# Quantitative Analysis of Noise in Speech on MFCC Parameters

Vinod Kumar Pandey and Sunil Kumar Kopparapu
TCS Innovation Lab, Tata Consultancy Services
Yantra Park, Thane 400 601, Maharastra, India
Phone: +91 (0) 22 6778 88096, Fax: +91 (0) 22 6778 2190
E-mail: {Vinod.Pande, SunilKumar.Kopparapu}@tcs.com

***Abstract*—Manifestation of noise in speech is inherent unless a conscious effort is made to minimize the disturbances in the surroundings while recording speech. The performance of a speech recognition system often degrades in the presence of noise. Mel frequency cepstral coefficients (MFCCs) are the most popularly used acoustic features in speech and speaker recognition applications. In this paper, we investigate the effect of additive noise and number of Mel filters on the extracted speech features, used in speech recognition. Experimentally we demonstrate that the mean and variance of the error in MFCC due to additive noise in the speech is related to the mean and variance of the noise and number of Mel filters.**

***Keywords- MFCC; Gaussian noise, Mel filter bank .***

## I. INTRODUCTION

The performance of speech and speaker recognition systems, trained with clean speech, severely degrade in noisy environments. The recognition accuracy in a matched train-test conditions is very high, however the recognition accuracy decreases drastically when there is a mismatch in test-train conditions [1]. Reasons for the degradation are due to (a) acoustic mismatch between the training and testing condition and (b) noisy environments. Mel-frequency cepstral coefficient (MFCC) feature [2] extraction is most often used as a front-end module in speech and speaker recognition systems [1], [3], [4]. However, MFCC is not very robust in the presence of additive noise [5]. Performance of the recognition system, in unmatched test-train environments, can be improved by either fine-tuning the feature extraction module or the classifier module.

Several techniques have been reported for handling environmental noise [4]-[7]. Boll [7] used a spectral subtraction technique for suppressing the additive background noise. In this technique, an estimate of the noise spectrum computed during silence region is subtracted from noisy speech. Performance of this technique heavily depends on accurately detecting speech pauses to estimate noise. Ming *et al.* [4] investigated the problem of speaker identification and verification in noisy conditions. They assume that speech signals are corrupted by environmental noise, but the exact knowledge about the characteristics of noise was assumed to be unknown. Based on experiments, they concluded that one set of features is not optimal across different environmental conditions. Laxmi Narayana and Kopparapu [8] investigated effect of noise on the MFCC parameters. The error in MFCC was estimated by taking the difference between MFCC of clean speech and MFCC of noisy speech. It was shown that the distribution of the error in MFCC parameters is related to the noise in speech when the noise is additive Gaussian.

This work is an extension of our earlier work [8]. In this paper we study the effect of number of Mel-spaced triangular filters on the MFCC error. We show through experiments that the distribution of the MFCC error is not only related to the additive noise in speech but also the number of the Mel-filters. The rest of the paper is organized as follows. Section 2 presents the method for estimating the error in the MFCC parameters. Section 3 gives the details of the experiments conducted to find the distribution of the error signal, followed by a conclusion in Section 4.

## II. EFFECT OF NOISE ON MFCC PARAMETERS

For computing MFCC features, speech signal is first windowed into overlapping frames, typically of length 20-30 ms, and the spectrum is computed using DFT for each frames. Next, the absolute value of the speech spectrum is passed through a filter bank of Mel-spaced triangular filters whose center frequencies are spaced along the perceptually motivated Mel frequency scale [4], [9]. The filter band output is log-compressed and the Mel frequency cepstral coefficients are estimated by applying Discrete Cosine Transform (DCT) of the filter bank spectrum (in dB).

Important parameters that define a Mel filter bank are (a) number of Mel filters, (b) minimum frequency and (c) maximum frequency. The Mel filters are spread over the whole frequency range from the minimum frequency to the maximum frequency. For speech, generally, minimum frequency greater than 100 Hz is used [2], [9] ensuring rejection of hum resulting from the AC power. The maximum frequency is chosen less than the maximum frequency. Usually, maximum frequency of 6.8 kHz is used for designing Mel filter [9]. The number of Mel filters can vary from 30 to 40. However, it has been suggested [9]

using 40, 36, and 31 Mel filters for analyzing the speech sampled at 16 kHz, 11 kHz, and 8 kHz, respectively.

As discussed earlier in introduction, MFCC is very sensitive to the additive noise. Several algorithms have been proposed for improving the noise robustness of the MFCC features [5], [6], [10]. Tyagi and Wellekens [5] multiplied the log compressed Mel filter outputs by an exponential function. They showed that the scaling reduces the influence of low-energy components caused by the additive noise. Lima *et al.* [6] used spectral normalization of the MFCC features for increasing the noise robustness. Skowronski and Harris [10] redesigned the Mel filters which have wider bandwidth and overlap more with neighboring filters compared to the conventional Mel filters. They showed that the redesigned filters are robust to additive noise.

In this paper, we investigate how the noise-in-speech and number of Mel filters affects the MFCC parameters computation. We use error in MFCC parameters, estimated by taking difference between MFCCs derived from clean speech and the MFCC from noisy speech, to characterize the noise. The noisy speech is generated by adding Gaussian noise of certain mean and variance to clean speech of different signal-to-noise ratio (SNR) values. Different numbers of filters are used to establish the relationship between the error in the MFCC parameters and the number of banks in the Mel filter.

## III. EXPERIEMNTS AND RESULTS

To characterize the additive noise, and its effect on MFCC parameters and number of Mel filter bands, we perform experiments on several recorded utterances, recorded from several subjects. In all our experiments we sample speech signal at 16 kHz with 16 bits. The speech signal is divided into frames of duration 25 ms with 50% overlap. The speech frames are multiplied by Hamming window and DFT is computed for each of the frames. The Fourier transform is warped according to the Mel scale. The log energies for the various number of Mel filters are calculated and further transformed to cepstral domain by applying the DCT. We used $k$ = 20, 30, 40 number of Mel filters spread from 100 Hz (minimum frequency) to a maximum frequency of 8 kHz. We retain the first 13 cepstral coefficients (excluding $0^{th}$ coefficient) for each of the frames.

Initially the MFCC parameters ($\Phi_{\underline{k}}$), for $k$ number of filter bank, are computed for the clean speech $x[n]$. Gaussian noise, $\gamma^{(\mu,\ \sigma 2)}[n]$, with different means ($\mu$) and variances ($\sigma^2$) are generated and added to the speech signal $x[n]$ for generating noisy speech $Y^{(\mu,\ \sigma 2)}[n]$

$$Y^{(\mu,\sigma^2)}[n] = x[n] + \gamma^{(\mu,\sigma^2)}[n] \tag{1}$$

The signal-to-noise ratio (SNR) of the noisy speech $Y^{(\mu,\ \sigma 2)}[n]$ is given by

$$SNR_{in} = 10\log_{10}\left(\frac{P_x}{P_\gamma}\right) = 20\log_{10}\left(\frac{A_x}{A_\gamma}\right) \tag{2}$$

where $P_x$ and $A_x$ are the average power and root mean square (RMS) amplitude of the speech signal $x[n]$ and $P_\gamma$ and $A_\gamma$ are the average power and RMS amplitude of the Gaussian noise $\gamma^{(\mu,\ \sigma 2)}[n]$. The MFCC parameters extracted from $Y^{(\mu,\ \sigma 2)}[n]$ is denoted by $\Phi^{\gamma}_{\underline{k}}$ for $k$ number of the filter bank. The error in the MFCC parameters, for $k$ number of filter bank, due to additive noise is

$$E_{\underline{k}} = \Phi^{\gamma}_{\underline{k}} - \Phi_{\underline{k}} \tag{3}$$

We assume that the distribution of error in the MFCC parameters is Gaussian. The mean $\mu_E$ and variance $\sigma_E^2$ of the error in MFCC parameter, for varying additive Gaussian noise $\gamma^{(\mu,\ \sigma 2)}[n]$ and different number of Mel filters ($k$) with the corresponding SNR values are tabulated in Table 1.

It can be observed from Table 1 that the error in MFCC parameters is independent of $\mu$ and depends only on the variance ($\sigma^2$) of the additive noise across different number of Mel filter bank. Observe that even a large value of $\mu$ has minimal influence on $\mu_E$. Values of $\mu_E$ are found to be different with different number of Mel filter bank. However, $\sigma_E^2$ depends on the variance of the additive noise and $k$ the number of filter banks.

The relation between the variance of the error in calculating MFCC ($\sigma_E^2$) and the variance of the additive Gaussian noise ($\sigma^2$) can be modeled as

$$\sigma_E^2 \approx (0.0049k - 0.037)\sigma^2 - (0.0045k - 0.065) \tag{4}$$

where $k$ is number of Mel filters and its value varies from 20 to 40.

## IV. CONCLUSION

The performance of a speech recognition system often degrades in noisy conditions. Mel frequency cepstral coefficients are the popularly used speech features in speech and speaker recognition systems. The effect of additive Gaussian noise-in-speech and number of Mel filter banks on the extracted MFCC parameters are investigated. We have shown experimentally that the variance in the error of extracted MFCC parameters is related to the variance of the additive Gaussian noise and number of Mel filters. Experimental results show that the mean of the additive Gaussian noise does not have much influence on the parameters of the error, irrespective of the number of the filter bank.

TABLE I. ERROR IN MFCC WITH VARYING NOISE-IN-SPEECH AND NUMBER OF DIFFERENT MEL FILTER BANK FOR NOISY SPEECH OF DIFFERENT INPUT SNR

| Input noise parameters | | $\mu_E$ | | | $\sigma_E^2$ | | | SNR |
|---|---|---|---|---|---|---|---|---|
| μ | $\sigma^2$ | 20 banks | 30 banks | 40 banks | 20 banks | 30 banks | 40 banks | |
| 0 | 1 | -0.01 | 0.00 | 0.00 | 0.04 | 0.05 | 0.06 | 33.65 |
| 0 | 2 | -0.15 | 0.00 | -0.01 | 0.09 | 0.17 | 0.19 | 30.64 |
| 0 | 3 | -0.02 | -0.01 | -0.02 | 0.16 | 0.24 | 0.35 | 28.80 |
| 0 | 4 | -0.02 | -0.01 | -0.03 | 0.22 | 0.41 | 0.52 | 27.63 |
| 0 | 5 | -0.03 | -0.02 | -0.03 | 0.28 | 0.53 | 0.69 | 26.70 |
| 1 | 1 | -0.01 | -0.01 | 0.00 | 0.04 | 0.06 | 0.07 | 28.90 |
| 2 | 1 | -0.01 | -0.01 | 0.00 | 0.04 | 0.06 | 0.07 | 26.62 |
| 3 | 1 | -0.01 | -0.01 | 0.00 | 0.04 | 0.06 | 0.07 | 23.60 |
| 4 | 1 | -0.01 | -0.01 | 0.00 | 0.04 | 0.06 | 0.07 | 21.30 |
| 5 | 1 | -0.01 | -0.01 | 0.00 | 0.04 | 0.06 | 0.07 | 19.50 |

## REFERENCES


[1] M. Benzeghiba, R. de Mori, O. Deroo, S. Dupont, T. Erbes, D. Jouvet, L. Fissore, P. Laface, A. Mertins, C. Ris, R. Rose, V. Tyagi, and C. Wellekens, "Automatic speech recognition and speech variability: A review," Speech Communication, vol. 49, pp. 763-786, 2007.

[2] S. B. Davis and P. Mermelstein, "Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences," IEEE Trans. Acoust., Speech, Signal Process, vol. 28, no. 4, pp. 357–366, 1980.

[3] M. R. Hasan, M. Jamil, M. G. Rabbani, and M. S. Rahman, "Speaker identification using Mel frequency cepstral coefficients," Proc. 3rd Int. Conf. on Electrical & Computer Engineering, Dec. 2004, pp. 28-30.

[4] J. Ming, T. J. Hazen, J. R. Glass, and D. A. Reynolds, "Robust speaker recognition in noisy conditions," IEEE Trans. Acoust., Speech, Signal Process, vol. 15, pp. 1711–1723, 2007.

[5] V. Tyagi and C. Wellekens, "On desensitizing the Mel-Cepstrum to spurious spectral components for robust speech recognition," Proc. IEEE ICASSP'05, vol. 1, 2005, pp. 529–532.

[6] C. S. Lima, A. C. Tavares, C. A. Silva, J. F. Oliveira, "Spectral normalization MFCC derived features for robust speech recognition," Proc. Int. Conf. on speech and Computers, 2004.

[7] S. F. Boll, "Suppression of acoustic noise in speech using spectral subtraction," IEEE Trans. Acoust., Speech, Signal Process., vol. 27, pp. 113-120, 1979.

[8] M. Laxmi Narayana and S. K. Kopparapu, "Effect of Noise-in-speech on MFCC Parameters", Proc. 9th WSEAS Int. Conf. on. Signal, Speech, and Image Processing (SSIP '09), 2009.

[9] CMU. http:// cmusphinx.sourceforge.net/ sphinx4/javadoc/ edu/ cmu/ sphinx/ frontend/ frequencywarp/melfrequencyfilterbank.html.

[10] M. D. Skowronski and J. G. Harris, "Increased MFCC filter bandwidth for noise-robust phonemerecognition," Proc. IEEE ICASSP'02, vol. 1, 2002, pp. 801–804.