# Advanced Feature Extraction for Robust Speech Recognition based on Maximizing the Sharpness of the Power Distribution

Dr. B. C Premkumar[1], Dr. Prasanna Kumar. C[2], Dr. Suresh D[3]
[1]*Professor, ECE Department, SCEM, Mangalore*
[2]*Associate Professor, ECE Department, AIET, Moodbidre, Mangalore*
[3]*Dr. Suresh. D, Professor, ECE Department, RNSIT, Bangalore*

***Abstract*** **-This paper presents a new robust feature extraction algorithm based on a modified approach to power bias subtraction combined with applying a threshold to the power spectral density. Power bias level is selected as a level above which the signal power distribution is sharpest. The sharpness is measured using the ratio of arithmetic mean to the geometric mean of medium-duration power. When subtracting this bias level, power flooring is applied to enhance robustness. These new ideas are employed to enhance our recently introduced feature extraction algorithm PNCC (Power Normalized Cepstral Coefficient). While simpler than our previous PNCC, experimental results show that this new PNCC is showing better performance than our previous implementation.**

***Index Terms*****— Robust speech recognition, physiological modeling, sharpness of power distribution, power flooring, auditory threshold.**

## I.INTRODUCTION

This project presents a new robust feature extraction algorithm based on a modified approach to power bias subtraction combined with applying a threshold to the power spectral density. The sharpness is measured using the ratio of arithmetic mean to the geometric mean of medium-duration power. When subtracting this bias level, power flooring is applied to enhance robustness. These new ideas are employed to enhance our recently introduced feature extraction algorithm PNCC (Power Normalized Cepstral Coefficient).

The introduction of hidden Markov models and statistical language modeling techniques has greatly improved the performance of speech recognition systems in clean environments. Nevertheless, speech recognition accuracy still degrades significantly in noisy environments.

Many algorithms have been proposed to address this problem and they have demonstrated significant improvement in performance for quasi-stationary noise. Unfortunately these same algorithms frequently do not show comparable improvements in more difficult transitory environments such as background music. In this paper we describe a new approach to power-bias subtraction that is based on maximization of the sharpness of the power distributions.

First, instead of matching the sharpness of the distribution of power coefficients to a training database, we simply maximize this sharpness distribution. We continue to use the ratio of the arithmetic mean to the geometric mean of the power co-efficients, which we refer to as the "AM-to-GM ratio", as this measure has proved to be a useful and easily-computed way to characterize the data.

Second, we apply a minimum threshold to these power values (which we call "power flooring)" because the spectro temporal segments representing speech that exhibit the smallest power are also the most vulnerable to additive noise. Using power flooring, we can reduce spectral distortion between training and test sets for these regions.

## II.LITERATURE SURVEY

The paper titled "Feature extraction for robust speech recognition using a power-law nonlinearity and power-bias subtraction" [1], gives information about power bias level selected as a level above which the signal power distribution is sharpest. The sharpness is

measured using the ratio of arithmetic mean to the geometric mean of medium-duration power. When subtracting this bias level, power flooring is applied to enhance robustness.

## III DESIGN METHODOLOGY

### REVIEW OF PNCC STRUCTURE

The structure of PNCC system is described in Fig. 1. Briefly, the PNCC procedure is as follows:

A pre-emphasis filter of the form H(z) = 1 - 0:97z-1 is to the input first. The Short time Fourier analysis follows using Hamming windows of duration 25.6 ms, with 10 ms between frames. Spectral analysis is accomplished by integrating the squared magnitude spectrum integration over frequency using weighting coefficients derived from the transfer functions of a 40- channel gammatone-shaped bank. The center frequencies of the gammatone filters are linearly spaced in the Equivalent Rectangular Bandwidth (ERB) scale between 200 Hz and 8000 Hz.

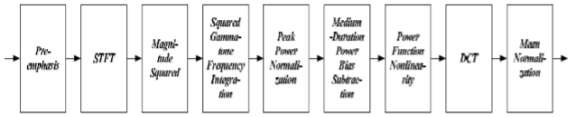


Fig 1: The structure of PNCC feature extraction

We obtain the short-time spectral power p(m; l) using the squared gammatone integration as shown below:

$$P_{org}(m,l) = \int_0^{\pi} |X(m;e^{j\omega})H_l(e^{j\omega})|^2 d\omega \qquad (1)$$

where Porg(m; l) is the short-time spectral power in the mth frame and the lth gammatone channel, Hl(ejw) is the frequency response of the l-th channel, and X(m; ejw) is the short-time spectrum of the m-th frame of the signal. The power is normalized using the peak power Ppeak (the 95th percentile of the short-time power) as shown below:

$$P(m,l) = p_0 \frac{P_{org}(m,l)}{P_{peak}} \qquad (2)$$

The p0 value may be considered as a constant scaling factor; its actual value is not important provided that the generated features are in the normal range for the speech recognition system in question. We use medium-duration power for the PBS processing, which is the running average of the short-time power P(m; l) as given below

$$Q(m,l) = \frac{1}{2M+1} \sum_{l'=l-M}^{l+M} P(m,l') \qquad (3)$$

We use M = 2 in our study based on speech recognition results obtained with different values of M. Using the PBS processing with power flooring, we obtain the processed power P'(m; l). After this PBS processing (with smoothing across channels), we apply the power-law nonlinearity (power to 1=15), and the result is applied to the Discrete Cosine Transform (DCT) as in the case of conventional MFCC.

### IMPLEMENTATION OF PBS

The objective of PBS is to apply a bias to the power in each of the frequency channels that maximizes the sharpness of the power distribution. This procedure is motivated by the fact that the human auditory system is more sensitive to changes in power over frequency and time than to relatively constant background excitation. The motivation of power flooring is twofold.

First, we wish to limit the extent to which power values of small magnitude affect, specifically to avoid values of Q(l) that are close to zero which cause the log value to approach negative infinity.

Second, since, small power regions are the most vulnerable to additive noise, we can reduce the spectral distortion caused by additive noise by applying power flooring both to the training and to test data.

## IV. RESULTS

The results described in this paper are also somewhat better than the previous results described in [1], which were obtained under exactly the same conditions. Improvements compared to the original implementation of PNCC were greatest at lowest SNRs and with background music. The improved PNCC algorithm is conceptually and computationally simpler, and it provides better recognition accuracy

## V. CONCLUSION & FUTURE SCOPE

We prefer to characterize improvement in recognition accuracy by the amount of lateral threshold shift provided by the processing. For white noise, PNCC provides an improvement of about 13db compare to MFCC.

ACKNOWLEDGMENT

We express our heart filled thanks to, Dr. M V Satyanarayana, Professor and Head, Department of Electronics and Communication Engineering, GSSSIETW, Mysore whose guidance and support goes beyond words.

We wholeheartedly thank our Guide, Sri. Jayanth J Associate Professor, Department of Electronics and Communication, GSSSIETW, Mysore, for having shared a genuine desire to make a positive contribution to address the challenges associated with every element of the project.

REFERENCES

[1] C. Kim and R. M. Stern, "Feature extraction for robust speech recognition using a power-law nonlinearity and power-bias subtraction," in INTERSPEECH-2009, Sept. 2009.

[2] P. J. Moreno, B. Raj, and R. M. Stern, "A vector Taylor series approach for environment-independent speech recognition," in Proc. IEEE Int. Conf. Acoust., Speech and Signal Processing, May. 1996.

[3] R. M. Stern, B. Raj, and P. J. Moreno, "Compensation for environmental degradation in automatic speech recognition," in Proc. of the ESCA Tutorial and Research Workshop on Robust Speech Recognition for Unknown Communication Channels, Apr. 1997.

[4] P. D. Patterson, K. Robinson, J. Holdsworth, D. McKeown, C. Zhang, and M. H. Allerhand, "Complex sounds and auditory images," in Auditory and Perception, Oxford, UK, 1992, pp. 429–446, Y. Cazals, L. Demany, and K. Horner, (Eds), Pergamon

[5] B. Raj, V. N. Parikh, and R. M. Stern, "The effects of background music on speech recognition accuracy," in *Proc. IEEE Int. Conf. Acoust., Speech and Signal Processing*, Apr. 1997 vol. 2, pp. 851–854.

[6] B. Raj and R. M. Stern, "Missing-Feature Methods for Robust Automatic Speech Recognition," *IEEE Signal Processing Magazine*, vol. 22, no. 5, pp. 101–116, Sept. 2005.

[7] H. Hermansky, "Perceptual linear prediction analysis of speech," *J. Acoust. Soc. Am.*, vol. 87, no. 4, pp.1738–1752, Apr. 1990.

[8] C. Kim, Y.-H. Chiu, and R. M. Stern, "Physiologically motivated synchrony-based processing for robust automatic speech recognition," in *INTERSPEECH-2006*, Sept. 2006, pp.1975–1978.

[9] C. Kim and R. M. Stern, "Feature extraction for robust speech recognition using a power-law nonlinearity and power-bias subtraction," in *INTERSPEECH-2009*, Sept. 2009.

[10] C. Kim and R. M. Stern, "Robust signal-to-noise ratio estimation based on waveform amplitude distribution analysis," in *INTERSPEECH-2008*, Sept. 2008, pp. 2598–2601.

[11] C. Kim ,K. Kumar and R. M. Stern, "Robust speech recognition using small power boosting algorithm," in *IEEE Automatic Speech Recognition and Understanding Workshop*, Dec. 2009 (accepted).

[12] C. Kim, K. Kumar, B. Raj, and R. M. Stern, "Signal separation for robust speech recognition based on phase difference information obtained in the frequency domain," in *INTERSPEECH-2009*, Sept. 2009.