ROBUST FEATURE EXTRACTORS FOR CONTINUOUS SPEECH RECOGNITION

M. J. Alam, P. Kenny, P. Dumouchel, D. O'Shaughnessy

CRIM, Montreal, Canada
ETS, Montreal, Canada
INRS-EMT, Montreal, Canada

ABSTRACT

This paper presents robust feature extractors for a continuous speech recognition task in matched and mismatched environments. The mismatched conditions may occur due to additive noise, different channel, and acoustic reverberation. In the conventional Mel-frequency cepstral coefficient (MFCC) feature extraction framework, a subband spectrum enhancement technique is incorporated to improve its robustness. We denote this front-end as robust MFCCs (RMFCC). Based on the gammatone and compressive gammachirp filter-banks, robust gammatone filterbank cepstral coefficients (RGFCC) and robust compressive gammachirp filterbank cepstral coefficients (RCGCC) are also presented for comparison. We also employ low-variance spectrum estimators such as multitaper, regularized minimum-variance distortionless response (RMVDR), instead of a discrete Fourier transform-based direct spectrum estimator for improving robustness against mismatched environments. Speech recognition performances of the robust feature extractors are evaluated in clean as well as multi-style training conditions of the AURORA-4 continuous speech recognition task. Experimental results depict that the RMFCC and low-variance spectrum-estimators-based robust feature extractors outperformed the MFCC, PNCC (power normalized cepstral coefficients), and ETSI-AFE features both in clean and multi-condition training conditions.

Index Terms**—** Robust feature extractor, speech recognition, multi-style training, aurora 4, multitaper

1. INTRODUCTION

The objective of automatic speech recognition (ASR) is to recognize human speech such as words or phonemes and sentences. Speech recognition systems may be divided into two modules: a front-end or feature extractor and a back-end or recognizer. The task of a feature extractor is to obtain a compact representation of a speech signal that compresses the relevant information into a small number of coefficients, e.g., Mel-frequency cepstral coefficients (MFCC) [1] and perceptual linear prediction (PLP) [2] coefficients. The back-end module recognizes the underlying content (i.e., text) of the input signal using the features extracted by the front-end. Conventional MFCC and PLP feature-based speech recognition systems perform well in ideal operating conditions where there is no mismatch between the training and test environments. A major impediment for deployment of speech recognition technologies is the degradation of recognition performance when differences exist between environments during the training and testing conditions. These differences, known as mismatched conditions, are due to corruption of speech signals by acoustic background noise, channel frequency response, different channels, and reverberation. Much research in the literature has been done to improve the robustness of speech recognition systems under mismatched conditions. The methods to compensate for the effects of environmental mismatch can be implemented at the front-end or at the back-end or both. Robust feature extractors are usually obtained either by appending a pre-processing step, like speech enhancement [3-5], or by incorporating algorithms in an MFCC or PLP computation framework such as power normalized cepstral coefficients (PNCC) [6], robust compressive gammachirp cepstral coefficients (RCGCC) [7], frequency masking [9], or by adding a post-processing step, like feature normalization techniques [8, 9], (e.g., cepstral mean normalization (CMN)) or by combining any two or all of the above mentioned steps [6-7, 10]. Most of the front-ends use, in addition to other techniques for environmental mismatch compensation, a feature normalization technique, at the least CMN, as a post-processing scheme. Additive noise reduction approaches usually have a tradeoff between the amount of noise reduction and speech distortion induced due to processing of a speech signal. At very low SNR the intensity of this induced distortion is high, thereby deteriorating the performance of the speech recognition systems.

In this work we present robust feature extractors that employ a subband spectrum enhancement technique based on *a posteriori* signal-to noise ratio (SNR) and as a post-processing scheme use a short-time feature normalization method to normalize the features. Depending on the method, a DFT-based direct spectrum estimator or regularized MVDR (RMVDR) spectrum estimator [19], used for estimating the speech power spectrum, two robust feature extractors, namely, robust MFCCs (RMFCC) and robust RMVDR cepstral coefficients (RRMCC) [16] are obtained. RMFCC is similar to robust compressive gammachirp filterbank cepstral coefficients (RCGCC) proposed in our previous work [7]. In RMFCC, a Mel-scale filterbank is used for auditory spectral analysis instead of a compressive gammachirp filterbank. Similarly, by incorporating a medium duration power bias subtraction (MDPBS) [6] into the MFCC framework, two front-ends, dubbed as normalized MFCCs (NMFCC) and normalized regularized MVDR cepstral coefficients (NRMCC) [16], are also presented. We also present multiple windowed spectrum estimation-based MFCCs, dubbed as multitaper MFCC (MMFCC), features for improving the performance of MFCC features in clean as well as multistyle training conditions.

In order to compare speech recognition performances of the above-mentioned front-ends, the following front-ends are chosen: MFCC, ETSI-advanced front-end (ETSI-AFE) [10], PNCC [6], and RCGCC [7]. Experimental results on the AURORA-4 corpus depict that the RMFCC, NMFCC, RRMFCC, NRMCC front-ends provide lower recognition WERs than the PNCC, ETSI-AFE, RCGCC, and RGFCC front-ends in both training conditions.

2. ROBUST FEATURE EXTRACTORS

Robust front-ends presented in this work are shown in fig. 1. The RMFCC feature extractor, similar to [7], incorporates a sigmoid shape suppression rule $W(n,m)$ based on the subband *a posteriori* signal-to-noise ratio (the ratio of noisy auditory spectra and the noise auditory spectra) $\gamma_{sb}(n,m)$ in order to enhance the auditory spectrum as:

$$W(n,m) = \frac{1}{1+e^{-\frac{(\gamma_{sb}(n,m)-c)}{a}}}, \qquad (1)$$

where n is the subband index, m is the frame index, $1/a$ and c are the slope and mean, respectively. Here, $a = c = 4.5$ is chosen experimentally [7].

For auditory frequency analysis of speech spectra, a mel-filterbank is used instead of a compressive gammachirp filter-bank. Similar to [6, 20] this feature extractor also uses power function nonlinearity with a coefficient of 1/15 to approximate the relationship between a human's perception of loudness and the sound intensity. For feature normalization a short-time mean and scale normalization (STMSN) [8] technique is used with a sliding window of 1.5 seconds. It, under mismatched conditions, helps to remove the difference of log spectrum between the training and test environments by adjusting the short-time mean and scale. Using a gammatone filterbank and following the same framework as the RMFCC we also present robust gammatone filterbank cepstral coefficient (RGFCC) features to provide a comparison of the Mel-, gammatone-, and compressive gammachirp-filterbanks in terms of speech recognition performances.

The only difference between the RRMCC and RMFCC front-ends are in the spectrum estimation. The RRMCC features are computed from regularized MVDR spectral estimates whereas DFT-based direct spectral estimates are used for computing RMFCC features.

The NMFCC front-end incorporates a medium-duration power bias subtraction (MDPBS) technique, originally proposed in [6], based on the arithmetic mean (AM)-geometric mean (GM) ratio, for background noise reduction. A cepstral mean normalization technique is used to normalize the features. The NRMCC front-end [16] can be considered a variant of the PNCC in which gammatone filterbank is replaced with the triangular-shaped Mel filterbank and a regularized MVDR (RMVDR) spectrum estimator [19] is used instead of a DFT-based direct spectrum estimator. As a post-processing scheme, cepstral mean normalization is used.

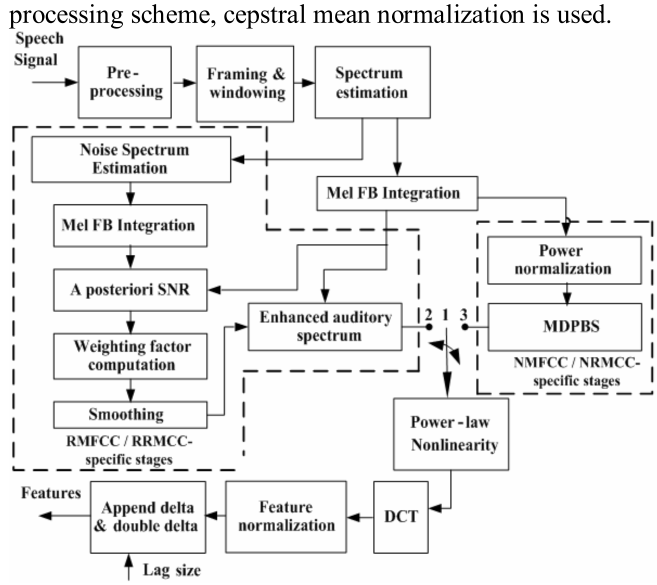

Fig. 1. Block diagram showing different steps of the robust mel-frequency cepstral coefficients (RMFCC), normalized MFCCs (NMFCC), robust regularized MVDR cepstral coefficients (RRMCC), and normalized regularized MVDR cepstral coefficients (NRMCC). Depending on the spectrum estimator, RMFCC/RRMCC features can be obtained when 1 is connected to 2. When points 1 & 3 are connected, NMFCC/NRMCC features can be obtained depending on the DFT-based spectrum estimation method/RMVDR spectrum estimator used.

3. REGULARIZED MVDR CEPSTRAL COEFFICIENTS (RMCC) FEATURE

When RMVDR spectrum estimator is used to compute the cepstral features instead of the DFT-based spectrum estimator we denote the features as the regularized MVDR cepstral coefficients (RMCC). RMCC was introduced in [19] and evaluated on the AURORA-4 corpus under clean training mode. Here we evaluate RMCC under multistyle training mode. In this section we provide a brief description about the RMVDR spectrum estimation method.

Similar to the MVDR spectrum estimator, the p-th order regularized MVDR spectral estimate can be parametrically written as

$$S_{RMVDR}\left(f\right)=\frac{1}{\sum_{k=-p}^{k=p}\mu_r\left(k\right)e^{-i2\pi fk}}, \tag{2}$$

where the parameter $\mu_r\left(k\right)$ of the regularized MVDR method can be obtained from a non-iterative computation using the regularized LP (RLP) coefficients a_q^r and the prediction error variance σ_e^r as:

$$\mu_r\left(k\right)=\begin{cases}\frac{1}{\sigma_e^r}\sum_{q=0}^{p-k}\left(p+1-k-2q\right)a_q^r a_{q+k}^{r*}, & \text{for } k\geq 0\\ \mu_r^*\left(-k\right), & \text{for } k<0.\end{cases} \tag{3}$$

The regularized predictor coefficients a_q^r are computed by adding a penalty measure $\psi\left(a^u\right)$, which is a function of the unknown predictor coefficients a^u, to the objective function of the LP method and therefore, minimizing the modified objective function of the following form [26-27]

$$\sum_n\left(s\left(n\right)+\sum_{q=1}^{p}a_q s\left(n-q\right)\right)^2+\lambda\psi\left(a^u\right), \tag{4}$$

where $s\left(n\right)$ is the current speech sample, regularization constant $\lambda>0$ controls the smoothness of the all-pole spectral envelope. RLP method helps to penalize the rapid changes in all-pole spectral envelope and therefore, produces a smooth spectral estimate keeping the formant positions unaffected [26]. The optimal values chosen for the model order p and regularization constant λ are 100 & 10^{-9}, respectively [19].

4. MULTIPLE WINDOWED MFCC (MMFCC) FEATURES

In the MMFCC front-end [17, 18, 21], cepstral features are computed from a multiple windowed or multitapered (e.g., Thomson window) spectrum estimate instead of the single windowed (e.g., Hamming) spectrum estimates as used in conventional MFCC. A multitaper spectrum estimator is given by:

$$\hat{S}_{MT}(m,k)=\sum_{p=1}^{M}\lambda\left(p\right)\left|\sum_{j=0}^{N-1}w_p\left(j\right)s\left(m,j\right)e^{-\frac{i2\pi jk}{N}}\right|^2, \tag{5}$$

where N is the frame length, w_p is the pth data taper ($p=1,2,...,M$) used for the spectral estimate $\hat{S}_{MT}(\cdot)$, which is also called the pth eigenspectrum, M denotes the number of tapers and $\lambda\left(p\right)$ is the weight corresponding to the pth taper. The tapers $w_p\left(j\right)$ are chosen to be orthonormal so that

$$\sum_j w_p\left(j\right)w_q\left(j\right)=\delta_{pq}=\begin{cases}1, & p=q\\ 0, & \text{otherwise.}\end{cases} \tag{6}$$

Unlike conventional tapers, the M orthonormal tapers used in a multi-taper spectrum estimator provide M statistically independent (hence uncorrelated) estimates of the underlying spectrum. The weighted average of the M individual spectral estimates $\hat{S}_{MT}\left(\bullet\right)$ then has smaller variance than the single-taper spectrum estimates by a factor that approaches $1/M$ [17]. The cepstral features computed from this spectrum will also have low variance [17, 18].

In this work we use the Thomson multitaper method. The tapers and taper wieghts in this method can be obtained using the following MATLAB function:

$$\left[w\ \lambda\right]=\text{dpss}(N,\ 3.5,\ M).$$

The optimum number of tapers for a continuous speech recognition task was found to be M = 4 for multistyle training and M = 6 for clean training condition.

5. EXPERIMENTS

To evaluate speech recognition performances of the front-ends considered in this work, experiments were carried out on the AURORA-4 LVCSR corpus [22, 23] using clean and multistyle training. Results are reported on the four different evaluation conditions mentioned in [23] and [24]. Word error rate (WER) is used as an evaluation metric.

5.1. Corpora used for performance evaluation

The AURORA-4 continuous speech recognition corpus, derived from the Wall Street Journal (WSJ0) corpus, has 14 test sets grouped into the following 4 groups [22-23]: (a) Test set A - clean/multistyle speech in training and clean speech in test, same channel (set 1), (b) Test set B - clean/multistyle speech in training and noisy speech in test, same channel (sets 2-7), (c) Test set C - clean/multistyle speech in training and clean speech in test, different channel (set 8), and (d) Test set D - clean/multistyle speech in training and noisy speech in test, different channel (sets 9-14). The number inside the brackets represents the test set number defined in the AURORA-4 corpus.

5.2. Experimental setup

For the continuous speech recognition task on the AURORA-4 corpus, all experiments employed state-tied crossword speaker-independent triphone acoustic models with 16 Gaussian mixtures per state. A single-pass Viterbi beam search-based decoder was used along with a standard 5K lexicon and bigram language model with a prune width of 250 [16, 23]. The HTK [24]-based recognizer is used for training and decoding tasks. For our experiments, we use 13 static features (including the 0th cepstral coefficient) augmented with their delta and double delta coefficients, making 39-dimensional feature vectors. The analysis frame length is 25 ms with a frame shift of 10 ms. The delta and double features were calculated using a 5-frame window. For all front-ends only static features were normalized by the feature normalization method and then dynamic features were computed from them.

5.3. Results and Discussion

Here, we considered both clean and multi-condition training to evaluate the performances of the following front-ends:
RMFCC: robust mel-frequency cepstral coefficients
NMFCC: normalized mel-frequency cepstral coefficients
RRMCC: robust regularized MVDR cepstral coefficients
NRMCC: normalized regularized MVDR cepstral coefficients
RMCC: regularized MVDR cepstral coefficients and
MMFCC: multitapered spectrum estimator-based mel-frequency cepstral coefficients.
For comparison the following front-ends are selected:
PNCC: power normalized cepstral coefficients
ETSI-AFE: The ETSI advanced front-end
RCGCC: robust compressive gammachirp filterbank cepstral coefficients
RGFCC: robust gammatone filterbank cepstral coefficients
MFCC: mel-frequency cepstral coefficients
Table 1 presents the WERs obtained by all feature extractors on the AURORA-4 corpus when the recognizer is trained on clean data and tested on all four test sets A, B, C, and D. It is observed from this table that low-variance spectrum estimation-based cepstral features, e.g., MMFCC and RMCC, help to reduce the WERs. As mentioned in [25], the variance in the feature vectors has a direct bearing to the variance of the Gaussians modeling speech classes. In general, reduction in feature vector variance increases class separability and, thereby, decreases recognition word error rates. Robust feature extractors RMFCC, RRMCC, NMFCC, and NRMCC provided better recognition accuracy (i.e., lower WERs) compared to other front-ends. The NRMCC performed the best among all front-ends presented in table 1.

		WER (%)				
		A	**B**	**C**	**D**	**Avg.**
1	**MFCC**	9.98	50.81	28.88	64.55	38.56
2	**MMFCC**	10.06	45.98	19.89	56.38	33.08
3	**RMCC**	**9.94**	45.75	21.77	59.37	34.21
4	**PNCC**	11.36	30.15	18.93	40.00	25.11
5	**ETSI-AFE**	11.41	30.42	20.48	38.49	25.20
6	**RCGCC**	11.10	31.13	19.06	40.75	25.51
7	**RGFCC**	11.06	31.53	19.30	40.61	25.63
8	**RMFCC**	11.27	29.15	18.90	38.05	24.34
9	**NMFCC**	11.97	27.37	17.97	**35.91**	23.31
10	**RRMCC**	10.71	28.80	18.12	37.12	23.69
11	**NRMCC**	11.38	**26.78**	**17.53**	**35.88**	**22.89**

Table 1. Word error rates (WERs) obtained by different feature extractors on the four evaluation conditions of the AURORA-4 task in clean training conditions (trained on clean data and evaluated on clean as well as noisy data). The lowest WERs are highlighted in boldface.

		WER (%)				
		A	**B**	**C**	**D**	**Avg.**
1	**MFCC**	14.62	23.84	19.19	31.47	22.28
2	**MMFCC**	15.15	24.12	18.17	30.50	22.02
3	**RMCC**	13.81	23.81	**17.53**	30.45	21.40
4	**PNCC**	15.29	25.81	17.86	31.33	22.57
5	**ETSI-AFE**	14.55	23.32	18.31	29.68	21.47
6	**RCGCC**	14.55	24.77	18.20	31.90	22.35
7	**RGFCC**	14.99	25.41	17.97	31.31	22.42
8	**RMFCC**	14.51	23.21	19.37	29.42	21.63
9	**NMFCC**	15.21	22.69	17.94	28.49	21.08
10	**RRMCC**	**13.37**	23.16	18.27	30.62	21.36
11	**NRMCC**	15.58	22.28	17.94	**28.23**	**21.01**

Table 2. Word error rates (WER) obtained by different feature extractors on the four evaluation conditions of the AURORA-4 task in multistyle training condition (trained on (clean + noisy) data and evaluated on clean as well as noisy data). The lowest WERs are highlighted in boldface.

When a multistyle training condition (i.e., training on (clean plus noisy) data and testing on all test data) is used, the WERs achieved by all the front-ends considered in this work are presented in table 2. The goal of multistyle training is to create matched training/test environments. Although it is expensive to obtain enough representation noisy data that can cover a wide range of noise types and signal-to-noise ratios, it is an effective method for mismatch compensation. On the average the MMFCC and RMCC features provided lower WERs compared to the MFCC in multistyle conditions. From table 2 it is evident that the RMFCC, RRMCC, NMFCC, NRMCC, RMCC, and ETSI-AFE yielded lower WERs. In multistyle training mode the performance of the RMCC is comparable to that of the RMFCC, RRMCC, NMFCC, NRMCC, ETSI-AFE. Among all the features the NRMCC performed the best (provided lowest WER) both in tables 1 & 2. Comparing the performances of the RCGCC, RGFCC and RMFCC front-ends from both tables, it can be said that the triangular-shaped Mel scale filterbank-based robust feature extractor (RMFCC) performed slightly better than the gammatone and compressive gammachirp filterbank-based robust features. Use of the RMVDR spectrum estimator (reduces spectral variance) yielded a slight reduction of WER (when comparing NRMCC versus NMFCC) over the windowed periodogram estimates. Robust feature extractors helped to reduce WER both in clean and multistyle training mode, though in multistyle training mode the reduction in WER is not as huge as observed in the clean training mode.

6. CONCLUSION

In this paper, we presented several robust feature extractors and their performances were evaluated and compared with the ETSI-AFE and PNCC under clean and multistyle training modes on the AURORA-4 corpus. It is found that the Mel filterbank-based robust extractors performed slightly better, in terms of WER, than the gammatone and compressive gammachirp filterbank-based features in additive background - channel mismatch environment. The normalized RMVDR-based cepstral coefficients (NRMCC) features outperformed all other features under clean as well as multi-style training modes.

REFERENCES

[1] S. Davis and P. Mermelstein, “Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences,” *IEEE Trans. Acoustics, Speech, and Signal Processing*, vol. 28, no. 4, pp. 357–366, August 1980.

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

[3] W. Zhu, D. O’Shaughnessy, “Using noise reduction and spectral emphasis techniques to improve ASR performance in noisy conditions,” Proc. ASRU 2003, US Virgin Islands, Nov. 2003.

[4] W. Zhu, D. O’Shaughnessy, “Incorporating frequency masking filtering in a standard MFCC feature extraction algorithm,” Proc. ICSP, pp. 617-620, Beijing, Aug.-Sep., 2004.

[5] J. Droppo, A. Acero, “Environmental robustness,” in springer handbook of speech processing, Benesy, J.; Sondhi, M. M. and Huang, Y. [Eds], pp. 653-679, 2008.

[6] C. Kim and R. M. Stern., “Feature extraction for robust speech recognition based on maximizing the sharpness of the power distribution and on power flooring,” In IEEE Int. Conf. on Acoustics, Speech, and Signal Processing, pp. 4574-4577, March 2010.

[7] M. J. Alam, P. Kenny, D. O'Shaughnessy, "Robust Feature Extraction for Speech Recognition by Enhancing Auditory Spectrum," Proc. INTERSPEECH, Portland Oregon September 2012.

[8] Alam, J., Ouellet, P., Kenny, P., O’Shaughnessy, D., “Comparative Evaluation of Feature Normalization Techniques for Speaker Verification,” *Proc NOLISP*, LNAI 7015, pp. 246-253, Las Palmas, Spain, November 2011.

[9] F. H. Liu, R. M. Stern, X. Huang, and A. Acero, “Efficient cepstral normalization for robust speech recognition," in *Proc. ARPA Human Language Technology Workshop '93*, (Princeton, NJ), pp. 69-74, Mar. 1993.

[10] ETSI ES 202 050, Speech Processing, Transmission and Quality aspects (STQ); Distributed speech recognition; advanced front-end feature extraction algorithm; Compression algorithms; 2003.

[11] M. Delcroix, T. Nakatani and S. Watanabe, “Static and Dynamic Variance Compensation for Recognition of Reverberant Speech With Dereverberation Preprocessing,” in Transactions of Audio, Speech, and Language Processing, 17(2), pp. 324-334, 2009.

[12] K. Kumar and R. M. Stern, "Maximum-likelihood-based cepstral inverse filtering for blind speech dereverberation," *IEEE International Conference on Acoustics, Speech, and Signal Processing*, March 2010.

[13] Vikramjit Mitra, H. Franco, M. Graciarena, A. Mandal, "Normalized Amplitude modulation features for large vocabulary noise-robust speech recognition," Proc. of ICASSP, pp. 4117-4120, 2012.

[14] M. J. Alam, P. Kenny, D. O'Shaughnessy, Smoothed Nonlinear Energy Operator-based Amplitude Modulation Features for robust speech recognition," Proc. of NOLISP, LNAI 7911, pp. 168-175, Springer, Heidelberg, 2013.

[15] Md. Jahangir Alam, Patrick Kenny, Douglas O'Shaughnessy, "A Novel Feature Extractor Employing Regularized MVDR Spectrum Estimator And Subband Spectrum Enhancement Technique," Proc. WOSSPA, Algiers, Algeria, May, 2013.

[16] Md. Jahangir Alam, Patrick Kenny, Douglas O'Shaughnessy, "Regularized MVDR Spectrum Estimation-based Robust Feature Extractors for Speech Recognition," Proc. INTERSPEECH, Lyon, France, 2013.

[17] Md. Jahangir Alam, Tomi Kinnunen, Patrick Kenny, Pierre Ouellet, Douglas O'Shaughnessy, "Multitaper MFCC and PLP Features for Speaker Verification Using i-Vectors," Speech Communication, vol. 55(2), pp. 237-251, February 2013.

[18] Md. Jahangir Alam, Patrick Kenny, Douglas O'Shaughnessy, "Low-variance Multitaper Mel-frequency Cepstral Coefficient Features for Speech and Speaker Recognition Systems," Cognitive Computation (Springer), December 2012.

[19] Md. Jahangir Alam, Patrick Kenny, Douglas O'Shaughnessy, "Speech Recognition Using Regularized Minimum Variance Distortionless Response Spectrum Estimation-Based Cepstral Features," Proc. ICASSP, Vancouver, Canada, May, 2013.

[20] C. Kim and R. M. Stern, “Feature extraction for robust speech recognition using a power-law nonlinearity and power-bias subtraction,” Proc. INTERSPEECH, Sept 2009.

[21] T. Kinnunen, R. Saeidi, F. Sedlak, K.A. Lee, J. Sandberg, M. Hansson-Sandsten, H. Li, “Low-Variance Multitaper MFCC Features: a Case Study in Robust Speaker Verification”, *IEEE Transactions on Audio, Speech and Language Processing*, vol. 20(7), pp. 1990--2001, September 2012.

[22] N. Parihar, J. Picone, D. Pearce, H.G. Hirsch, "Performance analysis of the Aurora large vocabulary baseline system," Proceedings of the European Signal Processing Conference, Vienna, Austria, 2004.

[23] S.-K. Au Yeung, M.-H. Siu, "Improved performance of Aurora-4 using HTK and unsupervised MLLR adaptation," Proceedings of the Int. Conference on Spoken Language Processing, Jeju, Korea, 2004.

[24] S. J. Young et al., HTK Book, Entropic Cambridge Research Laboratory Ltd., 3.4 edition, 2006. online: http://htk.eng.cam.ac.uk/.

[25] S. Dharanipragada, B. D. Rao, "MVDR based Feature Extraction for Robust Speech Recognition”, *Proc. ICASSP*, pp. 309-312, 2001.

[26] M. N. Murthi and W. B. Kleijn, “Regularized linear prediction all-pole models,” in *IEEE Speech Coding Workshop*, pp. 96–98, 2000.

[27] C. Hanilci, T. Kinnunen, F. Ertas, R. Saeidi, J. Pohjalainen, P. Alku, “Regularized All-Pole Models for Speaker Verification Under Noisy Environments”, *IEEE Signal Processing Letters* 19(3), 163--166, March 2012.