VisRecord

A real-time visualization to record daily life sound

14th January 2018

DT2140 Project MR4

Member: Guanghao Guo, Sihan Yuan, Yumin Hong, Yuxiang Liu

Supervisor: Mario Romero

Video URL: https://vimeo.com/251052814

Source file: VisRecord_MR4

Abstract

While visualization technology has been developing rapidly these years, now it can be used in various areas. This project focuses on sound visualization, which can be used to generate real-time graph by extracting characteristic from the captured sound. Also, a timeline corresponding to the volume and central frequency is also in the interface. Users can save visualization image of any episode in record history. This project is useful for recording the neglected sounds of everyday life and for memorizing sound through visual modality. We develop this visualization using processing and javascript (Mainly p5.js library). After that, we use two rounds evaluation to do iterative design and improvements, according to subjective and quantitative feedback.

Background

Visualization is the technique to create images, diagrams, or animations to communicate a message [1]. As human always discover and understand the world through interactive visual sensations, visualize the non-visual information is a more natural way for people to understand and communicate [2] .Nowadays visualization has been applied to many fields including science, education, engineering, interactive multimedia, medicine and so on. For example, using visualization of spatial position to present voices from unique locations around a listener's head to enhance the perception and cognition of auditory events for a variety of listening tasks [3].

There are many previous course-related projects, such as MEEBLE -- a meeting record software which can record the whole process of meeting and do some semantic analysis based on that such as frequently-used words and attending members; Music-tree[4] -- a colorful simulation tree based on input sounds such as piano sounds and many other instrument sounds, then transform it into different shapes of the tree; SPEAKASSO[5] is a web-based, interactive visualization generator that creates an artistic interpretation of your speech or conversation and SounDark[6], by team Hubris, is a thriller game set in a surreal environment where screaming is recommended.

The goal of the project is to capture the sound of daily life and make it meaningful. The real-time changing graph can be used as the auxiliary animation of music player or can be used as a tool for art education for early childhood since people perceive more information visually than other senses. One important application of this project can be the recording of daily life sound, transforming the noise or meaningless sound to a meaningful and beautiful graph.

As users can also check the previous graph and save it by clicking the mouse, the project combined visual, tactile and sound interaction, which is more like what we experience every day, and users have a sense of naturalness that derives from the use of interaction modes similar to the ones used in everyday human-human interactions[7], multimodality makes this system more lively.

Aim and design goal

This project we decide to do is a system that can keep the graphic record by speaking directly, or playing music, or any sound produced in daily life. The project will be a visualization that indicates specific sound features. Specifically, we plan to achieve following goals:

1. Real-time voice visualization: When user says something, there is an object on the screen that can change dynamically (size, color, line, thickness, shape, etc.) according to real-time human voice input, to some degree same as Siri. But we decided to design a more attractive graph.

2. Semantic and emotion detection: The system can detect semantic meaning if input sound is from person speaking. This result can use in interface design, to create emotional perception.

3. Image generation: After the speaking input, the system will print a final static image, according to his or her voice characteristic, utterance meaning and selected shape.

Implementation

1. Interaction Design

Our main methods of interaction are sound input and visual output, which realized sound visualization and made the interaction more natural and lively.

Every day we listen to different sound which is intangible, thus it is hard to memorize the sound as well as the simple daily life. What we did here is to present sound visually, making it intuitional that users get feedback immediately. The reason we visualize the sound, instead of making haptic feedback or using other multimodalities, is that we always discover and understand the world through interactive visual sensations, visualize the non-visual information is a more natural way for people to understand [2] .Besides the sound modality and visual modality, we also use the touch modality to allow the user to interact with the system, because the direct-manipulative property of touch is often viewed as natural, intuitive and easy to use [8].

There is also limitation relating to our project. For example, as the sound serves as input modality, it is impossible to use a sound control to make command, also the operation should be real time and user have to be near with the electronic devices they are using.

We use javascript and processing to develop these interaction, including sounds analysis, graph generation, and interaction with interface.

2. Graph Design

The entire image of visualization is combined from several parts: (1) Central semi-transparent circle; (2) Three-color rings of high, medium, low pitch; (3) Rounding spectrum waveform; (4) Backward rotary flower shape. The structure is illustrated in figure 1.

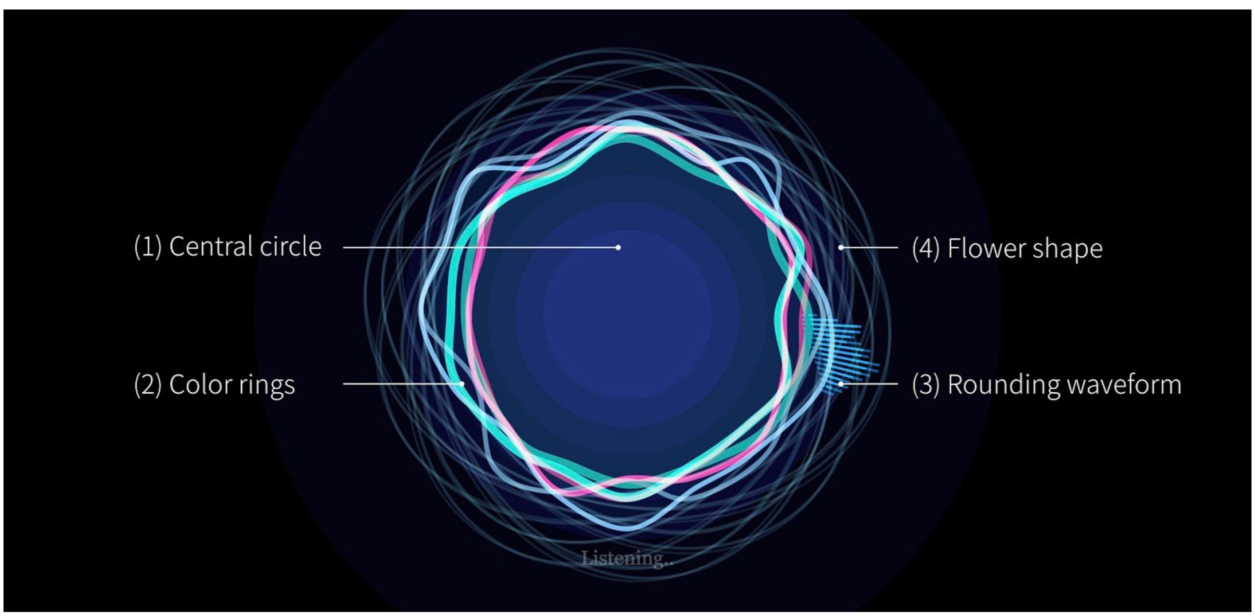

Figure 1. Structure of visualization

Sounds variable	Visual elements
Volume	Size of central semi-transparent circle, timeline element

High frequency	Size, color and thickness of the ring/Rotation degree
Medium frequency	Size, color and thickness of the ring/Vertical radius
Low frequency	Size, color and thickness of the ring/Horizontal radius
Centre frequency	Color of central semi-transparent circle / timeline element
Sound spectrum	Size and color of rounding waveform

Table 1. Mapping between sound and visuality

The whole mapping variables are described in table 1. The Central semi-transparent circle is mapped by volume. Radius changes according to volume. Transparent degree and the color are related to centre frequency (or mean frequency), which is calculated by Peeters G [9]. The higher centre frequency, the more untransparent and warmer color. This color mapping is shown in the bottom of figure.x.

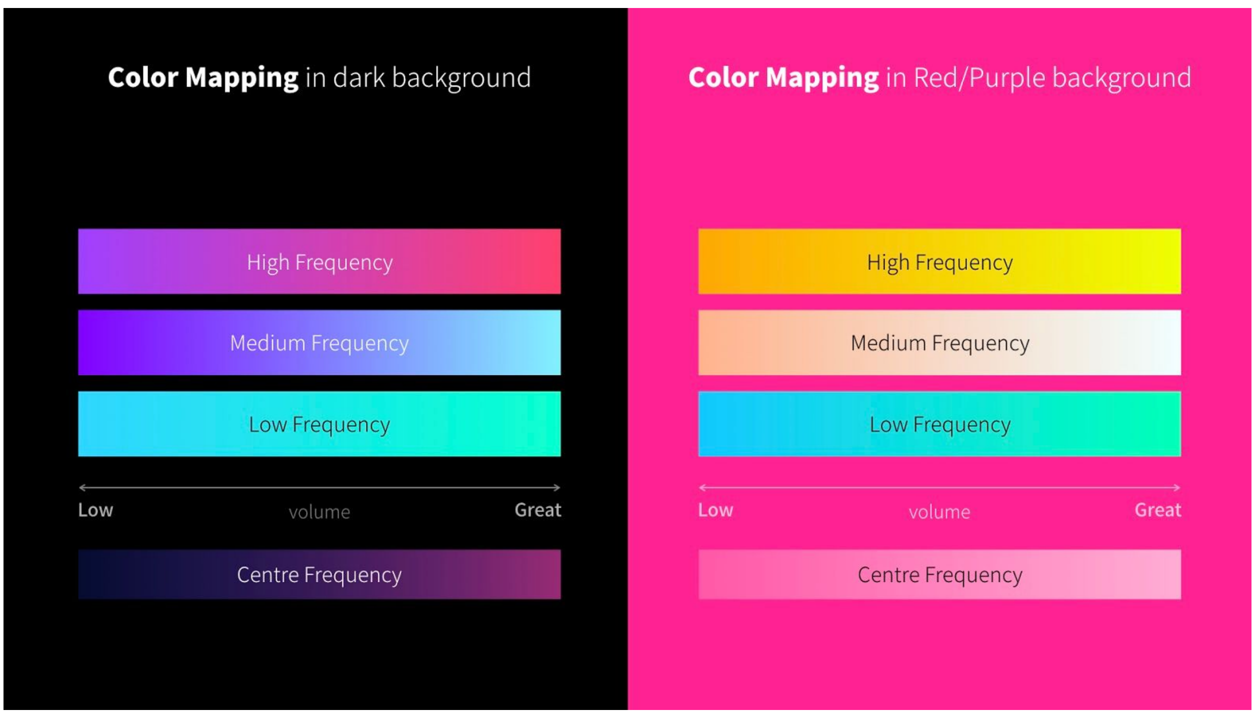

Figure 2. Color mapping

The three rings are mapped with high, medium, low frequency, which is the most remarkable in this visualization graph, because these three variables should be easy to perceive so that people can understand features of sound more comprehensively. The color mapping is shown in figure 2. Take dark background as an example, for each pitch, low volume is encoded with low brightness color, while great volume is encoded with more bright and saturated color. The thickness and “shake” of the ring are also variable according to frequency volume. Great high-frequency volume can be a good example for describing our ideas: High frequency can interpret as “sharp” and “thin” image on our mind, thus we use think and more “shake” line.

The backward flower shape is made from ellipse, which horizontal and vertical radius are encoded by low and medium frequency respectively. Its rotation degree is about high frequency. And because delay of human visual perception (framerate is 30, so human will perceive image as smooth animation rather than single continuous image). This shape generation is described in figure 3.

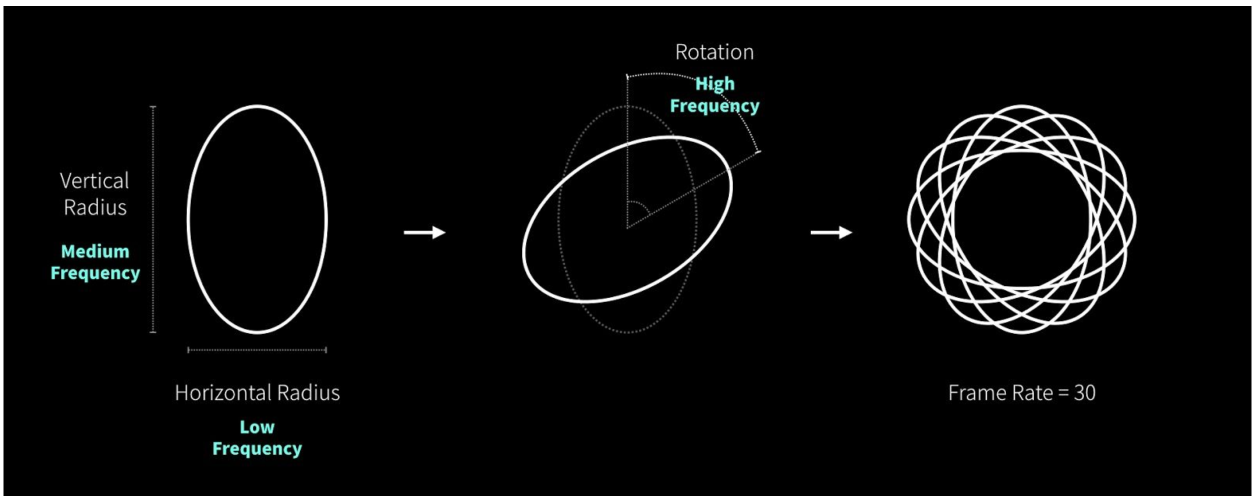

Figure 3. Backward flower shape generation process

Design goal of timeline is to record main features of instantaneous sound piece. So we use color to encode mean frequency of sound piece, and use size to encode volume. The greater mean frequency is, the bright color we use. In addition, this design should reflect not only volume and mean frequency of sounds piece, but also compactness and intenseness of period of sound. The timeline design are shown in figure.x.

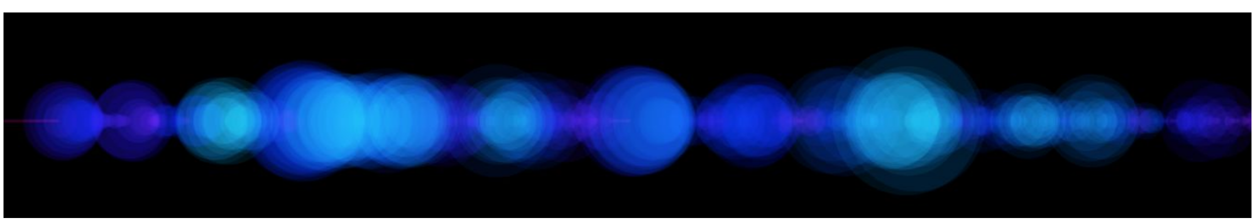

Figure 4. Timeline design

3. Image Record and Retrospect

After receiving the real-time images, we decided to add a new function to record the voice and generate a picture from the recorded voice so that users are able to save an superimposed picture made up from two frame images of a specific time.

The first step of realizing these functions is to record the sound, which means record the basic characteristics of the sound, such as volume, the energy of low frequency, middle frequency and high frequency as well as the value of spectral centroid. Thus we implement 5 arrays to store these values of every frame. Then we create a timeline which represent the time flow of the recording sound. We draw continuous circle to constitute the timeline, and map the value of volume and mean frequency to the color and size of the circle.

In order to use these stored values to reproduce the image which has been generated few seconds ago, we need to introduce a new interaction method to use the mouse. By simply clicking the mouse at the specific location in the timeline, we can get the value of basic sound characteristics which stores into the array and then use these values to reproduce a image by using a newly defined function just similar to the method of generating the real-time image. As the function we used can only save the current image shown on the screen, we create a new canvas so that the downloaded picture will not be overlapped by the real time image.

Evaluation

We have done two-round evaluation. The first round was evaluated by teammate personal assessment and feedback from presentation in Jan 11. The second round was in Jan 13, after we do iterative design and improvement according to feedback of 1st round.

In the first round evaluation, we have achieved basic functions according to design goals, including real-time voice visualization and graphic generation, and we find that this visualization can present totally different image according different sounds, figure 5 is some example images we save from our test. But these functions are too basic to make sense. In our idea, it is not a complete project. In presentation, the feedback of supervisors is that the design of mapping between sounds and visuality is not clear, and function is too simple. Therefore, followed supervisors' suggestion, we designed the timeline, and developed a new function to save image, additionally, made the improvement of visual design.

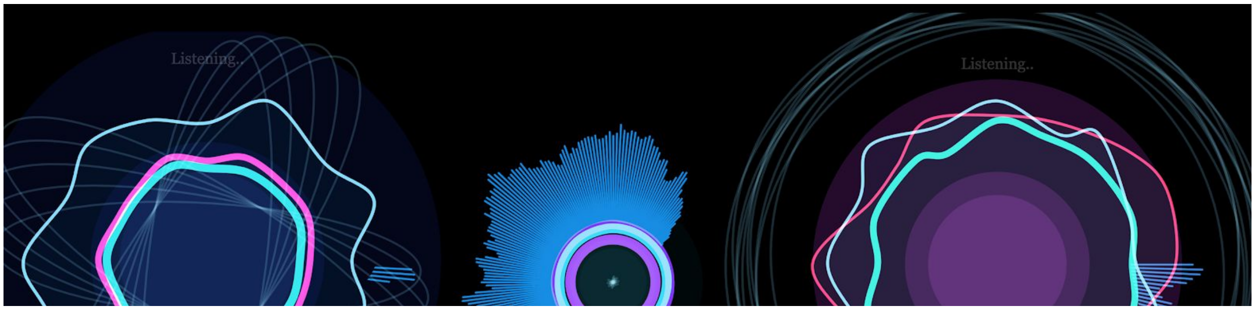

Figure 5. Examples of visualization result

In the second round, we divided evaluation into three parts: Visual present, semantic detection, image generation and save. However, the semantic part is fail because of technology troubles, so we can not evaluate this part. We have 3 test subjects from different background, use think aloud test method, asking users to think out loud as they are performing a task[10], and then ask them give us subjective feedback and quantitative score. We specify the evaluation question and set up score criterion according to table 2.

Score	**Part 1: Visual present**	**Part 2: Semantic detection**	**Part 3: Image generation**
A	Visualization has cool visual effect, can express the characteristics and emotion of sounds , and give real-time visual response.	Fail to implement	User can understand the timeline, and save sound image of any episode freely and easily.
B	Visualization has good visual effect, can present dimensions of sound exhaustively, and give real-time visual response.		User can recognize the timeline, and save sound image of any episode.
C	Visualization has good look, can present some dimensions of sound, and give appropriate visual response.		User can recognize the timeline, but has some trouble with saving sound image of any episode.

Table 2. Evaluation score criterion

In first part, every test subject is surprised with our visualization visual effect. “It’s really cool and gorgeous ” is subjective appraise of them. The most trouble thing they have is that “I can’t understand what these three circles represent for, even if I notice they are related”. For quantitative scores in this part is B, B, A-. They don’t think this visualization can present every dimension exhaustively, especially timbre, and they don’t understand exact visual element represent for. But because of cool visual effect, one of them give us A-.

In part 3, visual effect of timeline also receives big praise. What’s more, the timeline design is easy to understand, after they try several attempts to record different kinds of sounds. And there is not trouble when they attempt to save image. However, according to their feedback, on the one hand, the saved image is not gorgeous as real-time present, on the other hand, they can only check picture after they download, rather than checking firstly and decide whether to download, which is not friendly for user. For quantitative, they all give us B.

To sum up, feedbacks of visual effect, visual meaning, and functionality are different. Visual effect win great praise, while visual meaning still has space to improve. As for functionality, it can satisfy basic need actually, but we know it is stupid and crude as well, which can be increased in the future.

Discussion

We consider the sound visualization the most proud part of our project. The visualization we generate is really fabulous and gorgeous. As greenhand of coding, we have to learn a new programming language javascript and know how to take advantage of p5.js library to design creative interaction. Apart from that we learn to use design process to develop our product, namely take iterative design according to several evaluation do improvement. There always be challenges during the process, especially technique barriers. We have tried for semantic detection and camera-based interaction but failed.

After the presentation, we got some advice from our supervisor Mario. He suggested that it would be better to add one more function, such as recording, to make this product more usable. We continued working on this project after presentation, and successfully realized this function, even though it is limited.

The project itself has room for improving. Not only about functionality, but also design mapping. The future work can be divided into two parts. Firstly, sound analysis should include timbre, so that system could know what kind of sound it is. Timbre detection capability is not available in p5.js, even in javascript library. If we can detect timbre, the design mapping would be enhanced significantly. Secondly, we consider develop record function in high usability, for example, currently we develop our system in web, but people will walk around in different place, so it might be more useful if it is developed in mobile setting.

Responsibilities

Guanghao Guo: Did the part of sound processing, which extracts the basic characteristics of the sound including volume, pitches, and spectral centroid. Managed to write a mouse interaction function. When you click the mouse, you can obtain an array which contains the value of basic sound characteristics.

Sihan Yuan: Designed the visualization graph, but the design was not used in the project because the mapping is not very clear. Wrote the function which used the stored value of previous frame to draw the current picture of a specific frame and save the two superimposed frame graph.

Yumin Hong: Design and code sound visualization, including visual design, interaction design and visualization development. As the project idea generator, be responsible for leading discuss, project management and technology research. Finally, did an evaluation and make the video.

Yuxiang Liu: Did the background research and drew up the initial plan. Managed to set up the local testing server. Successfully did the semantic recognition with only a few word using p5.js but failed to integrate with the main function. Managed to reproduce the previous images with others, and managed to store the detected data in real-time for later use. Lastly, refined the design with other group members and finished the report.

Conclusion

Based on the design goal we set up, we have done several functions of the project: A multimodal interface includes real-time visualization, the interactive timeline, and interactive background. The sound visualization encodes several features of sounds into visual elements. The timeline presents the main features of historical sounds piece and compactness of sound period. Additionally, user can change the background color by touching the screen. The image recording and saving function mean that we can reproduce the visualization status anytime in the timeline, and store it as the image file (.png file) by easy mouse interaction. These results are shown in the video. We have done two rounds evaluation, and made design improvement according to test feedback. The most valuable for us is not only the project outcome itself, but also we learn a new language—Javascript, and learn how to use the p5.js library to create interesting stuff.

References

[1]En.wikipedia.org. (2018). *Visualization (graphics)*. [online] Available at: https://en.wikipedia.org/wiki/Visualization_(graphics) [Accessed 14 Jan. 2018].

[2]Lucente, Mark, Gert-Jan Zwart, and Andrew D. George. "Visualization space: A testbed for deviceless multimodal user interface." Intelligent Environments Symposium. Vol. 98. 1998. APA

[3]Kilgore R, Chignell M. Listening to Unfamiliar Voices in Spatial Audio: Does Visualization of Spatial Position Enhance Voice Identification[J]. Human Factors in Telecommunication, 2006.

[4]Anon, (2017). Music tree [online] Available at: http://www.cawards.se/project/music-tree/ [Accessed 29 Nov. 2017].

[5]Csc.kth.se. (2017). SPEAKASSO. [online] Available at: http://www.csc.kth.se/~acvds/info_vis/speakasso/ [Accessed 29 Nov. 2017]

[6]Hubris37.github.io. (2017). SounDark. [online] Available at: https://hubris37.github.io/Sonar/ [Accessed 29 Nov. 2017].

[7]Caschera, M. C., D'Ulizia, A., Ferri, F., & Grifoni, P. (2012). Towards evolutionary multimodal interaction. Lecture Notes in Computer Science (including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 7567, 608-616.

[8]Olwal, Alex. "An Introduction to Augmented Reality." An introduction to augmented reality (2009).

[9]Peeters, Geoffroy. "A large set of audio features for sound description (similarity and classification) in the CUIDADO project." (2004).

[10]Jaspers, Monique WM, et al. "The think aloud method: a guide to user interface design." International journal of medical informatics 73.11 (2004): 781-795.