The Effect of Using Different Transforms for Speech Signals in Wireless Commincations

Sujatha Choppara
Asst.Professor
Dept.of ECM
P.V.P. Siddhartha Institute of Technology, KANURU, VIJAYAWADA-7, Krishna District, Andhra Pradesh

Padma Yenuga
Asst.Professor,
Dept.of IT,
P.V.P. Siddhartha Institute of Technology, KANURU, VIJAYAWADA-7, Krishna District, Andhra Pradesh

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 effect of using different transform for converting the time domain speech signal into frequency domain before plotting spectrogram. . It is observed that resolution of speech spectrogram is transform dependent in wireless.

1. Introduction

In many practical situations, speech has to be recorded in the presence of undesirable background noise. As noise often degrades the quality/intelligibility. In many practical situations, speech has to be recorded in the presence of undesirable background noise. As noise often degrades the quality/intelligibility of recorded speech, it is beneficial to carry out noise suppression. In the literature, a variety of speech enhancement methods capable of suppressing noise has been proposed. In speech enhancement the graphical representation of speech is spectrogram plays vital role to examine speech quality. The quality of speech can be observed quickly using spectrogram. This is one of the important applications of spectrogram 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 effect of using different transform for converting the speech signal into frequency domain before plotting spectrogram.

Zenton Goh, Kah-Chye Tan, and B.T.G.Tan 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.

An experienced spectrogram reader has no trouble identifying the word "compute" from the visually salient patterns in the image above. To give one example, the vertical burst of energy followed by a red area at the bottom and lesser energy above at the extreme right of the spectrogram is a typical pattern for the sound 't' at the end of a syllable or word. The other speech sounds, or phonemes, in the word "compute", are equally distinct in their shapes; the initial unstressed syllable /kh ^ m/, the silence and bilabial burst of /pc ph/, and the stressed vowel /ju/ which represents the passage from a high front vowel to a high back vowel by the falling F2, and the proximity of the alveolar plosive by a subsequent rise in F2 toward the alveolar locus of 1800 Hz.
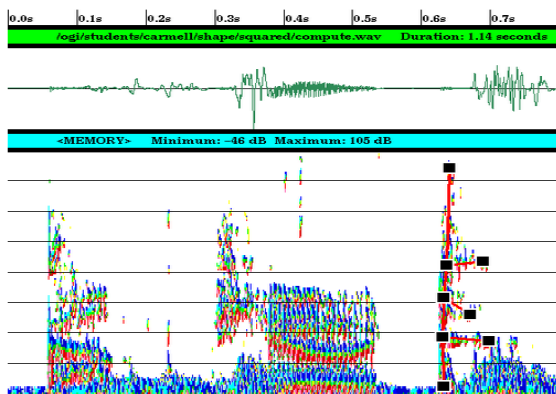
Fig: 1. The waveform and spectrogram for the word "compute" combined with a ruler which gives us a measure of the duration of the various components in the utterance.
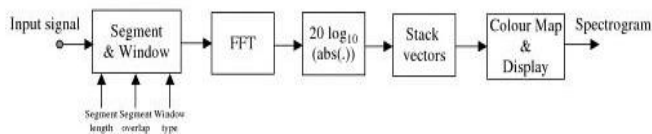
Fig:2. Block diagram illustration of spectrogram

2. Generation of Spectrogram

The use of spectrogram in speech enhancement is discussed in this paper. The additive noise model is described by the following equation,

$$y(t) = x(t) + n(t)$$

Where, y(t) is the observed noisy speech ,x(t) is the clean speech and n(t) is the additive background noise. The observed speech is then divided into overlapping frames of length of 256 samples in each frame .The amount of overlap is normally either 50% or 75%. In this paper, 75% overlapping is used throughout. The nth frame can be represented by a column vector described by the following equation:

$$f_L = [y(64L) \; y(64L + 1) y(64L + 2)…y(64L + 255)]^T$$

All indices used in this paper starts from zero. A speech block can be obtained by arranging a number of frames together to form a matrix. Suitable numbers of frames are found experimentally to be 8, 16 and 32.In this paper, the number of frames used is 16 throughout. Similarly each block overlaps its neighboring block by 75%. Then the speech block can be represented mathematically as a matrix, of size 256 by 16 as shown in the following equation:

$$b_n = [f_{8n} \quad f_{8n + 1} \quad f_{8n + 2} .... f_{8n + 15}]$$

This signal is windowed using Hamming window. Then the transform can be applied onto the speech block.

Spectrograms are usually created in one of two ways; approximated as a filter bank that results from a series of band pass filters, or calculated from the time signal using the short time fourier transform. These two methods actually form two different quadratic time frequency distributions, but are equivalent under some conditions.

The band pass filters method usually uses analog processing to divide the input signal into frequency bands; the magnitude of each filter's

output controls a transducer that records the spectrogram as an image on paper.

Creating a spectrogram using the STFT is usually a digital process. Digitally sampled data, in the time domain, is broken up into chunks, which usually overlap, and fourier transformed to calculate the magnitude of the frequency spectrum for each chunk. Each chunk then corresponds to a vertical line in the image; a measurement of magnitude versus frequency for a specific moment in time. The spectrums or time plots are then "laid side by side" to form the image or a three dimensional surface.

3. Applications

Early analog spectrograms were applied to a wide range of areas including the study of bird calls, with current research continuing using modern digital equipment and applied to all animals' sounds. Contemporary use of the digital spectrogram is especially useful for studying frequency modulation (FM) in animal calls. Specifically, the distinguishing characteristics of FM chirps, broadband clicks, and social harmonizing are most easily visualized with the spectrogram. A particularly interesting example for the use of the spectrogram is in analysis of the vocalizations of a pod of Dolphins.
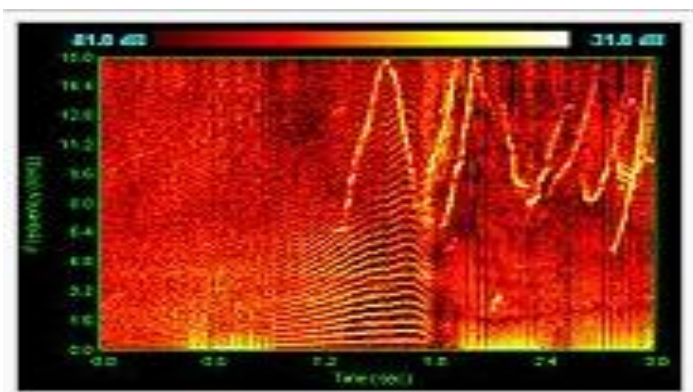
Fig:3. Spectrogram of Dolphin volcanizations

Spectrograms are useful in assisting in overcoming speech defects and in speech training

for the portion of the population that is profoundly deaf.

The studies of phonetics and speech synthesis are often facilitated through the use of spectrograms.

By reversing the process of producing a spectrogram, it is possible to create a signal whose spectrogram is an arbitrary image. This technique can be used to hide a picture in a piece of audio and has been employed by several electronic music artists.

4. RESULTS

Here spectrogram is plotted for different utterances of human speech male. The speech utterances are obtained from noise us database. The spectrograms of clean and noisy speech signals are:
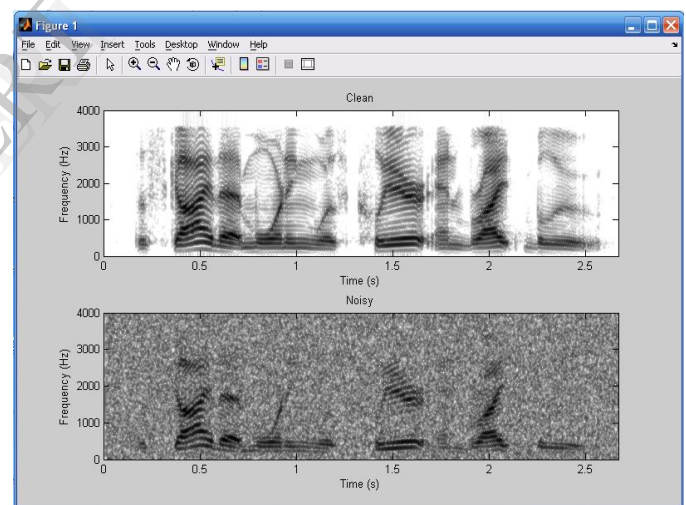
Fig:4 Periodogram spectrum
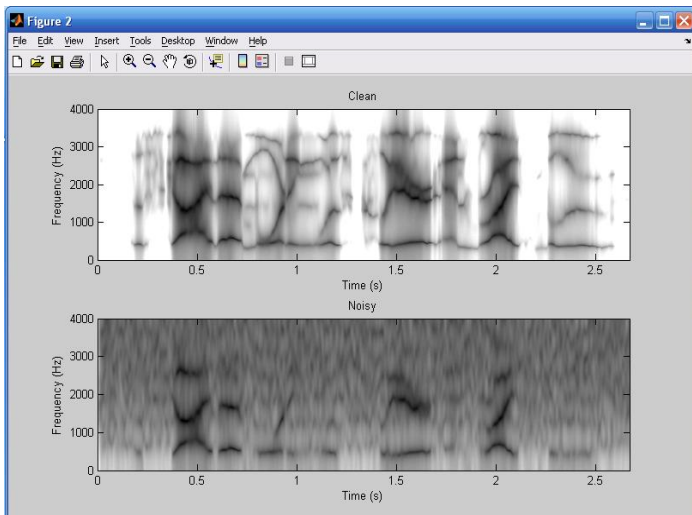
Fig:5. Autoregressive spectrum with pre-emphasis

5. CONCLUSIONS

From the results shown above we can conclude that the spectrograms plotted using STFT. From the visual inspection we can see the amount noise available in the speech signal. Thus the quality of input signal can be inspected from spectrogram The voiced and unvoiced regions are very well differentiated and the energy at different time instant in particular frequency bin can be observed very clearly in spectrogram plotted using periodogram spectrum .Whereas the in the spectrograms plotted using autoregressive spectrum with pre-emphasis the energy content, amount of noise and voiced/unvoiced region detection is much difficult. Thus plotting spectrogram helps to identify the speech cues in acoustic distinction of speech signals.

REFERENCES

1. S. Haykin, editor, Advances in Spectrum Analysis and Array Processing, vol.1, Prentice-Hall, 1991.

2. B. Boashash, editor, "Time-Frequency Signal Analysis and Processing – A Comprehensive Reference", Elsevier Science, Oxford, 2003; ISBN 0080443354

3. JL Flanagan, Speech Analysis, Synthesis and Perception, Springer- Verlag, New York, 1972

4. B. Boashash, "Estimating and Interpreting the Instantaneous Frequency of a Signal-Part I: Fundamentals", Proceedings of the IEEE, Vol. 80, No. 4, pp. 519-538, April 1992, doi:10.1109/5.135376

5. Behavioural case formulation and intervention by Peter Sturmey

6. Discrete time speech signal processing by Thomas F.Quatieri

7. Springer handbook of speech processing by Jacob Benesty

8. Introduction to Digital Speech Processing by Lawrence R.Rabiner and Ronald W.Schafer

9. Oppenheim, A.V., and R.W. Schafer, Discrete-Time Signal Processing, Prentice-Hall, Englewood Cliffs, NJ, 1989, pp. 713-718.

10. Rabiner, L.R., and R.W. Schafer, Digital Processing of Speech Signals, Prentice-Hall, Englewood Cliffs, NJ, 1978.