Vis Diary

Graphic diary using voice and speech visualization

1st December 2017

Project MR4

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

Supervisor: Mario Romero

Background

Many people keep the habit of writing diary to record their life. However, when they want to look back on their life, it is difficult for them to have an overview: unless they read the written words again and do quantitative analysis, they can not see how they feel everyday or what is the proportion of feeling happy or upset in a period of time.

Besides the difficulty of overview, many people who wants to reflect their life may consider it as troublesome to keep the habit of writing diary, which requires much time and effort to do it. In fact, many people are willing to record their life if there is an easy way to do it.

While reducing the complexity of writing diary is important to solve such problem, preserving the uniqueness of individual recording needs to be considered as well. If we simply transform the writing process to single graph, there will be a loss of personalization and differences between everyday life. As everyone is unique, we would like to develop a graphic diary for users through using their voice information, which includes voice attributes and semantic meaning.

As people read same sentence in different volume, tone, pitch, timbre… these variables could be used to express the pattern visually, thus form an unique 'logo' for everyone and everyday, which is a base of graphic diary.

Aims and Delimitation

This project we decide to do is a system that can keep graphic diary by speaking directly, and the diary will be a visualization that indicate human individuality and daily emotion. Specifically, we plan to achieve following goals:

1. Real-time voice visualization: When user say 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 decide to design a more attractive graphs.
2. Emotion recognition: This system can recognize emotion and understand what user's utterance. Some graphic attributes, such as color, will be changed according to what he says.
3. Gesture recognition: User could use hand gesture to change object's basic shape (Circle, rectangle…) when saying.
4. Graphic generation: After the speaking input, the system will print a final static graph, according to his or her voice characteristic, utterance meaning and selected shape. This logo is unique, to demonstrate the certain person in a certain day.

There are also some delimitation:

1. We only recognize the emotion part of semantic meaning, for example, we want to know if user is happy, upset or angry, rather than the whole meaning of utterances.
2. User can only speak one short sentence as input utterance, rather than a long description of his day experience.

Set up

To solve the problem and achieve our goals, we use some modalities to design our system: Voice attributes, utterance meaning, gesture and visual expression.

To begin with, we plan to focus on two different aspects of sound. Firstly, the features of sound itself. As the volume, tone, pitch and timbre of sound differs, we can take advantage of these features to graphs that represent people's individuality and different daily recording. Second, the utterance meaning. We want the system to understand the meaning of sentence, especially emotion meaning, said by the user and use it as an element to design the unique graph as a personal 'logo'. Therefore, we will refine key data from these two aspects to formulate visualization design.

Moreover, we can make the system smarter by introducing hand gesture to modify the generated graphic. As spoken word cannot be withdrawn, the hand gesture modification provides user the chance to edit the graphic basis.

It is more efficient for people to perceive and understand information visually, also the direct-manipulation interfaces is easier to control, and picture is often cited to be worth a thousand words and easier to use than is a textual description or a spoken report[1]. Thus, we transform the voice and gesture input to visual expression to reach the goal of graphic diary recording.

Basically, We divides this project implementation to 5 parts: Voice recognition, emotion recognition, voice visualization, gesture recognition and graphic generation.

Voice Recognition

When users say 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 just like Siri. But we decide to design a more attractive graphs. We decide to use P5.js to capture the voice and identify different characteristics such as pitches, voice, frequency and things like that. P5.js is a javascript library that starts with the original goal of processing, but it is more suitable to process the sound.we can also use some algorithm to realize voice recognition,such as pitch extraction algorithm[2].

Emotion Recognition

The graph can also shows the emotion of people, if applicants say some words with emotional tendency, such as happy, sad, frustrated or any other emotions, the graph can show with different color. For example, if you are in a good mood, the graph is more likely to be warm-colored, otherwise, would be cold-colored if you are blue. The technology behind this would be semantic analysis based on p5.speech, and mapping these words which has been recognized into different color of graph.

Voice Visualization

As humans discover and understand the world through interactive visual sensations[3], we want to make the voice visualized as well. Once the user's sound is recorded, the characteristics in the sound can be captured and transformed into visual patterns. Moreover, different characteristics refers to different graphic features. For example, the volume of sound controls the size of pattern, and the pitch of sound is responsible for sharpness of the graphic edge. This mapping makes the sound more intuitive.

Gesture Recognition

The novel device Leap Motion Controller provides an informative representation of hands, According to *Analysis of the accuracy and robustness of the leap motion controller*[4], the Leap Motion Controller provides a detection accuracy of about 200 µm. We plan to utilize tracking data through the API of Leap Motion Controller to recognize hand movement and gestures, which can be used to change object's basic shape (Circle, rectangle…) at run time.

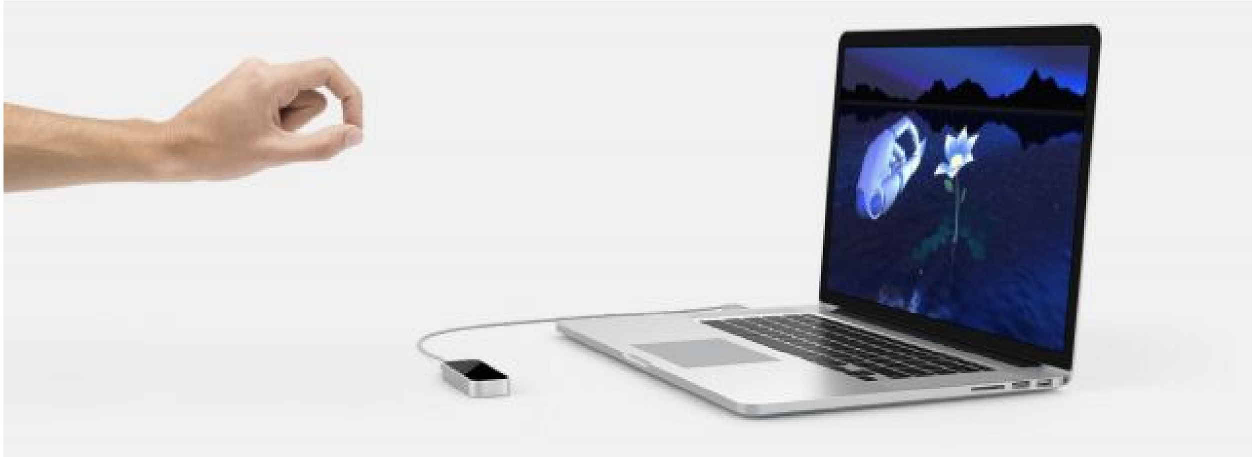

Figure 1: Leap Motion Usage. The small object in the middle is Leap Motion Controller connecting to the Mac on the right. Hand on top of the Leap Motion is tracked and interacted with virtual objects.

Graphic Generation

After input voice, the system will print a final static graph, according to his voice characteristic and selected shape which is factored out in the process of voice recognition and gesture recognition. This logo is unique.

Evaluation scheme

Our evaluation criteria are set according to the sub-goals:

1. Real-time voice visualization: Graph on the screen that can change dynamically (size, color, line, thickness, shape etc.) according to real-time human voice input.
2. Emotion recognition: Color can be changed according to the speaker's emotion, such as happiness, sadness.

3. Gesture recognition: User could use hand gesture to change object's basic shape (Circle, rectangle…) when speaking. The priority of this part is the lowest.
4. Graphic generation: The output of system is a picture of the graph.

Our evaluation scheme is based on two parts: Firstly we check if the system can run and achieve expected functions. The functional criterions are listed. Secondly, we will use user feedback and user satisfaction to evaluation.

Project plan

Phase 1: Learning techniques and developing separately

We divide the entire project into 5 subtasks, including voice recognition, emotion recognition, voice visualization, gesture recognition and finally graphic generation. Every task corresponds to certain members of the team. After individual work, integration needs to be done together to realize the whole project aim. Because we all don't have much experience in these techniques, so every member will learn corresponding technique and share it with others. The priority of gesture recognition is the lowest, thus if we don't have enough time and resources, we will give up this part.

Phase 2: Testing lean system

In this phase, we combine each parts to make sure the whole system can run successfully, though it might be primary and rough. Then we will do user test to see the usability problems and figure out improvement opportunities.

Phase 3: Refining the design and doing report.

After user test, we plan to do iteration design, according to user feedbacks, usability problem. In this phase, we will make visualization more beautiful and attractive.

Responsibilities

Guanghao Guo and Yuxiang Liu are responsible for voice recognition and gesture recognition. Sihan Yuan and Yumin Hong are responsible for voice visualization and graphic generation.

Time plan

- Dec 1: Submit Pre-study assignment.
- Dec 7: Finish phase 1. Finish voice recognition part and voice visualization part. For voice recognition, different attributes of real-time human voice should be recognized. For visualization, graphics should be changed according recognized voice data.
- Dec 20: Finish phase 2. Visualization design should be improved. Finish the part of gesture recognition.
- Jan 1: Finish phase 3. Have Refined all parts.

Risks analysis

1. Lack of knowledge of sound visualization related coding: We can start search for some technology-related materials and practice programming skills based on p5.js
2. Not being familiar with speech recognition technique: we can learn some basic techniques by watching some tutorial videos on the youtube given by The Coding Train.
3. Do not have experience of Leap motion development:find some example code on the github,understand that and try to write our own programs to realize our functionalities.

Related work

There are some works related to our project. MEEBLE is a meeting record software which can records the whole process of meeting and do some semantic analysis based on that such as frequently-used words and attending members. We can draw some experience from their methodology of semantic analysis. SPEAKASSO[5] is a web-based, interactive visualization generator that creates an artistic interpretation of your speech or conversation. Music-tree[6] is 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. The method of generating graph can be referred to by us. SounDark[7], by team Hubris, is a thriller game set in a surreal environment where screaming is recommended. The player wearing the VR-headset plays as a blind chicken that can only see through echolocation.Spectators, who can see the map from an isometric perspective, can choose to help the player by providing them with information about the maze layout .

There some scholars that do research about life data visualization. Yang Yang and Cathal Gurrin use smartphones to capture detailed lifelogs for individual, they create visualization tools to support user access to lifelogs. [8] Lifelogs offer rich voluminous sources of personal and social data for which visualisation is ideally suited to providing access, overview and navigation. Daragh Byrne present some guidelines and goals which should be considered when designing presentation modes for lifelog content [9].

References

[1] Robertson G G, Mackinlay J D, Card S K. Cone trees: animated 3D visualizations of hierarchical information[C]//Proceedings of the SIGCHI conference on Human factors in computing systems. ACM, 1991: 189-194.

[2] Azar J, Saleh H A, Al-Alaoui M A. Sound Visualization for the Hearing Impaired[J]. International Journal of Emerging Technologies in Learning, 2007, 2(1).

[3] Lucente M, Zwart G J, George A D. Visualization space: A testbed for deviceless multimodal user interface[C]//Intelligent Environments Symposium. 1998, 98.

[4] Weichert F, Bachmann D, Rudak B, et al. Analysis of the accuracy and robustness of the leap motion controller[J]. Sensors, 2013, 13(5): 6380-6393.

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

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

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

[8] Yang Y, Gurrin C. Personal lifelog visualization[C]//Proceedings of the 4th International SenseCam & Pervasive Imaging Conference. ACM, 2013: 82-83.

[9] Byrne D, Lee H, Jones G J F, et al. Guidelines for the presentation and visualisation of lifelog content[J]. 2008.

Presentation

A record for daily life

The product we built can record all kinds of sound such as the voice of baby crying, the noise of plane taking off, even the whispers around people and therefore, generating different kinds of fascinating pictures, and if you are not satisfied with the color which has been generated , you can adjust that by simply clicking the mouse. Imagine such a situation , you wake up by the sound of alarming, wash your face with water splitting out of the faucet, pour milk from the bottle, and dress up to attend school. All these sounds made in the process can be used to generate a series of fabulous pictures, and stored into the cloud. When you would like to recall the things happens yesterday, or even one week ago, these pictures can help you memorize what you have done. And you don’t have to worry about the privacy being invaded, because the product can only record pictures generated by the sound but not the content you talked.

Auxiliary animation for playing music

The product can also be acted as the auxiliary tools of the music, by using that, you could probably get visual appreciation of music works. If you are a fan of rock music, you must like this product, because the pictures generated by the rock music can be really changeable and colorful. If you attend classical music concert and cannot stand the boring rhythm , staring at the gorgeous pictures can also be a good option. What’s more, it can be also used as a auxiliary tools for people with hearing impairment. By looking at the pictures , they can detect if there is speeding cars driving by or even enjoy music concert.

Art education tool for early childhood

As now it is common to use technology of sound visualization in many different fields, the product is useful for early childhood education as well. Naturally people are able to perceive much more information through vision than hearing, thus, through the changing color, size and moving speed, children are able to improve their awareness about sound, making the invisible, abstract conception become concrete.

Technical details

Sound processing

The first step of sound visualization is sound detection and processing. We use processing 5 based on javascript to realize this function , because it has a large varieties of libraries to support sound processing . We can detect the volume and pitch of the sound. When detecting the pitch, fft method will be used