Novel Detection Algorithm of Speech Activity and the impact of Speech Codecs on Remote Speaker Recognition System

RIADH AJGOU[1], SALIM SBAA[2], SAID GHENDIR[1], ALI CHAMSA [1] and A. TALEB-AHMED[3]
[1] Department of sciences and technology
El-oued University
PO Box 789 39000 El-oued
ALGERIA
[2] Electric engineering department, LESIA Laboratory,
Med Khider University
B.P 145 R.P , 07000 Biskra
ALGERIA
[3] LAMIH Laboratory
A. TALEB-AHMED University
UVHC - Mont Houy - 59313 Valenciennes Cedex 9
FRANCE
riadh-ajgou@univ-eloued.dz, s.sbaa@univ-biskra.dz, said-ghendir@univ-eloued.dz,
chemsadoct@yahoo.fr. abdelmalik.taleb-ahmed@univ-valenciennes.fr

Abstract: - In this paper, we studied the effects of voice codecs on remote speaker recognition system, considering three types of speech codec: PCM, DPCM and ADPCM conforming to International Telecommunications Union - Telecoms (ITU-T) recommendation used in telephony and VoIP (Voice over Internet Protocol). To improve the performance of speaker recognition in a noisy environment, we propose a new speech activity detection algorithm (SAD) using "Adaptive Threshold", which can be simulated with speech wave files of TIMIT (Texas Instruments Massachusetts Institute of Technology) database that allows recognition system to be done under almost ideal conditions. Moreover, the speaker recognition system is based on Vector Quantization as speaker modeling technique and Mel Frequency Cepstral Coefficient (MFCC) as feature extraction technique. Where, the feature extraction proceed after (for testing phase) and before (for training phase) the speech is sending over communication channel. Therefore, the digital channels can introduce several types of degradation. To overcome the channel degradation, a convolutional code is used as error-control coding with AWGN channel. Finely, In our simulation with Matlab we have used 30 speakers of different regions (10 male and 20female), the best overall performance of speech codecs was observed for the PCM code in terms of recognition rate accuracy and runtime.

Key-Words: - PCM, DPCM, ADPCM, speaker recognition, SAD.

1 Introduction

Speaker recognition or voice classification is the task of recognizing people from their voices. Such systems extract features from speech signal, process them and use them to recognize the person from the voice [1]. There are various techniques to resolve the automatic speaker recognition problem [1, 2, 3, 4, 5]. Where most published works where the speech does not under phone network and digital channels in noisy environment. Few published works on phone network and digital channels and degradation [6, 7, 8, 9, 10]. Our aim is to provide a comprehensive assessment of speech codecs, considering codecs conforming to ITU-T (International Telecommunications Union - Telecoms) and that are used in Internet telephony and internal IP network in a remote speaker recognition system. Figure 1 shows a general diagram of a remote information system using a remote speaker (or speech) recognition approach, where after recognition, the server provides and transmits the required information to the client [7]. Moreover, the speaker recognition system is based on Vector Quantization (VQ) [11, 12] and Mel - Frequency Cepstral Coefficients (MFCC) algorithm [12, 6]. There are various kinds of

speech codecs available. The speech codecs generally complies industry standards like ITU-T. The main software of ITU standard are : G.711, G.722, G.723, G.726, G.728, G727, and G.729. We consider in this paper PCM, DPCM and ADPCM codecs and their effects on speaker recognition accuracy.

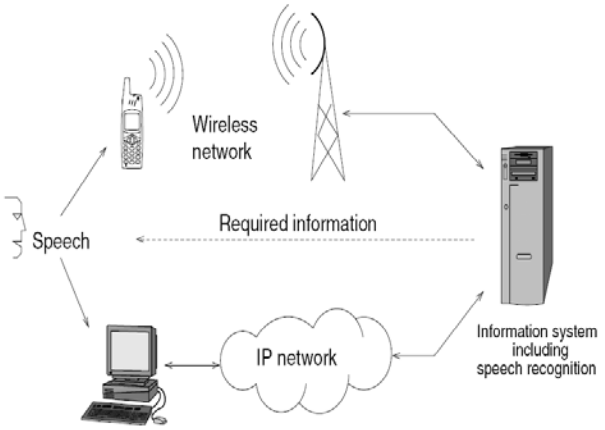

Fig. 1 General scheme of a speech-driven remote information system [7]

There are several possibilities for the implementation of a remote *speaker or speech recognition system over a digital channel*. In the first approach, usually known as *network speech recognition* (NSR), the recognition system resides in the network from the client's point of view [7]. In this case, the speech is compressed by a speech codec in order to allow a low bit rate transmission and/or to use an existing speech traffic channel (as in the case of mobile telephony). The recognition is usually performed over the features extracted from the decoded signal, although it is also possible to extract the recognition features directly from the codec parameters. Figure 2 shows a scheme of this system architecture. In the case where implementation is over an IP network, a VoIP codec can be employed [7]. The second approach known as *distributed speech (or speaker) recognition* (DSR) [7]. In this case, the client includes a local front end that processes the speech signal in order to directly obtain the specific features used by the remote server (back end) to perform recognition, thus avoiding the coding/ decoding process required by NSR [7]. The conceptual scheme of DSR is shown in Figure 3.

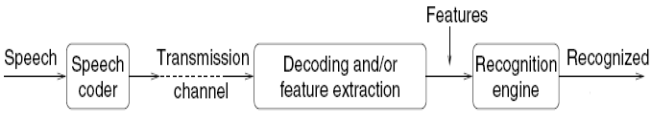

Fig 2. Scheme of a network speaker (or speech) recognition system [7].

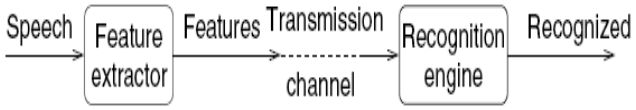

Fig 3. Scheme of a distributed speech recognition (DSR) system [7].

In our work we adopted network *speaker (speech) recognition* conforming to figure 4.

Our work is divided in five steps: in the first we illustrate speech codecs standard. In the second a system configuration is set including speaker recognition system over digital channel (AWGN). In the third we illustrate feature extraction used in recognition system (MFCC). In the forth, in order to improve the memory capacity and speaker recognition performance in noisy environment we propose a new robust speech activity detection algorithm (SAD) using "Adaptive Threshold". The key advantages of this algorithm is its simple implementation and its low computational complexity. This algorithm is based on energy and zero crossing rate. In the fifth, we introduce error correcting code methods that are necessary to improve immunity to noise of communication channels, in our work we considered convolutional code. In the sixth, a simulation results and discussion is done, where we started by the evaluation of speech activity detection algorithm (SAD), then a work is done about remote speaker identification accuracy using PCM, DPCM, and ADPCM with AWGN channel versus SNR. To illustrate the advantage of using channel code, we have use remote speaker identification system with and without convolutional code. Moreover, , we have done a comparative study of PCM, DPCM, and ADPCM codecs in term of runtime.

2 Speech codecs standard

Access to a voice server is not only made through the conventional telephone network, but voice can also be transmitted through wireless networks or *IP* networks. The main factors that determine voice quality are choice of codec, packet loss, latency and jitter. The number of standard and proprietary codecs developed to compress speech data has been

quickly increased. We present, only the codecs conforming to ITU-T.

2.1 PCM

G.711 [13,14] known as PCM codec used in VoIP and fixed telephony. VoIP standard describes two algorithms μ-law and A-law. The μ-law version is used primarily in North America. A-law version used in most other countries outside North America. Both algorithms code speech using 8 bits per sample which provides 50 % reduction in bandwidth use for original signal sampled with 16 bits at 8 kHz sample rate, i.e. reduction from 128 kb/s to 64 kb/s.

2.2 DPCM (G727)

Differential coding is a signal encoder that uses the baseline of PCM witch naturally appropriate in speech quantization. One of the first scalable speech codecs was embedded DPCM [15] Later in 1990, an embedded DPCM system was standardized by the ITU-T as G.727 [16]. The typical operating rate of systems using this technique is higher than 2 bits per sample resulting in rates of 16 Kbits/sec or higher [16].

2.3 ADPCM

G.726 [17, 14] algorithm provides conversion of 64 kb/s A-/μ-law encoded signal to and from a 40, 32, 24 or 16 kb/s signal using Adaptive Differential Pulse Code Modulation (ADPCM). The principle application of 40 kb/s is to carry data modem signals not speech. The most commonly used bitrates for speech compression is 32 kb/s which doubles the capacity compared to the G.711

3 Configuration of the proposed system (remote identificaton)

Speaker recognition can be classified into identification and verification. Speaker identification is the process of determining which registered speaker provides a given utterance. Speaker verification is the process of accepting or rejecting the identity claim of a speaker. The system that we will describe is classified as text independent speaker identification system.

The system we used for experiments included a remote text independent speaker recognition system which was set up according to the following block diagram in figure 4.

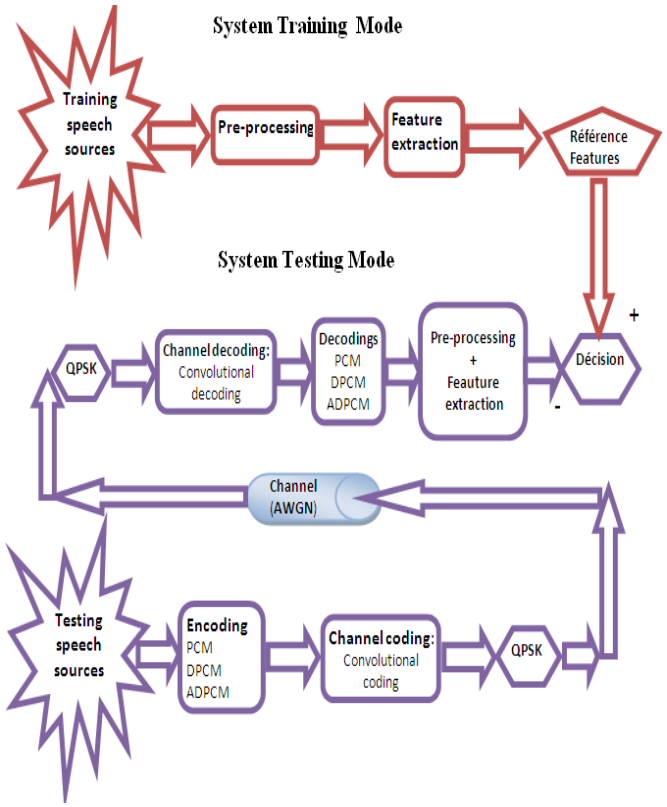

Fig. 4: Remote speaker recognition system.

3.1 Training Phase.

In the training stage, pattern generation is the process of generating speaker specific models with collected data. The generative model used in speaker recognition is Vector Quantization [11]. One of the most successful text-independent recognition methods is based on Vector Quantization (VQ). In this method, VQ codebook s consisting of a small number of representative feature vectors are used as an efficient means of characterizing speaker specific features. A speaker-specific codebook is generated by clustering the training feature vectors of each speaker. In the recognition stage, an input utterance is vector-quantized using the codebook of each reference speaker and the VQ distortion accumulated over the entire input utterance is used to make the recognition decision [11].

The system was trained using speakers from the TIMIT database where we have used 30 speakers from different regions (10 male and

20female). Speech signal passed through pre-processing phase (emphases + speech activity detection). We emphasize the speech using a high pass filter. We usually use a digital filter with 6dB/ Octave. In formula (1), μ is a constant which is taken 0.97 usually [12] [6].

$$y(z) = 1 - \mu z^{-1} \quad (1)$$

After emphasising phase, silence segments are removed by the speech activity detection algorithm so that *twenty four* mel-frequency cepstral coefficients are extracted and form the characterization of the models using Vector Quantization.

3.2 Testing phase

In this stage we have used speech codecs: PCM, DPCM and ADPCM therefore their coefficients are converted into a binary sequence. Before using QPSK modulation we introduce Convolutional code [7, 18] with a rate of ½. The coded signal is transmitted over AWGN channel. After demodulation (QPSK), convolutional decoding, and PCM, DPCM, or ADPCM decoding, the binary data is converted back to a synthesized speech file. Finally, cepstral coefficients are extracted (from the synthesized speech file) .

3.3 Decision Phase

Pattern matching is the task of calculating the matching scores between the input feature vectors (arrived from testing phase) and the given models. In our work we have used the Euclidean Distance (ED) classifier method. The Euclidean distance (ED) classifier has the advantage of simplicity and fast computational speed. The classification is done by calculation the minimum distance to decide which speaker out of all the training set and the most likely to be the test speaker.

Consider a class '***i***' with an m-component mean feature vector X, and a sample vector Y, given respectively by [19]:

$$\overline{X}_i = \left[\overline{X_{i1}}, \overline{X_{i2}}.....\overline{X_{im}}\right]^t \quad (2)$$

and [19]:

$$Y = \left[y_1, y_2....y_m\right]^t \quad (3)$$

The ED between class i and vector Y is given by [19]:

$$d(i,Y) = \left\|\overline{X_{i1}} - Y\right\| = \sum_{k=1}^{m} (\overline{x_{ik}} - yk)^2 \quad (4)$$

For a number of classes C, the decision rule for the ED classifier is that Y be assigned to class ***j*** if [19]:

$$d(i,Y) = \min\{d(i,Y)\}, \forall i \in C \quad (5)$$

4 Features extraction

The speaker-dependent features of human speech can be extracted and represented using the mel-frequency cepstral coefficients MFCC [6]. These coefficients are calculated by taking the cosine transform of the real logarithm of the short-term energy spectrum expressed on a mel-frequency scale [10]. After pre-emphasis and speech/silence detection the speech segments are windowed using a Hamming window. MFCC calculation is shown in figure 5.

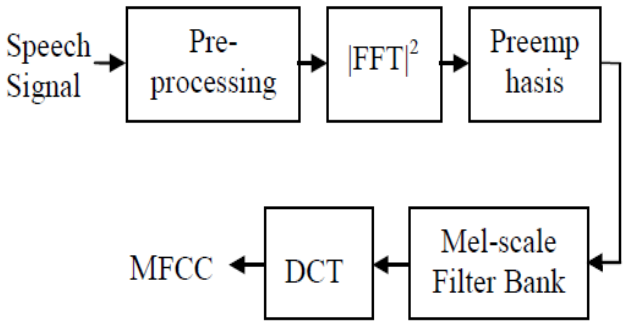

Fig 5. The process of calculating MFCC [20].

The TIMIT database files are sampled with a rate of 16000 samples/s, these files were downsampled to a rate of 8000 samples/s. In our speaker recognition experiments, the speech signal is segmented into frames of each 16 ms, generally it takes 128 points as a frame and the overlapped data is 64 points (8 ms). The Discrete Fourier Transform is taken of these

windowed segments. The magnitude of the Fourier Transform is then passed into a filter-bank comprising of twenty four triangular filters (corresponds to MFCC coefficients).

5 New speech activity detection algorithm (SAD) with adaptive threshold

Speech presence detection in a background of noise is the important pre-processing step in Automatic Speaker Recognition (ASR) systems[21]. Removing nonspeech frames from the speaker recognition system input stream effectively reduces the insertion error rate of the system. SAD perform the speech/nonspeech classification and background noise reduction process on the basis of speech features extracted from the frame under consideration. There are many SAD algorithms available [21, 22, 23, 24, 25].

The new SAD is based on two original works [25, 26]. In [25] the author had used the LPC residual energy and zero crossing rate to detect speech activity using adaptive threshold where this threshold is calculated for every frame introduced in comparison with previous calculated features of frames which means a probable mistakes for firs frames (the algorithm is initiated and spans up to a few frames 0-15 frames, which is considered as non-speech). The second author [26] used energy and zero crossing rate ratios to voiced/non voiced classification of speech using a fixed threshold.

Our new SAD is based on Energy and Zero crossing Rate (EZR) ratios using an adaptive threshold to detect speech activity and remove the silent and noise intervals. SAD operating with a rectangular window of 8ms. The procedure of calculating threshold is as follows:

1- Segmenting the whole speech signal (speaker's signal) in frames of 8ms with rectangular window and without overlapping.
2- Calculating the energy (E[m]) and zero crossing rate (ZCR[m]) for each frame and calculating E[m]/ZCR[m].
3- Calculating the maximum and minimum of EZR.
4- Calculate Threshold (formula 10).

The principle of EZR application explains itself by the fact that the energy of the speech activity is important while the rate of zero crossing rates is weak; therefore the value of EZR is important.

If a frame have an EZR superior to a threshold, this frame classified as speech, if the opposite the frame considered as nonspeech (the recognition system does not extract features from this frame). The threshold determination is estimated by the SAD algorithm in automatic and adaptive way [26]:

$$EZR[m] = \frac{\overline{E}[m]}{ZCR[m]} \tag{6}$$

Where ZCR [m] and $\overline{E}[m]$ present respectively the zero crossing rate and the average energy of a frame [27]:

$$\overline{E}(m) = \sum_{n=0}^{N-1} x^2(n).w(m-n) \tag{7}$$

Where: w is a rectangular window of length N (length of a frame) and x(n) is the frame signal with N samples. ZCR is defined as [28]:

$$ZCR(m) = \sum_{n=0}^{N-1} \left|\text{sgn}[x(n)] - \text{sgn}[x(n-1)\right| w(m-n) \tag{8}$$

Where sgn(.) is the signum function which is defined as [28]:

$$\text{sgn}[x(n)] = \begin{cases} +1, x(n) \geq 0 \\ -1, x(n) < 0 \end{cases} \tag{9}$$

SAD algorithm calculates EZRs of all frames (for a speaker's signal) and estimate threshold:

$$Threshold = \min(EZR) + \alpha * [DELTA] \tag{10}$$

$$DELTA = \max(EZR) - \min(EZR) \tag{11}$$

α: is a real number in the interval of]0,1[. In our simulation we fixed: α=0,35. We can resume our algorithm of speech activity detection in figure 6.

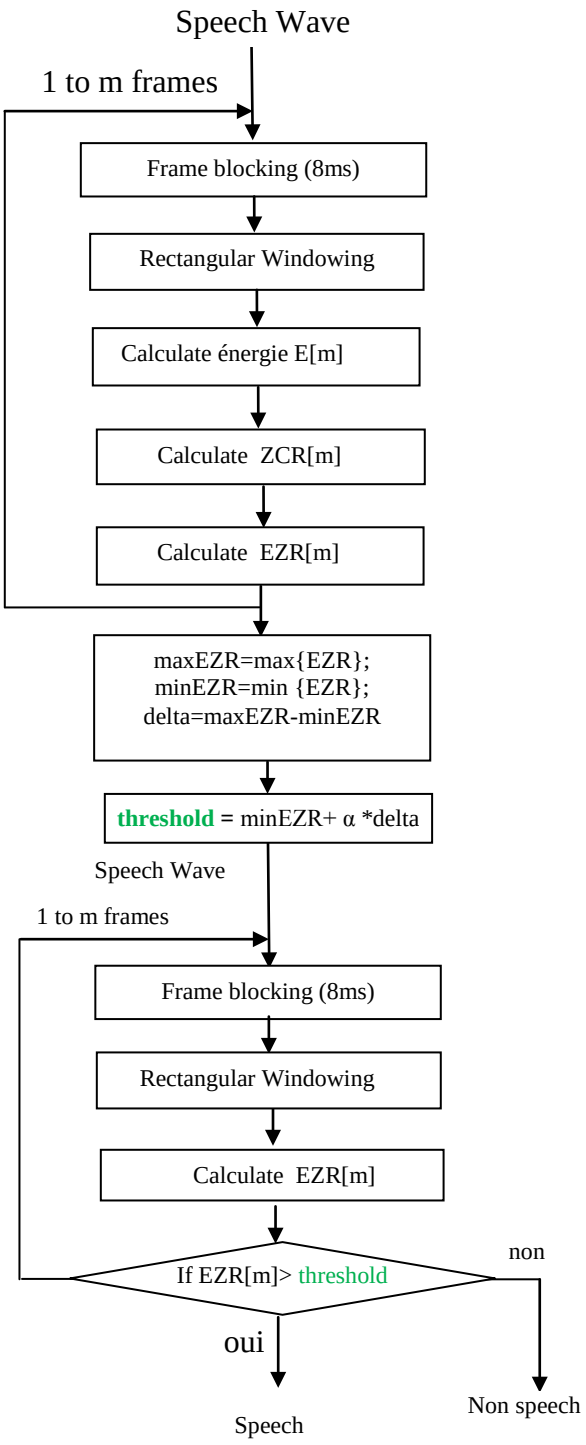

Fig. 6. Block diagram of the proposed speech/nonspeech detection algorithm.

6 Channel coding techniques

Channel coding is developed to maximize the recognition performance, There are different types of FEC (Forward Error Correction) techniques, namely Reed-Solomon and Convolutional codes [29]. The Viterbi algorithm is a method for decoding convolutional codes. A convolutional code with a code rate k/n also generates n output bits from every k input bits, as in the case of block codes. The difference is because the encoding of these k bits is not independent from the bits previously received but it has "memory." A general diagram of a convolutional encoder is shown in Figure 7. At each time unit, the encoder takes a k-bit input sequence, shifts it through a set of m registers, and generates an n-bit output by performing a linear combination (or convolution), in modulo-2 arithmetic, of the data stored in the registers. The integer m is called *constraint length*. When k = 1, we have a special case for which the input bitstream is continuously processed [7].

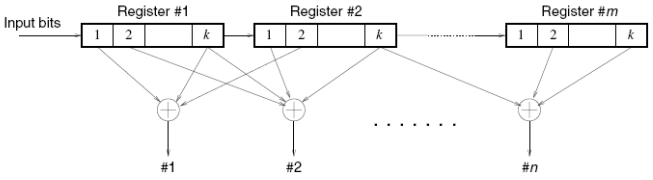

Fig. 7. Structure of a convolutional encoder[7].

In our work we use a convolutional code with ½ rate.

7 Simulation results and discussion

Experiments of speaker identification are led on TIMIT database. The *TIMIT* corpus of read speech has been designed to provide speech data for the acquisition of acoustic-phonetic knowledge and for the development and evaluation of automatic speech recognition systems [30]. Although it was primarily designed for speech recognition, it is also widely used in speaker recognition studies, since it is one of the few databases with a relatively large number of speakers. It contains 630 speakers' voice messages (438 Male and 192 Female), and each speaker reads 10 different sentences.

In our experiments we have chosen 30 speakers (10 male and 20 female). Moreover, in the training stage, we have used three utterances for each speaker.

The generative model used in speaker recognition system is Vector Quantization. We calculate the vector feature extraction from 24 coefficients MFCC.

The first test we evaluate our SAD algorithm, where the speech signal which is " *She had your dark suit in greasy wash water all year*" passed through the algorithm (α = 0.35). The figure 8 represents the original speech signal. The figure 9 represents the speech signal after it has been passed through SAD algorithm. The figure 10 illustrates a speech signal (clean speech) and its SAD counter. From figures 9 and 10, we observe the efficiency of our algorithm where silent segments are eliminated, therefore the memory capacity and recognition accuracy improved. Figures 11, 12 and 13, represent the SAD counter as function of SNR for 10dB, 5dB, 0dB respectively where it is clear that the SAD counter is effective, where Silent and background noise segments are eliminated. Further, Speech activity detection is robust down to SNR=5 dB. To observe the effect of the SAD algorithm on speaker identification rate, we have used our system of speaker recognition with and without SAD (not over digital channel). Figure 14 shows an identification rate with and without speech activity detection algorithm as function of SNR. It is clearly shown that this figure represents an improvement of identification rate accuracy when using the SAD algorithm in noisy environment. Further, the detector functioned accurately in low SNR environments

The second test is about channel errors effect on a remote speaker recognition system. Therefore, we use original and reconstructed wave files after transmition over AWGN channel, furthermore we use these files with speaker identification system. Table 1 shows a simulation results of identification rate accuracy using original and reconstructed speech wave files, where we observe performance degradation of speaker identification accuracy when using reconstructed files.

The third test consists to do the identification rate of speaker recognition system using : PCM, DPM and ADPCM code used in our Remote Speaker Recognition system where the figure 15 illustrates this study using AWGN channel in noisy environment, where we can conclude the efficiency of PCM code.

The fourth test is the runtime of each codec used in our work Table 2 shows a simulation results of: PCM, DPCM and ADPCM techniques in term of runtime using our speaker recognition system to distances, where we can observe that DPCM requires more time to execute than PCM and ADPCM, but PCM technique requires low runtime.

Although it is clear that the use of channel coding provides good results, the fifth test is about the effects of channel coding on the accuracy of speaker identification rate. Therefore, we evaluated the identification rate with and without convolutional code (AWGN + code) based on SNR considering PCM technique. Figure 16 shows a simulation result for the identification rate as a function of SNR with and without convolutional code where we conclude the effectiveness of the channel code.

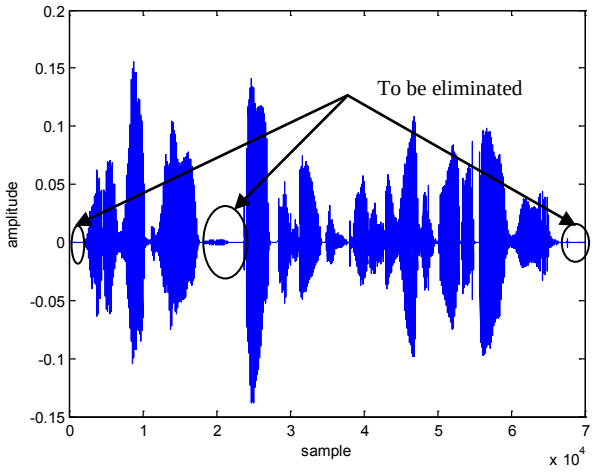

Fig.8. The original speech signal.

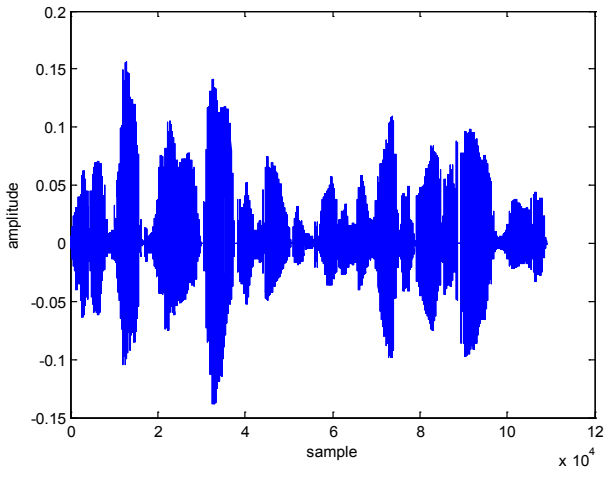

Fig. 9. Speech signal through activity detection algorithm (α=0.35).

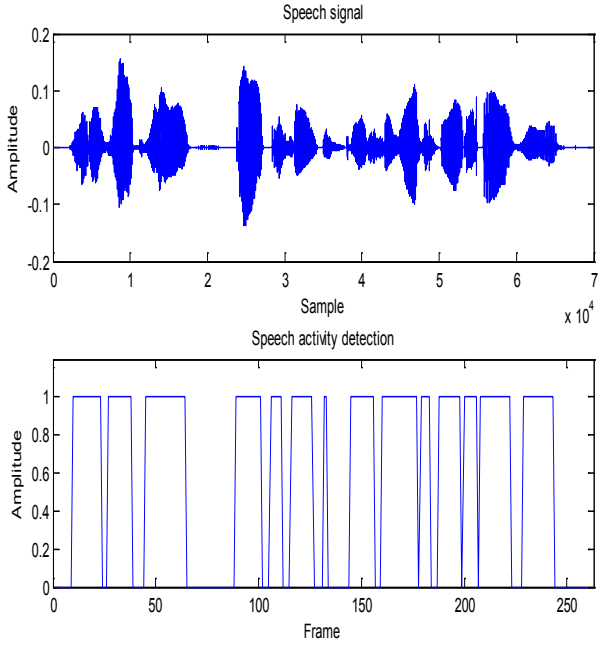

Fig. 10. Speech signal (clean signal) and SAD counter (below).

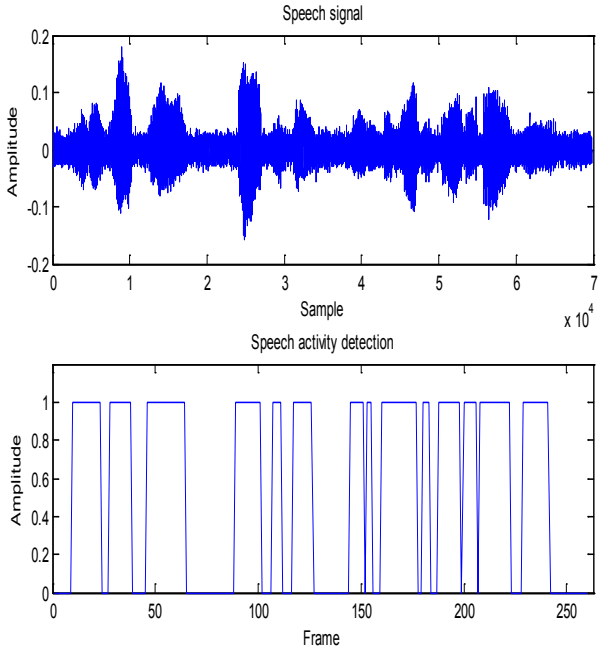

Fig . 12. Speech signal and SAD counter (below) at SNR=5dB

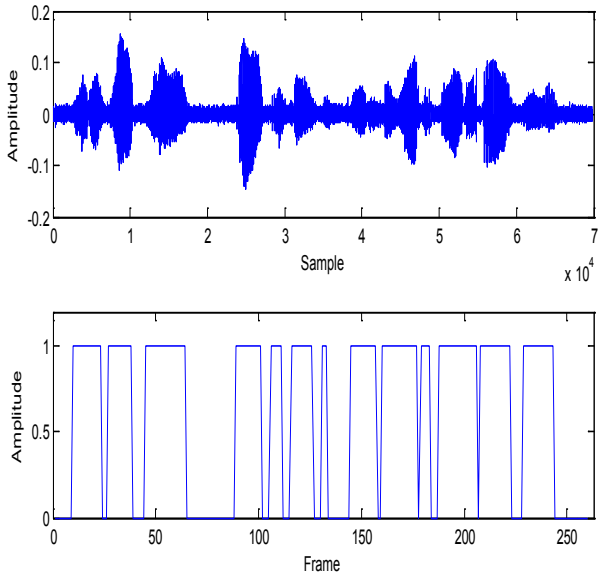

Fig. 11. Speech signal and SAD counter (below) at SNR=10dB.

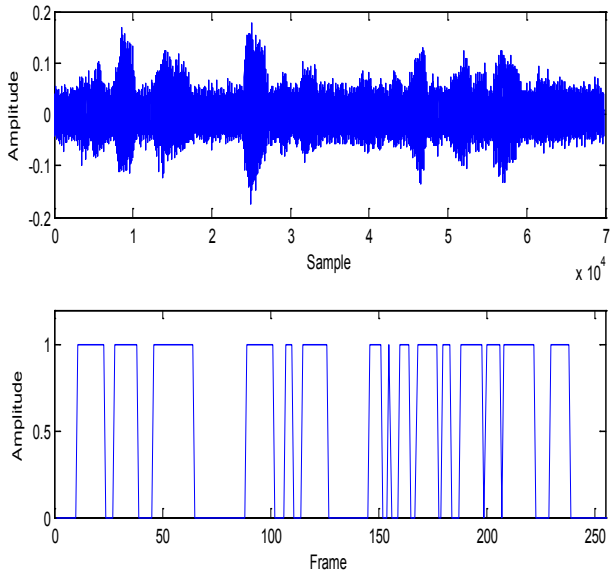

Fig. 13. Speech signal and SAD counter (below) at SNR=0dB

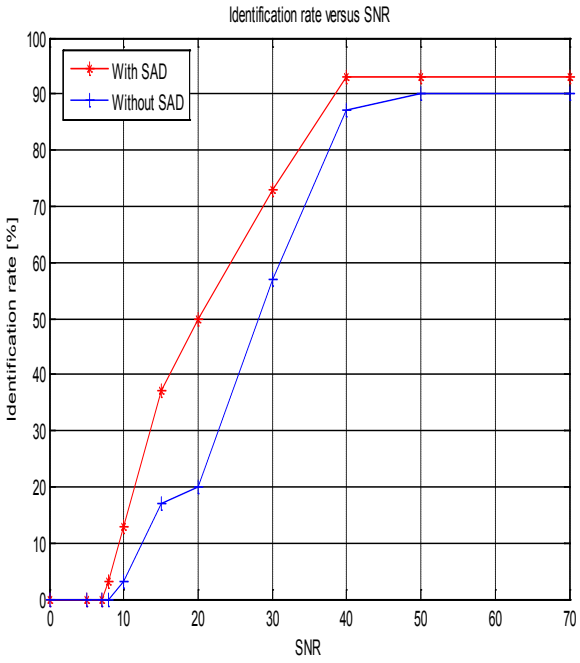

Fig 14. Identification rate with and without speech activity detection algorithm versus SNR

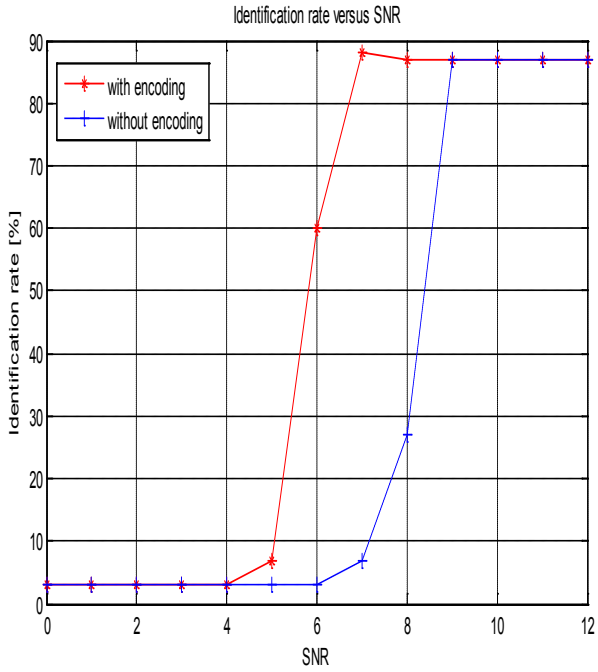

Fig. 16. Speaker identification accuracy with and without encoding (covolotional code).

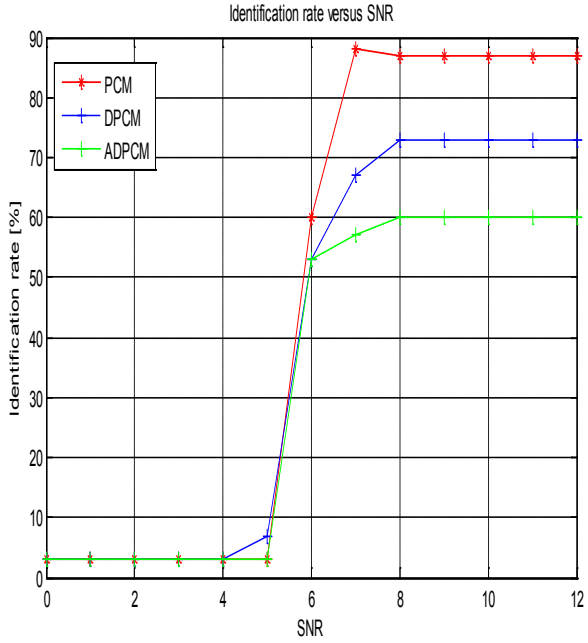

Fig. 15. Remote speaker identification accuracy using PCM, DPCM, and ADPCM versus SNR

Table1: Identification rate accuracy using original speech waveform and reconstructed speech after transmition over AWGN channel (SNR=100 dB).

	Speaker identification system	**Speaker identification over AWGN Channel**
Identification rate %	93	87

Table2: Runtime of : PCM, DPCM and ADPCM.

	PCM	**DPCM**	**ADPCM**
Elapsed time [sec]	100.84	764.24	554.32

8 Conclusion

In this work we have done a comparative study of speech codecs: PCM, DPCM and ADPCM in view of their effects on our recognition system performance of remote automatic speaker in noisy

environment. Therefore, a system configuration is set. Since, Speech activity detection (SAD) algorithm performs speech/nonspeech classification and background noise reduction process, we have developed a new (SAD) algorithm which improves memory capacity and identification rate accuracy.

Our proposed SAD algorithm that is based on energy and zero crossing rate, performs suitable counter of speech activity. Furthermore, it functioned accurately in low SNR environments (down to SNR=5 dB) and leads to a good identification accuracy.

In order to improve identification rate accuracy, the use of channel coding is necessary to make the remote system more robust against channel errors; therefore, we have chosen convolutional code. The best overall performance of speech codecs was observed for PCM code in terms of identification rate accuracy and runtime . Moreover, it's recommended using PCM technique as speech codec in remote speaker recognition system in VoIP applications.

References:

[1] Al-Sawalmeh, W., Daqrouq, K., Al-Qawasmi, A. R., & Hillal, T. A. (2009). The use of wavelets in speaker feature tracking identification system using neural network. *WSEAS Transactions on Signal Processing*, *5*(5), 167-177..

[2] FURUI, Sadaoki. Recent advances in speaker recognition. *Pattern Recognition Letters*, 1997, vol. 18, no 9, p. 859-872.

[3] LUNG, Shung-Yung. Feature extracted from wavelet eigenfunction estimation for text-independent speaker recognition. *Pattern recognition*, 2004, vol. 37, no 7, p. 1543-1544.

[4] Impedovo, D., & Refice, M. (2008). Frame length selection in speaker verification task. *WSEAS Transaction on Systems*, *7*(10), 1028-1037.

[5] D. Impedovo and M. Refice: "Optimizing Features Extraction Parameters for Speaker Verification", 12th WSEAS International Conference on SYSTEMS, 2008, pp. 498-503.

[6] SAHIDULLAH, Md et SAHA, Goutam. Design, analysis and experimental evaluation of block based transformation in MFCC computation for speaker recognition. *Speech Communication*, 2012, vol. 54, no 4, p. 543-565.

[7] PEINADO, Antonio et SEGURA, Jose. *Speech Recognition Over Digital Channels: Robustness and Standards*. John Wiley & Sons, 2006.

[8] HECK, Larry P., KONIG, Yochai, SÖNMEZ, M. Kemal, *et al.* Robustness to telephone handset distortion in speaker recognition by discriminative feature design. *Speech Communication*, 2000, vol. 31, no 2, p. 181-192.

[9] ALEXANDER, Anil, BOTTI, Filippo, DESSIMOZ, D., *et al.* The effect of mismatched recording conditions on human and automatic speaker recognition in forensic applications. *Forensic science international*, 2004, vol. 146, p. S95-S99.

[10] NEVILLE, Katrina, AL-QAHTANI, Fawaz, HUSSAIN, Zahir M., *et al.* Recognition of Modulated Speech over OFDMA. In : *TENCON 2006. 2006 IEEE Region 10 Conference*. IEEE, 2006. p. 1-3.

[11] Al-Sawalmeh, W., Daqrouq, K., Al-Qawasmi, A. R., & Hillal, T. A. (2009). The use of wavelets in speaker feature tracking identification system using neural network. *WSEAS Transactions on Signal Processing*, *5*(5), 167-177..

[12] XIE, Chuan, CAO, Xiaoli, et HE, Lingling. Algorithm of Abnormal Audio Recognition Based on Improved MFCC. *Procedia Engineering*, 2012, vol. 29, p. 731-737.

[13] RECOMMENDATION, G. 711:"Pulse Code Modulation (PCM) of voice frequencies". *ITU (November 1988)*, 1988.

[14] Silovsky, J., Cerva, P., & Zdansky, J. (2011, September). Assessment of speaker recognition on lossy codecs used for transmission of speech. In*ELMAR, 2011 Proceedings* (pp. 205-208). IEEE..

[15] N. S. Jayant, "Digital coding of speech waveforms-PCM, DPCM and DM quantizers," Proc. IEEE, vol. 62, pp. 621-624, May 1974.

[16] DONG, Hui, GIBSON, Jerry D., et KOKES, Mark G. SNR and bandwidth scalable speech coding. In : *Circuits and Systems, 2002. ISCAS 2002. IEEE International Symposium on*. IEEE, 2002. p. II-859-II-862 vol. 2.

[17] 40, 32, 24, 16 kbit/s Adaptive Differential Pulse Code Modulation (ADPCM), International Telecommunication Union Std. G.726 (12/90), Geneva 1990.

[18] MUÑOZ PORRAS, José María et IGLESIAS CURTO, José Ignacio. Classification of convolutional codes. *Linear*

Algebra and its Applications, 2010, vol. 432, no 10, p. 2701-2725.

[19] AMADASUN, M. et KING, R. A. Improving the accuracy of the Euclidean distance classifier. *Electrical and Computer Engineering, Canadian Journal of*, 1990, vol. 15, no 1, p. 16-17.

[20] NIDHYANANTHAN, S. S., & KUMARI, R. S. S. (2013). Language and Text-Independent Speaker Identification System Using GMM. *Wseas Trans. Signal Process*, *4*, 185-194..

[21] HATAMIAN, Shahin. Enhanced speech activity detection for mobile telephony. In : *Vehicular Technology Conference, 1992, IEEE 42nd*. IEEE, 1992. p. 159-162.

[22] PADRELL, Jaume, MACHO, Dušan, et NADEU, Climent. Robust speech activity detection using LDA applied to FF parameters. In : *Proc. ICASSP*. 2005.

[23] MACHO, Dusan, PADRELL, Jaume, ABAD, Alberto, *et al.* Automatic speech activity detection, source localization, and speech recognition on the CHIL seminar corpus. In : *Multimedia and Expo, 2005. ICME 2005. IEEE International Conference on*. IEEE, 2005. p. 876-879.

[24] LIANG, Zhang, YING-CHUN, Gao, ZHENG-ZHONG, Bian, *et al.* Voice activity detection algorithm improvement in adaptive multi-rate speech coding of 3GPP. In : *Wireless Communications, Networking and Mobile Computing, 2005. Proceedings. 2005 International Conference on*. IEEE, 2005. p. 1257-1260.

[25] HARSHA, B. V. A noise robust speech activity detection algorithm. In : Intelligent Multimedia, Video and Speech Processing, 2004. Proceedings of 2004 International Symposium on. IEEE, 2004. p. 322-325.

[26] KOTNIK, Bojan, HOGE, Harald, et KACIC, Zdravko. Evaluation of pitch detection algorithms in adverse conditions. In : Proc. 3rd international conference on speech prosody. 2006. p. 149-152.

[27] Charalampidis, D., & Kura, V. B. (2005, July). Novel wavelet-based pitch estimation and segmentation of non-stationary speech. In *Information Fusion, 2005 8th International Conference on* (Vol. 2, pp. 5-pp). IEEE.

[28] KATHIRVEL, P., MANIKANDAN, M. Sabarimalai, SENTHILKUMAR, S., *et al.* Noise robust zerocrossing rate computation for audio signal classification. In : *Trendz in Information Sciences and Computing (TISC), 2011 3rd International Conference on*. IEEE, 2011. p. 65-69.

[29] LU, Jun, TJHUNG, Tjeng Thiang, ADACHI, Fumiyuki, *et al.* BER performance of OFDM-MDPSK system in frequency-selective Rician fading with diversity reception. *Vehicular Technology, IEEE Transactions on*, 2000, vol. 49, no 4, p. 1216-1225.

[30] Garofolo, J. S., Lamel, L. F., Fisher, W. M., Fiscus, J. G., Pallett, D. S., and Dahlgren, N. L., "DARPA TIMIT Acoustic Phonetic Continuous Speech Corpus CDROM," *NIST*, 1993.