Voice Recognition: A Foundation

Andy Yong

Follow this and additional works at: https://mosaic.messiah.edu/honors

Part of the Electrical and Computer Engineering Commons, and the Physical Sciences and Mathematics Commons

Permanent URL: https://mosaic.messiah.edu/honors/167

Recommended Citation

https://mosaic.messiah.edu/honors/167

Sharpening Intellect | Deepening Christian Faith | Inspiring Action

Messiah College is a Christian college of the liberal and applied arts and sciences. Our mission is to educate men and women toward maturity of intellect, character and Christian faith in preparation for lives of service, leadership and reconciliation in church and society.
"I grant permission to Dr. Lamarr Widmer for the use of my paper Voice Recognition: A Foundation, for educational purposes."

"I give permission for public access to my Honors paper, and for Messiah College to maintain and provide access in electronic format for online use. In granting this limited permission, my intention is not to surrender my intellectual property rights, but to promote scholarship by providing electronic copies for personal study and not-for-profit education use. By providing this permission, I also certify that, to the best of my knowledge, my work does not infringe upon any third-party copyright."

Andy Yong ___________________________ 14th May, 2013

name ___________________________ date
Andy Yong
Senior Mathematics Seminar
Dr. Widmer
May 12, 2013

Voice Recognition: A Foundation

Context:
Technology has seen such an exponential growth in the past few years that it is almost inconceivable to think that the World Wide Web was only created just 23 years ago. New technology is being created and innovated everyday to make daily activities more convenient. One particular subset of new technology that has been gaining much attention is voice recognition.

While voice recognition has been around for quite some time, it did not gain a lot of attention until Apple released the iPhone 4s which included the famous, Siri. Siri is the name of the voice recognition program installed in iPhones following the 4s. Siri was not very different from other voice recognition programs in terms of functionality, however, it stood out in terms of accuracy and usability. Siri’s success brought voice recognition programs out of its previous purpose as a tool for people who needed it for accessibility reasons into a helpful assistant for any regular person.

Speaker-Dependent vs. Speaker-Independent Systems:

There are two kinds of voice recognition systems: speaker-dependent and speaker-independent. A speaker-independent system utilizes a library of phonemes, words and logic queries to match speech to text. A phoneme is a basic unit of a phonetic language that make words when combined with other phonemes. A speaker-dependent system differs
from speaker-independent systems as it also has the capability to learn and improve on its accuracy. Speaker-dependent systems is able to learn a specific user’s voice, accent, typical vocabulary, pronunciation, sentence structure tendencies, and speech patterns depending on how detailed the system is programmed.

As such, speaker-dependent systems are generally more efficient because they can hone in on a particular preset for a particular user. This allows the system to gain a higher accuracy when interpreting speech. For example, a speaker-dependent system may learn that a user is British through corrections and inflection recognition; which allows the system to bypass all other accents in the analysis stage. A speaker-dependent system can also learn frequency signatures to determine which user is using the system in order to pull up the “dictionary” for a particular user.

**System Accuracy:**

Accuracy of both voice recognition systems is dependent on multiple factors. These factors can be divided into two categories: physical and systemic. Wind, noise and loudness are some examples of physical factors. Systemic factors include the quality of the filters, how noise-reduction is implemented, phoneme recognition, sentence logic and frequency differentiation. While all these factors play a significant role in determining the accuracy of the recognition process, the primary focus of this project will be on frequency differentiation.

This paper will explore the concept of using the frequency domain to recognize differences between two separate input signals. Furthermore, this paper will discuss the use of the different Fourier analyses to obtain frequency representations of audio recordings. Finally, using those two discussions, this paper will strengthen the argument that differential equations is indeed an important tool in the design of voice recognition systems.
Fourier’s Contribution:

In order to examine frequencies for comparison, we would have to analyze the input in the frequency domain. One would have to use some form of Fourier analysis to assess the frequency domain. Jean Baptiste Joseph Fourier was born on the 21st of March 1768 in Auxerre, France. Fourier was a mathematician and physicist who worked on the famous propagation of heat in solid bodies problem. He was also the initiator for the study of the Fourier analyses.

There are four kinds of Fourier analysis: Fourier Series, continuous Fourier Transform, discrete-Time Fourier Transform, and discrete Fourier Transform. For the purpose of the project, we only need to discuss the Fourier Series and discrete Fourier Transform. According to Fourier, all periodic waves can be decomposed into the sum of simple sine and cosine waves. The formulas for 2π-periodic functions using the Fourier Series (F1, F2 & F3) and a visual example (G1) are shown below.

\[a_n = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \cos(nx) \, dx, \quad n \geq 0 \quad (F1)\]

\[b_n = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \sin(nx) \, dx, \quad n \geq 1 \quad (F2)\]

\[(S_N f)(x) = \frac{a_0}{2} + \sum_{n=1}^{\infty} [a_n \cos(nx) + b_n \sin(nx)], \quad N \geq 0 \quad (F3)\]
The Fourier Series does more than just recreate complicated periodic functions, when it is used on a function in the time domain, the result is the frequency representation of the function. This result is what we want to use to analyze two waves because the difference in the composition of two functions can help us examine signatures or unique details that make one distinct from or similar to another. However, there is a problem with using the Fourier Series. The Fourier Series is a solution to a mathematics problem, which may not be realistic when it comes to actual implementation. Most input signals for voice recognition systems are finite and not periodic. Consequently, the Fourier Series cannot be applied; which brings us to the introduction of the discrete Fourier Transform.

Discrete Fourier Transform Enters:

The discrete Fourier Transform (DFT) converts a finite list of samples of a function into the coefficients of a finite combination of complex sinusoids, ordered by their
frequencies. In simpler terms, the discrete Fourier Transform converts a sampled function from its original domain to the frequency domain. Since this project is about voice recognition, our input signal be in the form of audio signals, where the original domain would be the time domain. The formula for the discrete Fourier Transform of an N-periodic sequence of complex numbers from a sequence of N complex numbers \( x_0, \ldots, x_{N-1} \) is shown below (F4).

\[
X_k = \sum_{n=0}^{N-1} x_n \cdot e^{-i \frac{2\pi k n}{N}} \quad (F4)
\]

Fast Fourier Transforms:

It is pretty clear that the formula for the calculation of the discrete Fourier Transform is not straightforward. So, while the DFT is a great way to access the (finite) frequency domain of a (finite) function, we need the fast Fourier Transform to make calculations less tedious. The fast Fourier Transform (FFT) may sound like a form of Fourier analysis, however, it is simply an algorithm. The FFT is a particular method of computing the discrete Fourier Transform much more rapidly and efficiently than other available algorithms. A good way to think about the FFT is to think of it as a calculator for the DFT. The best part about the FFT is that it is an algorithm that is programmed into many mathematical applications like MATLAB. This project will use the FFT algorithm in MATLAB to plot graphs for observations.

Assumptions and Definitions:

Now that the platform is set. We will now ensue the details of the actual project. As I have already mentioned, the primary focus of this paper is to analyze frequency differentiation; thus, we will assume the ideal “clean” input and disregard all physical
factors. To further expand on the primary objective, this project also seeks to explore the possibility of the existence of frequency signatures for different objects. For the purpose of this paper, we will define a frequency signature as the shape or contour of the frequency representation of a particular audio sample without paying too much attention to amplitude.

The reason why we do not need to pay attention to amplitude is because we would like to be able to pair two separate recordings of the same object disregarding how close the microphone is to the object (assuming the microphone is neither too far away from the object that it would not obtain enough data nor too close that would cause the audio file to "clip"). Our definition for “clip” -- an audio file "clips" when a recording device attempts to record an input beyond its maximum capacity, hence causing a distortion in the recording.

Procedures:

Now that we have set some definitions and assumptions, here are the procedures of the project:

a. Collect audio recordings of various objects.

b. Load the .wav files into MATLAB.

c. Plot the time graphs.

d. Utilize the FFT function.

e. Plot the frequency graphs.

f. Analyze the similarities and differences.

The objects collected were 3 different kinds of puzzles: 3x3 Rubik’s cube, 4x4 Rubik’s cube and a Megaminx (images shown below). The audio recordings were recorded using a Shure PG42USB condenser microphone on Audacity. Two separate recordings are done for each object to check for consistency. The graphs of the 3 different recordings are presented below.
MATLab Code:

```matlab
>> [X, fs] = wavread('X.WAV');

>> N = length(X);

>> res = fs/N;

>> freq = 0:res:fs-res;

>> Xfft = abs(fft(X));

>> plot(freq,Xfft);

>> plot(X);
```

Top Left: 3x3 Rubik's Cube
[Image](http://cdn1.fiverrcdn.com/photos/1558754/medium/537-1431-thickbox.jpg?1363070937)

Top Right: 4x4 Rubik's Cube
[Image](http://www.smarterpuzzles.com/images/img_4x4cube_large.jpg)

Bottom Left: Megaminx
[Image](http://www.jaapsch.net/puzzles/images/megaminx.jpg)
Comments:

While the two time domain representations barely resemble each other, the frequency graphs both have similar frequency signatures. The contour of the two samples are similar in that they both have comparable looking spikes at the 3k and 5k hertz region.
Comments:

The difference in the time domain for these samples are significantly different. However, the frequency domain shows a similar range and contour. There is a valley like shape between the 1k and 4k hertz region.
Comments:

These two time domain graphs are the most unalike in comparison to the other two objects. Regardless of the differences, the frequency signature can be considered identical if the magnitudes were disregarded.
Conclusion:

In a nutshell, we can see that the frequency domain does help differentiate the different objects. Thus, we can draw a similar conclusion about voice types. This project is very dependent on a user’s eyeballing skills. However, when implemented into a computer program, it will be able to judge more closely using ratios of the spikes and contours with relation to given samples. Furthermore, we know that Fourier analysis significantly contributes to the ability to make such differentiations; as it would be impossible to tell two objects apart if only the time domain graphs were given (as demonstrated in the examples above). Ultimately, we can see how differential equations play a role in the development of Voice Recognition systems.
Bibliography:


