Using Sound Cue to Improve Time Perception Accuracy in Data Sonification

by

Shuishi Fang

Project submitted in partial fulfillment of the requirements for the degree of
Master of Science in Human-Computer Interaction

Rochester Institute of Technology

B. Thomas Golisano College of
Computing and Information Sciences

Department of Information Sciences and Technologies

Oct 16, 2019

Abstract

People’s perception of time is significantly less accurate compared to the perception of pitch while listening to a sonified data chart. I propose a referential sound cue to improve people’s perception over time when listening to a data sonification. Within-subject experiments were conducted to test this design approach, and the result shows such improvement is sound. This work also shows people’s perception over time changes based on the properties of the sonified data chart.

Contents

1 Introduction

Data charts, as a way to express complex data simple and straight forward, have become more and more important. An estimate pointed out that "in 2010 alone, we will generate 1,200 exabytes—60 million times the content of the Library of Congress." And the number continues to grow [17]. For users who are blind or low vision or situationally visual impaired, the lack of alternative ways to represent data poses a great accessibility issue, since they can't see the data charts as sighted people can.

Although the Web Content Accessibility Guideline suggests representing dataset to blind and low vision users using screen reader by reading out the numeric table directly [6], usability testing revealed such an approach to be "time-consuming" and "take enormous cognitive effort"[27]. As an alternative, the history of creating accessible data charts can be traced back to the 90s [18] as data visualization techniques developed rapidly at the end of the last century [17].

Researchers tried to allow visually impaired users to perceive data charts effectively. There are mainly four categories of accessible data chart: sonification, haptic, tactile, and natural language. Sonification [24] [14] [13] [4] [5] [26] maps the vision variables into acoustic variables as pitch, timbre, loudness, and audio icon. Haptic [19] [21] approach maps them into different levels and types of vibrations. Tactile [2] [16] [23] approach allows users to touch the data chart in a similar way as the braille keyboard and screens. While natural language [10] [11] [12] [8] [27] [20] communicates the key attributes of a data chart with users through descriptive sentences. Finally, hybrid methods [9] [27] combines multiple modalities and remedy the shortcomings of each.

In all those approaches, I think the data sonification is most promising. The only hardwares data sonification needs is a sound card and a headphone, which is cheap and universally available. And other approaches all have flaws, which I don't think can be compensated: Natural language approach can't explain something complex in a single way, so the cognitive cost will be too high under complex data sets (like a DNA comparison chart we use in biology research). Haptics devices are too expensive, too heavy and potentially very fragile due to all those small mechanical parts it has.

When trying to sonify the two axes of a data chart, time naturally is utilized to represent the x-axis, as historically, people mostly use the x-axis to represent time [17]. However, some experiments [4] show people's perception of time is not very accurate (Figure 1). And although large in number, most of the modern sonification methods are designed for combining usage with visualized data. By utilizing their perception over sound, sighted users find it easier to reveal abnormalities in a data set, which can't be perceived by eyes [3] [25]. Sometimes data sonification is also used to represent real-time data. For example, in the application of EEG, researchers report that anomaly is more easily captured by hearing instead of watching [25].

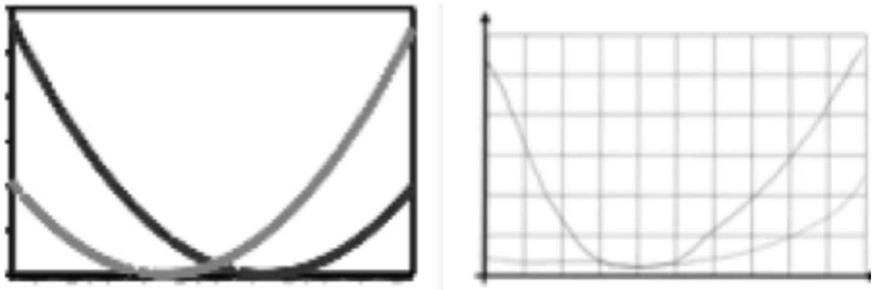


Figure 1: The original chart(left) and a typical participant drew chart(right) [4]

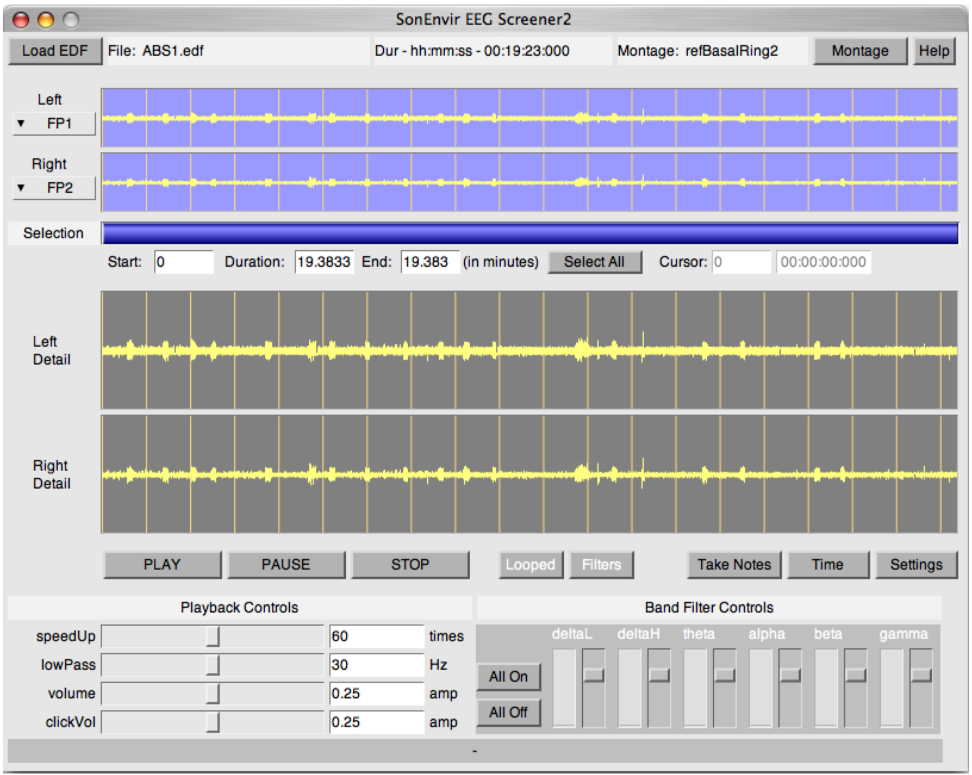


Figure 2: An example of EEG real-time monitor application using data sonification

Thus, researchers never actually address this time accuracy problem as it is less important to their applying area of sonification.

Nowadays, neurologists and psychologists argue that the perception of time is not controlled by certain independent brain parts, but the result of some “sequence of events” mechanism. And experiments show that people’s sense of time is different for each sense, and synchronized by some mechanism often, thus not very reliable [22].

In this work, I design a new approach to improve people’s time perceiving accuracy when accessing data non-visually. New approach involves playing a referential sound every second, creating something similar to a grid in a user’s perception of time to make it more possible for comparison of different parts of a sonic chart in mind. I evaluated this new approach of sonification using a within-subjects experiment, which I will explain in detail during Chapter 6: Evaluation.

In July 2019, I invited 14 participants and conducted one session with each of them. Participants were given test data sonification and were asked to draw down what they heard on an iPad provided. Overall, I received 56 samples over four different charts.

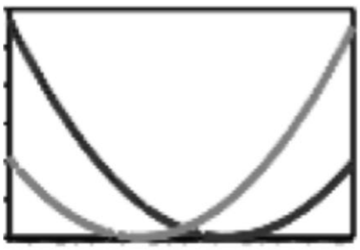
The contribution of this work includes a new way of improving time perceiving during data sonification, and experimental evaluation of this approach. Also, this work expands the “sonify area”, so both positive and negative numbers can be represented in data sonification. Moreover, through this work, a potential property of such sound sonification for future probing was found.

2 Related Work

Different attempts have been made to code visual variables into acoustic variables by using timbre, pitch, loudness, and audio icons.

Other researches show the capacity of acoustic modality. For example, one research shows there might be three or four acoustic sounds people could track simultaneously, and other works determined how many audio icons (audio icon is an approach of non verbal sound communication using metaphors to relate a sound with their virtual referent, like use the sound of door slam to indicate a user leaving the chatting room)[15] can people recognize at a time [14].

The earliest research of using time to represent x-axis is done by Mansur in 1984, and later in Brown and Brewster’s work “Drawing by ear” [4], the researchers sonified a series of line charts of two data series, using time to represent x-axis, and pitch to represent y-axis. The researchers had a different such sonification for each major mathematic attribute (Figure 3). The researchers patch the two different data series in each sonification into different ears of a headphone, provided a few control functions for participants to probe the sonification, like forward and backward, and overview. Twelve sighted participants probed the sonification for two minutes, and were asked to draw down what they heard. Then their drawn pictures were compared with the original data charts. The researchers coded the original curve by the attributes, like a curve dropping down, a line going up, etc. At the end of each session, each participant completed the NASA TLX scale [26] to determine their subjective workload.



Line 1 (black)	Line 2 (grey)	Interaction
1. Curve	5. Straight line	7. One intersection point
2. Generally decreasing	6. Continuously increasing	8. Intersection at halfway point
3. Level section at start		9. Maxima roughly equal
4. Slight increase at end		10. Minima roughly equal

Figure 3: Attributes of a curve [4]

On the other hand, Abu and others' work on Microsoft Excel Accessibility [1] used similar ways to "Draw by Ear" to examine how much information about the data set a participant could absorb. Moreover, they showed a set of tasks contextually related to data usage. They argue that there are three kinds of cognitive processes happening while a sighted user is perceiving a visualized data chart: First is "Question-Answering", which uses data to perceive certain facts in the dataset. Second is "Knowledge Synthesis", which is using the data set to create an idea. The third one is "Knowledge Acquiring" which focuses on the generating of long-term memory about certain ideas created by a visualized data set. And in their work, they also measured the ability of 9 sighted participants to reproduce the data chart to see how well their alternative data representing works. Unlike Brown and Brewster's work – which only compared how many attributes of the different curves participants captured – Abu and his colleagues also tested participants' ability to identify the crossing points and mutual positions of different lines within the same chart. They also measured the time participants took to find out that fact.

3 Research Goal and Research Question

As the new approach aims to improve accuracy of people's perception of time, I listed two major questions to evaluate its effectiveness.

Research Question 1:

Does including a referential sound slice time equally in sonification improve people's accuracy of the axis represented by time?

Research Question 2:

Does including such a referential sound in sonification deteriorate people's accuracy of the other dimension represents the y-axis?

4 Data Sonification, Software Selection, Chuck, Earlier Iteration

Bearman and Brown [3] mentioned five different special designed systems that were previously used by researchers to generated sonification in 2012. They are SuperColider, Pd, Max/MSP, Csound, and ChucK.

After comparing all those tools, I choose ChucK (http://chuck.cs.princeton.edu) as the tool kit to generate the sonification. The ChucK gives me full control over the sounds I generate. I can modify their basic properties as I wish to. Other sounds generate software mainly use musical instruments as their basic sound unit. Thus I would only be able to adjust the loudness and notes of those musical instruments, like a composer creating a music. ChucK uses a real-time sound synthesis mechanism and allows coding like syntax with only a little difference to usual coding languages like c++ or java. Using operator =>, one could give a variable a certain value. “X + 20 => Y” is the equivalence of “Y = X + 20” in most program languages. And there is a sound generating object called “DAC”, so by saying “object sound => dac”, the object “sound” will be played, and the property of this object, includes its frequency, gain, etc, will determine how this object impact what the DAC generates. Many sample codes are available here: http://chuck.cs.princeton.edu/doc/learn/.

With the convenience provided by Chuck, I am also able to separate the mathematics function I used to generate the sound and the transfer functions putting a certain number into a certain sound. So later, it is easy to change any part of it during design iteration. Also, the coding like syntax system makes it easier for me to share them in this report.

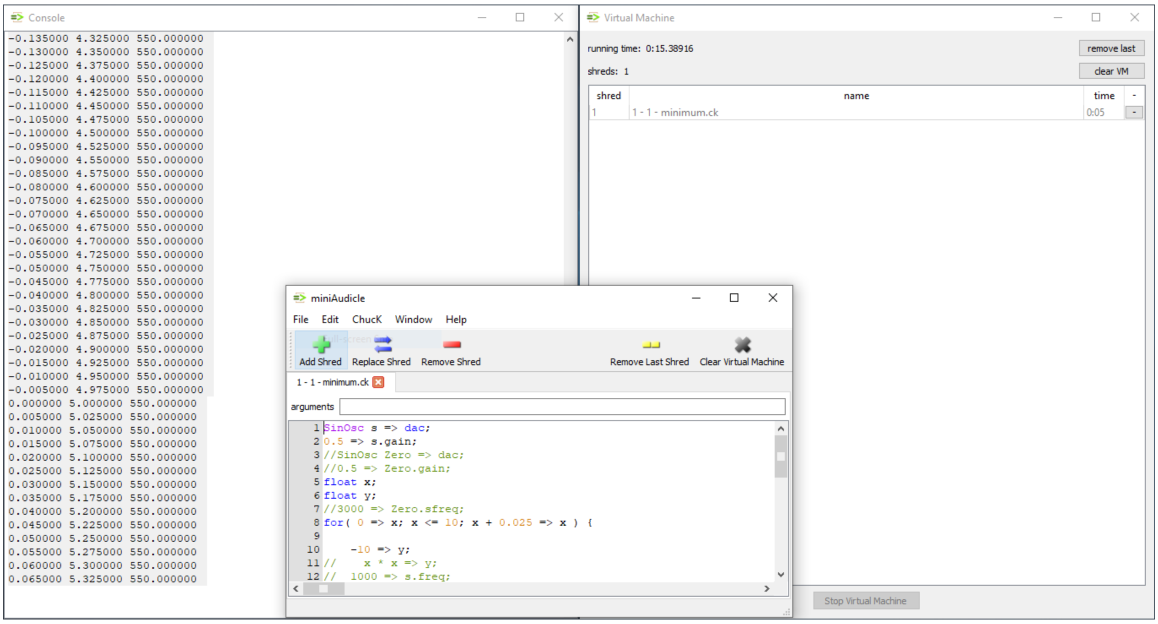


Figure 4: ChucK interface. The left window is a console for debugging. The middle windows is the main interface for coding. And the right window allows tracking active sound files

To sonify a data chart, I need to map the data chart properties into sound properties first. A typical data chart has two dimensions, the x-axis and the y-axis. It has an enclosure as well: Although the author has the freedom to put different legends over the axes so a certain distance to the axis could represent any different actual number, the drawing happens within the enclosure. Moreover, a Cartesian coordinate system has four quadrants, so both x-axis and y-axis could have positive and negative numbers.

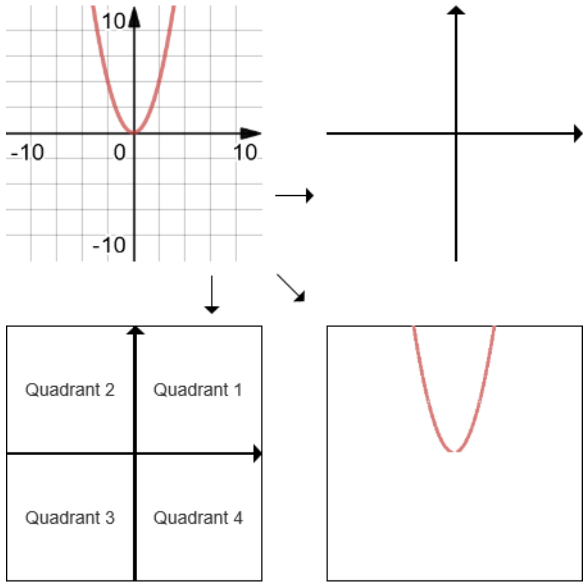


Figure 5: The elements of a data chart using Cartesian coordinate system (top left): Two axes(top right), the enclosure(bottom right), and four quadrants(bottom left).

4.1 Iterations

The earlier stage of the project was to create a reasonably good transferring system mapping the properties of a data chart into different properties of sounds. Only when most people could at least make sense of what's going on in the sonification they heard can we measure the accuracy of their perception. For each iteration round, I invited participants to listen to the sound, do the drawing tasks, and give me feedback on how to improve the sound.

In the end, I made five rounds of iteration, as listed below, all of which use a 10 seconds time period as the x-axis: For each iteration, I've test it with a few participants: if the sounds worked badly, I modified them base on feedbacks from only one or two participants, if the sounds show promises, I tried them with three or four participants to find it's potential flaw.

4.1.1 1st round:

Use the pitch to represent the number on the y-axis. At first, I tried to utilize as much of the audible frequency range as possible to represent the y-axis, up to 13000hz as some adults can no longer hear higher frequency. But participants' feedbacks show that those high-frequency sound make them nervous and sometimes nauseous. So later I lowered the highest frequency I was using to 5500hz, which represents the highest number could show up on the y-axis. While the positive number starts at 3500hz and increasing up to 5500hz, the negative numbers start at 2500hz and decreasing as low as 500hz. And the original point is 3000Hz. Such a setting creates a gap between the positive numbers and the negative numbers. As a result, if a line crosses the zero, the listener could hear a “Tzu” sound, if they know what it is, they couldn’t miss it.

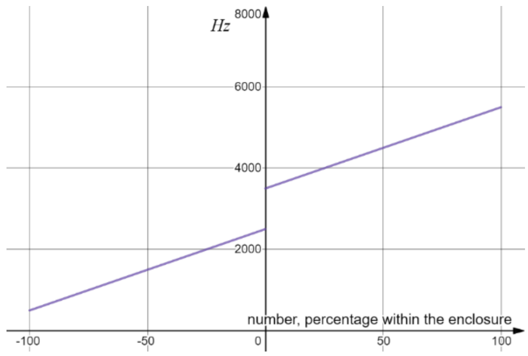


Figure 6: transfer number into pitch

The ChucK code:

```
SinOsc s => dac;
0.5 => s.gain;
SinOsc Zero => dac;
0.5 => Zero.gain;
float x;
float y;
3000 => Zero.sfreq;
for( 0 => x; x < Math.PI * 4; x+Math.PI/400=>x ) {
  25::ms => now;
  Math.sin(x) => y;
  if (y > 0) {
      y *2000 + 3500 => y;
  }
  else if ( y < 0) {
      y *2000 + 2500 => y;
  }
  y => s.sfreq;
  250 / y => s.gain;
  <<< Math.sin(x) , x , y , 250 / y >>>;
}
```

Later pilot runs show that people still feel a little uncomfortable with high frequencies I used as positive numbers, and more than one said it just “sounds like a siren”.

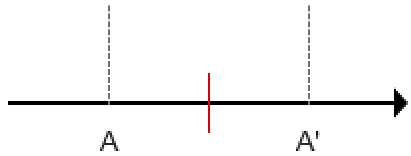


Figure 7: A and A' have same distance towards the original point, suggests the same absolute value.

Also, I realize such a mapping is hard for the listener to tell a number and its negative has the same distance to the origin point on a digital axis. While this is a natural and useful feature in a visualized data chart.

4.1.2 2nd round:

In this round, I tried a different approach by using gain (loudness) to represent numbers in the y-axis (Figure 8). Also since this time I'm not using pitch to represent the number, it is not going all the way up to the high-frequency zone, which eliminates the uncomfortableness and nauseating. "Gain" is the property used in the ChucK system to control the loudness of a certain sound generator. However, it can't be transferred to the commonly used unit decibel as other controllers, like the volume control, could modify the final outcome.

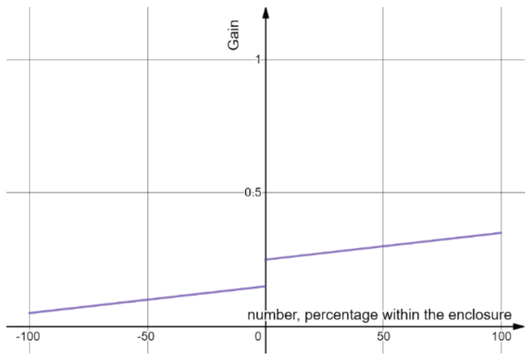


Figure 8: transfer number into gain (loudness)

The ChucK code:

```
SinOsc s => dac;
SinOsc Zero => dac;
float x;
float y;
float t;
for( 0 => x; x < Math.PI * 4; x+Math.PI/100=>x ) {
    0.2 => Zero.gain;
    50::ms => now;
    Math.sin(x)  => y;
```

```
        if (y > 0){
          y / 10 +  0.25 => y;
        } else if (y < 0){
          y / 10 +  0.15 => y;
        } else {
                0.20 => y;
        }
        <<<  x , y , t>>>;
        y  => s.gain;
        <<< s.gain >>> ;
    }
```

Maybe because the final outcome of loudness could be inference by other factors easily, participants didn't react very well to the change of sound. They seemed to consider that a too random thing to pay attention to. Both participants just get confused and said "I'm not sure what's the sound is about." Altought they know the sound is " first getting louder, then quieter."

4.1.3 3rd round:

Since the 2nd round didn't provide desired effects, I switched back to use pitches. I checked some online music generating software. And although I wasn't able to reproduce any of the music instrument sounds, I noticed that most musical instruments have a very narrow pitch range (see https://blog.landr.com/eq-cheat-sheet/). For example, a violin has four fundamental notes from 195.998hz to 659.225hz. http://onlinetonegenerator.com/tuning.html (Figure 9).

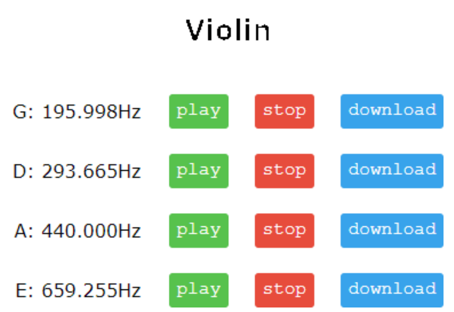


Figure 9: the fundamental notes of a violin. Click the green play button on the webpage to hear those notes.

So, I narrowed the pitch range I use accordingly. And in this iteration, I was trying to make sure people could tell a positive number and its negative counterpart, have the same distance to the original point. As we could easily tell in any data chart.

Now the highest frequency I'm using is 550hz, which represents the highest absolute number on the y-axis, and the lowest is 150hz, which represents the original point. And in the later iterations, participants no longer reporting "anxiety" nor "nauseating". Also, I used different "gain" to represent positive numbers (gain set as 0.5), negative numbers (gain set as 0.4) and zero (gain set as 0.3). (Figure 10)

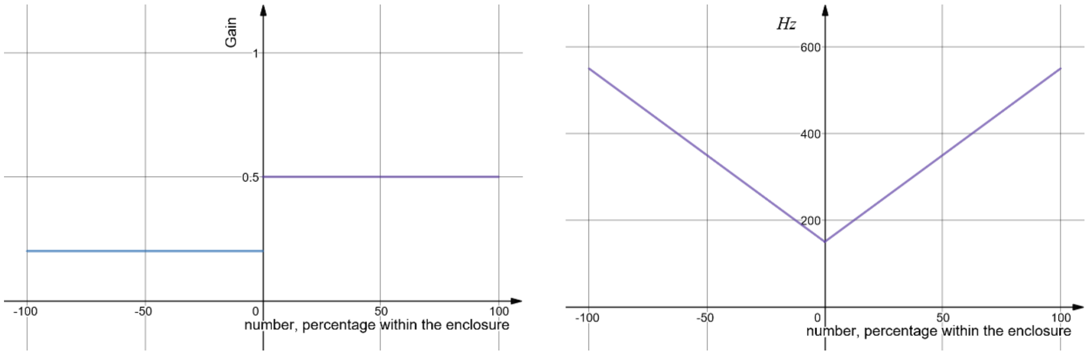


Figure 10: transfer number into pitch, but in a much lower range, and loudness

The ChucK code:

```
SinOsc s => dac;
0.5 => s.gain;
float x;
float y;
for( 0 => x; x < Math.PI * 4; x+Math.PI/400=>x ) {
    25::ms => now;
    Math.sin(x) => y;
  if (y > 0) {
    y *400 + 150 => y;
    0.5 => s.gain;
  }
  else if ( y < 0) {
    - y *400 + 150 => y;
    0.2 => s.gain;
  }else{
    0.35  => s.gain;
  }
  y => s.freq;
  <<< Math.sin(x) , x , y >>>;
}
```

The result shows that this sound is quite promising, as the picture below shows: participants made out that the number is getting bigger than smaller, then get bigger again. It shows that at least the sound is giving the participants a steady impression, even they are not 100% agree with each other on how that impression should be interpreted.

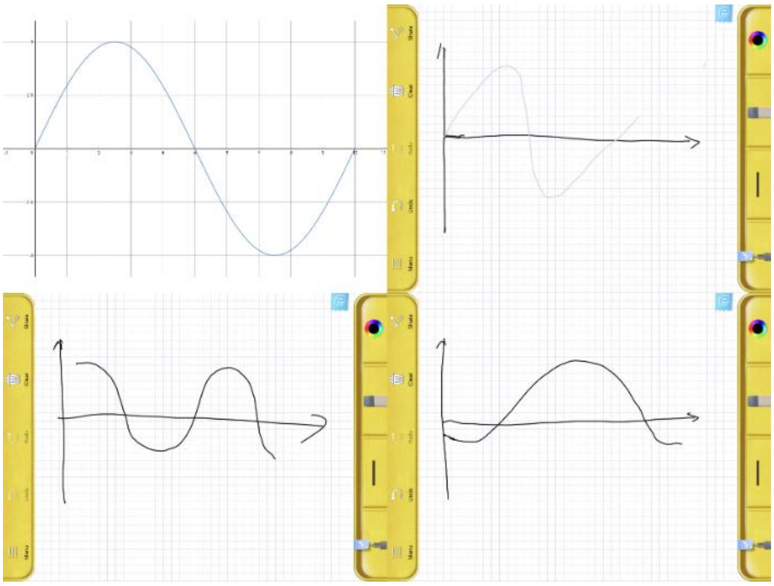


Figure 11: pilot test for iteration round 3, left upper chart shows the data used to generate sound, and the other three are participants' drawings.

4.1.4 4th round:

I also probed more unorthodox approaches. In this iteration, I created two different versions of dotted sounds. (Unlike the continuous sound I created before, dot sounds are separated sound with silent moments between each other). One version has the same time interval between each sound, and using frequency and gain to mark the number, similar to the sound in iteration 3. And another version of the dotted sound focusing on the time lapse: when the value has a high absolute value, the dotted sound is rapid, and vice versa.

ChucK code for dotted sound 1:

```
SndBuf i => BiQuad f => Gain g => dac;
.995 => f.prad;
1 => f.eqzs;
1.5 * 3.14 => float v;
.2 => f.gain;
"special:glot_pop" => i.read;
1.0 => i.rate;
while( true )
{
   float mark1;
```

```
    float mark2;
    float mark3;
    0 => i.pos;
    660.0 + v*v*80.0 => mark1;
    mark1 => f.pfreq;
    v + .05 => v;
    0.3 + v*v*.2 => mark2;
    mark2 => g.gain;
    80::ms => now;
    <<<mark1, mark2, mark3>>>;
}
```

ChucK code for dotted sound 2:

```
SndBuf i => BiQuad f => Gain g => dac;
.995 => f.prad;
1 => f.eqzs;
1.5 * 3.14 => float v;
.2 => f.gain;
"special:glot_pop" => i.read;
1.0 => i.rate;
0.0 => float timePassed;
while( true )
{
    float mark1;
    float mark2;
    float mark3;
    0 => i.pos;
   660 => f.pfreq;
    Math.sin(v) => mark2;
    if(mark2>=0){
        200 - 100 * mark2 => mark3;
        mark3::ms => now;
        0.4 => g.gain;
    }else{
        200 + 100 * mark2 => mark3;
        mark3::ms => now;
        0.1 => g.gain;
    }
```

```
        timePassed + mark3 => timePassed;
        v + mark3 * (3.115926 / 5000) => v;
        <<<mark1 , mark2, mark3, timePassed >>>;
    }
```

However, listening to the sounds from iteration 4th, participantsare basically confused. Most drawings in Figure 12 and 13 shown below have little resemblance to the original data chart. Thus dotted sound are discarded as a potential solution.

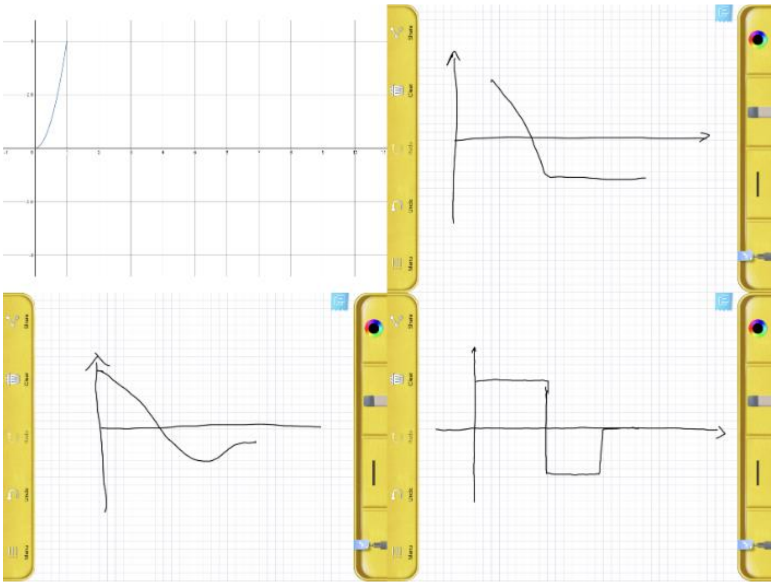


Figure 12: pilot test for dotted sound 1, left upper chart shows the data chart, and the other three are participants' drawings.

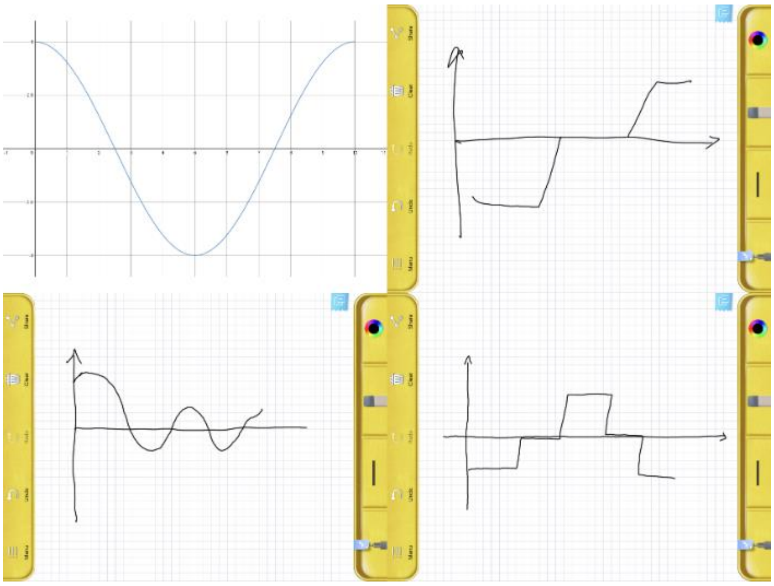


Figure 13: pilot test for dotted sound 1: left upper chart shows the data chart, and the other three are participants' drawings.

4.1.5 5th round:

Based on the result and feedback from the last round of pilot tests, I discard the dotted sound two as it shows little promise in the user test. Dotted sound one seems good for dotted plots. However, since the test in planning is about continuous data, so I didn't probe that direction further. Instead, I changed it into the second cue later used in the final test.

I've also tried something called duplicated code or repeating code, which is usually done in data visualization to make it easier for eyes to capture [17]. The design in the 3rd iteration sets the positive numbers to a 0.5 gain, zero to a 0.4 gain, and the negative numbers to a 0.3 gain (Figure 14 left). This time, the positive number starts at 0.4 gain, where the number is slightly larger than 0, and ends in 0.6 gain for the largest positive number within the enclosure. And for negative numbers, it starts at 0.3 gain where the number is slightly smaller than 0 and ends in 0.1 gain for the smallest negative number within the enclosure (Figure 14 right).

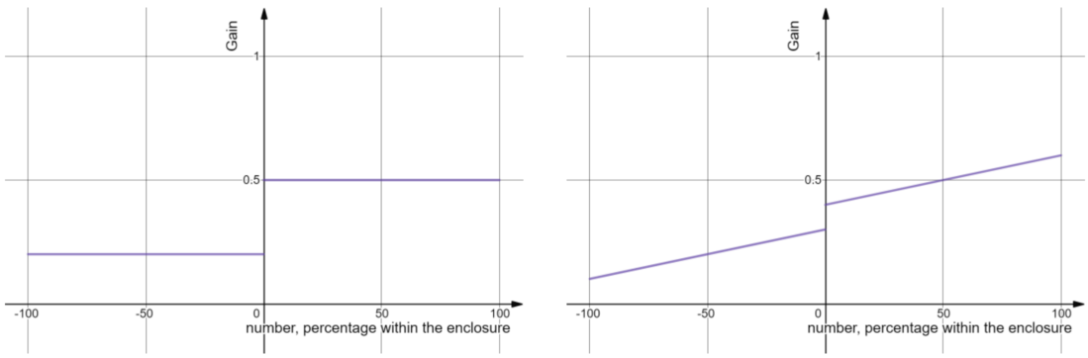


Figure 14: The difference between iteration 3 (left) and iteration 5 (right) in loudness.

So when the absolute value increases, a positive number will increase its pitch and loudness, and a negative number will increase its pitch but decrease its loudness. By doing so, I wish to give positive and negative numbers a unique sound property to each other. So people could discern a positive or a negative number even without having to listen to the standard sample sound and compare them.

Also in this final round of iteration, I've corrected a few bugs from previous codes and cleaned the code up a little bit. For example, as I always use y = Sin(x) to test the transfer mechanism, I didn't notice for y values larger than 1, sounds are generated as well (similar to drawing out of the enclosure in visualized data chart) which disrupts the entire design, and sometimes even providing illegal inputs for the sound generator. Also, the 0 usually doesn't have its own proper frequency but using the frequencies stored in the variable from the last loop. And I separated the mathematics function to generate the curve and the y-axis number – sound transferring function, so in the future when needed, I can use the same code to read data files and generate curves directly.

The ChucK code:

```
SinOsc s => dac;
0.5 => s.gain;
float x;
float y;
for( 0 => x; x <= 10; x + 0.025 => x ) {
   -10 => y;
   y / 10 => y;
   if(y >= -1 && y <= 1){
      if (y > 0) {
         0.4 + y * 0.2 => s.gain;
         y *400 + 150 => y;
      }
      else if ( y < 0) {
         0.3 + y * 0.2 => s.gain;
         - y *400 + 150 => y;
      }else{
         150 => y;
         0.35  => s.gain;
      }
      y => s.freq;
   }else{
      0 => s.gain;
   }
   25::ms => now;
   <<< x * 0.2 - 1, x , y >>>;
}
```

However, when I used the code in the fifth iteration in the final test, it seems even worse than the ones in the 3rd iteration. Participants often create charts resemble the original ones but have some other kind of flaws: for example: some give me drawings mirroring the correct one (Figure 15 left). And some can’t tell the difference between an upper cureve, a straight line, and a lower curve (Fingure 15 middle). And sometimes they know it is the highest the data gets in a chart, but can’t tell the correct absolute value of the sound on the y-axis (Figure 15 right).

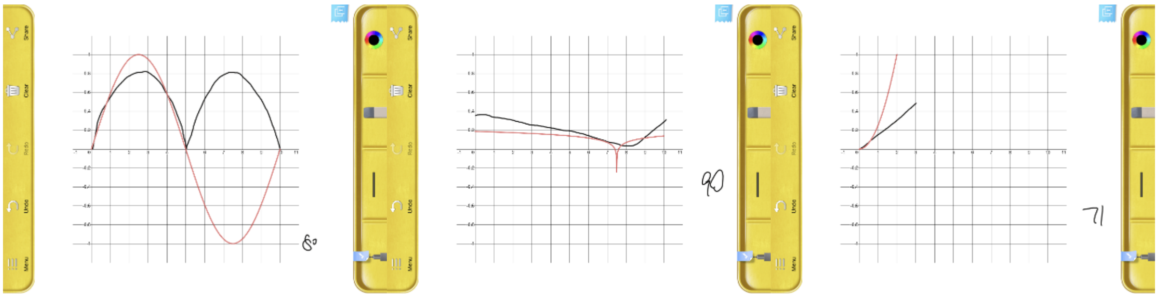


Figure 15: the original (red curve) and the participant drawing (dark curve)

On the other hand, over the 5 rounds of iteration, I still can't let people figure out the correct y-axis number by playing the sound alone. Maybe it is due to the lack of exposure to this kind of sound. After all, even for sighted people and data visualization, it took a while in our high school to learn that. Or maybe the current design still missing something. If time allows it, I would like to do more iterations to address this problem.

Another problem never solved is that people can't tell the difference among an upper curve, a straight line, and a lower curve from the sound alone. About this problem, I think modifying the timbre of a sound could be the approach to try. However, since ChucK does not provide the ability to adjust the timbre, this remains unexplored.

Other than continually improving the sound properties mapping, I've also done iterations on the materials I prepared for the participants, and the methods I used to ask my questions. At first, I provided them with pen and paper but found people tend to change their drawing a lot during the procedure: sometimes change what they want to draw, sometimes just refine the drawing slightly. So later I provide Apple pencil and iPad instead, to make things easier for them. On the other hand, in the beginning, I provide simple grid paper, but found out participants would draw coordinate systems differently. And the results can hardly be compared with each other. So later, I provided a coordinate system, then found the participants were not aware of the drawing enclosure, and thus made it hard for me to analyze data in the future as well. So, in the end, I modify the way to represent the coordinate system even more, to make the result more comparable (Figure 16).

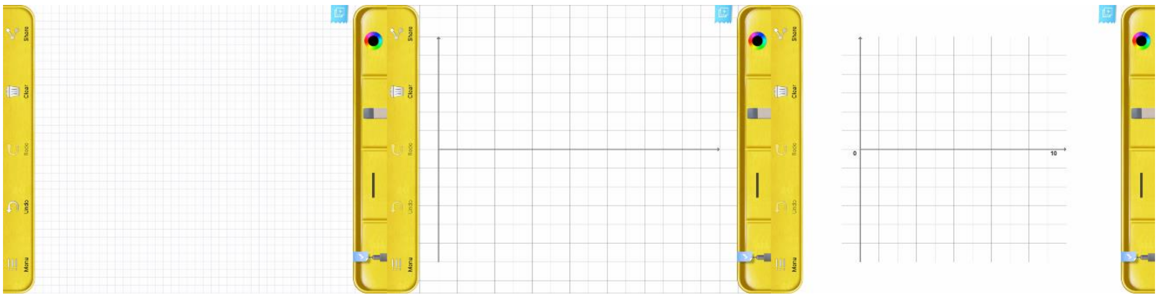


Figure 16: 3 versions of the provided drawing paper, grid paper without axes (left), grid paper with axes (middle) and paper with axes, and only have grid within the enclosure (right).

The questions I've been asking were also modified throughout the procedure.

5 Implementation

The final system includes a PC with Windows 10 system (which could be changed into a Mac or a Linux system with ease as ChucK is a cross-platform solution), and ChucK software installed. A predefined sound will be played through the computer's built-in speaker or a connected headphone. Also, an iPad with a specified application, “Sketch Book”, installed was used to record participants' reactions toward the sound.

A major difference between my implementation and the one in the original research is that the original research allows participants to “probe” the sonification: The researchers provided a control which participants could get 2 seconds back and forth and control the sound by themselves. While I'm able to do the same with ChucK, I didn't think it is a good idea.

Unlike a picture or a chart that a user could “stay” somewhere as the medium is holding still. The sound, on the other hand, is a medium keep changing to make sense. So you can't “stay” somewhere in the sound, while you could jump from one place to another, you are still under the influence of the time (and the sound) flowing.

Another major difference between visual and sound modalities is that in visual, it is easy to locate your focus somewhere, and confirm it is the right “somewhere” of the medium by glancing at the medium. It's some topographical awareness one doesn't have when facing a sound. While encountering a sound file with a duration, you have no way to tell which part of it you are by just listening to it, unless you are already very familiar with the content.

Bring this down to user interaction, if a user is at somewhere of a sound, let's say 7th second and want to jump back 2 seconds and listen again what the fifth-second sounds like. First there will be a 0.7 second of delay, due to how the human body is built, there are neuro latency and reflex in effect. And then, the user triggers control, which leads the user to 5.7 seconds. Yet the user might think it is still 5 seconds, and mapping the content to a wrong place and get a wrong understanding of what's going on.

I think allowing such a vague (a control takes a user back and forth based on user's will, and the impact of reality combine together) interaction takes the form of a solid one (a control which takes me back and forth based on my will) is usually the recipe of chaos and confusion. So, I didn't implement this function in my work. Instead, I allow my participants to listen to the sound as many times as they wish.

The ChucK code:

```
"C:/Users/.../Desktop/Data Sonification/July/Final Test/" => string fileLocation;
90  => int s;
if(s == 10){
   Machine.add( fileLocation + "1 - 1 - minimum.ck" );
} else if (s == 11){
   Machine.add( fileLocation + "second.ck" );
   Machine.add( fileLocation + "1 - 1 - minimum.ck" );
} else if (s == 20){
   Machine.add( fileLocation + "1 - 2 - 0.ck" );
} else if (s == 21){
   Machine.add( fileLocation + "second.ck" );
   Machine.add( fileLocation + "1 - 2 - 0.ck" );
} else if (s == 30){
   Machine.add( fileLocation + "1 - 3 - maximum.ck" );
} else if (s == 31){
   Machine.add( fileLocation + "second.ck" );
   Machine.add( fileLocation + "1 - 3 - maximum.ck" );
} else if (s == 40){
   Machine.add( fileLocation + "1 - 4 - linear" );
} else if (s == 41){
   Machine.add( fileLocation + "second.ck" );
   Machine.add( fileLocation + "1 - 4 - linear.ck" );
} else if (s == 60){
   Machine.add( fileLocation + "3 - 1.ck" );
} else if (s == 61){
   Machine.add( fileLocation + "second.ck" );
   Machine.add( fileLocation + "3 - 1.ck" );
} else if (s == 70){
   Machine.add( fileLocation + "3 - 2.ck" );
} else if (s == 71){
   Machine.add( fileLocation + "second.ck" );
   Machine.add( fileLocation + "3 - 2.ck" );
} else if (s == 80){
   Machine.add( fileLocation + "3 - 3.ck" );
} else if (s == 81){
   Machine.add( fileLocation + "second.ck" );
   Machine.add( fileLocation + "3 - 3.ck" );
```

```
} else if (s == 90){
   Machine.add( fileLocation + "3 - 4.ck" );
} else if (s == 91){
   Machine.add( fileLocation + "second.ck" );
   Machine.add( fileLocation + "3 - 4.ck" );
}
```

By manipulating the variable “int s”, the ChucK will play the sound accordingly: If the s ended with 1 (11, 21, 31, etc.), the sound played with the referential sound cue. And if the s ended with 0, the sound played alone. In this way, the referential sound and the data sonification could be sync perfectly. And the detail of the code is shielded from the participants.

There are 8 different data charts provided to the participants. 4 are sounds to help participants being familiar with the sounds, and another 4 are test sounds which participants should draw down on the provided sheet. The detail of those sounds will be described in the Evaluation.

6 Evaluation

To evaluate this design approach I proposed, I designed a within-subject experiment to compare the difference of participants’ time perceiving accuracy between sonification charts with or without referential sound.

6.1 Methods

Things being tested here are participants’ accuracy in perceiving the data sonification. And I tested it by measuring how well a sighted participant could reproduce the image. I matched the attributes captured by participants during the reproducing with the original data chart -- the data chart describes the original data used to produce the sound. For example, in the picture below (figure 17), the sound slowly dropping for 7 seconds, then dropped and then increased rapidly in the 7th second, and then increased slowly in the last two seconds. Also, three important changing happen in the chart: twice when the number crosses the zero, and once when the number starts to increase again.

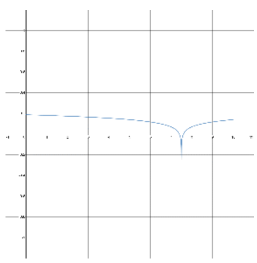


Figure 17: Test sound

During the test, eight different sonified data charts were shown to the participants: The first group of four are sample sounds, aiming to get the participants familiar with this idea of sonification, and trying to show them how such sounds are working. The first three represent the lowermost value of the enclosure, the zero, and the uppermost value. The fourth one showing a straight line starts from the lowermost value, and increase all the way up to the uppermost value over the 10 seconds, steadily. For each of the four sonification, both versions (with or without the referential sound cue) were shown to the participants. And in the procedure, an explanation of how the transferring takes place was provided: Participants were told that “the zero have the lowest pitch, and the pitch increasing along with the absolute value”. And the difference between a positive value and its negative counterpart is that the positive ones are louder. The second group of four sonification are testing sound. They are showing to the participant in a random order, and for each of them, only one version of it (either with or without the referential sound cue) was shown to the participants. Participants can probe that sonification as many times as they prefer, then they were asked to draw down what they think they heard. Sometimes participants draw out of the enclosure. Then I would explain the enclosure rule again and ask them to redraw the chart.

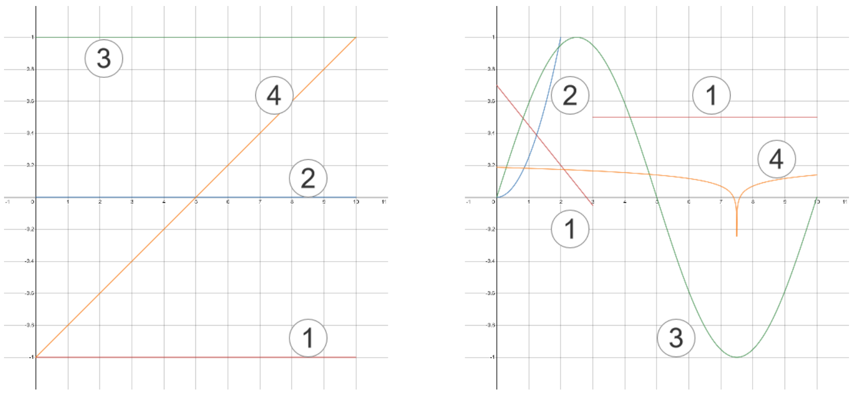


Figure 18: the original data chart for the sample sounds (left) and the test sounds (right).
Coded by color: red is No.1, blue is No.2, green is No.3, and orange is No.4

As for collecting data, each test will contain 1 or 3 “turning point”, where the attribute of the sound would change dramatically. And I’ll measure the deviation of the turning point between the played data set and participants drawing. For example, in the picture below, the turning point in the data set is at the end of the 3rd second, and the participant drawing has a turning point at the end of the 4th second. So the deviation being recorded is 1 second (figure 19).

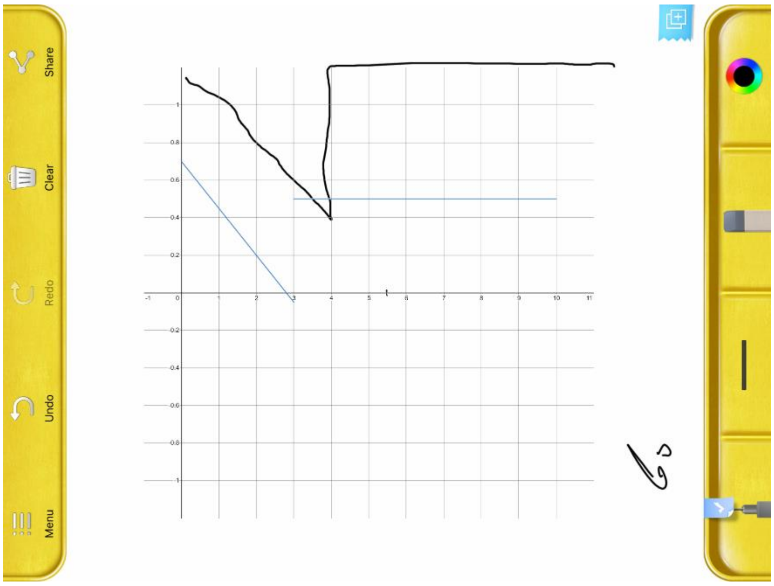


Figure 19: recording data, example, the original (blue curve) and the participant drawing (dark curve)

Also, some samples are discarded as they show that participants didn't make sense of what the data sonification is about (figure 20).

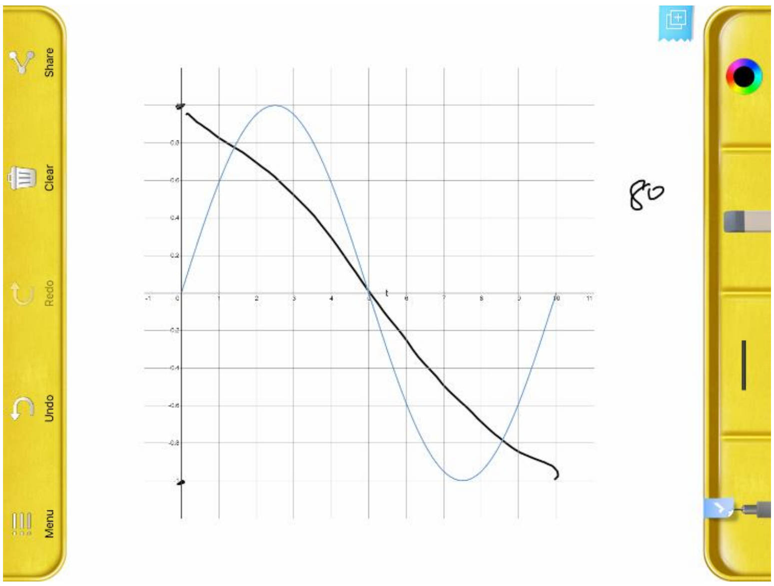


Figure 20: discarded data, example, the original (blue curve) and the participant drawing (dark curve)

Overall, I choose the unbalanced two-way factorial ANOVA to test the data, because there are two samples with different sample numbers and probably different square variance as well, and both with/without the reference sound, and the difference among test pictures could cause particpant's performance changing, thus need to be tested.

During July 2019, I invited 14 participants. All participants age from 19 to 30, 6 of them are female. All participants have an education level higher than high school, 3 of them have a Ph.D. degree. All are familiar with the idea of data charts, and 2 of them have systematical training as statistic scientists. I conduct one session with each of them. In the end, I received 56 samples over 4 different charts, and 9 samples are discarded.

6.2 Findings

The original research question 1:

Does including a referential sound slice time equally in sonification improve people's accuracy of the axis represented by time?

The result of the two-way factorial anova (Figure 21) shows that there are significant difference between different data charts (the rows). However, not significant enough to support a difference between with or without the referential sound cue (columns).

Two Factor Anova (via Regression)						
ANOVA				Alpha	0.05	
	SS	*df*	*MS*	*F*	*p-value*	*sig*
Rows	25.99936	3	8.666453	9.589008	3.69E-05	yes
Columns	2.572543	1	2.572543	2.846393	0.097457	no
Inter	11.63899	3	3.879664	4.29266	0.008748	yes
Within	47.90089	53	0.90379			
Total	89.56318	60	1.49272			

Figure 21: Result for question 1.

The original research question 2:

Does including such a referential sound in sonification deteriorate people's accuracy of the other dimension represents the y-axis?

This question can't be answered because there are some problems with the test sound, with or without the referential sound cue, participants tend to perceive the data sonification wrong, as discussed before. So current data can't really answer this question.

On the other hand, some trending I didn't expect before also seems to exist. The result from factorial anova suggests there are significant difference between different pictures, and by observing the data, it seems that participants performing much better in picture 3 than in other pictures. Data shows a very small average deviation, and very small variance of deviation too.

MEAN			
	With Sound Cue	Without Sound Cue	
Picture 1	1.555555556	1.25	1.402778
Picture 2	0.9375	2.8125	1.875
Picture 3	0.357142857	0.241111111	0.299127
Picture 4	0.375	0.583333333	0.479167
	0.806299603	1.221736111	1.014018
VARIANCE			
	With Sound Cue	Without Sound Cue	
Picture 1	1.277777778	1.175	1.17381
Picture 2	0.602678571	3.424107143	2.816667
Picture 3	0.226190476	0.146961111	0.17239
Picture 4	0.053571429	0.141666667	0.09478
	0.78125	2.252671429	1.49272

Figure 22: The difference between different pictures.

So, I proposed the research question below:

Research Questions 3:

Is it possible, with or without the test sound cue, the participants will appear to perform better in time perceiving accuracy if the turning point happens to be at the middle point of the full duration?

For example, in test sound 3 (Figure 23), since it is a variance of math function y = Sin(x), there are three turning points locating at 2.5 second, 5 second, and 7.5 second. Thus, one slice the entire duration into two halves, the other two turning points slice the small 5 seconds duration into two halves as well.

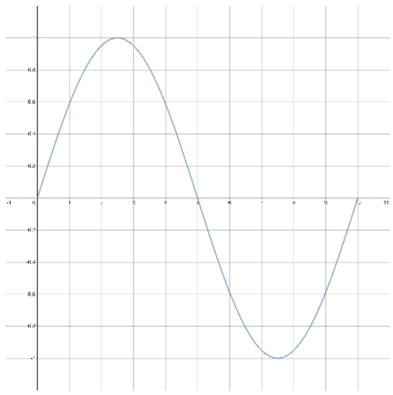


Figure 23: Picture 3, or sound 3.

So, I compared the data from test sound 3 with the rest of the data using Welch-t test, and the result $t_{statistic}$ is 5.20, while the $t_{0.05/2,32}$ is 1.697. The result strongly rejected the null hypothesis which suggests there is no difference between the two. This is a very interesting and unexpected result. And this is also the reason why in research question 1 I also calculated results that excluded all data from test sound 3.

7 Discussion

Since this research is in its pretty early stage, I might be able to reveal some potential trends, although the experiment was not tailored to validating any of them. For example, the probing on research question 3 above suggests that people's perception over time is not indiscriminating to all the data. And inferring from that, I would suggest it is quite possible that people "tend to think" that such a "turning point" locates where it could separate the chart in half. For example, if the actual ratio is 4:6, people may still modify it in their mind as 5:5. Of course, that is going to be another day's research.

During the research, the iterative design process only found one kind of referential sound cue. But probably other sounds also exist which could be used as a referential sound. And it is possible that the attributes of the referential sound, the loudness, the pitch and the timbre of that sound could potentially interfere with the participant's perception with the sonified data sets. So, finding more useable referential sound and evaluate them could also be a future research direction.

Also, there are still other approaches of sonification I haven't tried yet. For example the use of sound icon to represent a series of data. For example, when sonifying a climographs (graph which describe climate of a certain location, usually have two series of data: how much rain/snow falls and how hot/cold it is.) Maybe we could represent the data by sound icons: use the sound of rain to represent the level of precipitation, maybe when it is low we heard the sound of a drizzle, and when it is high we heard the sound of a storm, etc. Also we could find an sound icon for temperature as well, also it might not be very intuitive. The things is, when sonifying data, the main thing we are doing is to find ways to represent different level of intensity, and there are ways other than changing the loudness and pitch level to do that. The sound of a storm is louder than the sound of a drizzle, however there are more to it, and despite those difference, our mind might still be able to associate them with one single entity. Another example is if we could use the sound of a young boy to represent young people, and the sound of an older man to represent older people, take ten of such sample, can we create a sonification using different portion of such sounds, to represent the age structure of an area? Say if we are use the resonate of 100 sounds to represent this data, and there are 10% people who are 0~9, then we use a 5 years old boy's sound to represent these young people, thus 10 of the 100 sounds are this 5 years old boy's sound. Something like that. That is an interesting topic to explore.

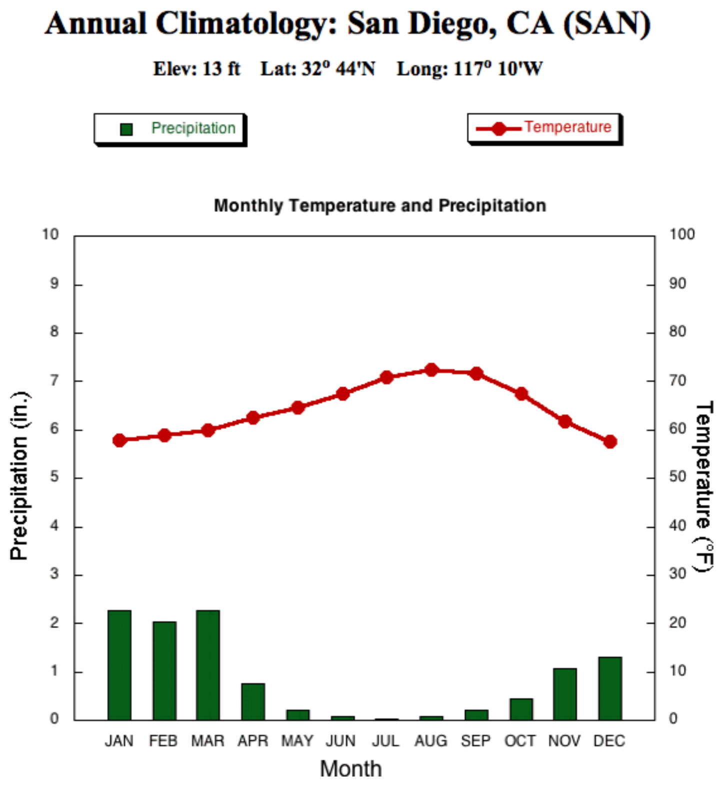


Figure 24: An example of climographs

Just like any other skills, training could be an important part of making sense of a data sonification. It turns out It is hard for people to tell from the sound where it should be on the y-axis. Thus, more than half of my sample can only be considered “Make Sense” instead of “Right” when compared with the original data chart. If the participant can’t figure out the shape correctly, to be honest, I doubt how much guessing work there is when they draw the data chart down. Thus, an important part of future research will be trying to determine participants' improvement over time and compare it with people’s performance when encountering data visualization.

Also, due to logistic reasons, I recruited sighted, instead of blind and low vision participants to perform the task, so they could redraw the data chart they heard and provide data for analyzing. Blind and low vision users are not able to regenerate that chart in another form, then I would be limited to asking indirect questions, and unable to probe details of participants’ perception. However, this choice of participants could potentially skew the result because sighted users would have more experience with data charts, thus have a higher graph literacy [7] then blind and low vision users. Without testing the design with blind and low version users, only the potential, not the promise of an accessible data chart approach, could be determined by this work.

8 Limitations

One problem I found during the late stage of the research is that the sound cue I provide started with, and ends with the sound. The sound itself lasts 10 seconds, but the cue appears 11 times. It's highly possible that a participant counts the appearance of the sound cue but thinks he or she is counting time. So when the participant heard the first sound cue, he or she counts it as 1 (second), but it is 0 (second). Thus, it creates an extra 1-second deviation in the data.

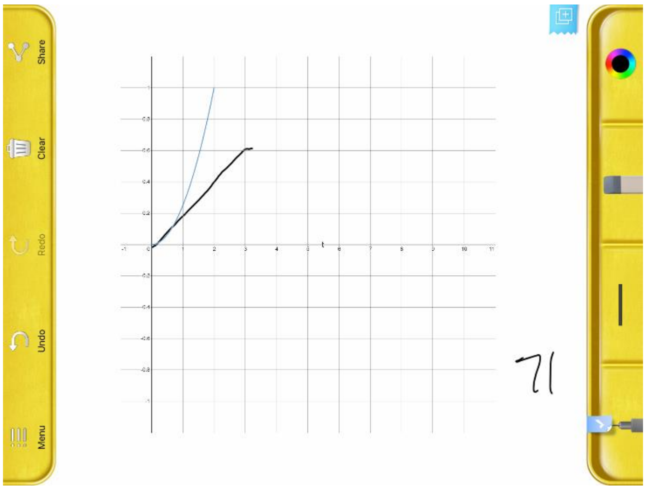


Figure 25: what cause by the mechanism

I find some samples that seems to be affected by the mechanism: the turning point happens in the first half of the duration, and there was little room for ambiguity as this sound only plays two seconds, then turn silence for 8 seconds)

And after I find these problem, for all later participants who seems to show such behavior, I asked them: How are you using the sound? And they would tell me: I count the sound of course. And then I asked: The counting you are doing, does it start with 0, or does it start with 1? And they told me they start with 1.

I changed those data (participants who testified they start count from 1, and does have a 1 second deviation in their drawing) into a 0-second deviation and ran the numbers again. And this time, the result show a much higher significance level, not to mention there are more data come from participants who did this before but can't be identified now. Thus, it seems to me that if I recruit another batch of participants, and make the sound cue starts at the end of second 1, I'll get a much significant result then.

Two Factor Anova (via Regression)						
ANOVA				Alpha	0.05	
	SS	*df*	*MS*	*F*	*p-value*	*sig*
Rows	20.17367	3	6.724558	7.314955	0.000342	yes
Columns	5.499256	1	5.499256	5.982076	0.017806	yes
Inter	19.17503	3	6.391677	6.952849	0.000496	yes
Within	48.72232	53	0.919289			
Total	95.30154	60	1.588359			

Figure 26: Result for question 1, after changing the verified data into 0 seconds deviation.

Another regrettable limitation comes from the sound itself. Since I can't get the original sound from papers like "Drawing by Ear", I had to create the sonification myself. It seems to me that the sound I've created so far is not as good as what those earlier researchers had created. And as a result, sometimes that participants were drawing down a chart while they can't make sense of the sound. That would influence the result as well.

The idea of "precision" itself could also be a factor. When provided with a paper with grids representing 1 second, you could only anticipant participants having a sense of precision at second level, or half a second level at most. In the test chart 1, I put two potential turning points very close with one another (when the first part of the line cross zero, and about 0.1 seconds later when it turns into a flat line which is very different from the first part). And potentially, different participants are referring to different turning points, which could impact the data accuracy as well.

9 Conclusion

This work shows how well people's perception over time is when encountering a data sonification. Also, during the work, a number – sound transferring system has been designed, tested and improved over time. Moreover, a design approach to improve people's perception over time was proposed and tested by a within-subject experiment. And the experiment result supports that the proposed design approach could help improve people's perception over time in data sonification.

If probed further, this work could produce a usable new way of conveying data in a short time, which could have its use in the area of accessibility. For example, providing users who are blind or low vision a way of absorbing data chart, like how a sighted user looking at a data chart.

10 Acknowledgments

I would like to thank Patrick Scully for his suggestion and for introducing basic knowledge regarding sound attributes during the early stage during this work. All participants for volunteering to be part of this research. And Kristen Shinohara and Mick Mcquaid for advising me throughout the procedure and helping me improve the work. Also, I would like to thank all the wonderful faculty members and classmates in the HCI program for giving me a very educational experience and a really good time during the last two years.

References

[1] Iyad Abu Doush, Enrico Pontelli, Tran Cao Son, Dominic Simon, and Ou Ma. 2010. Multimodal Presentation of Two-Dimensional Charts: An Investigation Using Open Office XML and Microsoft Excel. *ACM Trans. Access. Comput.* 3, 2 (November 2010), 8:1–8:50. DOI:https://doi.org/10.1145/1857920.1857925

[2] Maryam Azh, Shengdong Zhao, and Sriram Subramanian. 2016. Investigating Expressive Tactile Interaction Design in Artistic Graphical Representations. *ACM Trans. Comput.-Hum. Interact.* 23, 5 (October 2016), 32:1–32:47. DOI:https://doi.org/10.1145/2957756

[3] Nick Bearman and Ethan Brown. 2012. Who's sonifying data and how are they doing it? A comparison of ICAD and other venues since 2009. (June 2012). Retrieved November 26, 2018 from https://smartech.gatech.edu/handle/1853/44402

[4] Lorna M. Brown and Stephen A. Brewster. 2003. Drawing by ear: Interpreting sonified line graphs. (July 2003). Retrieved November 8, 2018 from https://smartech.gatech.edu/handle/1853/50453

[5] Robert F. Cohen, Rui Yu, Arthur Meacham, and Joelle Skaff. 2005. PLUMB: Displaying Graphs to the Blind Using an Active Auditory Interface. In *Proceedings of the 7th International ACM SIGACCESS Conference on Computers and Accessibility* (Assets '05), 182–183. DOI:https://doi.org/10.1145/1090785.1090820

[6] Michael Cooper, Ben Caldwell, Loretta Guarino Reid, and Gregg Vanderheiden. 2008. Web Content Accessibility Guidelines (WCAG) 2.0. Retrieved April 30, 2018 from https://www.w3.org/TR/WCAG20/

[7] Benjamin Kenneth Davison. 2013. Universal graph literacy: understanding how blind and low vision students can satisfy the common core standards with accessible auditory graphs. (April 2013). Retrieved November 29, 2018 from https://smartech.gatech.edu/handle/1853/47621

[8] Leo Ferres, Gitte Lindgaard, and Livia Sumegi. 2010. Evaluating a Tool for Improving Accessibility to Charts and Graphs. In *Proceedings of the 12th International ACM SIGACCESS Conference on Computers and Accessibility* (ASSETS '10), 83–90. DOI:https://doi.org/10.1145/1878803.1878820

[9] Leo Ferres, Gitte Lindgaard, Livia Sumegi, and Bruce Tsuji. 2013. Evaluating a Tool for Improving Accessibility to Charts and Graphs. *ACM Trans. Comput.-Hum. Interact.* 20, 5 (November 2013), 28:1–28:32. DOI:https://doi.org/10.1145/2533682.2533683

[10] Leo Ferres, Avi Parush, Shelley Roberts, and Gitte Lindgaard. 2006. Helping People with Visual Impairments Gain Access to Graphical Information Through Natural Language: The iGraph System. In *Computers Helping People with Special Needs* (Lecture Notes in Computer Science), 1122–1130.

[11] Leo Ferres, Petro Verkhogliad, Gitte Lindgaard, Louis Boucher, Antoine Chretien, and Martin Lachance. 2007. Improving Accessibility to Statistical Graphs: The iGraph-Lite System. In

Proceedings of the 9th International ACM SIGACCESS Conference on Computers and Accessibility (Assets '07), 67–74. DOI:https://doi.org/10.1145/1296843.1296857

[12] Leo Ferres, Petro Verkhogliad, Livia Sumegi, Louis Boucher, Martin Lachance, and Gitte Lindgaard. 2008. A Syntactic Analysis of Accessibility to a Corpus of Statistical Graphs. In *Proceedings of the 2008 International Cross-disciplinary Conference on Web Accessibility (W4A)* (W4A '08), 37–44. DOI:https://doi.org/10.1145/1368044.1368053

[13] John H. Flowers, Dion C. Buhman, and Kimberly D. Turnage. 2005. Data Sonification from the Desktop: Should Sound Be Part of Standard Data Analysis Software? *ACM Trans. Appl. Percept.* 2, 4 (October 2005), 467–472. DOI:https://doi.org/10.1145/1101530.1101544

[14] Ronja Frimalm, Johan Fagerlönn, Stefan Lindberg, and Anna Sirkka. 2014. How Many Auditory Icons, in a Control Room Environment, Can You Learn? (June 2014). Retrieved November 26, 2018 from https://smartech.gatech.edu/handle/1853/52084

[15] Stavros Garzonis, Simon Jones, Tim Jay, and Eamonn O'Neill. 2009. Auditory Icon and Earcon Mobile Service Notifications: Intuitiveness, Learnability, Memorability and Preference. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '09), 1513–1522. DOI:https://doi.org/10.1145/1518701.1518932

[16] Cagatay Goncu and Kim Marriott. 2008. Tactile Chart Generation Tool. In *Proceedings of the 10th International ACM SIGACCESS Conference on Computers and Accessibility* (Assets '08), 255–256. DOI:https://doi.org/10.1145/1414471.1414525

[17] Jeffrey Heer, Michael Bostock, and Vadim Ogievetsky. 2010. A Tour Through the Visualization Zoo. *Commun. ACM* 53, 6 (June 2010), 59–67. DOI:https://doi.org/10.1145/1743546.1743567

[18] Satoshi Ina. 1996. Computer Graphics for the Blind. *SIGCAPH Comput. Phys. Handicap.* 55 (June 1996), 16–23. DOI:https://doi.org/10.1145/231674.231677

[19] Caroline Jay, Robert Stevens, Roger Hubbold, and Mashhuda Glencross. 2008. Using Haptic Cues to Aid Nonvisual Structure Recognition. *ACM Trans. Appl. Percept.* 5, 2 (May 2008), 8:1–8:14. DOI:https://doi.org/10.1145/1279920.1279922

[20] Edward Kim and Kathleen F. McCoy. 2018. Multimodal Deep Learning Using Images and Text for Information Graphic Classification. In *Proceedings of the 20th International ACM SIGACCESS Conference on Computers and Accessibility* (ASSETS '18), 143–148. DOI:https://doi.org/10.1145/3234695.3236357

[21] Angélique Montuwy, Béatrice Cahour, and Aurélie Dommes. 2017. Questioning User Experience: A Comparison Between Visual, Auditory and Haptic Guidance Messages Among Older Pedestrians. In *Proceedings of the 12th Biannual Conference on Italian SIGCHI Chapter* (CHItaly '17), 14:1–14:10. DOI:https://doi.org/10.1145/3125571.3125572

[22] Timothy Muller and Anna C Nobre. 2014. Flow of Time: Perceiving the passage of time: neural possibilities. *Ann N Y Acad Sci* 1326, 1 (October 2014), 60–71. DOI:https://doi.org/10.1111/nyas.12545

[23] Denise Prescher, Jens Bornschein, and Gerhard Weber. 2017. Consistency of a Tactile Pattern Set. *ACM Trans. Access. Comput.* 10, 2 (April 2017), 7:1–7:29. DOI:https://doi.org/10.1145/3053723

[24] Tony Stockman, Louise Valgerour Nickerson, and Greg Hind. 2005. Auditory graphs: A summary of current experience and towards a research agenda. (July 2005). Retrieved November 27, 2018 from https://smartech.gatech.edu/handle/1853/50097

[25] A. Väljamäe, T. Steffert, S. Holland, X. Marimon, R. Benitez, S. Mealla, A. Oliveira, and S. Jorda. 2013. A review of real-time eeg sonification research. (July 2013). Retrieved November 26, 2018 from https://smartech.gatech.edu/handle/1853/51645

[26] David Worrall. 2018. Sonification: A Prehistory. (June 2018). DOI:https://doi.org/10.21785/icad2018.019

[27] Hong Zou and Jutta Treviranus. 2015. ChartMaster: A Tool for Interacting with Stock Market Charts Using a Screen Reader. In *Proceedings of the 17th International ACM SIGACCESS*

Conference on Computers & Accessibility (ASSETS ’15), 107–116. DOI:https://doi.org/10.1145/2700648.2809862