PLAI 600 Final Paper

Kayhan Qaiser, 260507016

1 Summary

The act of listening is not purely a mechanical process. Though listening is dependent on physical sound waves, the perception of that sound is dependent on the internal state of the mind of the listener.

In this project a listener is played a clip of distorted audio and is asked to identify the sentence within. The sentence sounds like gibberish and is unidentifiable. The listener is then told the correct sentence and is asked to listen to the clip again. This time all the listener can hear is the original sentence. Or put another way, it is impossible to 'un-hear' the sentence. A demo of the experiment can be found at http://kayhanqaiser.com/2015/12/09/listening-for-information/.

This experiment demonstrates how what we hear strongly depends on our internal state of mind. Interestingly, it may be possible to train a listener to identify any transformed sentence by training them on the distortion. If this is possible it could serve as a simple cryptography system.

2 Method

The programming for this project was done in C++. Processing was done on the pcm files of the audio samples. The main library used was KissFFT (keep it simple stupid fast Fourier transform).

The audio file was put through a Fourier transformation and then the pitch information as well as phase information were changed.

Initially the project relied on computer generated speech however this was incredibly hard to understand. I then switched to human speech which provided much

better results and gave the type of distortion that was required. Below is selected code that performed the two transformations for the final demo.

```
// shift pitch
int shift = 20;
kiss_fft_cpx * tmpBin = new kiss_fft_cpx[N/ 2 + 1];

for(size_t i = 0; i < N / 2 + 1; i++) {
  int j = (i-shift)%N;
  if (i < 3 || i > N/2 + 1 - 3)
  {
    j = i;
  }
  tmpBin[i] = phi[j];
}
phi = tmpBin;


// decrease phase
for(size_t i = 0; i < N / 2 + 1; i++) {
  if (i < 3)
  {
    continue;
  }
  phi[i].i *= 1/((double)(i)/5);
}
```

All audio recordings went through the same transformations and were tested on subjects individually and in a group setting.

3 Results

This section will cover some of the broad lessons and tentative hypotheses that I have drawn from the experiment.

The most obvious question that must be answered is what is the success rate for subject? Or, did the subject correctly understand the distorted sentence once he had heard the original? Almost all subjects reported understanding the sentence clearly. Approximately 10% of subjects reported not hearing the correct sentence in the distorted audio.

Interestingly, many of the subjects who reported failure also did not speak English as their first language. This does not mean their English was weak, but rather that they all had noticeable accents. It is important to remember that the original audio was recorded in a neutral English Canadian accent.

This phenomenon may have to do with how the transformation affects the audio. The transformation was simply a pitch and phase shift. So, although the words didnt sound quite like words, a lot of information was still preserved. For example intonation, pause and rhythm were all still present. These are all aspects of what linguists call prosody, or how we say units of speech. If one of the things preserved

in the distorted audio was prosody structures, then it makes sense that people who had accents were not able to understand.

In further tests I observed that numbers and months were very easily identified by participants. This may be because we are very sensitive to this information because of the number of times we hear them in our daily lives.

Interestingly some of the sentences were understood in very different ways by participants. In other words participants heard two different meanings from the same distorted audio. This suggests that the distorted audio may have more information than the original.

4 Further Experiments

One experiment which I would like to do is to try and train a participant to always understand the distortion. This would involve having him hear distorted paragraphs of text while reading the original text. The hypothesis is that he will eventually be able to understand any distorted audio he hears.

There is an analogy here to understanding a very complex accent. After being exposed to a foreign accent we eventually learn to understand it. Although this is an extreme example of an accent it may still be possible. If so, it could provide a simple encoding system. Audio could be played on a public system but only the people who have trained on the distortion will be able to understand it.

5 Further Information

I'd like to thank Marc Burns for his help with creating the distortion for this project as well as all the participants that I tested.

Another demo of the auditory illusion described above can be found below. This demo relies on a different transformation and was created by Scientist Jayatri Das. It can be found at https://soundcloud.com/whyy-the-pulse/an-audio-illusion.