Liner Prediction and DCT based De-noising of Speech Signal

Diksha Sharma¹, Rupinder Kaur²

¹, ²DIET, Kharar, Punjab, India

Abstract: The proposed algorithm adopts LP residual as one of the sparse representation of speech, considering it is feasible and advantageous, as analysed in the previous section. To make full use of the sparsity of speech, DCT coefficients are also included to contribute as a measurement. The proposed algorithm aims to recover the clean speech, whose LP residual and DCT coefficients are both sparse, via solving an optimization problem under a series of constraints.

Keywords: DCT, Linear Prediction, Speech Enhancement

1. Introduction

Speech processing is a diversified area of research, primarily due to the fact that speech processing will always be needed in such important areas as telecommunication technologies. For example, speech processing in real-time and transparent speech transmission sets many challenges for researchers. The requirements for the quality of processed speech are increasing constantly. Since speech has many aspects, distinguishing between different utterances is not enough; all the other information related to e.g., identification of the speaker, the speaker’s disposition or other emphases included in the speech signal should also be delivered to the receiver. In phonetics, the characteristics of speech, including intonation, enunciation and length of the phoneme, i.e., quantitative, are all called the prosody of the speech. Depending on the application and the required quality, all the perceptually important characteristics of speech must be maintained during processing. Future wireless or mobile communication applications, for example, will no longer tolerate speech quality which is inferior to the toll quality.

2. Related Works

Yong Xu, Jun Du, Li-Rong Dai, and Chin-Hui Lee; (2014) “An Experimental Study on Speech Enhancement Based on Deep Neural Networks.” Here we presents a regression-based speech enhancement framework using deep neural networks (DNNs) with a multiple-layer deep architecture [1]

Shenoy, R.R.; Seelamantula, C.S.(2014) Frequency domain linear prediction based on temporal analysis. This states that Frequency-domain linear prediction (FDLP) is widely used in speech coding for modeling envelopes of transients signals, such as voiced and unvoiced stops, plosives, etc. FDLP fits an auto regressive model to the discrete cosine transform (DCT) coefficients of a sequence. [2]

Ante Jukić, Toon van Waterschoot, Timo Gerkmann, Simon Doelo1(2014) Speech Dereverberation with Multi-Channels Linear Prediction And Sparse Priors For The Desired SIGNAL. This states that the quality of recorded speech signals can be substantially affected by room reverberation. [3]
years, sparse representation is adopted to improve the quality of noise corrupted speech. [10]

Yongjun He, Jiqing Han, Shiwen Deng, Tieran Zheng, Guibin Zheng; (2012) “A Solution to Residual Noise in Speech De-noising with Sparse Representation,” Sparse representation has been extensively investigated in signal processing community. [11]

Van Hamme, H.; Cranen, B.; Boves, L.; (2010) “Compressive Sensing for Missing Data Imputation in Noise Robust Speech Recognition,” An effective way to increase the noise robustness of automatic speech recognition is to label noisy speech features as either reliable or unreliable (missing), and to replace (impute) the missing ones by clean speech estimates. [12]

Sigg, C.D., Dikk, T., Buhmann, J.M.; (2010) “Speech Enhancement with Sparse Coding in Learned Dictionaries,” The enhancement of speech degraded by non-stationary interferers is a highly relevant and difficult task of many signal processing applications. [13]

Anis Ben Aicha; (2005), “Adaptive Noise Smoothing for Speech Enhancement Filters,” Noise estimation is one of the main keys for successful speech enhancement processing. Traditional noise estimators based on inter-frame smoothing of the estimated power spectrum (PSD) of the background noise leads to artificial artifacts on the residual noise well known as ‘musical noise’. [14]


Loizou, Philipos, Kamath, Sunil; (2002) “A multi-band spectral subtraction method for enhancing speech corrupted by colored noise.” The spectral subtraction method is a well-known noise reduction technique. Most implementations and variations of the basic technique advocate subtraction of the noise spectrum estimate over the entire speech spectrum. [16]

3. Algorithm

The presented work is implemented in following steps:

- Speech Signal Acquisition
  \[ x(n) = \sum_{k=1}^{N} \omega(n) \cos \left( \frac{(2n-1)(k-1)}{2N} \right), \quad k = 1, \ldots, N \]
  \[ \omega(n) = \frac{1}{N} \quad k = 1 \]
  \[ \omega(n) = \frac{2}{N} \quad 2 \leq k \leq N \]

- The residual noise is extracted using linear predictor coefficients and de-noised sound is received after taking the difference of noisy signal and residual noise. The snr before and after is computed in order to measure the performance of the algorithm.

Results and Conclusion

The presented algorithm has been implemented on different audio signal of different frequency. The signal to noise ratio before and after the applying of the proposed algorithm is shown in result table below. The improvement in signal to noise ratio proves the fairness of the presented algorithm when de-noising the audio signal.

Table 1: SNR before and after De-noising

<table>
<thead>
<tr>
<th>Sr. No.</th>
<th>Test Signal Freq.</th>
<th>SNR Before De-noising (db)</th>
<th>SNR After Denoising (db)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5</td>
<td>4.004</td>
<td>8.878</td>
</tr>
<tr>
<td>2</td>
<td>10</td>
<td>4.026</td>
<td>8.560</td>
</tr>
<tr>
<td>3</td>
<td>20</td>
<td>4.053</td>
<td>8.261</td>
</tr>
<tr>
<td>4</td>
<td>30</td>
<td>3.955</td>
<td>8.138</td>
</tr>
<tr>
<td>5</td>
<td>40</td>
<td>3.794</td>
<td>8.422</td>
</tr>
</tbody>
</table>
Flow Chart

References


Author Profile

Ms. Diksha Sharma is pursuing her M.Tech in ECE from DIT, Kharar, Mohali, INDIA. Her field of interest is in digital signal processing based system design and application.