

Frequency Dependent Noise Flooring Parameter (FDNFP) For Speech Enhancement

Anuja Chougule¹, Dr.Mrs.V.V.Patil²

P.G. Student, Department of Electronics Engineering, Dr.JJMCOE, Jaysinghpur, Maharashtra, India 1

Professor and HOD, Department of Electronics, Dr.JJMCOE, Jaysinghpur, India²

ABSTRACT: An enhancement method for single-channel speech degraded by additive noise is proposed. A spectral weighting functions derived by constrained optimization to suppress noise in the frequency domain. Two design parameters are included in the suppression gain, namely, the frequency-dependent noise flooring parameter (FDNFP) and the gain factor. The FDNFP controls the level of admissible residual noise in the enhanced speech. Enhanced harmonic structures are incorporated into the FDNFP by time-domain processing of the linear prediction residuals of voiced speech. Further enhancement of the harmonics is achieved by adaptive comb filtering derived using the gain factor with a peak-picking algorithm.

KEYWORDS: Speech Enhancement, FDNFP,

I. INTRODUCTION

The enhancement of single-channel speech degraded by additive noise has been extensively studied in the past and remains a challenging problem because only the noisy speech is available. Techniques have been proposed to exploit the harmonic structure of voiced speech for enhancing the speech quality [1]. In the work of [1] and [2], voiced speech is modeled as harmonic components plus noise-like components, and enhancement is performed by estimating the harmonic components while reducing the additive noise in the noise-like components. [3].

Here we propose a new method that enhances the harmonics of voiced speech without ascribing to any underlying speech models. The harmonic speech structure obtained through short-time Fourier analysis is enhanced by applying a combination of time and frequency domain-based criteria, which are applicable for white as well as for colored additive noise conditions.

The proposed method improves speech quality by suppressing the noise in the frequency domain with the use of a spectral weighting function. Two design parameters are introduced into the proposed suppression gain, namely the frequency-dependent noise-flooring parameter (FDNFP) and the gain factor. The FDNFP shapes the residual noise in the frequency domain such that the harmonic structure of clean speech is preserved. To further enhance the harmonics of voiced speech, adaptive comb filtering is performed. The performance of the enhancement method was evaluated by the modified bark spectral distance (MBSD), ITU-Perceptual Evaluation of Speech Quality (PESQ) scores,[4] composite objective measures and listening tests. Experimental results indicate that the proposed method outperforms spectral subtraction; a main signal subspace method applicable to both white and colored noise conditions and a perceptually based enhancement method with a constant noise-flooring parameter, particularly at lower signal-to-noise ratio conditions. Our listening test indicated that 16 listeners on average preferred the proposed approach over any of the other three approaches about 73% of the time.

II.LITERATURE REVIEW

J. Jensen and J. H. L. Hansen, has been given their ideas in the paper. This paper presents a sinusoidal model based algorithm for enhancement of speech degraded by additive broad-band noise. In order to ensure

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 2, February 2015

speech-like characteristics observed in clean speech, smoothness constraints are imposed on the model parameters using a spectral envelope surface (SES) smoothing procedure. [2]

Y. Stylianou, given various concepts regarding harmonic plus noise model (HNM) which divides the speech signal in two sub bands: harmonic and noise, is implemented with the objective of studying its capabilities for improving the quality of speech synthesis in Investigations show that HNM is capable of synthesizing all vowels and syllables with good quality. [4]

Y. Hu and P.C. Loizou, given various ideas for removing noise from audio signals requires a non-diagonal processing of time-frequency coefficients to avoid producing “musical noise”. State of the art algorithms perform a parameterized filtering of spectrogram coefficients with empirically fixed parameters. [5]

III. SPEECH ENHANCEMENT

We discuss the introduction about speech enhancement which includes- what is mean by speech enhancement? , why it is required? Actually speech enhancement improves the quality of speech signal. Speech enhancement aims to improve speech quality by using various algorithms. It may sound simple, but what is meant by the word *quality*? It can be at least

- 1) clarity and intelligibility,
- 2) pleasantness, or
- 3) compatibility

with some other method in speech processing.

Intelligibility and pleasantness are difficult to measure by any mathematical algorithm. Usually listening tests are employed. However, since arranging listening tests may be expensive, it has been widely studied how to predict the results of listening tests. No single philosopher’s stone or minimization criterion has been discovered so far and hardly ever will. The central methods for enhancing speech are the removal of background noise, *echo suppression* and the process of artificially bringing certain frequencies [5].

When the background noise is suppressed, it is crucial not to harm or garble the speech signal. Another thing to remember is that quiet natural background noise sounds more comfortable than more quiet unnatural twisted noise. If the speech signal is not intended to be listened by humans, but driven for instance to a speech recognizer, then the comfortless is not the issue. It is crucial then to keep the background noise low. into the speech signal. We shall focus on the removal of background noise.

IV. PROPOSED METHOD

In many real world applications the speech enhancement is necessary for getting clear speech signal instead of speech signal with added background noise. Enhancing of speech degraded by noise, or noise reduction, is the most important field of speech enhancement, and used for many applications such as mobile phones, VoIP, teleconferencing systems, speech recognition, and hearing aids . The proposed method improves speech quality by suppressing the noise in the frequency domain with the use of a spectral weighting function. Two design parameters are introduced into the proposed suppression gain, namely the frequency-dependent noise-flooring parameter (FDNFP) and the gain factor[6].

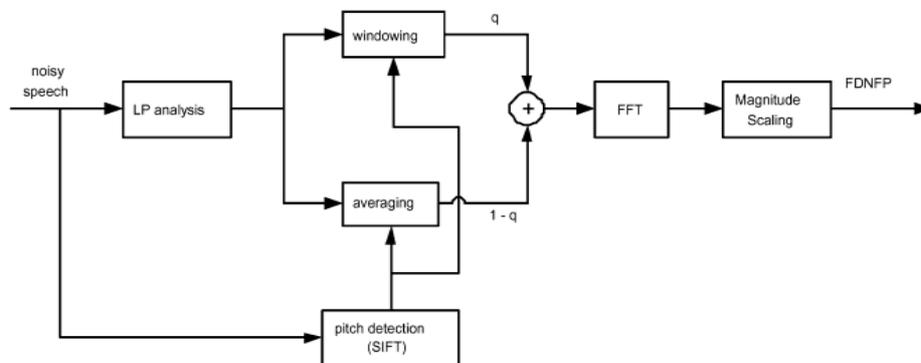


Fig.1. Computation of the frequency-dependent noise flooring parameters using LP analysis.

Fig.1 depicts the steps for computing the FDNFP. For voiced speech, a linear prediction (LP) analysis is performed on the noisy speech. In our implementation, the classical autocorrelation method is used to derive the LP parameters. The model order is set to 15. The LP residual signal is processed in parallel by two different methods to enhance the excitation peaks.[7] The first method attenuates the signal amplitudes between excitation peaks by windowing the LP residual signal with a Kaiser window series. The duration of each window is set to be equal to the pitch period.

The centers (peaks) of the windows are aligned in time with the peaks of excitation pulses. The purpose of windowing is to enhance the amplitude contrast between peaks and valleys of the excitation pulses. By averaging the LP residuals over several pitch periods, the periodic components will therefore be enhanced while the uncorrelated random components will be suppressed. In order to provide the necessary pitch information for the aforementioned windowing and averaging process, a pitch detection algorithm is run in parallel to determine the pitch period of the current frame [9]

IV.I LINEAR PREDICTION

In speech coding applications, the LPC parameters are extracted frame-wise from the speech signal. The LPC parameters are quantized prior to their transmission. Most commonly, memory less quantizers using 20 to 40 bits are employed to encode the LPC parameters at each frame update[10]. Linear predictive coding (LPC) is a tool used mostly in audio signal processing and speech processing for representing the spectral envelope of a digital signal of speech in compressed form, using the information of a linear predictive model. It is one of the most powerful speech analysis techniques, and one of the most useful methods for encoding good quality speech at a low bit rate and provides extremely accurate estimates of speech parameters.

LPC starts with the assumption that a speech signal is produced by a buzzer at the end of a tube (voiced sounds), with occasional added hissing and popping sounds (sibilants and plosive sounds). Although apparently crude, this model is actually a close approximation of the reality of speech production. The glottis (the space between the vocal folds) produces the buzz, which is characterized by its intensity (loudness) and frequency (pitch).The vocal tract (the throat and mouth) forms the tube, which is characterized by its resonances, which give rise to formants, or enhanced frequency bands in the sound produced. Hisses and pops are generated by the action of the tongue, lips and throat during sibilants and plosives.

LPC analyzes the speech signal by estimating the formants, removing their effects from the speech signal, and estimating the intensity and frequency of the remaining buzz. The process of removing the formants is called inverse filtering, and the remaining signal after the subtraction of the filtered modeled signal is called the residue. The numbers which describe the intensity and frequency of the buzz, the formants, and the residue signal, can be stored or transmitted somewhere else.

LPC synthesizes the speech signal by reversing the process: use the buzz parameters and the residue to create a source signal, use the formants to create a filter (which represents the tube), and run the source through the

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 2, February 2015

filter, resulting in speech. Because speech signals vary with time, this process is done on short chunks of the speech signal, which are called frames; generally 30 to 50 frames per second give intelligible speech with good compression. LPC is frequently used for transmitting spectral envelope information, and as such it has to be tolerant of transmission errors. Transmission of the filter coefficients directly (see linear prediction for definition of coefficients) is undesirable, since they are very sensitive to errors. In other words, a very small error can distort the whole spectrum, or worse, a small error might make the prediction filter unstable.

There are more advanced representations such as log area ratios (LAR), line spectral pairs (LSP) decomposition and reflection coefficients. Of these, especially LSP decomposition has gained popularity, since it ensures stability of the predictor, and spectral errors are local for small coefficient deviations.

Applications of LPC-

1. LPC is generally used for speech analysis and resynthesis. It is used as a form of voice compression by phone companies, for example in the GSM standard. It is also used for secure wireless, where voice must be digitized, encrypted and sent over a narrow voice channel; an early example of this is the US government's Navajo I.
2. LPC synthesis can be used to construct vocoders where musical instruments are used as excitation signal to the time-varying filter estimated from a singer's speech. This is somewhat popular in electronic music. Paul Lansky made the well-known computer music piece not just moreidlechatter using linear predictive coding.[1] A 10th-order LPC was used in the popular 1980s Speak & Spell educational toy.
3. LPC predictors are used in Shorten, MPEG-4 ALS, FLAC, SILK audio codec, and other lossless audio codecs.
4. LPC is receiving some attention as a tool for use in the tonal analysis of violins and other stringed musical instruments [11].

IV.II PITCH DETECTION

A **pitch detection algorithm (PDA)** is an algorithm designed to estimate the pitch or fundamental frequency of a quasiperiodic or virtually periodic signal, usually a digital recording of speech or a musical note or tone. This can be done in the time domain or the frequency domain or both the two domains.

For pitch detection there are many algorithms. –

- 1) Modified auto correlation method using clipping (AUTOCLIP)
- 2) Cepstrum method (CEP)
- 3) Simplified inverse filtering technique
- 4) Data reduction method (DARD)
- 5) Parallel processing method (PPROC)
- 6) Spectral equalization LPC method using Newton's transformation (LPC)
- 7) Average magnitude difference function (AMDF)

Out of these algorithms we will use SIFT for Pitch Detection. Now we will see the detail information about SIFT.

The simplified inverse filter tracking (SIFT) is an algorithm for classification of the voicing of speech segments and to estimate the pitch period of the speech labelled as voiced. Since both time- and frequency-domain approaches are used for the actual pitch detection; the SIFT algorithm is referred to as a hybrid pitch detector. This means that this algorithm influences the spectral properties using the inverse filter and extracts the pitch period information from the short-time autocorrelation.

The pitch period is obtained by interpolating the autocorrelation function in the neighbourhood of the peak of the autocorrelation function. A voiced-unvoiced decision is made on the basis of the amplitude of the peak of the autocorrelation function. The threshold used for this test is a normalized value of 0.4 for the autocorrelation peak.

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 2, February 2015

IV.III WINDOWING & AVERAGING

After being partitioned into frames each frame is multiplied by a window function prior to the spectral analysis to reduce the effect of discontinuity introduced by the before process by attenuating the values of the samples at the beginning and at the end of each frame, commonly used window is called Hamming. The purpose of windowing is to enhance the amplitude contrast between peaks and valleys of the excitation pulses. By averaging the LP residuals over several pitch periods, the periodic components will therefore be enhanced while the uncorrelated random components will be suppressed.

V. DISCUSSION

In this paper, speech enhancement method which aims at emphasizing harmonics is presented. For that FDNFP method is suggested which is based on Linear Prediction and Pitch Detection. FDNFP is used to emphasize the harmonics of voiced speech as well as to control the frequency-dependent level of admissible residual noise. For voiced speech, the periodicity in the linear prediction residual signal is detected and enhanced and then transformed to the frequency domain to be used as the FDNFP. The magnitudes of the FDNFP are scaled to some small values in order to suppress the level of residual noise in the enhanced speech. Windowing & Averaging is to enhance the amplitude contrast between peaks and valleys of the excitation pulses.

REFERENCES

- [1] J. Hardwick, C. D. Yoo, and J. S. Lim, "Speech enhancement using the dual excitation model," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, 1993, pp. 367–370.
- [2] J. Jensen and J. H. L. Hansen, "Speech enhancement using a constrained iterative sinusoidal model," *IEEE Trans. Speech Audio Process.*, vol. 9, no. 7, pp. 731–740, Oct. 2001.
- [3] M. E. Deisher and A. S. Spanias, "HMM-based speech enhancement using harmonic modeling," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, 1997, pp.
- [4] Y. Stylianou, "Applying the harmonic plus noise model in concatenative speech synthesis," *IEEE Trans. Speech Audio Process.*, vol. 9, no. 1, pp. 21–29, Jan. 2001.
- [5] Y. Hu and P. C. Loizou, "Incorporating a psycho acoustical model in frequency domain speech enhancement," *IEEE Signal Process. Lett.*, vol. 11, no. 2, pp. 270–273, Feb. 2004.
- [6] S. Wang, A. Sekey, and A. Gersho, "An objective measure for predicting subjective quality of speech coders," *IEEE J. Sel. Areas Commun.*, vol. 10, no. 5, pp. 819–828, Jun. 1992.
- [7] R. Martin, "Noise power spectral density estimation based on optimal smoothing and minimum statistics," *IEEE Trans. Speech Audio Process.*, vol. 9, no. 5, pp. 504–512, Jul. 2001.
- [8] S. F. Boll, "Suppression of acoustic noise in speech using spectral subtraction," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-27, no. 2, pp. 113–120, Apr. 1979.
- [9] A. Erell and M. Weintraub, "Estimation of noise-corrupted speech dftspectrum using the pitch period," *IEEE Trans. Speech Audio Process.*, vol. 2, pp. 1–8, Jan. 1994.
- [10] A.-T. Yu and H.-C. Wang, "New speech harmonic structure measure and it application to post speech enhancement," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, 2004, pp. 729–732.
- [11] T. F. Quatieri and R. G. Danisewicz, "Cochannel speaker separation by harmonic enhancement and suppression," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 38, no. 1, pp. 56–69, Jan. 1990.
- [12] A. Erell and M. Weintraub, "Estimation of noise-corrupted speech dftspectrum using the pitch period," *IEEE Trans. Speech Audio Process.*, vol. 2, pp. 1–8, Jan. 1994.