disconnected" towards the top of the interactive area, and to appear more "full" and "apparent" at the bottom. In regard to the filters, one can imagine a zipper on a pair of pants or a jacket which is closed at the top and open at the bottom. Sonically however, there is no apparent support for this metaphor linking the gesture to the sound. The aspects of the sound which change by moving in the y-direction were largely the result of experimentation.

The MIDI notes mapped in steps of a G blues scale in combination with the simple synth sound summarizes the main idea behind the sound; a playful and lighthearted mix of concepts, fitting the game aspect of the project. The blues scale and the detunening were added to give the sound a "western" characteristic. While the synth, the filter sweeps and the delay were intended to give the sound a "retro futuristic" quality.

4.7 Robotic Voice by Markus Wesslén

This sound design is based around the idea of creating a sound that most humans already know and has learnt to recognize. The human voice is a perfect example of sound that fulfills that criteria since all humans with intact hearing has spent their whole lives learning to hear and interpret intricate details of the human voice and this sound design tries to emulate some aspects of the human voice with the goal to create a sound that can be recognized fast and that varies in ways which is easily distinguishable.

A simplified version of the system generating the human voice can be described as consisting of two main parts, the vocal cords and the vocal tract. The vocal cords vibrates when air is pushed between them, creating sound closely resembling a sawtooth wave, and the vocal tract filters the sound generated by the vocal cords. The frequency response of this filter has a few peaks which are independent of the fundamental frequency generated by the vocal cords. These peaks are called formants. In the case of the voice, a certain configuration of the frequencies of these formants are interpreted as a certain vowel sound by us humans. By pinpointing only two formants, almost all of the vowel sounds in the English language can be encoded. [17]

This sound design uses concepts from the described simplified human voice model to synthesize a sound resembling different vowels of a human voice by applying a simple source-filter synthesis model. The source, representing the vocal cords, is a sawtooth oscillator at a low frequency and the filter applied to that signal consists of two resonating second order low pass filters in chain with cutoff frequencies higher then the fundamental frequency of the source, representing the two first formants of the vocal tract filter.

The gesture mapping take inspiration from a classic visualization of the first two formants of a voice where the frequency of the first and second formant are displayed in a two dimensional coordinate system on the x and y axis respectively. This is translated directly to the two dimensional control surface of our game where the x axis is mapped to the cutoff frequency of the first formant filter and the y axis to the cutoff frequency of the second formant.

To create constant bandwidths for the resonating second order low pass filters, the Q-values had to be controlled based on their current cutoff frequencies according to this formula:

$$Q = \frac{cutoff\ frequency}{bandwidth}$$

The chosen bandwidths were 38 Hz for the first formant and 45 Hz for the second formant.

To eliminate the risk of two points on the control surface sounding roughly the same, a subtle mapping of the source pitch was added to the x axis as well. This is a risk since two points on the control surface can have the same values for the cutoff frequencies of the filters, only switched between them, which can generate similar sounds. This pitch mapping ensure unique sounds over the whole control surface.

5. RESULTS

From the pilot study and the within group evaluation, across all seven sound designs, 4.7 was rated the most enjoyed one, 4.1 the least difficult, and 4.2 the most difficult (see Figure 3).

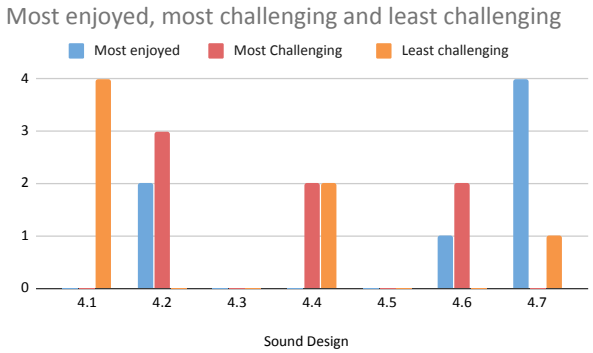

Figure 3. Most enjoyed, most challenging and least challenging sound design, across all seven. The vertical axis shows the number of votes out of seven.

Apart from the results on difficulty in Figure 3, we wanted a more objective measure of how difficult each task and sound design was, hence the question for each individual sound design: *How many times did you try the task before getting it right?* With questionnaire options of various ranges as viewed in the legend of the graph (see Figure 4), in order to reach an estimate of number of trials, each range was multiplied with a factor from the formula:

$$Multiplication factor = \frac{max\ number\ in\ range}{2}$$

As such, the max number for *10+* is estimated to 15, and *I never got it right* to 20.

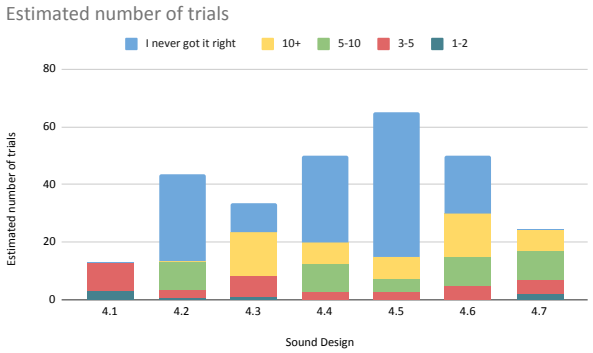

Figure 4. An estimated number of trials for each sound design. The vertical axis one shows number of trials.

Furthermore, an average rate of enjoyment, gesture-sound mapping and attentiveness for each individual sound was calculated for each sound design (see Figure 5).

A possible correlation was also found between ratings of enjoyment and mapping (see Figure 6).

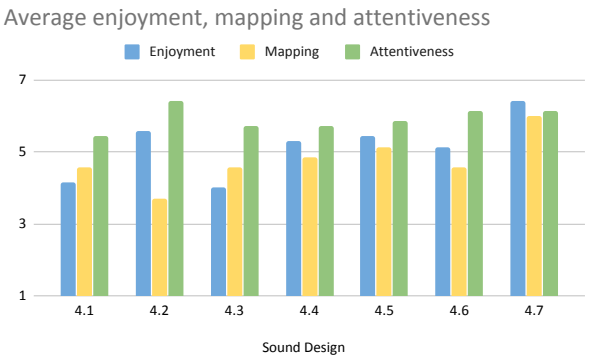

Figure 5. Average enjoyment, gesture-sound mapping and attentiveness for all sound designs.

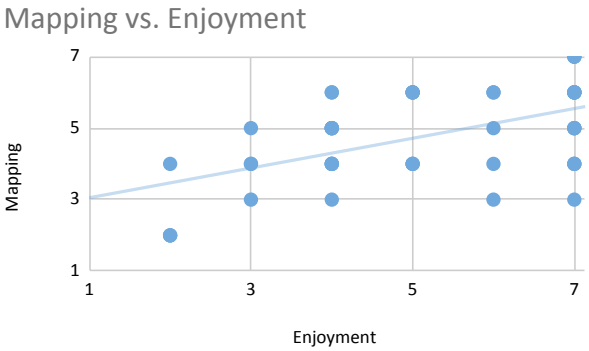

Figure 6. The correlation between enjoyment and mapping. Each point is a a pair of ratings for enjoyment and mapping.

No correlation was found between enjoyment and difficulty.

6. DISCUSSION

In order to keep the possibilities open when designing the sounds for the game, each designer was allowed to set the margins in the comparison function themselves. Which means that a slight gestural change in one sound design, with narrow acceptance margins, might make the difference of a gesture succeeding or not, while in another sound design, the same slight change, might make no difference in the comparison, and the gesture will succeed either way.

This however, introduced problems in the pilot study, when evaluating the different sound designs against each other, as the differences between the seven sound designs were no longer only the sound itself, but also within what margins each designer deemed an attempt acceptable. For instance, it is possible that the hardest sound to replicate was not sound design 4.5, as presented in the Results Section (see Figure 4), but rather that this was the sound design with the strictest acceptance margin in the comparison function.

While the game was developed with the intention of use on handheld devices, we still wanted it to work well across different platforms and screen sizes, hence the interface was made reactive. Different devices and screen sizes also

proved to complicate the design of the interactive area. It was decided that the interactive area should be a square and should scale accordingly to the screen size (see Figure 7). This allowed players to utilize the most of their screen, but may also have made the game considerably harder or easier across different devices.

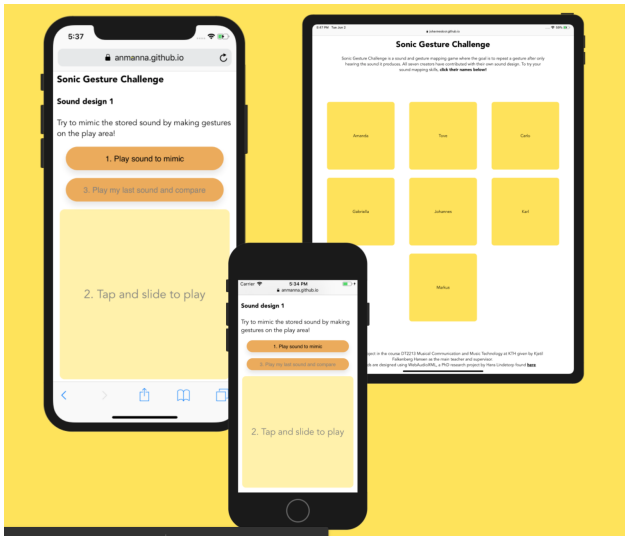

Figure 7. The *Sonic Gesture Challenge* on different devices.

The pilot study and the following evaluation were done in uncontrolled environments, and the questionnaire left room for individual interpretation of the questions. We later realized that our self-assessment of attentiveness was based on very different criteria, and the question and the following results do not provide any evidence of whether the game had any effect on active listening. Similarly, the reasoning behind the ratings of mapping also varied a lot across the respondents. Meaning that the enjoyment-mapping correlation might turn out to be arbitrary. We also do not know whether the correlation points to, the higher the enjoyment rating, the higher the mapping rating, or vice versa.

Each of us assessing our own sound design is also likely to have influenced the results, as you are more likely to succeed with the task, when you are the one adding the gesture you are to repeat. While the estimated number of trials (see Figure 4) provided useful insight, as a complement to the overall rated difficulty, the number of trials should have been properly measured, without estimation.

Having tested the game on a group of hearing persons, and developed the sound designs with the same group in mind, the game, and the sound designs in particular, will need to see some changes in perceived complexity, in order for the game to be accessible for people with varying of hearing acuity. The complexity of the sound designs in this application can be controlled through isolation of sound control parameters, e.g. only working with gesture controlled changes in two dimensions, the horizontal axis and time, rather than all three horizontal axis, vertical axis and time. Adding visual cues, e.g. where within the touch area the gesture starts and ends might also be a useful addition to the game.

6.1 Discussion of 4.1, Freq Blender by Amanda Andrén

All participants got sound design 4.1 right with a maximum of five tries, thus making it and 4.7 the only sound designs that all participants got right. It scored in the lower half for enjoyment and for the gesture mapping. That fits with our theory that mapping is correlated to the enjoyment, in some way. 4 out of 7 also picked this sound design as the least challenging. Although attentiveness is not really a measurement of difficulty, it can be an indicator of it. Sound design 4.1 scored low on attentiveness but it was also the least challenging. This might suggest that the lack of challenge experienced from the user lowers the attentiveness because it is perceived as “too easy” and thus not require the user to focus as much.

Even though no correlation was found between the difficulty and the enjoyment, the fact remains that sound design 4.1 was picked as the least challenging and it scored low on enjoyment. Because we never looked at how challenging each design was, and only looked at the extremes, it is hard to draw any decisive conclusions on that front. That being said, sound design 4.1 does not really allow the user to generate any new and interesting sounds like other sound designs in this evaluation that scored high on enjoyment. This might have resulted in the user not experiencing this sound design as enjoyable.

The comparison function might also have contributed to this being the least challenging sound design and the easiest one to get right. The comparison function was changed to allow a more generous margin of error for the user’s gesture. It was possible to get the gesture right with only being “close enough”. This might have been confusing and consequently lowered both the enjoyment score and the mapping score.

6.2 Discussion of 4.2, Polyphonic Loop by Tove Grimstad Bang

Sound design 4.2, Polyphonic Loop, came out second most enjoyed both from the average individual ratings (see Figure 5) as well as across all sound designs (see Figure 3). While it was rated the most challenging sound design across all seven, it only came out fourth on the estimated number of trials, which points to the sound design being perceived as more difficult than it perhaps was. The changes in the comparison function, made the task very forgiving, perhaps too much so. Narrowing down the margins and acceptance rate of the task might have yielded more coherent ratings across the subjective perceived difficulty and the more objective number of trials.

Furthermore, 4.2 was the sound design with the highest average rating of attentiveness, and the lowest rating of gesture-sound mapping. 4.2 was the only one including a sample. There is no physical or gestural link between the sample alone and the gesture, other than a mere touch (Pointer Down and Pointer Up). One can hold a finger completely still on the touch area, and the sample keeps playing, so other than the low pass filter applied to

the sample, the gesture-sound mapping is indeed limited. However, it is worth mentioning, that from the enjoyment-mapping correlation, 4.2 is an outlier, with a high enjoyment rating and a low mapping rating as opposed to the other sound designs.

While the use of a synthesis on top of a sample was made with the intention of inciting gestures in the users, through exploring harmonies and gestures, the low rating on mapping might point to the sound design not living up to the intention. However, the high rating of attentiveness might point to the sound design being able to grab the user's attention in a positive way.

The stored gesture up for imitation was perhaps too long, and with quite a complex musical output from a very simple gestural input, might have been efficient in terms of grabbing the user's attention, but is likely to have taken a toll on the mapping [6, 7, 18].

6.3 Discussion of 4.3, Air Whistle by Carlo Barone

A first significant outcome about this sound design is deducted from figure 4: is it noticeable how it was considered on average pretty challenging, since no one but the designer got it with less than three attempts, two testers needed more than ten attempts before getting it right and one never got it. This defines the need of improving the feasibility features already applied, since the aim of the enlarged error margin and the relative easiness of the gesture was such one. Perhaps, a diverse point of start of the gesture, or a different tempo could apply in this sense.

Secondly, the enjoyment rate turned out to be pretty low. This piece of data is significant, and can be interpreted in many ways: since the sound design itself does not include pleasant or soft sounds, on a mere physical enjoyment side, this result shows an accomplishment, whereas, in terms of task enjoyment, it shows necessity of improving, because this sound was designed in this way for being enjoyed, not being to difficult for the aforementioned strategies applied, hence not causing frustration or sense of incapability of accomplishing the task.

Thirdly, the mapping average rate is in the middle range, showing how the sound-gesture correlation in the design partially accomplished its scope. In this sense, perhaps, the enlarged margin might have damaged such outcome, since some gestures could have been very different from the meant one.

Lastly, the attentiveness level was in the middle range as well. It is reasonable to suppose that the low enjoyment has affected such parameter, but no certain correlation in this case has been found.

6.4 Discussion of 4.4, Spring Harmony by Gabriella Dalman

Spring Harmony scored in the middle compared to the other sound designs on the attentiveness, connection between sound and gesture as well as enjoyment. When listing which sound design was the most and least challenging Spring Harmony was selected as least challenging by two out of eight people and also selected as most challenging by two of eight people. It is the only design that some people though where the least and most challenging sound. We did not explore why this sound was both difficult and easy for different people, but it could be very interesting to know what knowledge or skill the people who found it to be easy had that the people who found it to be difficult did not have. One assumption is that because the sound was focused on interval training and hearing harmonies, the participants who play an instrument might be more familiar with listening for the intervals while the participants who do not play music are not as familiar with it. In a previous evaluation of the music puzzle it was shown that the participants who had more experience with music performed better with the task than participants who had less experience with music. [1] This could be the case with this sound design as well.

6.5 Discussion of 4.5, Cinematic Chaos by Johannes Loor

Only 2 out of 7 people got 4.5 *Cinematic Chaos* right, one of them being the designer, but no one pointed to this design being the most challenging one across all 7. This could be a result of how the questionnaire was ordered, making it difficult for participants to remember which sound they struggled with or possibly a matter of interpretation of the word challenging. It is in any case hard to tell what factors were in play when deciding on the hardest one. The sound was also not voted as least challenging by any of the seven people evaluating (see figure 3). This should indicate that the challenge level lay somewhere in the middle of the scale but is somewhat contradicted by how the majority of testers never succeeded in reproducing the sound.

Cinematic Chaos was not picked as most enjoyed by any of the participants but even so the average enjoyment was not rated the lowest amongst the sound designs, see figure 5. Both enjoyment, mapping and attentiveness got a rating between 5.1-5.9, which on a scale of 1-7 is leaning towards the upper end. Why these three variables got quiet similar scores while other designs varied more, like 4.2, is unclear but future research with a larger user base, that excludes the designers, would be of great interest to further investigate the matter.

Even if the attentiveness score for 4.5 was not the highest nor the lowest amongst the seven, it is by it self fairly high. This could be due to the "complexity" of the sound, meaning that it was composed with several sound sources. Having a range of sounds added together, creating complex harmonies, could impact the focus required from the listener in order to grasp the elements building up the sound. However, in this case all sound designs, no matter the above mentioned complexity level, received a similar attentiveness score. This is probably due to complexity not being the only, or even most import, factor to consider related to attentiveness. Factors like pitch, duration, rhythm etc could have different impact on attentiveness depending on the user and these varied plenty between the different designs.

Regarding the design it self I, the designer, am pleased with how the sound turned out but looking back the gesture

should have been given a bit more focus. Even if the mapping variable was given a relatively high score compared to most of the other designs, there is no clear connection between dragging one's finger/cursor diagonally across the screen and the sound it makes. The gesture mostly came from how the mapping was done and not the other way around. However, in my opinion the most frilling and memorable part of the THX-intro is the movement and impact of the bass. To sit in the theater and have my body rattle along with the lower octaves at the end of the intro, is what I associate this sound with the most. Considering this, the gesture of slowly moving from the top to the bottom whilst guiding that bass to its destination, could be seen as somewhat connected.

6.6 Discussion of 4.6, Galaxy blues by Karl Simu

The Galaxy blues design placed itself neither in the high end nor the low end across all of the four ratings: attentiveness during the task, number of trials before getting it right, enjoyment and gesture-sound mapping. There are a number of possible reasons for these somewhat mediocre results. The left to right mapping of pitch in steps to the x-axis is easy to follow and as noted, follows the mapping of an piano. For those reasons, the x-axis pitch mapping could have been interpreted as having a good connection between mapping and sound, but possibly somewhat uninteresting and unenjoyable. The y-axis mapping of the bandpass filters Q value, ie letting more of the sound pass towards the bottom of the interactive area, as well as the wetness of the delay were most likely interpreted differently. As noted, there is no apparent link between an more open filter towards the bottom and a less open one towards the top of the interactive area. One could even argue that since a more open filter lets more of the signal pass ie. making the volume higher, the Galaxy blues design goes against the common notation of turning volume "up" and "down". As the Q value of the filters was also mapped exponentially, moving only in the y-direction in the upper half of the interactive area made sonically little difference. This resulted in that navigating oneself in the y-direction was hard and possibly further making the design unejoyable, as well as having less of an connection between sound and gesture. Two evaluators also named the Galaxy blues design as the most challenging design. In addition to the possible reasons for this related to mapping, it may also be related to the arguably difficult stored gesture which they were to imitate and how difficult the comparison parameters were set. Another challenge in the making of this design was setting the position margin of the comparison function. Since the position margin for both the X and Y position were the same, a design decision had to be made. The x-axis had clear discrete steps in terms of MIDI notes. It therefore was decided that the comparison should follow the x axis, i.e. if a player got a note wrong, the comparison always fails them. However for some parts of the interactive area, moving the pointer the *distance* of a MIDI step in the X direction, in the Y direction was arguably not notable enough. Therefore the comparison was arguably also too hard. Lastly, it is also possible that the sound itself was not very enjoyable or contribute to an increased attentiveness.

6.7 Discussion of 4.7, Robotic Voice by Markus Wesslén

4.7 obtained overall higher scores on enjoyment, attentiveness and mapping and was also the last sound design in the evaluation. It was also comparably easy and only 4.1 got a lower score on number of trials.

A possible explanation of why this sound design gets such high scores on enjoyment and many votes as most enjoyed is the fact that it was last and also very different and maybe surprising for many users. This sound uses a less musical approach than the other sound designs and instead of mainly modifying pitches it puts focus on modifying the timbre of the sound. It does also manage to resemble a voice, which in many ways is a very complex sound, although it is still easily controllable with only the two input parameters. All these factors are possibly part of the surprise this sound design delivers.

The fact that the mapping score is high is possibly connected to the intuition of the mapping of the formants to a 2D plane. The model has no built in intuition but since it's the standard way to plot formants in literature and most test subjects are familiar with this literature, this intuition do maybe exist with this specific test group.

Lastly the high attentiveness score is likely a product of the high enjoyment and mapping scores as well as the task not being to hard.

7. CONCLUSION

The game, *Sonic Gesture Challenge*, is an audio game aiming to promote active listening through gestural control and interaction with sound. Through the external activity of the gestural interaction, the listener is encouraged to notice subtle differences in the sound and engage in active listening. The game was implemented using the new JS library WebAudioXML, which served as a great tool for rapid game creation, deployment and sound design, but showed some inadequacy in terms of space for artistic freedom in designing sounds.

Through a pilot study and evaluation of seven different sound designs applied to the game, there seems to be no connection between the perceived enjoyment and difficulty of the sound designs.

With further testing with both hearing and hearing impaired persons, we believe that the game provides an entertaining alternative and an accessible tool to engage users in active listening and hearing training.

8. REFERENCES

[1] K. F. Hansen and R. Hiraga, "The effects of musical experience and hearing loss on solving an audio-based gaming task," *Applied Sciences*, vol. 7, no. 12, p. 1278, 2017.

[2] W. T. Anderson, "Mindful music listening instruction increases listening sensitivity and enjoyment," *Update:*

Applications of Research in Music Education, vol. 34, no. 3, pp. 48–55, 2016.

[3] H. Lindetorp. Github wiki webaudioxml v1.0. [Online]. Available: https://github.com/hanslindetorp/WebAudioXML/wiki

[4] K. F. Hansen, Z. Li, and H. Wang, "A music puzzle game application for engaging in active listening," in *97th Information Science and Music (SIGMUS) Research Conference*. Information Processing Society of Japan, 2012, pp. 1–4.

[5] F. M. Diaz, "Listening and musical engagement: An exploration of the effects of different listening strategies on attention, emotion, and peak affective experiences," *Update: Applications of Research in Music Education*, vol. 33, no. 2, pp. 27–33, 2015. [Online]. Available: https://doi.org/10.1177/8755123314540665

[6] S. Jordà, "Digital instruments and players: part i—efficiency and apprenticeship," in *Proceedings of the 2004 conference on New interfaces for musical expression*, 2004, pp. 59–63.

[7] N. Orio, N. Schnell, and M. M. Wanderley, "Input devices for musical expression: Borrowing tools from hci," in *Proceedings of the 2001 Conference on New Interfaces for Musical Expression*, ser. NIME '01. SGP: National University of Singapore, 2001, p. 1–4.

[8] H. Lindetorp. Github wiki/pointer webaudioxml v1.0. [Online]. Available: https://github.com/hanslindetorp/WebAudioXML/wiki/Pointer

[9] ——. Github wiki/sequencer webaudioxml v1.0. [Online]. Available: https://github.com/hanslindetorp/WebAudioXML/wiki/Sequencer

[10] I. (www.jshaw.co.uk) of Freesound.org. Orchestral strings, warm, a.wav. [Online]. Available: https://freesound.org/people/InspectorJ/sounds/402656/

[11] I. Acoustics. Midi note numbers and center frequencies. [Online]. Available: https://www.inspiredacoustics.com/en/MIDI_note_numbers_and_center_frequencies

[12] R. I. Godøy, E. Haga, and A. R. Jensenius, "Exploring music-related gestures by sound-tracing: A preliminary study," 2006.

[13] I. DIVIL. (2011) Thx intro hd quality. [Online]. Available: https://www.youtube.com/watch?v=PomZJao7Raw

[14] H. Lindetorp. (2020) Simple fm-synthesis. [Online]. Available: https://codepen.io/hanslindetorp/pen/bGVrVmo

[15] G. Reid. (2000) An introduction to frequency modulation. [Online]. Available: https://www.soundonsound.com/techniques/introduction-frequency-modulation

[16] ——. (2000) More on frequency modulation. [Online]. Available: https://www.soundonsound.com/techniques/introduction-frequency-modulation

[17] D. E. Hall, *Musical acoustics*. Brooks/Cole Publishing Company, 1991.

[18] T. Kvifte and A. R. Jensenius, "Towards a coherent terminology and model of instrument description and design," in *Proceedings of the 2006 Conference on New Interfaces for Musical Expression*, ser. NIME '06. Paris, FRA: IRCAM — Centre Pompidou, 2006, p. 220–225.