Automatic Speech Recognition System Based on Hybrid Feature Extraction Techniques Using TEO-PWP for in Real Noisy Environment

Wafa Helali, Zied Hajaiej and Adnen Cherif

Research Unite of Processing and Analysis of Electrical and Energetic Systems,
Faculty of Sciences of Tunis, University Tunis El-Manar, 2092 Tunis –TUNISIA

Abstract
Automatic speech recognition presents an interesting research area that has always attracted researchers to the general public. It is now giving rise to an important set of applications of a very varied nature and difficulty, involving millions of people around the world every day. In this paper, a model of speech recognition system in noisy environment is developed and analyzed. The proposed model relies on several hybrid feature extraction methods. Indeed, Teager-Energy Operator, Perceptual Wavelet Packet (TEO-PWP), Mel Cepstrum Coefficient (MFCC) and Perceptual Linear Production (PLP) are combined to construct a robust HMM based system. TIMIT database, which consist of both clean and noisy speech files recorded at different level of Speech-to-Noise Ratio (SNR, has been used for the system test. Results and observations are performed to prove the effectiveness of the proposed system relying on speech recognition rates.
Keywords:
Teager-Energy Operator TEO-PWP; Enhancement Speech; MFCC; PLP; RASTA-PLP; HMM.

1. Introduction

Nowadays, automatic speech recognition field has progressed considerably by revealing innovative algorithms and techniques for speech processing. Indeed, any speech recognition system must have the ability to analyze and detect a string of words or phonemes from the speech signal. Automatic speech processing opens up new perspectives given the considerable difference between manual and voice control. The use of natural language in the human/machine interface puts the technology within the reach of all by reducing the constraints of the use of control devices.

Although progress is still to be made on complex recognition systems, it should be noted that the recognition of small vocabularies is almost perfect, which is largely enough for everyday voice processing tools. Thus, the error rate and learning time of recognition systems are steadily decreasing to reach results close to 95%. This rate is obviously variable according to the difficulty of the language. Indeed the machine sometimes has trouble avoiding some language traps. With the advent of sophisticated systems, recognition speech rate improvement is the subject of all scientific research. Indeed, several methods of classification and parameterization have emerged to achieve this objective.

Many works in the literature have shown interests in development, analysis and improvement of speech recognition systems relying on different classification techniques based on parameterization enhancement.

Recently, Wavelet transform has attracted the attention of many researchers in a lot of research areas such as compression, detection, signal and image de-noising, and pattern recognition. In [1], the author has proposed a powerful tool for de-noising signals based on wavelet shrinkage which operates by thresholding wavelet coefficients. The drawback of this approach is situated in no always exist possibility to separate the components corresponding to the target signal from those of noise using a simple thresholding. Improved algorithms based on shrinkage wavelet threshold have been more recently proposed in [2].

However, they suffer two problems: Inability in maintaining the signal continuity and the signal loss in speech information. On the other hands, ideal binary mask (IBM) method has proved its efficiency by increasing speech intelligibility, as a goal of binary time-frequency (T-F) masking approach [3]. Indeed, both separation and estimation of the target signal have been carried out from the residual called interference noise and not a very good separation has been reached of such voiced and unvoiced components [4-5].

Compared to previous works, this paper develops efficient isolated words recognition system based on HMM classification technique using combined feature extraction algorithms as PLP, RASTA-PLP, GF, MFCC and TEO-PWP.

Indeed, the speech recognition rate is evaluated using even MFCC and MFCC/wave atoms combination features. The MFCC based wave atoms parameterization presents high performance in speech recognition systems even though other techniques. So, the adopted feature method extraction using combined MFCC and wave atoms techniques proves its effectiveness in both clean and noisy environment with less complexity and accurate results.

Manuscript received October 5, 2019
Manuscript revised October 20, 2019

The remaining of this paper is developed as follows: section2 is devoted to the description of speech pre-processing steps. Section 3 gives details about different feature extraction techniques. Section 4 presents the adopted speech recognition system with used techniques. obtained results are analyzed and discussed in section 5. Finally, section 6 gives the conclusion and expected future works.

2. Related Works

The literature is enriched by several scientific works that develop and analyze speech recognition systems. Some works aim to improve these systems in the basis of parameters optimization via hybridization and combination between different extraction techniques. In fact, MFCC, LPC and PLP seem to be the best known parameterization approaches. Among the most recent researches in the field of automatic speech recognition, mention may be made of [6] which presented a deep Belief Networks (DBNs) to extract discriminative information from larger window of frames in speech signal. The objective desired by work was the exploration of the DBNs efficiency in learning features that are more invariant to the deep fluctuation in speech signal. The work in [7] proposes reduced combinational features for emotional speech recognition in the basis of wavelet coefficient, LPCC (linear prediction cepstral coefficient) and MFCC (mel-frequency cepstral coefficient). The feature sets are also combined and tested in the classifier. The proposed feature extraction method is applied to detect and classify five emotions as angry, fear, happy, disgust and neutral. However, authors in [8] aim to enhance a speech recognition system that combines audio and visual speech information in noisy environments. They use Gabor filters to extract feature in the front-end stage of both modules to capture robust spectro-temporal features. The performance obtained from the resulted Gabor Audio Features (GAFs) and Gabor Visual Features (GVFs) is compared to the performance of other conventional features such as MFCC, PLP, RASTA-PLP audio features and DCT2 visual features. The work in [9] investigates a variety of features based on a linear auditory filterbank, the Gammatone filterbank. Envelope features are derived from the envelope of the subband filter outputs. The proposed speech recognition system relies on a standard HMM based recognizer under both clean training and multi-condition training is conducted on a Chinese mandarin digits corpus.

3. Speech Pre-Processing

Making feature domain techniques more consistent is the main goal in diverse environmental conditions. In fact, there are two categories of feature domain methods: the first one is destined to modify the test features and to make them better matching the acoustic conditions for the trained models such as code word dependent cepstral normalization (CDCN) [10], speech enhancement methods [11-12], feature normalization methods, the stereo-based piecewise linear compensation for environment (SPLICE) [13-14], multivariate Gaussian based cepstral normalization (CMN) [15, 16, 17]; the second category is used to make a special robust speech feature representation which can be employed for both training and testing at the aim to make a reduction in the sensitivity in various acoustic conditions.

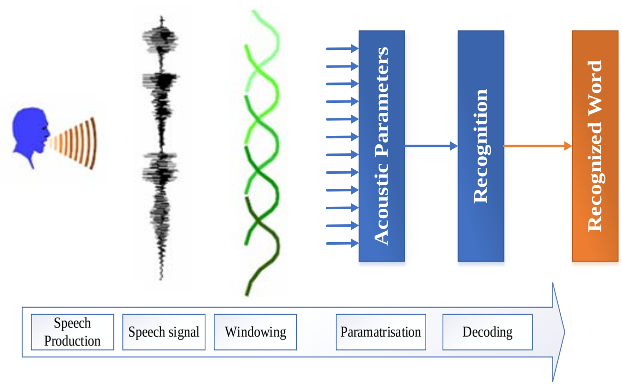

Fig. 1 Automatic Speech Recognition System

The objective in last category is to obtain robust feature extractors by integrating a pre-processing step such as incorporating algorithms for computing MFCC or PLP features, like frequency masking [18], amplitude modulation-based cepstral features [19-20], Power-Normalized Cepstral Coefficients (PNCC), speech enhancement [18, 21, 22, 23], or by adding feature normalization techniques [15-16] as a post-processing step, like cepstral mean normalization (CMN) or by combining both steps [24].

In the front-end part of Automatic Speech Recognition (ASR) systems, the acoustic features are extracted from the input speech signal via feature extraction algorithms, known as the signal modeling. Recognition result generation in back-end part known as statistical modeling is performed by matching obtained features with a reference model using a template or classification techniques [25], such as Vector Quantization (VQ), Hidden Markov Model (HMM), Dynamic Time Warping (DTW), and Artificial Neural Network (ANN).

Feature Extraction techniques

To get the extracted features, new hybrid feature extraction techniques are applied which are combinations of precedent feature extraction methods such as MFCC, PLP, RASTA-PLP, AMS, and GF with TEO-PWP method. The design of each above-mentioned feature has been performed for further generating 13 coefficient parameters.

The condition for training and testing data has a main function to design an acoustic model which therefore has a big impact on the performance of automatic speech recognition system [26].

Indeed, many works have been performed to model the human audio processing but the success percentage was limited for improving the robustness of the speech recognition front-end using PLP [27], dynamic spectral sub-band centroids [28], RASTA [29], or the auditory-based features [30]. Nevertheless, the MFCC features have kept its place as the most widely used features for ASR applications in virtue of its good discrimination capabilities and its low computational complexity

In [31], it has been shown that the Teager energy cepstrum coefficients (TECCs) has been succeeded to outperform MFCCs coefficients under mismatched testing/training conditions for noisy recognition tasks. Indeed, TECCs have been employed with Teager-Kaiser as alternative energy estimation instead of the human hearing-inspired filter banks, i.e., Gamma tone filters, and the square amplitude energy operator [32].

4. The Proposed Method

The speech recognition system under study consists of different important block diagrams. The system follows several steps to finally lead to an acceptable speech recognition rate. Indeed, feature extraction needs to be optimized to achieve a better signal clustering which is performed via the HMM classifier (see Fig.1). So, this work proposes a combination of existing feature extractions techniques in order to achieve a good performance.

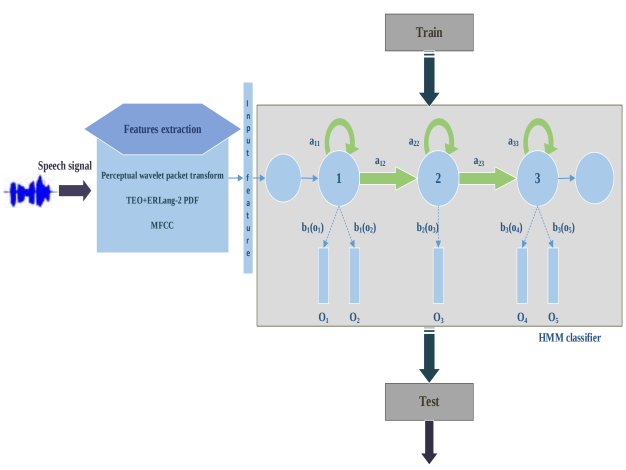

Fig. 2 Proposed speech recognition system

The feature extraction algorithm follows several steps (see Fig.3):

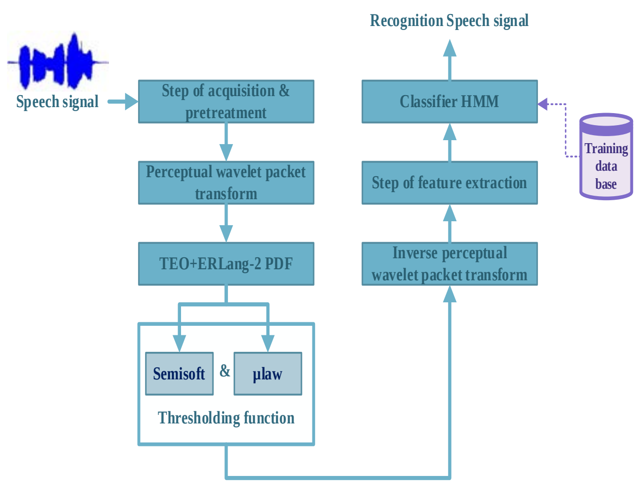

Fig. 3 Bloc diagram of our proposed method

A. PWP transform and TE operation

The TE operator has been applied on PWP coefficients of noisy speech instead of applying it directly on the noisy speech. Also, modeling TE operated PWP coefficients of noisy speech using an Erlang-2 PDF in order to determinate an appropriate sub-band-adaptive threshold, constitute one of main contributions of this paper.

Moreover, the proposed method is significantly fast and suitable for real-time enhancement of noisy speech since the Erlang-2 PDF is a single parameter which relies only on the variance of the modeled random variable and does not need computation of any additional parameter.

The change in the characteristics of the proposed thresholding function in this paper depends on the probabilities of speech presence and absence in a sub-band. Therefore, the proposed custom thresholding function is employed at the aim to obtain an enhanced speech by applying the calculated thresholding on PWP coefficients of noisy speech.

To make difference between speech and noise, PWP transform does not offer enough frequency resolution [33]. Let's consider $W_{k,m}$ the m^{th} PWP coefficient in the k^{th} subband, the expression of TE operated coefficient $t_{k,m}$ corresponding to $W_{k,m}$ can be given as follows:

$$t_{k,m} = T(W_{k,m}) \quad (1)$$

Where the definition of discrete TE operator $T(W_{k,m})$ is given as [34]:

$$T\left(W_{k,m}\right) = W_{k,m}^2 - W_{k,m+1}W_{k,m-1} \quad (2)$$

B. Erlang-2 PDF for modeling of TE operated PWP coefficients

In this paper, a common PDF was used to model the TE operated PWP coefficients of noisy speech and noise. The differentiating value between the speech and noise PWP coefficients in a sub-band is employed to found out the exact threshold on the basis of the theory of entropy between these coefficients.

In fact, it is so difficult to realize the actual PDF of speech PWP coefficients or it's t_k (where $t_k = t_{k,1}, \dots, t_{k,M}$, and M is the total number of PWP coefficients in k^{th} subband) because of the time-varying nature of speech signals. For this reason, the formulation of a PDF for t_k can be replaced by formulating its histogram and by this manner we can obtain a PDF that is closely similar to the histogram. To approximate the histogram, Student t PDF represents a good choice for this task but it has the drawback of time-consuming in determining the proper degree of freedom which makes it not recommendable for real-time speech enhancement. Finding a PDF that relies only on the variance of t_k of the noisy speech or noise, modeling and closely similar to the histogram is also a challenge to be resolved. Erlang-2 PDF is such a method which can be used for this task, as it depends only on the variance of the model it fits.

In [35], it has been shown that Erlang-2 PDF has significantly improved the speed of speech enhancement procedure and it has not degraded its performance in comparison to student PDF. Combining µ-law and semisoft thresholding functions represents the main idea to design a custom thresholding function which has proved its efficiency in speech enhancement in comparison to thresholding function based on combination of modified hard and semisoft function. The change in the characteristics of the proposed thresholding function in this paper depends on the probabilities of speech presence and absence in a sub-band.

Therefore, the proposed custom thresholding function is employed at the aim to obtain an enhanced speech by applying the calculated thresholding on PWP coefficients of noisy speech.

Thresholding in wavelet, wavelet packet, or PWP domain [33-36], it is not reasonable to consider a unique threshold for all sub-bands in [37]. In this paper, a common PDF was used to model the TE operated PWP coefficients of noisy speech and noise. Indeed, the symmetric K-L divergence is expressed as follows:

$$SKL(p,q) = \frac{KL(p,q)+KL(q,p)}{2} \qquad (3)$$

Where p and q represent 2-PDFs

5. Results and Analysis

To test the performance of the proposed feature extraction algorithm, simulation results are meeting. Indeed, signal decomposition at 4 levels using perceptual wavelet packet "PWP" is shown in Fig.4. However, the used Mel filter sub-bands is given by Fig.5.

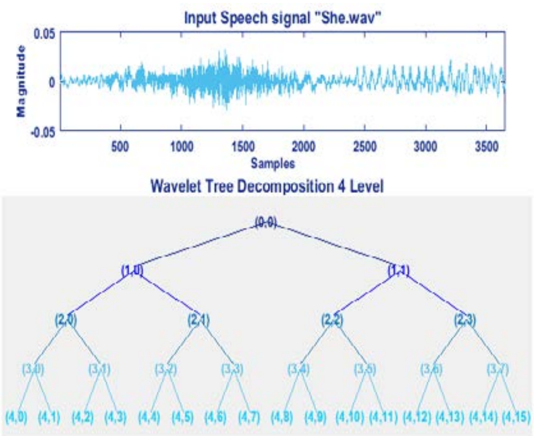

Fig. 4 PWP decomposition of original input signal

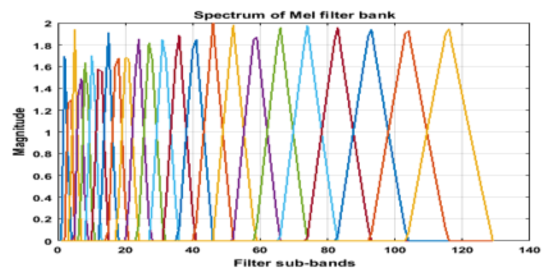

Fig. 5 Mel Filter sub-bands

To prove the interest of using Teager energy operator approach compared to the traditional signal energy detection method, a simulation test was done on a voice sample "She.wav". The detection results compared with the two methods are shown in Fig. 6.

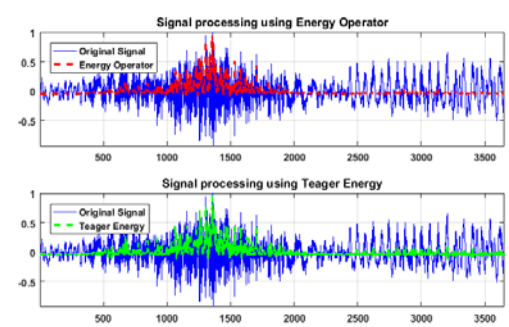

Fig. 6 Signal analysis using Energy Operator and Teager Energy

In each sub-band, the entropy between TE and the PWP coefficients for the noisy speech is not the same as for TE and the PWP coefficients of the noise as the power of the speech and the noise are different (see Fig. 7).

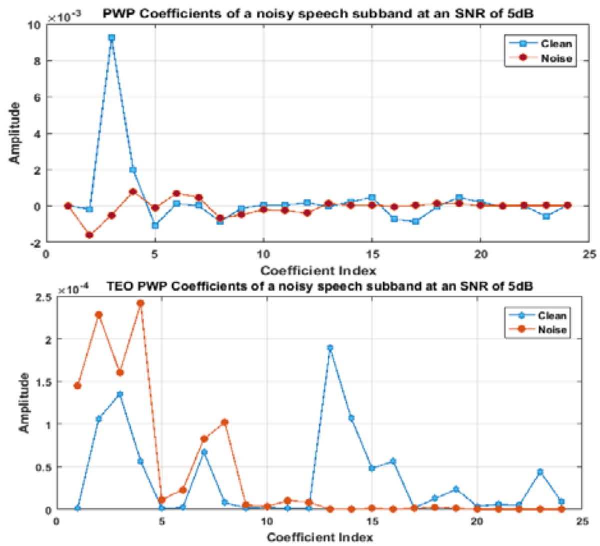

Fig. 7 PWP Coefficients of a noisy speech with/without TEO

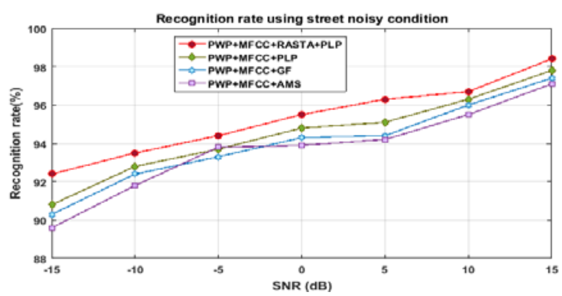

Fig.8. Recognition rate in clean and street noisy speech conditions using different combinations of feature extraction techniques

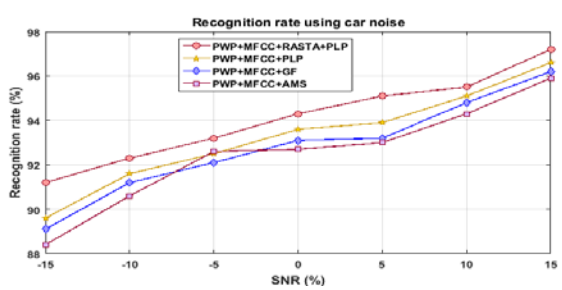

Fig.9. Recognition rate in clean and car noisy speech conditions using different combinations of feature extraction techniques

In Fig.8 and Fig.9, we can see that the proposed method using PWP, MFCC, and RASTA-PLP as feature extraction technique has been succeeded to give the best recognition ratio in both clean speech condition (100%) and noisy speech condition (98.8% at 15dB of SNR in street noise condition) in comparison to other combinations of feature extraction techniques. As it is shown in figure 2, a satisfied recognition rate (92.8%) has been obtained using the combination of PWP, MFCC, and RASTA-PLP at low level of SNR (-15dB) in street noisy condition. Comparable recognition rates have been also obtained in all clean and noisy speech conditions using the combination of PWP, MFCC, and PLP. For the combinations of PWP, MFCC and Gama tone, and PWP, MFCC and AMS, the recognition rates are closely similar at different levels of SNR in clean and noisy speech conditions. Moreover, we can say that the recognition rates for the proposed model using the combination of PWP, MFCC, and RASTA-PLP in street noise were better than those obtained in car and multi-talker noises. Furthermore, the addition of RASTA-PLP feature besides PWP and MFCC features has a significantly improvement in recognition rates in clean and noisy speech conditions in comparison to Gama tone and AMS features.

Table 1: Comparison between feature extraction techniques

Feature Extraction Techniques	**Clean Speech**	**Noisy Speech (15 dB)**
LPCC	89.83 %	88.77 %
PLP	99.95 %	98.59 %
MFCC	83.19 %	82.3 %
RASTA-PLP	92.50 %	91.67 %
MFCC, LPCC and RASTA	99.12 %	98.27 %
PLP, LPCC, and RASTA	98.93 %	97.99 %
MFCC, PLP and LPCC	98.79 %	97.86 %

In general, we can say that the combination of PWP, MFCC, and RASTA-PLP has shown a good capacity in achieve good recognition rates in clean speech condition (100%) and noisy speech condition (98.8 and 92.8 at -15dB and 15 dB of SNR, respectively) in comparison to other works using other combinations of feature extraction techniques with different databases as it is shown in table 1. Also, this combination of feature extraction techniques has helped to outperform and reach comparable results in terms of performances (recognition rate) in both clean and noisy speech conditions in comparison to the state-of-the art systems such as HMM in [38] and [39].

6. Conclusion and Perspectives

This paper has shown interests in HMM-based speech recognition system that combines several promising feature extraction techniques such as, PWP, MFCC and RASTA-PLP; PWP, MFCC, and PLP; PWP, MFCC and GF; and PWP, MFCC and AMS. According to obtained results, the combination between performed PWP, MFCC, and RASTA-PLP recognition features achieves the best ratio (100%) in clean speech. Moreover, the adopted recognition method proves its ability and effectiveness in street noise by reaching a rate value of 98.8% at high level of SNR (15dB) and 92.8% at low level of SNR (-15dB). As noticed, the use of PWP, MFCC, and PLP feature extraction methods can also reach a best performance of 98.2% at15dB SNR level in street noise. However, the recognition approach combining PWP, MFCC and RASTA-PLP, and PWP, MFCC and PLP seems to be the better method in terms of recognition rates in both clean and noisy speech conditions. As future work, we will investigate and develop a real-time speech recognition platform in the basis of the adopted feature extraction algorithm in an embedded system board like raspberry pi 3 or other performing electronic card as stm32, Arduino… The choice of the target can be selected according to test simulations necessities.

References

[1] Donoho DL, "Denoising by soft thresholding", IEEE Trans. on Information Theory 1995; 41(3):613–627; http://dx.doi.org/10.1109/18.382009.

[2] Zhu JF, Huang YD. "Improved threshold function of wavelet domain signal de-noising. In: Proc. ICWAPR," 2013; 14–17.

[3] Li N, Loizou PC, "Factors influencing intelligibility of ideal binary-masked speech: Implications for noise reduction," J Acoust Soc Am 2008; 123(3):1673–1682; PMID:18345855; http://dx.doi.org/10.1121/1.2832617.

[4] Sun J, Tang Y, Jiang A, Xu N, Zhou L, "Speech enhancement via sparse coding with ideal binary mask," in Signal Processing (ICSP 2014) 12th International Conference on IEEE 2014; 537–540; http://dx.doi.org/10.1109/ICOSP.2014.7015062.

[5] Lee G, Na SD, Cho JH, Kim MN, "Voice activity detection algorithm using perceptual wavelet entropy neighbor slope," Biomed Mater Eng 2014; 24(6):3295–3301; PMID:25227039.

[6] Mahboubeh Farahat,"Noise Robust Speech Recognition Using Deep Belief Networks", International Journal of Computational Intelligence and Applications, Vol. 15, No. 1 (2016) 1650005 (17 pages).

[7] Hemanta Kumar Palo, Mihir Narayan Mohanty, "Wavelet based feature combination for recognition of emotions", Ain Shams Engineering Journal 9 (2018) 1799–1806.

[8] Ali S.Saudi, Mahmoud I.Khalil, Hazem M.Abbasb, "Improved features and dynamic stream weight adaption for robust Audio-Visual Speech Recognition framework", Digital Signal Processing, article in press, 2019.

[9] Hui Yin, Volker Hohmann, Climent Nadeu, Acoustic features for speech recognition based on Gammatone filterbank and instantaneous frequency, Speech Communication 53 (2011) 707–715.

[10] A. Acero, "Acoustical and environmental robustness in automatic speech recognition," PhD thesis, ECE, Carnegie Mellon University, 1990.

[11] J. S. Lim and A. V. Oppenheim, "Enhancement and bandwidth compression of noisy

[12] speech," Proc. IEEE, vol. 67, pp. 1586-1604, Dec. 1979.

[13] R. J. Mcaulay and M. L. Malpass, "Speech enhancement using a soft-decision noise

[14] suppression filter," IEEE Trans. Acoustics, Speech and Signal Processing, vol. 28, pp. 137-145, Apr. 1980.

[15] L. Deng, J. Droppo, and A. Acero, "Dynamic compensation of HMM variances using the feature enhancement uncertainty computed from a parametric model of speech distortion," in IEEE Trans. Speech and Audio Processing, vol. 13, pp. 412-421, May 2005.

[16] Li Deng, Alex Acero, L. Jiang, Jasha Droppo, and Xuedong Huang, "High-Performance Robust Speech Recognition Using Stereo Training Data," in Proc. ICASSP, Institute of Electrical and Electronics Engineers, Inc., Salt Lake City, Utah, May 2001.

[17] Alam, J., Ouellet, P., Kenny, P., O'Shaughnessy, D., "Comparative Evaluation of Feature Normalization Techniques for Speaker Verification," in Proc NOLISP, LNAI 7015, pp. 246- 253, Las Palmas, Spain, November 2011.

[18] 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.

[19] O. Viikki and K. Laurila, "Cepstral domain segmental feature vector normalization for noise robust speech recognition," in Speech Communication, vol. 25, pp. 133-47, 1998.

[20] 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.

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

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

[23] ETSI ES 202 050, Speech Processing, Transmission and Quality aspects (STQ); Distribut-ed speech recognition; advanced front-end feature extraction algorithm; Compression Algo-rithms; 2003.

[24] Këpuska, V. and Klein, T. (2009), "A Novel Wake-Up-Word Speech Recognition System, Wake-Up-Word Recognition Task, Technology and Evaluation. Nonlinear Analysis: Theo-ry, Methods & Applications," 71, e2772-e2789. http://dx.doi.org/10.1016/j.na.2009.06.089

[25] Veisi, H. and Sameti, H. (2013), "Speech Enhancement Using Hidden Markov Models in Mel-Frequency Domain," Speech Communication, 55, 205-220. http://dx.doi.org/10.1016/j.specom.2012.08.005.

[26] H. Hermansky, "Perceptual linear predictive (PLP) analysis of speech," J. Acoust. Soc. Amer., vol. 87, pp. 1738–1752, 1990.

[27] J. Chen, Y. Huang, Q. Li, and K. K. Paliwal, "Recognition of noisy speech using dynamic spectral subband centroids," IEEE Signal Process. Lett., vol. 11, no. 2, pp. 258–261, Feb. 2004.

[28] 31. H. Hermansky and N. Morgan, "RASTA processing of speech," IEEE Trans. Speech Audio Process., vol. 2, no. 4, pp. 578–589, Oct. 1994.

[29] B. Mak, Y. Cheung-Tam, and Q. Li, "Discriminative auditory-based features for robust speech recognition," IEEE Trans. Speech Audio Process., vol. 12, no. 1, pp. 27–36, Jan. 2004.

[30] D. Dimitriadis, P. Maragos, and A. Potamianos, "Auditory Teager energy cepstrum coeffi-cients for robust speech recognition," in Proc. Eurospeech' 05, Sep. 2005.

[31] D. Dimitriadis, A. Potamianos, and P. Maragos, “A comparison of the squared energy and Teager–Kaiser operators for short-term energy estimation in additive noise,” IEEE Trans. Signal Process., vol. 57, no. 7, pp. 2569–2581, Jul. 2009.
[32] Islam, M. T., Shahnaz, C., Zhu, W.-P, Ahmad, M. O., 2015,“Speech enhancement based on student modeling of teager energy operated perceptual wavelet packet coefficients and a cus-tom thresholding function”, IEEE ACM Transactions on Audio, Speech, and Lan- guage Processing 23 (11), 1800–1811.
[33] Sanam, T., Shahnaz, C., 2012, “Enhancement of noisy speech based on a customthreshold- ing function with a statistically determined threshold,” International Journal of Speech Tech-nology 15, 463–475.
[34] Sanam, T. F., Shahnaz, C., 2012, “A combination of semisoft and -law thresholding func- tions for enhancing noisy speech in wavelet packet domain,” in: Electrical & Computer En-gineering (ICECE), 2012 7th International Conference on. IEEE, pp. 884–887.
[35] Md.Tauhidul Islam, C.Shahnaz,W-P.Zhu, M.Omair Ahmad, “Modeling of Teager Energy Operated Perceptual Wavelet Packet Coefficientswith an Erlang-2 PDF for Real Time En-hancement of Noisy Speech,” Preprint submitted to Journal of LATEX Templates, Febru- ary 9, 2018.
[36] Veton Z. Këpuska, Hussien A. Elharati“Robust Speech Recognition System Using Conven-tional and Hybrid Features of MFCC, LPCC, PLP, RASTA-PLP and Hidden Markov Model Classifier in Noisy Conditions,” In Journal of Computer and Communications, 2015,3, 1-9.
[37] D.Dimitriadis, S.Member, P.Maragos, “On the Effects of Filterbank Design and Energy
[38] Computation on Robust Speech Recognition,” in IEEE Transactions On Audio, Speech, And Language Processing, Vol. 19, No. 6, August 2011.
[39] Md .Alam, P.Kenny, D.O’Shaughnessy, “Robust Feature Extraction based on an Asymmet-ric Level-Dependent Auditory Filterbank and a Subband Spectrum Enhancement Tech-nique,” INRS-EMT, University of Quebec, Montreal, Quebec, Canada, 2013.