Final Project

B05902120 / Yu-Ting, TSENG

Jan 15, 2019

Introduction

The paper was called *An Experimental Study on the Phase Importance in Digital Processing of Speech Signal*. It was written by Lazar Tesic, Boban Bondzulic, Milenko Andric, Boban Pavlovic.

In the course, Professor has mentioned the most commonly used digital processing approaches of speech signal, a short time Fourier transform. The signal was represented with complex valued coefficients, which was composed with magnitude and phase. In other words, in this domain, signal can be represented by "magnitude" and "phase". Mostly people is convinced that magnitude is more important. The paper is doing deeply research of how important is the role that "phase" has played. To prove this idea, the paper set an experiment to illustrate the phase importance through "phase-only" image reconstruction.

Basic Idea

People have tried several ways to justify the importance of phase. The first is to view it from a statistical point. It shown that magnitudes distortion will not have great influence on signals, however, a random distortion over phases can dramatically change the signals.

The figure on the top show the waveforms of original source signal and the bottom shows the reconstructed signal from magnitude and phase respectively. Reconstruction is performed by inverse discrete Fourier transform, which is mentioned during the class. When constructing from magnitude spectrum, we assumed that the phase spectrum equals to zero; on the other hand, when constructing from phase spectrum, we supposed that the magnitude spectrum equals one. According to the figures, it is obviously that the waveform reconstructed from magnitude spectrum is completely different from the source signal, on the contrary, the signal somehow preserved in the phase spectrum.

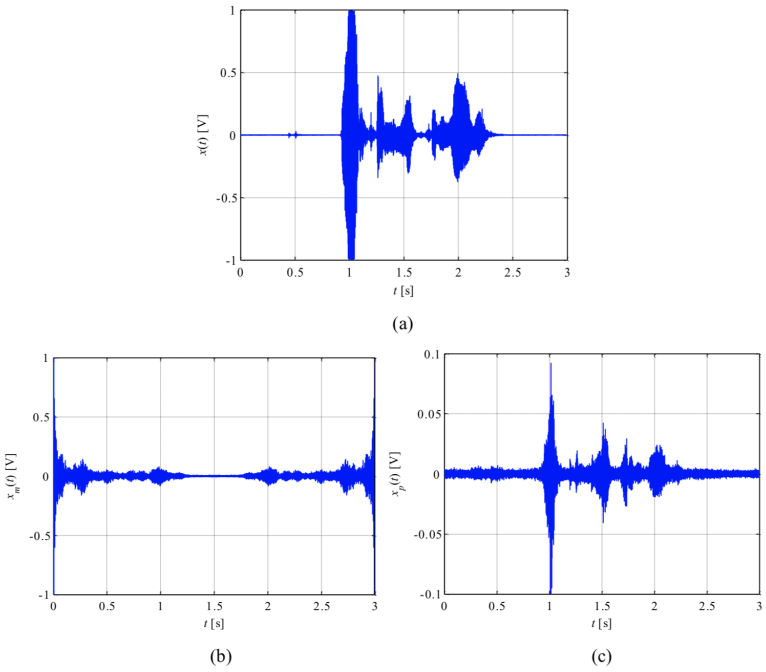

Figure 1
Waveforms of the: (a) source signal, (b) signal reconstructed from the magnitude spectrum of source signal and (c) signal reconstructed using the phase spectrum of source signal

Most of the paper study speech signals via subjectively quality assessment consider coding algorithms and speech enhancement under noisy condition. Some people create a database designed to evaluate the performance of speech recognition algorithms in noisy conditions. By this database, the paper has shown that coding algorithm might impact the accuracy of speech recognition, so does altering the phase spectrum.

Another research about phase is to study the relation between uncertainty phase and word error rate. It has been shown that at an SNR of -10dB, having random phases might cause a WER of 63% whereas a WER is 24% if the phase isn't modified. With the SNR of 0dB, random phase results in WER of 25%, and 11% in contrast. It gives a result that at low SNRs the effect of phase on the accuracy of recognition can be greatly significant.

In order to deeply study over importance of phase, a phase spectrum is used for quantifying speech quality. Some people proposed phase distortion derivation measure (PDD). The database is contribute by dysphonia speakers. They have presented that PDD is highly correlated with medical doctors, which means that phase deviation objective metric is a reliable quality estimator.

Main Research

More common way is to realize four phases: 1. gathering source speech sentences, 2. forming the speech database, 3. subjective tests running, 4. results arrangement and analysis. The authors used a filter and software for signal acquisition, and there are introduction in detailed that I will omit this part. They first asked three native speakers to read six sentences create the database. The signals then be filtered to reserve the magnitude response and phase response respectively, and project over two different cut-off frequencies. Then the subject in order listen to the speech and give each sentence no matter degraded or not a score.

The outcomes of the tests are presented through mean opinion scores (MOS), which is the most commonly used method of generalizing a subjective score. We can find that MOS is a democratic measure since it treats each subjective vote equally and the only true mean opinion.

$$MOS_i = \frac{1}{N_S}\Sigma_{n=1}^{N_S} SQ(n,i), \text{ where}$$

i is the index of degraded sentence;
$SQ(n,i)$ is the subjective grade given by the n-th listener to i-th sentence;
N_S is the number of listeners who graded i-th sentence.

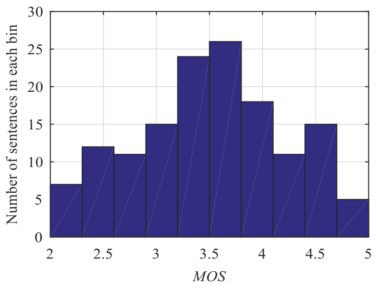

Figure 6
MOS scores histogram with 10 equally spaced bins

Without doubt, deviation is an indispensable norm; in this case, it is evaluated directly from the individual subjective scores for sentence, and the MOS

$$\sigma_i = \sqrt{E[SQ_i^2] - (E[SQ_i])^2} = \sqrt{E[SQ_i^2] - MOS_i^2}, \text{ where}$$

$E[X]$ is expected value or an average score of the random variable.

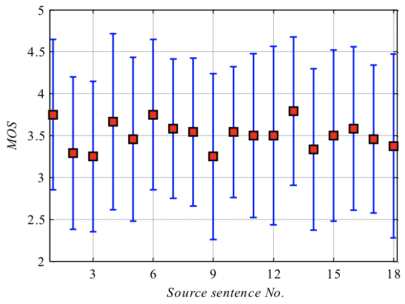

Figure 7

MOS values with respective reliability intervals of subjective scores of test sentences which are produced from the same source sentence

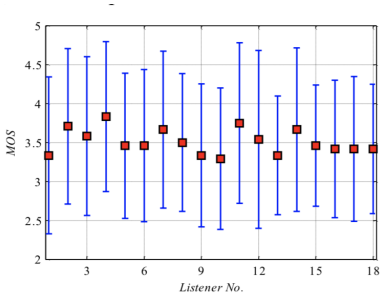

Figure 8

MOS values with respective reliability intervals of subjective scores given to sentences by one listener

From the other most critical viewpoint that the authors want to make sure is which filter has the highest score. And the figures below shows that the third one might have less error recognition, whose signal has same phase spectrum as the original source. This findings again prove the importance of the phase.

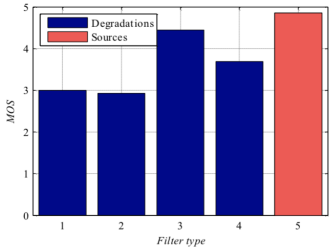

Figure 10

MOS values of subjective scores given to the sentences with the same type of degradation

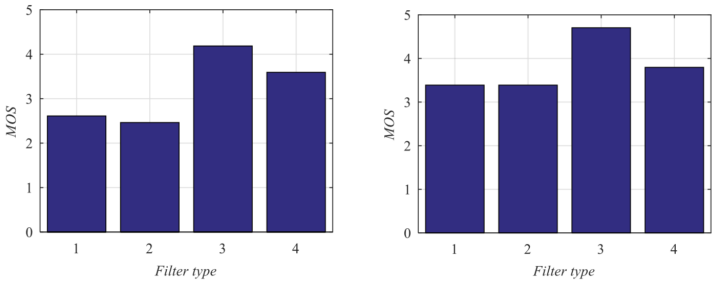

Figure 11

MOS values of subjective scores given to the test sentences with the same type of degradation: (a) for the cutoff frequency f_g=1 kHz and (b) for the cutoff frequency f_g=2 kHz

As a conclusion, the study evaluate phase importance in the perceptual quality of speech communication. In the modeling and processing of the time-frequency signal representation, phase can't be ignore. Worth mention is that phase also plays an important role in different applications, speech enhancement, automatic speech, speech recognition and speech synthesis are included.

Reference

1. Ni, X.S., Huo, X.: Statistical interpretation of the importance of phase information in signal and image reconstruction, Statistics & Probability Letters, 2007, Vol. 77, No. 4, pp. 447-454

2. Paliwal, K., Wojcicki, K., Shannon, B.: The importance of phase in speech enhancement, Speech Communication, 2011, Vol. 53, No. 4, pp. 465-494

3. Shi, G., Shanechi, M. M., Aarabi, P.: On the importance of phase in human speech recognition, IEEE Transactions on Audio, Speech and Language Processing, 2006, Vol. 14, No. 5, pp. 1867-1874

4. Koutsogiannaki, M., Simantiraki, O., Degottex, G., Stylianou, Y.: The importance of phase on voice quality assessment, In Proceedings of the 15th Annual Conference of the International Speech Communication Association (INTERSPEECH), Singapore, 2014, pp. 1653-1657