Speech Analytics using Spectrogram Excel-GUI Analyzer

Engr. Ranil Montaril, MSECE, Bolinas, Jordan Michael, Damandaman, Gilly Lea, Garcia, Edil Gester, Garcia, Jedidiah, Pusancho, Rose Mae, Verceles, John Lester, Zubiri, Ceasar Jem Rick

***Abstract*—An important aid in analysis & display of speech is sound spectrogram. It represents time-frequency intensity display of short time spectrum. The quality of speech can be studied by visual inspection of spectrogram. This is one of the important applications of spectrogram in speech processing especially in speech enhancement. Another application of spectrogram is in isolating voiced and unvoiced regions. But to conclude from visual inspection the clarity of spectrogram is also important. Before plotting the spectrogram, the time domain speech signal is converted to frequency domain. The transform domain used plays vital role in resolution of spectrogram. Generally, Fast Fourier Transform is used to convert the time domain signal into frequency domain signal. This paper discusses the step by step procedure of using this transform for converting the time domain speech signal into frequency domain before plotting spectrogram that is used for speech analytics.**

***Index Terms*—Fast Fourier Transform (FFT), Spectrogram, Speech Analytics**

1 Background of the Study

Human speech, along with most sound waveforms, is comprised of many frequency components; the human ear is capable of detecting frequencies between 20Hz and 20,000Hz, although most linguistic information seems to be "concentrated" below 8kHz, according to many researchers. Using this information, it's possible to design a graphical device to represent a speech waveform. [1] To be functional, it needs to show the magnitudes of frequency components, and should be displayed on the time axis (since speech signals vary in frequency composition with time). As luck would have it, such a graphical device is available: the spectrogram.

A speech spectrogram shows the Fourier Transform of a signal as it varies with time. The magnitude of the frequency components is generally either represented as changing colors (along a set color scale) or varying shades of black for a grayscale plot.

In the field of time–frequency signal processing, it is one of the most popular quadratic Time-Frequency distribution that represents a signal in a joint time-frequency domain. Also known as spectral waterfalls, sonograms, voiceprints, or voicegrams, spectrograms are used to identify phonetic sounds. [2] The most common format is a graph with two geometric dimensions: the horizontal axis represents time, the vertical axis is frequency; a third dimension indicating the amplitude of a particular frequency at a particular time is represented by the intensity or colour of each point in the image.

On the other hand, linguistics researchers have sought for years to characterize the sounds humans utter to communicate. The presently-accepted "fundamental unit" of speech is the *phoneme*; every meaningful sound is comprised of one or more of these units, and every language spoken by humans utilizes a unique set of the possible phonemes. It is the aim of phonetics researchers to develop methods for uniquely identifying phonemes in speech. Additionally, digital speech recognition depends on the ability to uniquely-identify single words or phrases; time and frequency domain analysis of speech signals are tools of choice for researchers in the field.

However, looking at the time-domain representation of speech signals can be frustrating; the waveform for a single word or phrase can vary significantly between utterances, even when performed by the same speaker. Consider, for example, the multitude of ways the word can be pronounced. Since researchers generally desire a consistent relationship between input and output, it is clear that a speech waveform must be transformed somehow to obtain a more descriptive relationship.

In particular, the Fast Fourier Transform (FFT) algorithm provides an easy-to-implement way of transforming a finite-time signal into its frequency-domain representation. Applying this approach to speech waveforms reveals a tremendous amount of information.

In this paper, the proponents attempt to generate a visual representation of speech which is the spectrogram with the aid of Excel in FFT computation and grayscale conversion and Matlab in spectrogram plotting.

- *J.M. Bolinas, G.L. Damandaman, E.G. Garcia, J. Gracia, R.M. Pusancho, J.L. Verceles and C.J.R. Zubiri are students of the Department of Computer Science, College of Computer and Information Sciences at the Polytechnic University of the Philippines, Sta. Mesa, Manila, Philippines.*
  Email: {jordanmichaelbolinas, gillylead, edilgesterg, jedidiahgarcia2203, rosemaepusancho , johnlesterverceles, jemzubiri} @gmail.com
- *R.M. Montaril is a professor of the Department of Computer Science*
  Email: rmmontaril@pup.edu.ph

2 Statement of the Problem

This paper aims to develop a spectrogram Excel-GUI analyzer. Specifically, it seeks to answer the following questions:

1. What is the effect of overlap in generating spectrogram?

3 Applicable Related Studies

Spectrographic analysis of speech is one of the most widely used techniques for studying the acoustic-phonetic characteristics of different phonemes in a language. It is an extension of the short-term spectral analysis, and primarily involves representation of the 3-D spectral information obtained by computing the magnitude spectrum over short overlapped window segments, i.e., 2-D spectral content varying with respect to time. The 3-D spectral information is represented on a 2-D plane with the X-axis representing time, Y-axis representing frequency, and the third dimension denoting the log-magnitude of the sinusoidal frequency components is converted to a proportional intensity or gray value. The resulting representation is referred to as a *spectrogram*. [3]

Zenton Goh, Kah-Chye Tan, and B.T.G.Tan [2] examined the spectrograms of typical clean speech, noisy speech, and enhanced speech. The horizontal axis of the spectrogram denotes time, vertical axis frequency, and the spectral magnitude is shown with gray shade (darker shade indicates larger value). It is observed that a large portion of the spectrogram is practically blank (i.e., unshaded) and the speech energy is concentrated in a few isolated regions. The voiced portion of speech is characterized by dark parallel "stripes" whereas unvoiced portion is characterized by gray patches. Some parallel stripes are horizontal while some are slanting up or down, indicating a change in the pitch of the speech signal. When white Gaussian noise amounting to the clean speech, the blank region of the spectrogram become shaded, and some of the stripes corresponding to voiced speech disappear. With an appropriate spectral subtraction, obtained an enhanced speech with spectrogram and observed a significant reduction of the unwanted short stripes. By observation of spectrogram [2] concluded about speech quality.

Annabi-Elkadri [4] discussed that choosing an appropriate window length for spectral analysis is not a straightforward process. A narrow window provides a low frequency resolution, approximating only roughly the spectral envelope, whereas a wider window provides a high frequency resolution and can even show the harmonics in the spectrum. The drawback to analysing a greater part of the signal can lead, however, to a lower temporal resolution, thus masking or distorting rapid acoustic landmarks occurring in speech. This suggests using a wide window for long steady-state vowels and a narrow window when investigating stop bursts in which the higher frequencies are more important.

Speech spectrogram reading involves interpreting the acoustic patterns in the image to determine the spoken utterance. One must selectively attend to many different acoustic cues, interpret their significance in light of other evidence, and make inferences based on information from multiple sources. [5] The evidence, obtained from spectrogram reading experiments indicates that the process can be modeled with rules. Formalizing spectrogram reading entails refining the language used to describe acoustic events in the spectrogram, selecting a set of relevant acoustic events that distinguish among phonemes, and developing rules which map these acoustic attributes into phonemes. Phoneme segmentation by an expert system utilizing spectrogram reading strategy and knowledge can detect phonemes in a spectrogram and determines their boundaries as well as their coarse categories.

4 Software Design Architecture

4.1 Applicable Equation

Discrete Fourier Transform (DFT) can be computed efficiently using a fast Fourier transform (FFT) algorithm. The discrete Fourier transform (DFT) is a specific kind of Fourier transform, used in Fourier analysis. It transforms the time domain function into frequency domain representation. FFT algorithms are so commonly employed to compute DFTs that the term FFT is often used to mean DFT in colloquial settings. DFT can be defined as,

For length N input vector x, the DFT is a length N vector,

$$X(k) = \sum_{j=1}^{N} x(j)\omega_N^{(j-1)(k-1)} \tag{1}$$

$$x(j) = \left(\frac{1}{N}\right)\sum_{k-1}^{N} X(k)\omega_N^{-(j-1)(k-1)} \tag{2}$$

where,

$$\omega_N = e^{(2\pi i/N)} \tag{3}$$

4.2 Functional Block Diagram

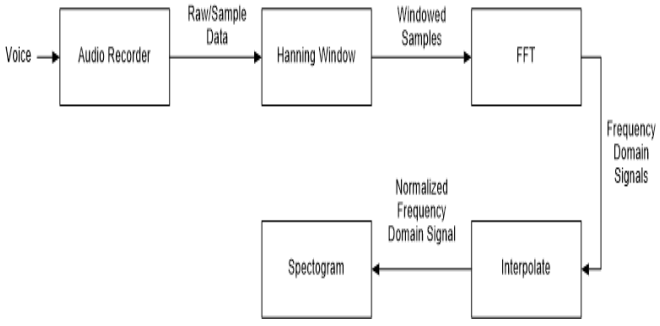


Figure 1: *Spectrogram Excel-GUI Analyzer Block Diagram*

First, voice/speech is recorded in Matlab and then to be converted into numerical value before exporting it in a spreadsheet in Excel. The raw/sample data will undergo windowing process using the Hanning window function. The DFT (Discrete Fourier Transform), a variation of FFT (Fast Fourier Transform) will then be computed for each window. Then the windowed FFT values will then be interpolated and will be converted to grayscale value. The array output of Excel will be imported to Matlab and spectrogram will be finally plotted.

4.3 State Diagram

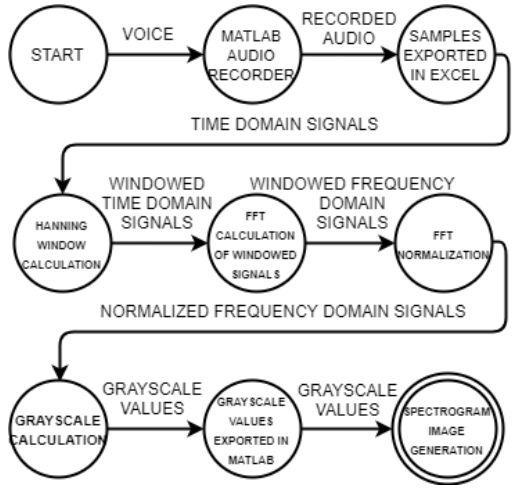


Figure 2: *Spectrogram Excel-GUI Analyzer State Diagram*

The state diagram shown above denotes the step by step procedures in spectrogram generation with the aid of Matlab and Microsoft Excel. The process starts by recording an audio (speech), then next is exporting the raw data in spreadsheet subject for computation. It ends with spectrogram image generation.

5 Simulation Test Results

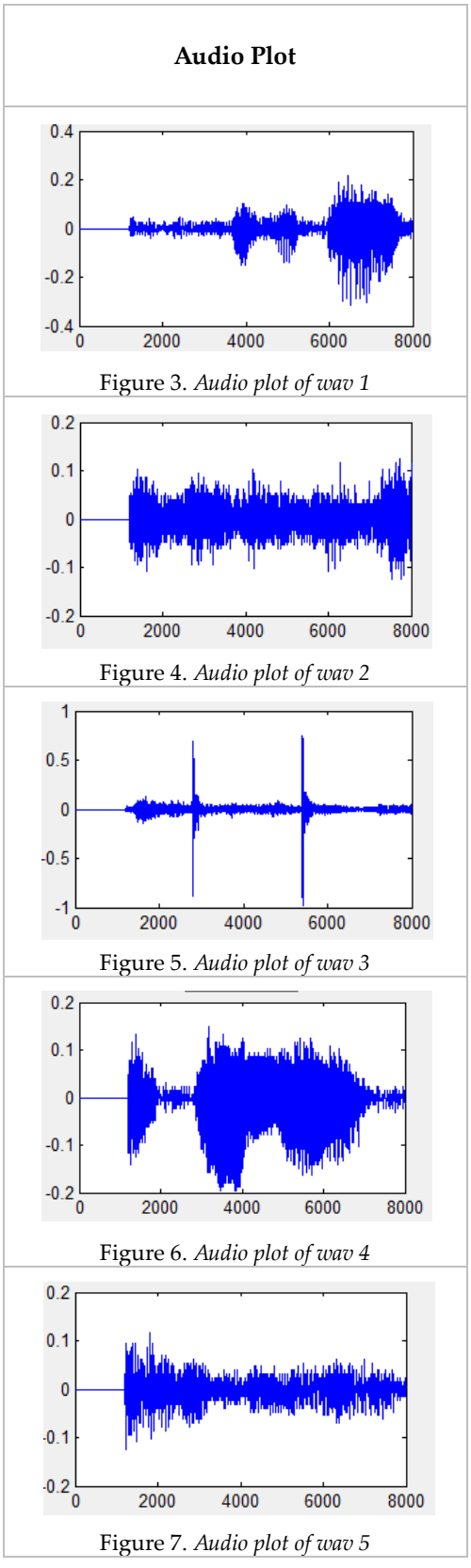


Figure 3. *Audio plot of wav 1*

Figure 4. *Audio plot of wav 2*

Figure 5. *Audio plot of wav 3*

Figure 6. *Audio plot of wav 4*

Figure 7. *Audio plot of wav 5*

Figure 8. *Spectrogram of wav 1*

Figure 9. *Spectrogram of wav 2*

Figure 10. *Spectrogram of wav 3*

Figure 11. *Spectrogram of wav 4*

Figure 12. *Spectrogram of wav 5*

6 Conclusion

The effect of the overlapped window creates a minor difference to the actual signal because of the roll of effect while the non-overlapped or rectangular window have leakage because no overlapping is done.

The output of our system developed in Excel have major difference since the window size of the Matlab is significantly larger and have a bigger hamming window used and handles the samples more efficiently in terms of performance.

7 Recommendation

In our case, we used 32-point FFT which has an impact on the generated spectrogram. We recommend using bigger FFT size and a larger hamming window for a higher resolution of spectrogram, so as for formants to be more clearly seen.

References

[1] "Project Rhea," [Online]. Available: https://www.projectrhea.org/rhea/index.php/Speech_Spectrogram. [Accessed 17 October 2017].

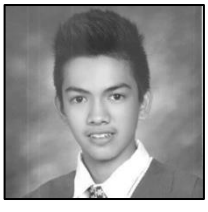
[2] R. R. Mergu and D. S. K. Dixit, "Multi-Resolution Speech Spectrogram," *International Journal of Computer Applications*, vol. 15, no. 4, pp. 28-32, 2011.

[3] "Speech Signal Processing Virtual Lab," [Online]. Available: http://cse16-iiith.virtual-labs.ac.in/exp06/indexie.html. [Accessed 17 October 2017].

[4] N. Annabi-Elkadri, "Comparative Studies for Speech Analysis based on Multiresolution Spectrograms," *International Journal of Latest Research in Science and Technology*, vol. 3, no. 2, pp. 51-54, 2014.

[5] R. R. Mergu and S. K. Dixit, "A New Paradigm for Plotting Spectrogram," *Journal of Information Systems and Communication*, vol. 3, no. 1, pp. 158-161, 2012.

Curriculum Vitae



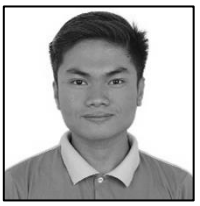
Jordan Michael M. Bolinas is currently taking BSc in Computer Science at the College of Computer and Information Sciences, Polytechnic University of the Philippines, Manila. He graduated from City of Mandaluyong Science High School (CMSHS) on 2014. He received the Loyalty Award and Certificate of Merit when he graduated from elementary. His research interests include Web Development and Expert Systems. He became a scholar of the Department of Science and Technology (DOST).

Gilly Lea E. Damandaman is a student of Bachelor of Science in Computer Science from the Polytechnic University of the Philippines. She has a knowledge on web programming and java programming and has a full potential in working as part of a team.

Jedidiah P. Garcia is currently studying Bachelor of Science in Computer Science in Polytechnic University of the Philippines. He graduated high school at Assumpta Technical High School, San Simon, Pampanga. He has a knowledge in C, Java and Web Programming and he is a member of Harambeats Developers Group since 2016. He has interest in Computer and Visual Graphics and Artificial Intelligence. He loves visual design and also good at Adobe Photoshop, Illustrator After Effects and Sketch Up.

Edil Gester T. Garcia currently lives at Sta. Mesa, Metro Manila NCR Philippines taking up his Bachelor's degree in Computer Science since 2014 having a decent GPA in the department anticipating to attend his commencement exercise hopefully next year, 2018. He is interested in Artificial intelligence specifically Neural nets and swarm intelligence, c++ and objective C program in hopes for creating windows based application as well as appleOSX. He is also included in varsity chess team representive his former university and successfully gained a gold medal way back 2012 and hopes to do it again at his present university. He has attended seminars in machine learning and ethical hacking, thesis writting and fundamentals of research.

Rose Mae S. Pusancho is currently taking Bachelor of Science in Computer Science at Polytechnic University of the Philippines in Sta. Mesa, Manila. Before she is a consistent honor student and finished her High School as First Honorable Mention at Ungos National High School in Real, Quezon. Now, she is pursuing her career in the field of Information Technology. She focuses on Computer Science Research and planning to explore more. Her interests include Artificial Intelligence, Robotics, Web Development and Natural Language Processing. She is willing to learn as she continues her studies. She is a current scholar of the Department of Science and Technology (DOST).



John Lester F. Verceles finished his secondary education at Mangaldan National High School (Special Science Class), Pangasinan as the 3rd Honorable Mention and is currently a student of Bachelor of Science in Computer Science under the Computer Science Research (CSR) track at the Polytechnic University of the Philippines. He is a learner of dedication and passion who is willing to delve more into the emerging trends that would impact to the society in the present time. He has experienced skills in web programming, matlab scripting, database management and object-oriented programming. He is a scholar of the Department of Science and Technology (DOST).

Ceasar Jem Rick D. Zubiri is a persistent, calm, and collected person. Though tardy, he complies to commands given to him and compensates with a different effort. He is currently working as a software developer at Getz Clinical developing web and windows applications. Though a master of none, he can be versatile when needed be.