Available online at www.elixirpublishers.com (Elixir International Journal)

Electrical Engineering

Elixir Elec. Engg. 41 (2011) 5771-5773

ISSN: 2229-712X

Enhancement of reverberant speech using various transforms

Philip Roshan Smith R and Shalini M.
Gnanamani College of Engineering, Namakkal.

ARTICLE INFO

Article history:
Received: 10 September 2011;
Received in revised form:
16 November 2011;
Accepted: 2 December 2011;

Keywords
Reverberant Speech,
Short-Time Speech,
FFT, DCTand DWT.

ABSTRACT
In this paper we propose a new approach of processing speech signal which is degraded by the reverberant components. This method is based on the analysis of short [2ms] segment of data to enhance the region of the speech signal and analyzing the speech components without degradation due to reverberation, using various methods [FFT, DCT& DWT] and comparison of their results.

Introduction

Degradations in speech are caused by additive noise and reverberation. In this paper we consider enhancement of speech under reverberant conditions. The focus is on the degradation of speech caused in a speakerphone situation. Speech from a speakerphone contains both the direct component and the reverberant component. The speech signal is enhanced in order to enhance the signal in the direct component, wherever possible, so that the resulting processed speech is perceived as less reverberant and thus increasing the comfort level for listening. Several methods have been proposed for enhancement of speech degraded by unvoiced unwanted components.

Normally degraded (additive noise or reverberant) speech is processed assuming that the degradation has long term stationary characteristics relative to speech. For reverberant speech, the reverberation effects are captured by estimating the impulse response of the room environment from long (500-1000ms) segments of speech. The unwanted unvoiced speech signal components are passed through an inverse filter for the room response to filter the speech.

The main problems in these approaches for processing degraded speech are that the estimates of the characteristics of the degradations are not good enough to remove their effects in short segments of speech. This is because the level of degradation in terms of Signal to Reverberant component Ratio (SRR) is different for different segments of speech. Moreover, the emphasis in many of these approaches seems to be on the degradation and not on speech. There appears to be a need to look at the problem of enhancement of reverberant speech with more focus on the direct component of speech at the receiving microphone. In processing it is necessary to increase the contribution of the direct component relative to the reverberant component.

In such an attempt there will be more focus on speech than on the degradation in the process of enhancement. There are segments of speech where reverberant component dominates over the direct component. For such segments there is no point in attempting to enhance the speech part. On the other hand, if regions, where the direct speech signal component is significantly higher compared to the reverberant component, could be identified, then by enhancing speech in such regions the annoyance due to reverberation could be reduced in some segments at least. The levels of the higher reverberation regions, if identified, could be reduced. In the regions where there is only reverberant component, such as silence regions, the levels could be reduced to very low values.

Perception of the overall speech is significantly influenced by the high signal energy regions, thus giving an impression of enhancement of de- graded speech. Thus the criterion for improvement is not based on all speech segments, but only on high direct path signal component regions.

The method proposed in this paper is different from the existing methods, as there is more emphasis on the characteristics of speech and also the analysis segments are much shorter (1-3ms) compared to the normal 10-30ms frames used in speech analysis based on quasi stationary assumption.

In particular, the importance of processing the linear prediction (LP) residual signal is emphasized, since most of the conventional approaches tend to ignore the details of the residual signal.

Characteristics of reverberant speech

In this section we will examine the characteristics of reverberant speech to determine clues for processing the speech for enhancement. The effects of reverberation can be seen by comparing the signal waveforms for clean and reverberant speech signals. The clean speech has clear damped sinusoidal-like pattern within each pitch cycle, whereas the reverberant speech is smeared within each cycle . The smearing of signal within each pitch cycle is more prominent when the gross envelope of the signal waveform. Some features of reverberation effects can be seen more clearly in the LP residual waveform.

The residual signal is computed for a segment of 2ms at every sampling instant, using a 5thorder autocorrelation LP analysis. The residual signal for reverberant speech signal has a significant direct component of the signal in the reverberant speech. This shows that there are segments in the reverberant speech where the direct component is significantly higher than the reverberant component. Due to the decaying nature of the overall signal amplitudes, the reverberation effects of the preceding speech dominate over the direct component.

Tele: +91-9500693030
E-mail addresses: drshalini2014@gmail.com

The residual signal is mainly due to the reverberation. Whenever the direct component of speech is higher than the reverberant component, the LP residual signal at the epochs is well behaved with significant energy around the instants of glottal closure. It is such regions that we need to identify, so that the signals in those regions can be processed to enhance the direct component over the reverberant component.. First of all it is necessary to identify these three different regions in reverberant speech. For this purpose the normalized error *(q)* of clean and reverberant speech is computed at every sampling instant using a 5th order autocorrelation LP analysis on a frame of size 2ms. The normalized errors for both clean and reverberant speech appear similar in the high SRR regions. But overall the normalized error for reverberant speech is lower than that for the clean speech, closer examination of the normalized error plot reveals that within each pitch cycle the error is maximum just before the region of glottal closure. This is because the residual signal amplitude is high in this region. The important point to be noted is that the enhancement needs to be done differently in different segments due to variation of short-time characteristics of speech in temporal and spectral domains.

Methods and implementations

The Discrete Fourier transform is an alternative way of Fourier representation for finite duration sequences. The DFT is a sequence function of continuous random variables which corresponds to samples equally spaced in a frequency. DFT plays a central role in various DSP and speech processing algorithms. The choice of the window in short-time speech processing determines the nature of the measurement representation. A long window would result in very little changes of the measurement in time whereas the measurement with a short window would not be sufficiently smooth.

Two representative windows are demonstrated, Rectangular and Hamming. The latter has almost twice the bandwidth of the former, for the same length. Furthermore, the attenuation for the Hamming window outside the pass band is much greater. Effects of the choice of window length are demonstrated. As the length increases, short-time energy becomes smoother, as expected. It should be noticed that the measurement is not taken for every sample. Due to the low pass character of the window, short-time energy is actually band limited to the bandwidth of the window, which is much smaller than 16 KHz. actually for the lengths we are interested in; it is less than 160 Hz. So, we could calculate the energy every 50 samples, that is with 320 Hz frequency for 16kHz speech sampling frequency. The property of the short-time autocorrelation to reveal periodicity in a signal is demonstrated. Notice how the autocorrelation of the voiced speech segment retains the periodicity. On the other hand, the autocorrelation of the unvoiced speech segment looks more like noise.

In general, autocorrelation is considered as a robust indicator of periodicity. Representation of speech in the frequency domain is demonstrated. The importance FFT as a computational tool becomes also clear. Demonstration of the calculation of the spectrum using long windows, either Hamming or rectangular. Demonstration of the calculation of the spectrum using short windows, either Hamming or rectangular. Demonstration of the covariance method for linear prediction. Compared to the autocorrelation method, the difference of the covariance method is that it fixes the interval over which the mean - square prediction error is minimized and speech is not taken to be zero outside this interval. Stability of the resulting model cannot be guaranteed but usually for sufficiently large analysis interval, the predictor coefficients will be stable. The error autocorrelation and spectrum are calculated as a measure of its whiteness. Linear predictive analysis of speech is demonstrated for various model orders. Notice how the model frequency response becomes more detailed and resembles the speech spectrum as the model order increases. The prediction error steadily decreases as the order of the model increases. There is an order however further from which any increases lead to minor decreases of the prediction error.

The choice of the order basically depends on the sampling frequency and is essentially independent of the LPC method used. Usually, the model is chosen to have one pole for each kHz of the speech sampling frequency, due to vocal tract contribution and 3-4 poles to represent the source excitation spectrum and the radiation load. So for 16kHz speech a model order of 20 is usually suitable. The Discrete fourier Transform plays an important role in the analysis , design and implementation of digital signal processing algorithms and speech processing algorithms. The computation of DFT can be done using many efficient algorithms.

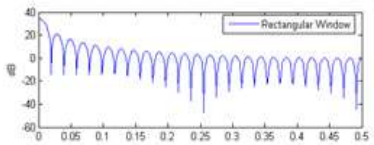
Fast Fourier Transform [FFT] algorithm provides more efficiency in computation of the N values of the DFT. When we require values of the DFT over the portion of a frequency range $0< w <2\pi$ FFT provides efficient computation of all the values of DFT. The computation is decomposed in to smaller DFT computations for dramatic efficiency results.

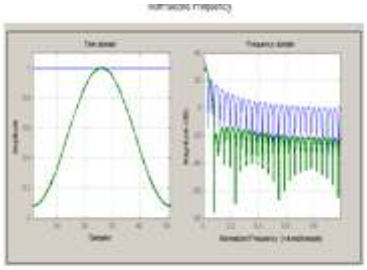
This process of decomposing the sequences in to smaller decomposition sequences is called as Decimation in time algorithms. The total number if computations are same for the decimation in time and decimation in frequency algorithms. The duality inherent in most Fourier transforms is evident when we compare the convolution and windowing theorems.

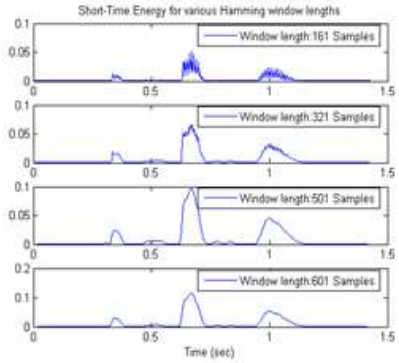
The Fourier transform is a sum while the inverse Fourier transform is integral with a periodic integrand where this duality is complete the convolution sum is equivalent to multiplication of corresponding sequences of the fourier transform. The simplest method of FIR filter design is Window method.

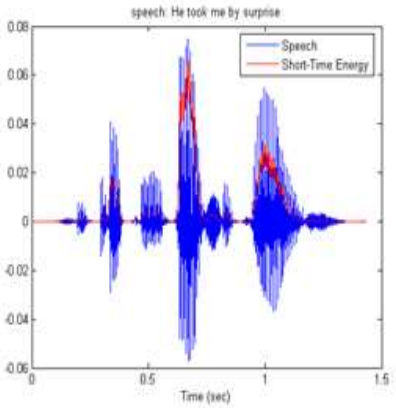
Results and implementations

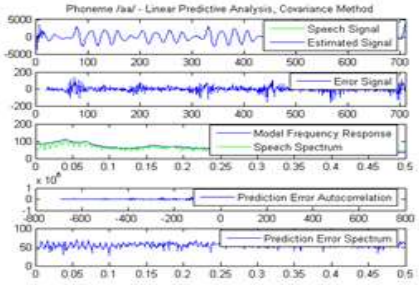
Short-time speech measurements, windows











Conclusion

In this paper we presented a various approach for processing reverberant speech signals. The methods followed here are based on the knowledge that the speech signal components spread over range in the splitted segments by identifying the high quality of speech components. Speech quality is ensured in these processes at the different segment levels. The resulting signal shows reduction of unvoiced components without affecting the voiced component in the signal.

References

[1] B. Yegnanarayanamoorthy, P. Satyanarayanamoorthy, " Enhancement of Reverbrant speech using LP Residual" 1998.

[2] Alan V.Oppenheim and Ronald W.Schafer and John R.Buck, Discrete Time Signal processing. Prentice Hall,1998.

[3] J.G. Proakis, DigitAL Communications. Singapore: McGraw-Hill,1989.

[4] In Speech enhancement (J.S.Lim,ed.), New Jersey: Prentice Hall,1983.