Week (11) - "Investigating Audio Data Visualization for Interactive Sound Recognition"

Khulood Alkhudaidi

Interactive machine learning GUI's are commonly investigated in the context of image and text. This paper aims to study four audio visualization techniques for interactive sound recognition GUI. The researchers suggested a GUI shown in Figure (1). On the left of the interface, the user can define different categories to classify the sounds. In the middle, the audio samples are visualized in 2D space in a hierarchical clustered manner considering the similarity in the feature space. Finally, on the right of the interface, the user can check the details of the chosen audio. Their goal is to find an intuitive visualization technique to visualize the audio samples.

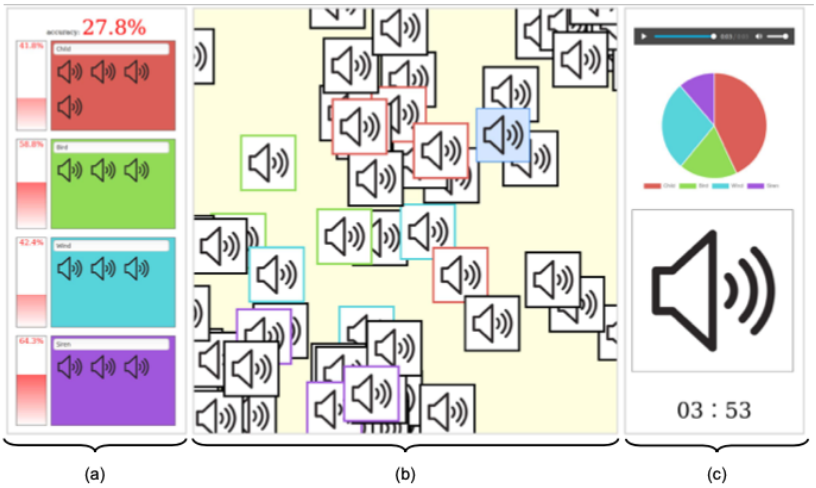

Figure (1)

For interacting with this GUI, users can add a box and name it to create categories on the left panel where the accuracy is displayed too. Also, in the middle panel, they can zoom in/out and drag the unlabeled data horizontally and vertically. They can drag and drop the audio to the corresponding defined class, and they can remove it from a class by dragging it out. Users can listen to the sound from the right panel, which also displays the probability scores. Whenever the user adds audio to a class, the model gets trained and updated. This process is shown in Figure (2).

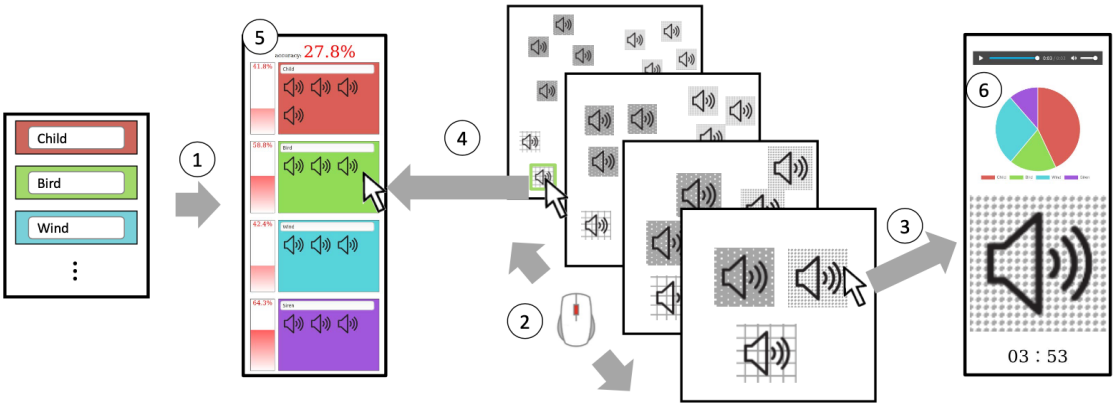

Figure (2)

The researchers tested replacing the speaker icons with four different visualization cues, and determined the pros and cons of each technique. The techniques are:

- *Thumbnail:* frames taken from a video clip if the sound is from a video.
- *Spectrogram:* a popular way to visualize sound characteristics.
- *Semantic:* a text to describe the sound.
- *Audio-to-Image Retrieval:* a resulting image from the process of converting audio to images.

They combined Thumbnails with Spectrograms because they complement each other. Figure (3) shows the clusters of different approaches in the experiments.

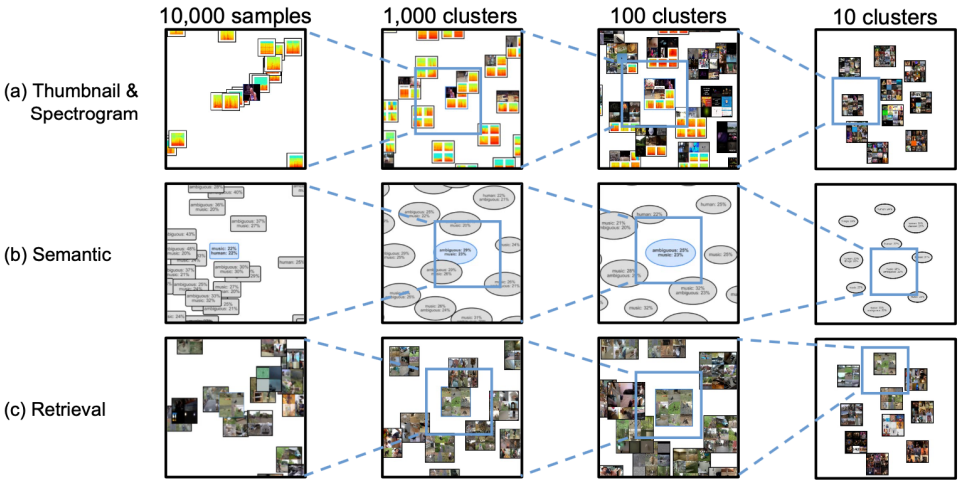

Figure (3)

The participants' interaction logs were recorded, and they were asked to fill a questionnaire to assess the workload.

The result of the experiments is:

- Thumbnail & spectrogram approach was the best.
- Semantic and audio-to-image were good to help users in exploring a large size of the audio sample.
- Retrieval allowed users to get an overview of the sample distribution.

The researchers' stated that each approach has pros and cons, and some techniques work as complementary to others. Therefore they recommended using more than one approach when needed to enhance the visualization.

I agree with using more than techniques because when I thought about it, the best approach is Thumbnails & spectrogram since thumbnails help me know the content of the sound, and spectrogram gives me an idea about the characteristics of the sound. Also, I agree with the researchers' prediction about the high potential of this application for people with hearing loss.