

Robust Features for Speech Recognition using Temporal Filtering Technique in the Presence of Impulsive Noise

Hajer Rahali, Zied Hajaiej and Nouredine Ellouze

National Engineering School of Tunis (ENIT)
Laboratory of Systems and Signal Processing (LSTS)
BP 37, Le Belvédère, 1002 Tunis, Tunisia
E-mail: hajer.rahali@enit.rnu.tn, zied.hajaiej@enit.rnu.tn, N.ellouze@enit.rnu.tn

Abstract—In this paper we introduce a robust feature extractor, dubbed as Modified Function Cepstral Coefficients (MODFCC), based on gammachirp filterbank, Relative Spectral (RASTA) and Autoregressive Moving-Average (ARMA) filter. The goal of this work is to improve the robustness of speech recognition systems in additive noise and real-time reverberant environments. In speech recognition systems Mel-Frequency Cepstral Coefficients (MFCC), RASTA and ARMA Frequency Cepstral Coefficients (RASTA-MFCC and ARMA-MFCC) are the three main techniques used. It will be shown in this paper that it presents some modifications to the original MFCC method. In our work the effectiveness of proposed changes to MFCC were tested and compared against the original RASTA-MFCC and ARMA-MFCC features. The prosodic features such as jitter and shimmer are added to baseline spectral features. The above-mentioned techniques were tested with impulsive signals under various noisy conditions within AURORA databases.

Index Terms—Auditory filter, impulsive noise, MFCC, RASTA filter, ARMA filter, HMM\GMM.

I. INTRODUCTION

Speech parameterization is an important step in modern automatic speech recognition systems (ASR). The speech parameterization block is used to extract from the speech waveform the relevant information for discriminating between different speech sounds. The information is presented as a sequence of parameter vectors. In this paper, two acoustic features are found: MFCC, RASTA-MFCC and ARMA-MFCC. Generally, these methods are based on three similar processing blocks: firstly, basic short-time Fourier analysis which is the same for both methods, secondly, auditory based filterbank, and, thirdly, cepstral coefficients computation. In addition, temporal filtering is also useful to enhance speech features, such as ARMA filtering [15] and RASTA filtering [16]. ARMA is evidently good for noisy speech recognition. This study applies an auto-regression moving average filter in the

cepstral domain. ARMA filter is essentially a low-pass filter, smoothing out any spikes in the time sequences of speech features. In clean speech the spikes might contain important information about speech utterance, while in noisy speech those spikes are most likely caused by noise [15]. The RASTA method belonging to the second category was proposed to extract robust speech features for recognition by processing temporal trajectories of frequency band spectrum using a band-pass filter [16]. The principle of RASTA method comes from the human auditory perception which indicates the relative insensitivity of human hearing to slowly and quickly varying auditory stimuli. Thus, the RASTA band-pass filter is designed with an IIR filter with a sharp spectral zero at the zero frequency in the modulation frequency domain. MFCC are used extensively in ASR. MFCC features are derived from the FFT magnitude spectrum by applying a filterbank which has filters evenly spaced on a warped frequency scale. The logarithm of the energy in each filter is calculated and accumulated before a Discrete Cosine Transform (DCT) is applied to produce the MFCC feature vector. There are many similarities between these methods. The difference however lies in the shape of the filterbank. In this paper we present the proposed modifications of MFCC method, and it will be shown that the performance of RASTA-MFCC and ARMA-MFCC, are also compared to MODFCC which integrate a new model. In the current paper, prosodic information is first added to a spectral system in order to improve their performance. Such prosodic characteristics include parameters related to the fundamental frequency such as the jitter and shimmer. The MODFCC shows consistent and significant performance gains in various noise types and levels. For this we will develop a system for automatic recognition of isolated words with impulsive noise based on HMM\GMM. We propose a study of the performance of parameterization techniques including the prosodic features proposed in the presence of different impulsive noises. Then, a comparison of the performance of different used features was performed in order to show that it is the most robust in noisy environment. The sounds are added to the word with different signal-to-noise SNRs (20 dB, 15 dB, 10 dB, 5 dB, 0 dB and -5 dB).

Note that the robustness is shown in terms of correct recognition rate (CRR) accuracy. The evaluation is done on the AURORA database.

This paper is organized as follow; in the next section we describe the proposed modifications of MFCC. An experimental study performed to compare the performance of the different parameterization methods in various acoustic environments is described in section 3. Finally, the major conclusions are summarized in section 4.

II. NEW PROPOSED TECHNIQUE

A. Gammachirp auditory filter

The gammachirp filter is a good approximation to the frequency selective behavior of the cochlea [4]. It is an auditory filter which introduces an asymmetry and level dependent characteristics of the cochlear filters and it can be considered as a generalization and improvement of the gammatone filter. The gammachirp filter is defined in temporal domain by the real part of the complex function:

$$g_c(t) = at^{n-1} e^{-2\pi Bt} e^{j2\pi f_r t + j c \ln t + j c \varphi}. \quad (1)$$

With

$$B = b \cdot \text{ERB}(f_r) = b \cdot (24.7 + 0.108 (f_r)). \quad (2)$$

With: $t > 0$, N : a whole positive defining the order of the corresponding filter. f_r the modulation frequency of the gamma function. φ the original phase. a an amplitude normalization parameter, “ c ” a parameter for the chirp rate. “ b ” a parameter defining the envelope of the gamma distribution and $\text{ERB}(f_r)$ Equivalent Rectangular Bandwidth. When $c=0$, the chirp term, $c \ln(t)$, vanishes and this equation represents the complex impulse response of the gammatone that has the envelope of a gamma distribution function and its carrier is a sinusoid at frequency f_r . Accordingly, the gammachirp is an extension of the gammatone with a frequency modulation term.

The energy of the impulse response $g_c(t)$ is obtained with the following expression:

$$E_{n,B} = \|g_c\|^2 = \langle g_c, g_c \rangle = a^2 \frac{\sigma(2n-1)}{(4\pi B)^{2n-1}}. \quad (3)$$

With $\sigma(n)$ is the n -th order gamma distribution function. Thus, for energy normalization is obtained with the following expression:

$$A_{E_{n,B}} = \sqrt{\frac{4\pi B^{(2n-1)}}{\sigma(2n-1)}}. \quad (4)$$

The Fourier transform of the gammachirp in “(1)” is derived as follows [5].

$$|G_c(f)| = \frac{a|\sigma(n+jc)|}{\sigma(n)} * \frac{\sigma(n)}{\left[2\pi\sqrt{(b\text{ERB}(f_r))^2 + (f-f_r)^2}\right]^n} e^{c\theta}. \quad (5)$$

$$|G_c(f)| = a_\sigma |G_T| * e^{c\theta(f)}. \quad (6)$$

$$\theta(f) = \arctan\left(\frac{f-f_r}{b\text{ERB}(f_r)}\right). \quad (7)$$

$|G_c(f)|$ is the Fourier magnitude spectrum of the gammatone filter, $e^{c\theta(f)}$ is an asymmetric function since is anti-symmetric function centered at the asymptotic frequency. The spectral properties of the gammachirp will depend on the $e^{c\theta(f)}$ factor; this factor has therefore been called the asymmetry factor. The degree of asymmetry depends on “ c ”. If “ c ” is negative, the transfer function, considered as a low pass filter, where c is positive it behave as a high-pass filter and if “ c ” zero, the transfer function, behave as a gammatone filter. In addition, this parameter is connected to the signal power by the expression [6]:

$$c = 3.38 + 0.107 P_s. \quad (8)$$

Figure 1 shows a block diagram of the gammachirp filterbank. It is a cascade of three filterbanks: a gammatone filterbank, a lowpass-AC filterbank, and a highpass-AC filterbank [7]. The gammachirp filterbank consists of a gammatone filterbank and an asymmetric compensation filterbank controlled by a parameter controller with sound level estimation.

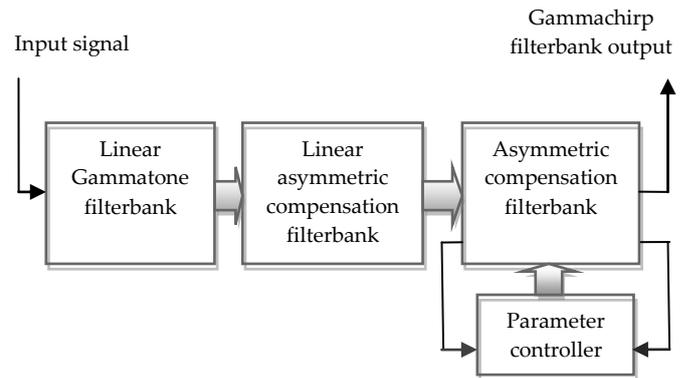


Fig. 1. Structure of the Gammachirp filterbank.

This decomposition, which was shown by Irino in [8], is beneficial because it allows the gammachirp to be expressed as the cascade of a gammatone filter with an asymmetric compensation filter. Figure 2 shows the framework for this cascade approach.

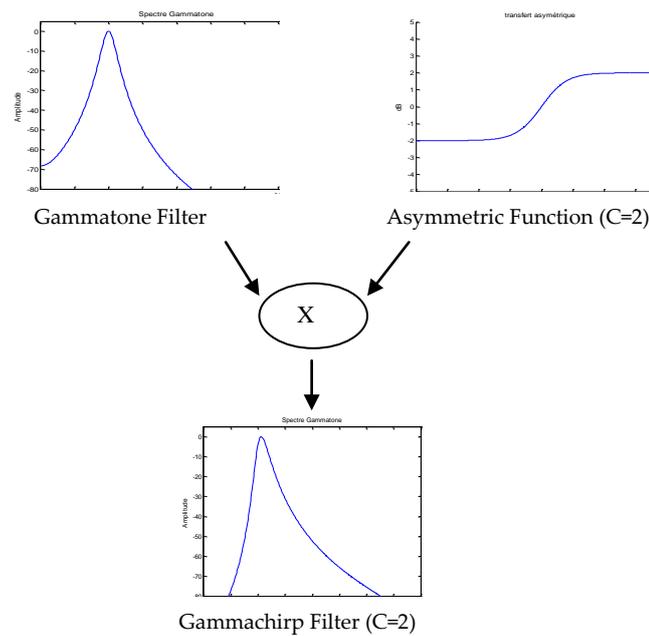


Fig. 2. Decomposition of the gammachirp filter.

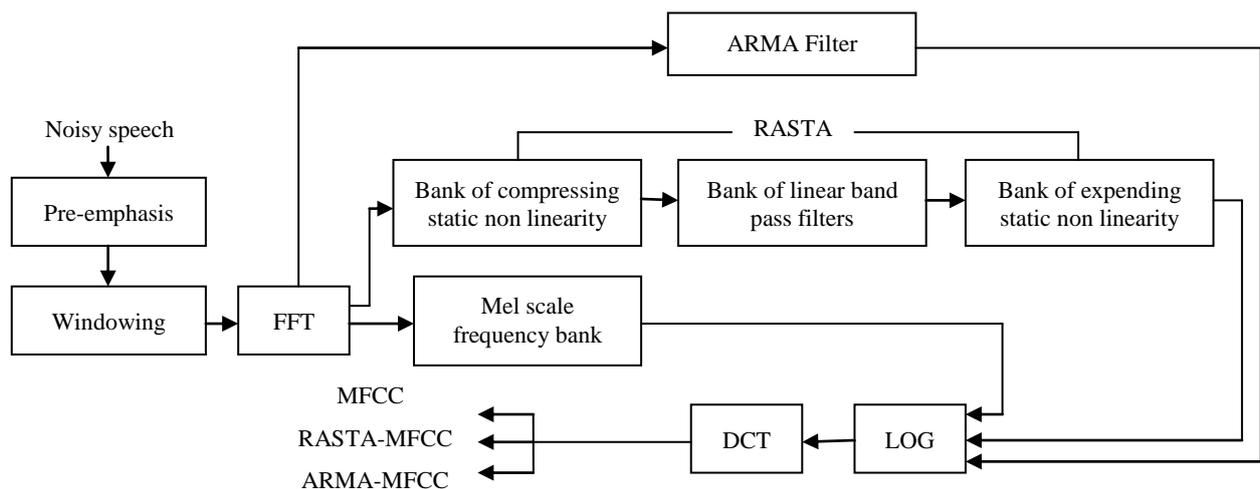


Fig. 3. Block diagram for extracting MFCC, RASTA-MFCC and ARMA-MFCC.

B. Implementation of MFCC

Figure 3 shows a block diagram for extracting MFCC, RASTA-MFCC and ARMA-MFCC. The steps of RASTA-MFCC are as follows. First, the Hamming window is applied to the pre-emphasized signal and then, it is processed by short-time Fourier transform (STFT).

In the next step, we have used logarithmic amplitude transformation as a compressing static non linearity in step one of RASTA. The RASTA-MFCC is derived using a band-pass filter where more slowly and quickly changing parts for each spectral component are suppressed. The expanding static non linearity in step 3 of RASTA was an antilogarithmic transformation. In addition, we could consider that ARMA filter is essentially a low-pass filter, while RASTA filter is a band-pass filter, which its low-pass portion helps in

temporal smoothing of speech features and its high-pass portion helps in alleviating the effect of channel distortion noise. Therefore, in this paper we integrate temporal filter with gammachirp filterbank to improve the recognition performance in noisy conditions.

C. Proposed modifications of MFCC

In this paper, we present some modifications of the standard MFCC feature extraction method. The proposed modifications are presented in the following section. A schematic diagram of the proposed technique is shown in figure 4. In this proposed algorithm MODFCC an application of pre-emphasis is applied to the speech signal before the short term spectral analysis. In the second step, the digitized noisy speech is segmented into overlapping frames, each of length 20 ms with 10 ms overlap, in speech processing a Hamming window are mostly used.

Next, the FFT is taken of these segments. Afterwards, the segmented signal is filtered using the non linear model of the external and middle ear. The next processing step applies a filterbanks. Many different types of filterbanks exist but for MODFCC features the gammachirp filterbank is used. In this study, our objective is to introduce new speech features that are more robust in noisy environments. We propose a robust speech feature which is based on the method with gammachirp filterbank. The output signal of the outer and middle ear model filter is applied to a gammatone filterbank. On each sub-band we calculate the sound pressure level P_s (dB) in order to have the corresponding sub-band chirp term C . Those values of chirp term C corresponding to each sub-bands of the gammatone filterbank lead to the corresponding gammachirp filterbank. The proposed auditory use filter that is smoother and broader than the Mel filterbank. The main differences between the proposed filterbank and the typical one used for MFCC estimation are the type of filters used and their corresponding bandwidth. In this

paper, we experiment with one parameter to create a family of gammachirp filterbanks: the number of filters. The free parameter in gammachirp filterbank as noted above is the number of filters. By increasing the number of filters they become narrow but with a small number of filters the loss of information is introduced. A new filterbank is presented where the width of the filters is fixed to 226 filter and the number of them is equal to the number of spectral coefficients (in our case we used 265 filters). The RASTA and ARMA filter removes variations in the signal that are outside the rate of change of speech by filtering the log-spectrum at each frequency band. Both very slow and very fast changes in sound are ignored by the human ear, so RASTA and ARMA processing attempts to filter these components out. The filter also helps to eliminate noise due to channel variation in the data. That is why we use the temporal filtering technique to process the cepstral coefficients, and then we get the features coefficients which we need.

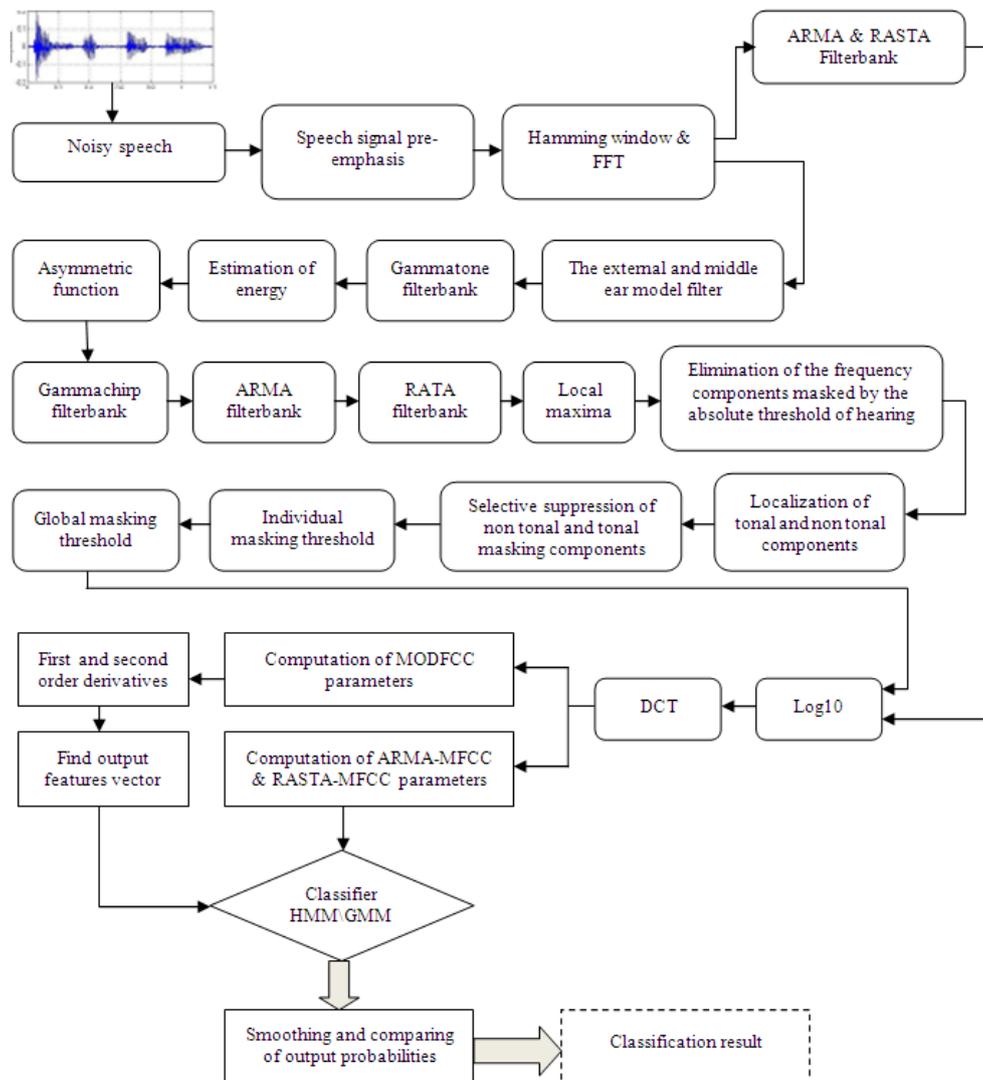


Fig. 4. The structure of MODFCC features extraction

In the next stage, we calculate tonal and non tonal components. This step begins with the determination of the local maxima, followed by the extraction of the tonal components (speech) and non tonal components (impulsive noise), in a bandwidth of a critical band. If frequency exceeds neighboring components within a bark distance by at least 6 dB then it will be treated as “speech” otherwise it will be considered as “noise”. The selective suppression of tonal and non tonal components of masking is a procedure used to reduce the number of maskers taken into account for the calculation of the global masking threshold. The tonal and non tonal components remaining are those which are above the hearing absolute threshold. Individual masking threshold takes into account the masking threshold for each remaining component. Then, the MFCC are applied to these channels to extract features characteristics. MFCC as previously stated has the advantage that they can represent sound signals in an efficient way because of the frequency warping property. In this way, the advantages of this technique are combined in the proposed method. For the final acoustic modeling we extended the modified MFCC-cepstral representation with derived delta and delta-delta features. The following features were extracted: 12 ARMA-MFCC, 12 RASTA-MFCC, and 39 MODFCC. The energy of the frame, first (Δ) and second temporal derivatives ($\Delta\Delta$) extracted from each enumerated parameter. At the end, we have 9 distinct feature vectors that can be categorized into three categories according to its length. Firstly, feature vector has length 13: (12 MODFCC and Energy). Secondly, feature vector has length 26: (12 MODFCC, Energy, and 13 Δ). Thirdly, feature vector has length 39: (12 MODFCC, Energy, 13 Δ , and 13 $\Delta\Delta$). We note that the addition of delta-cepstral features to the static 13 dimensional MODFCC features strongly improves speech recognition accuracy, and a further (smaller) improvement is provided by the addition of double delta-cepstral features. The feature vector constructed on the basis of MODFCC is applied in the statistical classifier. This classifier is based on Gaussian mixture model (GMM) and Hidden Markov Model (HMM). Following the above modifications a new acoustic features (named MODFCC are derived. In the next section, we investigate the robustness and compare the performance of the proposed features to that ARMA-MFCC and RASTA-MFCC with the different prosodic parameters by artificially introducing different levels of impulsive noise to the speech signal and then computing their correct recognition rate.

III. EXPERIMENTAL AND RESULTS

A. AURORA task

AURORA is a noisy speech database, designed to evaluate the performance of speech recognition systems in noisy conditions. The AURORA task has been defined by the European Telecommunications Standards Institute (ETSI) as a cellular industry initiative to standardize a robust feature extraction technique for a distributed speech recognition framework. The initial ETSI task uses the TI-DIGITS database down sampled from the original sampling rate of 20 kHz to 8 kHz and normalized to the same amplitude level [13]. Two different noises (Explosion and door slams) have been artificially added to different portions of the database at signal-to-noise (SNR) ratios ranging from clean, 20 dB to -5 dB in decreasing steps of 5 dB. The training set consists of 8440 different utterances split equally into 20 subsets of 422 utterances each. Each split has one of the three noises added at one of the seven SNRs (Clean, 20 dB, 15 dB, 10 dB, 5 dB, 0 dB and -5 dB). The test set consists of 4000 test files divided into four sets of 1000 files each. Each set is corrupted with one of the three noises resulting in a total of (2 x 1000 x 7) 14,000 test utterances.

B. Experimental setup

To evaluate the suggested techniques, we carried out a comparative study with different baseline parameterization techniques of ARMA-MFCC and RASTA-MFCC implemented in HTK. The features extracted from clean and noisy database have been converted to HTK format using “VoiceBox” toolbox [4] for Matlab. In our experiment, there were 10 HMM models trained using the selected feature MODFCC, ARMA-MFCC and RASTA-MFCC. Each model had 5 by 5 states left to right. The features corresponding to each state occupation in an HMM are modeled by a mixture of 12 Gaussians. In all the experiments, 39 vectors are used as the baseline feature vector. Jitter and shimmer are added to the baseline feature set both individually and in combination.

C. Results and discussion

The performance of the suggested parameterization methods is tested on the AURORA databases using HTK. We use the percentage of word accuracy as a performance evaluation measure for comparing the recognition performances of the feature extractors considered in this paper. Comparison of phoneme recognition rates is shown in the table 1, 2, 3 and 4.

Table 1. Word accuracy (%) using different parameterization techniques

Features SNR(dB)	Explosions							Door slams						
	∞	20	15	10	5	0	-5	∞	20	15	10	5	0	-5
ARMA-MFCC (24 filters)	72.48	62.05	60.20	58.43	46.39	33.60	33.34	74.24	60.52	55.98	44.06	40.19	35.32	32.76
RASTA-MFCC (24 filters)	73.05	63.85	61.89	59.07	47.98	34.25	34.06	75.94	61.16	56.58	45.36	41.72	36.92	33.36
MODFCC (24 filters)	74.45	64.25	62.27	60.44	48.38	35.67	35.34	76.34	62.56	57.98	46.76	42.12	37.32	34.76
MODFCC (265 filters)	75.55	65.35	63.39	61.54	49.48	36.75	36.44	77.44	63.66	58.76	47.86	43.22	38.42	35.86

Table 2. Word accuracy (%) of RASTA-MFCC using the prosodic features

Features SNR(dB)	Explosions							Door slams						
	∞	20	15	10	5	0	-5	∞	20	15	10	5	0	-5
RASTA-MFCC (Baseline)	85.45	82.25	78.29	78.44	66.38	50.65	33.34	84.34	80.56	78.98	77.76	70.12	55.32	42.76
RASTA- MFCC+Jitter	88.05	84.85	80.89	80.04	68.98	53.25	35.94	86.94	83.16	81.58	80.36	72.72	57.92	45.36
RASTA- MFCC+Shimmer	88.45	85.25	81.29	81.44	69.38	53.65	36.34	87.34	83.56	81.98	80.76	73.12	58.32	45.76
RASTA- MFCC+Jitter+ Shimmer	89.55	86.35	82.39	82.54	70.48	54.75	37.44	88.44	84.66	82.76	81.86	74.22	59.42	46.86

Table 3. Word accuracy (%) of MODFCC using the prosodic features

Features SNR(dB)	Explosions							Door slams						
	∞	20	15	10	5	0	-5	∞	20	15	10	5	0	-5
MODFCC (Baseline)	92.43	90.17	88.20	85.74	77.21	71.54	50.07	90.35	89.96	88.98	87.70	78.99	70.22	68.43
MODFCC+Jitter	95.05	92.77	90.80	88.34	79.81	74.74	52.67	92.95	92.56	91.58	90.30	81.59	72.82	71.03
MODFCC+Shimmer	95.43	93.17	91.20	88.74	80.21	74.54	53.07	93.35	92.96	91.98	91.70	84.59	75.82	74.03
MODFCC+Jitter+ Shimmer	96.53	94.27	92.30	89.84	81.31	75.64	54.17	94.45	94.06	93.08	92.80	85.69	76.92	75.13

Table 4. Recognition rate (%) of MODFCC using dynamic properties

Features SNR(dB)	Explosions							Door slams						
	∞	20	15	10	5	0	-5	∞	20	15	10	5	0	-5
MODFCC (13)	88	86.94	80.98	75.74	71.65	65.34	43.87	82.54	79.67	77.54	75.70	65.87	54.02	33.74
MODFCC+ Δ (26)	89.23	86.22	82.57	80.76	79.65	70.45	59.54	90.95	89.56	88.98	85.30	80.45	72.45	66.87
MODFCC+ Δ + $\Delta\Delta$ (39)	94.56	92.17	91.20	90.09	89.56	85.67	77.87	94.21	92.87	91.98	90.10	81.90	78.10	72.76

The new filterbank with fixed filter width and a large number of filters has also been applied and tested with the ARMA-MFCC and RASTA-MFCC methods. It can be seen (table 1) that the recognition accuracy improves slightly for a relative value of 1 %. From this table, it can be observed that improvement (3 % relative increase of recognition accuracy) is achieved with the new MODFCC method of parameterization over the RASTA and ARMA methods. Tables 2 and 3 present the performance of two voice features in presence of various levels of additive noise. We note that the MODFCC features that are extracted using the gammachirp containing frequency-domain noise and speech detection exhibit the best CRR. Also, it is observable that the performance of the two features decreases when the SNR decreases too, that is,

when the speech signal becoming more noisy. Jitter and shimmer are added to the baseline feature set both individually and in combination. The absolute accuracy increase is 2.6 % and 3.0 % after appending jitter and shimmer individually, while there is 4.1 % increase when used together. As we can see in these tables, the identification rate increases with speech quality, for higher SNR we have higher identification rate, the MODFCC based parameters are slightly more efficiencies than standard RASTA-MFCC for noisy speech (94.27 % vs 86.35 % for 20 dB of SNR with jitter and shimmer) but the results change the noise of another. From the above table 4, it can be seen that the recognition rates are above 90 %; this is recognition rates are due to the consideration of using 39 MODFCC features.

IV. CONCLUSION

The proposed features called MODFCC have been shown to be more robust than ARMA-MFCC and RASTA-MFCC in noise environments for different SNRs values. Adding jitter and shimmer to baseline spectral and energy features in an HMM/GMM based classification model resulted in increased word accuracy across all experimental conditions. The results gotten after application of this features show that this method give acceptable and better results by comparison at those gotten by other methods of parameterization.

REFERENCES

- [1] J. O. Smith III, and J.S. Abel, "Bark and ERB Bilinear Transforms," *IEEE Tran. On speech and Audio Processing*, Vol. 7, No. 6, November 1999.
- [2] H.G. Musmann, "Genesis of the MP3 audio coding standard," *IEEE Trans. on Consumer Electronics*, Vol. 52, pp. 1043 – 1049, Aug. 2006.
- [3] H. G. Hirsch, and D. Pearce, "The AURORA Experiment Framework for the Performance Evaluations of Speech Recognition Systems under Noisy Condition," *ISCA ITRW ASR2000 Automatic Speech Recognition: Challenges for the Next Millennium*, France, 2000.
- [4] M. Brookes, "VOICEBOX: Speech Processing Toolbox for MATLAB," Software, available [Mar, 2011].
- [5] E. Ambikairajah, J. Epps, and L. Lin., "Wideband speech and audio coding using gammatone filter banks," *Proc. ICASSP'01*, Salt Lake City, USA, May2001, vol.2, pp.773-776.
- [6] M.N. Viera, F.R. McInnes, and M.A. Jack, "Robust F0 and Jitter estimation in the Pathological voices," *Proceedings of ICSLP96*, Philadelphia, pp.745–748, 1996.
- [7] Salhi. L., "Design and implementation of the cochlear filter model based on a wavelet transform as part of speech signals analysis," *Research Journal of Applied Sciences* 2 (4): 512-521, 2007 Medwell-Journal 2007.
- [8] WEBER F, MANGANARO L, PESKIN B, and SHRIBERG E., "Using prosodic and lexical information for speaker identification," *Proc. ICASSP*, Orlando, FL, May 2002.
- [9] J. W. Pitton, K. Wang, and B. H. Juang, "Time-frequency analysis and auditory modeling for automatic recognition of speech," *Proc. IEEE*, vol. 84, pp. 1199–1214, Sept. 1996.
- [10] E. Loweimi and S. M. Ahadi, "A new group delay-based feature for robust speech recognition," in *Proc. IEEE Int. Conf. on Multimedia & Expo, Barcelona*, pp. 1-5, July 2011.
- [11] Irino. T, E. Okamoto, R. Nisimura, Hideki Kawahara and Roy D. Patterson, "A Gammachirp Auditory Filterbank for Reliable Estimation of Vocal Tract Length from both Voiced and Whispered Speech," *The 4th Annual Conference of the British Society of Audiology*, Keele, UK, 4-6, Sept, 2013.
- [12] 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.
- [13] Daniel PW Ellis and Byunk Suk Lee, "Noise robust pitch tracking by subband autocorrelation classification," in *13th*

Annual Conference of the International Speech Communication Association, 2012.

- [14] D. Povey, and L. Burget, "The Subspace Gaussian Mixture Model—A Structured Model for Speech Recognition," *Computer Speech & Language*, vol. 25, no. 2, pp. 404–439, April 2011.
- [15] C.-P. Chen, J. Bilmes and K. Kirchhoff, "Low-Resource Noise-Robust Feature Post-Processing on Aurora 2.0," pp. 2445-2448, *Proceedings of ICSLP 2002*.
- [16] H. Hermansky and N. Morgan, "RASTA Processing of Speech," *IEEE. Trans. on Speech and Audio Processing*, Vol.2, No.4, Oct. 1994.

Authors' Profiles



Hajer Rahali received the Electronic Engineer Diploma in 2011 at the National School of Engineer of Tunis (ENIT-Tunisia). She received the diploma of Master degree in automatic and signal processing (ATS) from the National School of Engineers of Tunis (ENIT). Currently, she works a teacher of physics sciences and she prepares his

doctorate thesis focused on the Analysis, Identification, and Classification of speech using auditory model.



Zied Hajaiej received the MS degree in electrical engineering (signal processing) in 2004, from Ecole Nationale d'Ingénieurs de Tunis (ENIT), Tunisia. He is currently working towards the Ph.D. degree in electrical engineering (signal processing) in ENIT. His research involved speech recognition. Since September 2006 he has been an Assistant in the Physics Department at Faculté des Sciences de Bizerte, Tunisia, where he teaches electronics, VHDL.



Nouredine Ellouze received the Ph.D. degree in 1977 from l'Institut National Polytechnique at Paul Sabatier University (Toulouse-France), and Electronic Engineer Diploma from ENSEEIHT in 1968 at the same University. In 1978, Dr. Ellouze joined the Department of Electrical Engineering at the National School of Engineer of Tunis (ENIT—Tunisia), as Assistant Professor in statistic, electronic, signal processing and computer architecture. In 1990, he became Professor in signal processing, digital signal processing and stochastic process. He has also served as Director of Electrical Department at ENIT from 1978 to 1983, General Manager and President of the Research Institut on Informatics and Telecommunication IRSIT from 1987 to 1990, and President of the Institut during 1990–1994. He is now Director of Signal Processing Research Laboratory LSTS at

ENIT, and is in charge of Control and Signal Processing Master degree at ENIT. Dr. Ellouze is IEEE fellow since 1987; he directed multiple Masters and Thesis and published over 200 scientific papers both in journals and proceedings. He is Chief Editor of the scientific journal *Annales Maghrebines de l'Ingénieur*. His research interests include neural networks and

fuzzy classification, pattern recognition, signal processing and image processing applied in biomedical, multimedia, and man machine communication.

How to cite this paper: Hajer Rahali, Zied Hajaiej, Noureddine Ellouze, "Robust Features for Speech Recognition using Temporal Filtering Technique in the Presence of Impulsive Noise", *IJIGSP*, vol.6, no.11, pp.17-24, 2014. DOI: 10.5815/ijigsp.2014.11.03