Speech/Non-speech Discrimination for Broadcast Radio Using Entropy Energy Modulation

Debabi Turkia and Cherif Adnen

Laboratory Analysis and processing of electrical and energy systems, Faculty of Sciences of Tunis, FST, Tunisia

Summary
Humans have a remarkable ability to classify sound signals into classes: music, speech, applause, laughter, etc. Faced with an excessive abundance of multimedia documents, we propose in this paper to develop a new configuration of multimedia documents based on entropy and entropy energy for automatic segmentation. A sound classification plays an important role in rich and varied applications, ranging from indexing audio documents to protecting copyright and archiving the diversity of radio and television channels. Given the diversity of requirements of these potential applications in the sound of classification, our object is to choose a generalist approach to the classification of sound documents that can easily adapt to classes defined according to its particular application. The proposed approach is based on entropy energy modulation. The classical problems of sound classification are summarized by three classes: the classification into music/speech, man/woman and action/non-action. Our application concerns the segmentation of a sound track in speech/non-speech.

Key words:
Indexing, broadcast, classification, entropy, speech discrimination.

1. Introduction

In recent years, sound classification is often used in many areas such as: biomedical applications [1, 2], mobile phones [3, 4] and biology [5]. Particularly, speech-music separation techniques are very useful in automatic speech recognition [17] dedicated to multimedia field, an effective navigation in this field is essential for generalized access and use of new sources of information.

The literature is enriched by many works that treats many speech-music separation methods. In [6], Chungsoo et al proposed an implementation of a speech/music [13, 14] classifier based on SVM by enhancing temporal locality in support vector references. Hence, this method is tested by applying it to a speech codec and it proved good results in term of the number of memory accesses, overall execution time, and energy consumption. Authors in [7] presented a single microphone speech-music separation based on mixture models. This method was evaluated with Poisson and complex Gaussian observation models and it yielded better results than the standard non-negative matrix factorization (NMF) method. The work given by [16] depicted a speech/music classification exploiting conditional maximum a posterior criterion. This improved SVM-based speech/music classification outperformed the speech/music classification rule in SVM.

In this framework, we have proposed an effective speech/non-speech discrimination solution based on the entropy energy modulation described by a new method for parameterizing entropy energy and its phase.

The article is organized as follows. Section 2 is devoted the related work with mathematical formulation for entropy and energy of the signal. The feature extraction is described in section 3 in which we proposed a new setting: modulation of entropy energy for discrimination speech/non-speech with their different steps, while in section 4 we report and discuss experimental results. Some concluding remarks are given in section 5.

2. Related Work

Entropy and energy of the signal are defined by the following equation:

2.1 Entropy of The Signal

Entropy has been considered as a transformation which is defined as the degree of disorder of the information contained in a system.

For a discrete scalar random variable X with $\{x_1,\ldots, x_N\}$ and the distribution of probabilities $\{p_1,\ldots, p_N\}$ which measures its disorder, we associate the entropy defined by (Eq. 1) [8]:

$$H(X) = -\sum_{i=1}^{n} p_i \operatorname{Ln} p_i \quad (1)$$

2.2 Energy of Signal

For a signal x(t), the total energy is defined as (Eq. 2) [9]:

$$E_x = \int_{-\infty}^{+\infty} |x^2| \quad (2)$$

Manuscript received April 5, 2019
Manuscript revised April 20, 2019

3. Feature Extraction: New Settings for Discrimination Speech/non-speech: Modulation of Entropy

The block diagram of the proposed automatic segmentation audio documents system is summarized by the following Figure ("Fig1").
The different steps of the system are clarified in the succeeding paragraph.

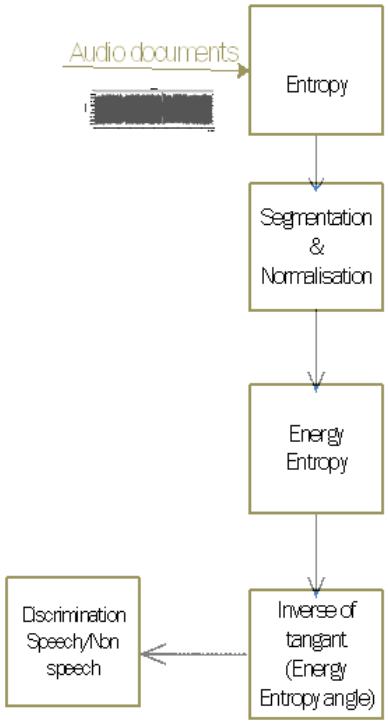

Fig. 1 Automatic segmentation system.

3.1 Data Base

We used a database of the radio variety ESTER: Evaluation of Transcription Systems Enriched with Radio Broadcasts [10], the corpus provided within the framework of this database for the development of the phase I systems, it is composed of signals with a total duration of 20h:04 m:01s, coming from IF broadcasts. It contains speech, music, songs and various sounds such as laughter, applause…
Our work is divided into 3 steps:
STEP1: Discrimination with entropy energy: After signal acquisition of audio documents, the application of entropy is used and is split into fixed frames (projecting entropy on the x-axis) and then normalization signal is exerted. The energy of the entropy is applied to find the angle of the signal and one leads to discrimination [12, 15] Speech/Non-Speech (S/N S).
STEP2: Discrimination with angle entropy energy: We use the inverse of a tangent (angle of the energy of the entropy): The discrimination (S/N S) is summarized with a threshold that is applied automatically.
STEP3: Discrimination with entropy energy modulation: Finally, both of energy of the entropy and its phase are used to show their discrimination (S/N S).

3.2 Discrimination with Entropy Energy

The first step is the acquisition of the signal of the sound documents, after the entropy is applied, then the segmentation of the signal into fixed frames by projecting it on the x-axis and its normalization and in the end the entropy energy is calculated.
The tested signal (Autoroute_Info_2009.wav) is from the database: ESTER.
Figure ("Fig2") shows an average plot of the signal, the segmentation and the normalization of entropy energy. In which, the entropy energy is anticipated on the x-axis to indicate unequivocally the correlation between entropy energy stage and speech. In what pursues figure below the original signal and the entropy energy of the signal which demonstrates that this entropy energy distinguished speech than other type of the signal (non-speech).

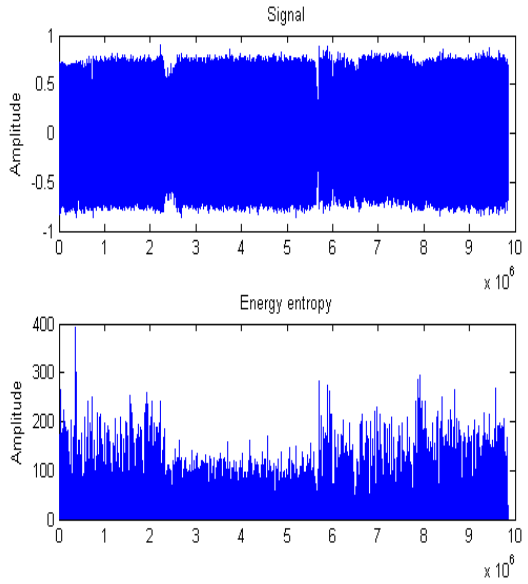

Fig. 2 Segmentation and normalization of entropy energy (Autoroute_Info (_2009).wav).

Figure ("Fig3") shows that the entropy energy increases with speech than with other sort of signal (Non-speech: music, song, applause, laugh …), to discriminate between speech and non-speech, a threshold is used that

is calculated from the maximum and the mean of entropy energy.

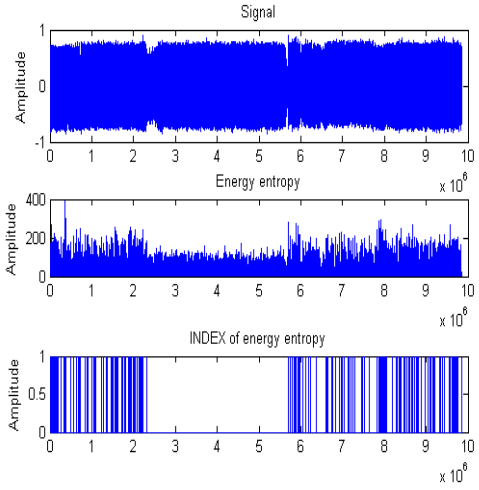

Fig. 3 Signal discrimination with entropy energy.

Figure (“Fig4”) below demonstrates speech/non-speech separation visibly with entropy energy technique.

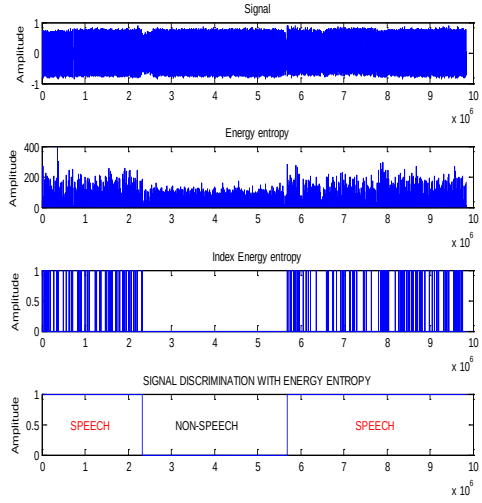

Fig. 4 Speech and non-speech automatic discrimination with entropy energy.

3.3 Discrimination with Angle Entropy Energy

To find the phase of the signal, the inverse of the tangent is applied called the entropy energy angle. The following Figure (“Fig5”) illustrates a typical plot of the original signal and the angle entropy energy. As observed, this figure proves that music has an angle of entropy greater than that of speech.

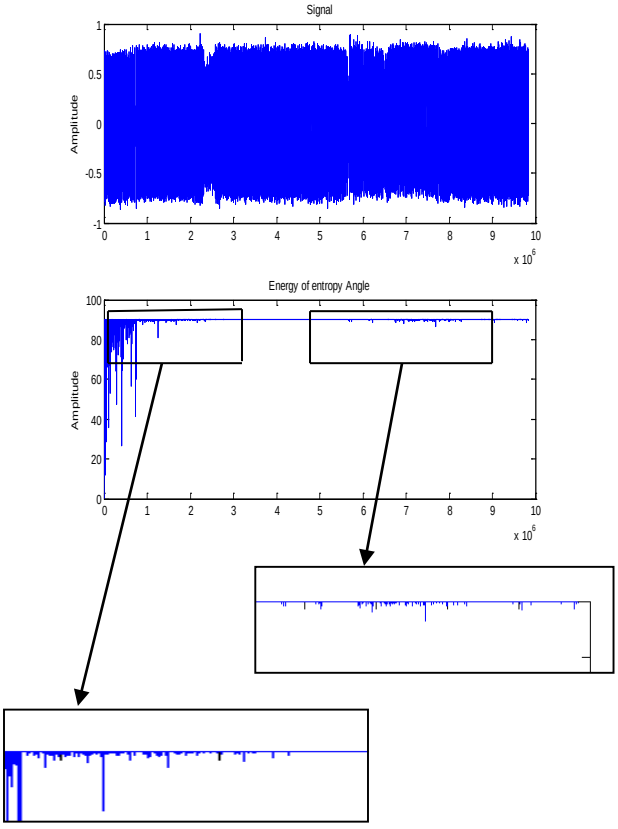

Fig. 5 Angle entropy energy.

The angle entropy energy demonstrates that it admits specific characteristics for speech. This speech/non-speech discrimination (S/N S) is obtained with thresholding applied automatically, all this is observed in the Figure (“Fig6”).

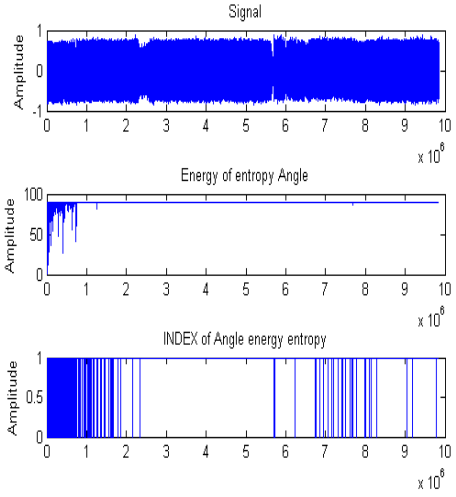

Fig. 6 Signal discrimination with angle entropy energy.

Figure (“Fig7”), as watched, illustrates a typical plot of the signal, the angle entropy energy and the discrimination(S/N S).

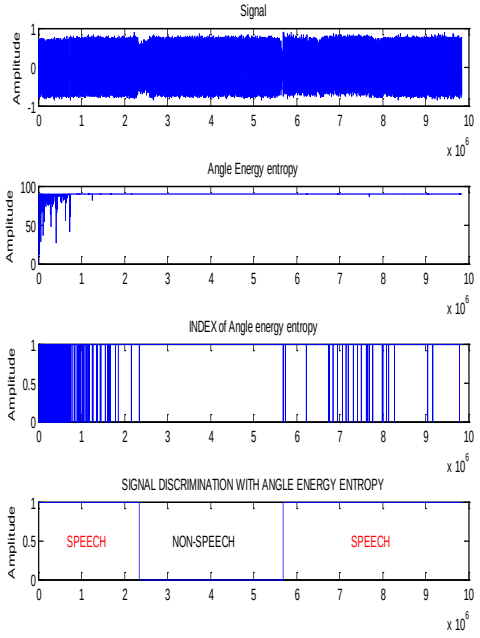

Fig. 7 Speech and non-speech automatic discrimination with entropy energy angle.

3.4 Discrimination with Entropy Energy Modulation

Figure (“Fig8”) uses both new parameters: The energy of the entropy and its phase to show their discrimination.

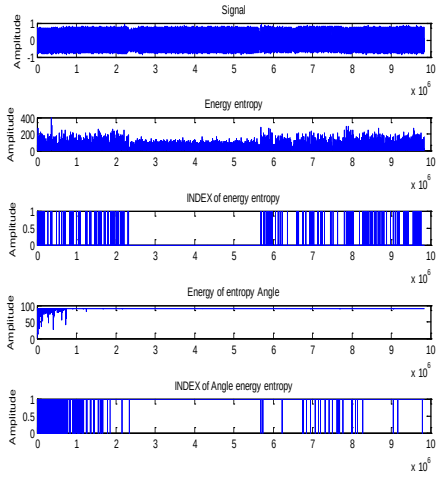

Fig. 8 Signal discrimination with entropy energy and his angle.

To obtain a better discrimination speech/non-speech, we use the combination of the two new parameters: entropy energy and angle entropy energy, Figure (“Fig9”) shows this discrimination with entropy energy modulation.

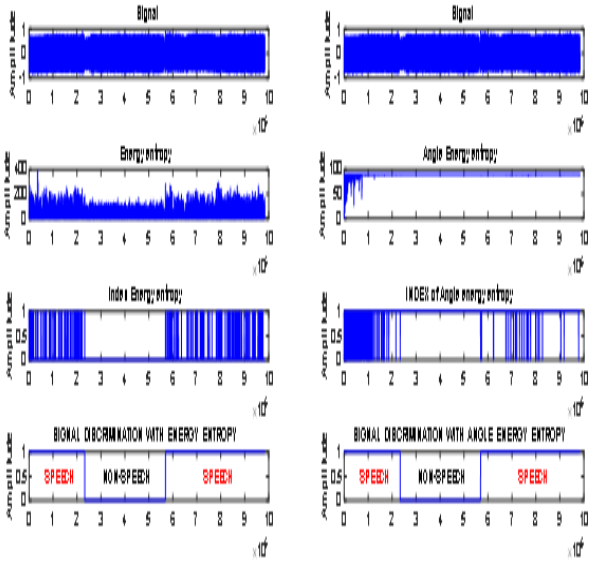

Fig. 9 Speech and non-speech automatic discrimination with entropy energy modulation.

4. The Experimental Results

4.1 Working Environment

The **MATLAB** environment is the basis of the simulation of the proposed algorithm whose system performance is evaluated from an **ESTER** database composed of a French radio database: it is a rich database in information (speech, music, song, etc.) duration 20h:04 m:01s, frequency = 16 kHz.

4.2 Simulation Results Analysis

The tested signal is composed with 100 audio documents of the database: ESTER(16h:31m:47s), which is a BD very difficult to study, constituted of speech segments in discussion (meeting), songs and music, these documents present a speech classification rate 99%(error of ~10m). According to [11], (“Table1”) summarizes speech classification.

Table 1: Speech Classification rate

Features	Speech classification rate (%)
Spectral Center(SC)	86.2
Spectral Frequency(SF)	85.2
Spectral Ratio(SR)	86.4
Zero Crossing Rate(ZCR)	87.7
Mel-Frequency Cepstral Coefficient(MFCC)	94.3
Entropy(E)	97
Energy entropy(Ee)	98
Angle energy entropy(α)	97
Modulation energy entropy(Mee)	99

The following Figure ("Fig10") demonstrates clearly the efficiency of our automatic discrimination method compared with others methods.

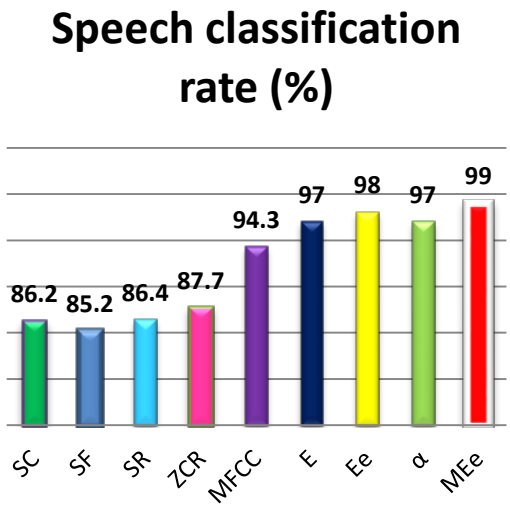

Fig. 10 Percentage of speech classification rate with different methods.

5. Conclusion

In this article, two new audio discrimination parameters based on entropy energy are presented. This developed system has a discrimination tool based on: Calculation of entropy energy and that of its phase to separate SPEECH / NON-SPEECH.

In our study, measurements and simulations are automatically determined by optimal parameters (threshold); on the performance discrimination system of the ESTER audio database, which is composed of speech (meeting), music and songs. We can mention as perspective of our work it is the indexation of the speakers present in a sound document by using classifiers GMM, SVM etc...

In conclusion to have a better discrimination, it is necessary to combine the two methods: to use the modulation of energy of entropy.

Acknowledgments

I would particularly like to express my thanks to my dear colleague from the laboratory Ms. **Sihem Nasri**, who helped me to complete this article.

References

[1] Saki F, Kehtarnavaz N. "Real-time hierarchical classification of sound signals for hearing improvement devices", Appl Acous 2018; 132(2018): 26-32.

[2] Sen I, Saraclar M, P. Kahya Y. "A Comparison of SVM and GMM-Based Classifier Configurations for Diagnostic Classification", IEEE Trans Biomedical Engineering 2015; 72(7): 1768-67.

[3] Lim C, Chang J.H. "Enhancing support vector machine-based speech/music classification using conditional maximum a posteriori criterion", IET Signal Processing 2012; 6(4): 335-340.

[4] Saki F, Sehgal A, Panahi I, Kehtarnavaz N. "Smartphone-based real-time classification of noise signals using subband features and random forest classifier", In:Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Shanghai, China, March 2016:p. 2204–8.

[5] González-Hernández FR, Sánchez-Fernández LP, Suárez-Guerra S, Sánchez-Pérez LA. "Marine mammal sound classification based on a parallel recognition model and octave analysis", Appl Acous 2017; 119(2017): 17-28

[6] Chungsoo L, Seong-Ro L, Joon-Hyuk C. "Efficient Implementation of an SVM-Based Speech/Music Classifier by Enhancing Temporal Locality in Support Vector References", IEEE Trans Consumer Electronics 2012; 58(3): 898-904.

[7] Demir C, Saraçlar M, Cemgil AT. "Single-Channel Speech-Music Separation for Robust ASR With Mixture Models", IEEE Trans Audio Speech Lang Process 2013; 21(4): 725-36.

[8] Doignone C. "Signal processing" Course, University of Louis Pasteur de Strasbourg, France (2008-2009).

[9] Mohammad-Djafari A. "Entropy in signal processing", Signal and System Laboratory (cnrs-sup'elec-ups), France, (2001).

[10] "ESTER Campaign: Evaluation of Transcription Systems Enriched with Radio Programs Phase I Evaluation Plan ", http://www.afcpparole.org/ester/.

[11] A. Masoumeh Velayatipour, B. Mohammad Mosleh, "A Review on Speech-Music Discrimination Methods", 2014;IJCSNS.

[12] EL-MALEH, Khaled, KLEIN, Mark, PETRUCCI, Grace, et al. "Speech/music discrimination for multimedia applications", In: 2000 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings (Cat. No. 00CH37100), IEEE, 2000, p. 2445-2448.

[13] SCHEIRER, Eric D. et SLANEY, Malcolm. "Multi-feature speech/music discrimination system", U.S. Patent No 6,570,991, 27 may 2003.

[14] PANAGIOTAKIS, Costas et TZIRITAS, Georgios. " A speech/music discriminator based on RMS and zero-c

rossings", IEEE Transactions on multimedia, 2005; vol. 7, no 1, p. 155-166.
[15] WANG, W. Q., GAO, W., et YING, D. W. "A fast and robust speech/music discrimination approach", In : Fourth International Conference on Information, Communications and Signal Processing, 2003 and the Fourth Pacific Rim Conference on Multimedia. Proceedings of the 2003 Joint, IEEE, 2003; p. 1325-1329.
[16] CHOU, Wu et GU, Liang. "Robust singing detection in speech/music discriminator design", In : 2001 IEEE International Conference on Acoustics, Speech, and Signal Processing, Proceedings (Cat. No. 01CH37221), IEEE, 2001; p. 865-868.
[17] CULLINGTON, Helen E. et ZENG, Fan-Gang. "Comparison of bimodal and bilateral cochlear implant users on speech recognition with competing talker, music perception, affective prosody discrimination and talker identification", Ear and hearing, 2011; vol. 32, no 1, p.16.

Debabi Turkia Received Diploma of Advanced Studies (DEA) degree from the National Engineering School of Tunis (ENIT) and she is currently pursuing his Ph.D. in laboratory analysis and processing of electrical and energy systems, Faculty of Sciences of Tunis, FST, Tunisia.

Adnen Cherif Received the engineer, master and doctorate degrees from the National Engineering School of Tunis (ENIT), in Tunisia, he is a senior Professor Doctor at the Science Faculty of Tunis and responsible in laboratory analysis and processing of electrical and energy systems, Faculty of Science of Tunis, FST, Tunisia. His fields of interest, concern digital signal processing, energetic systems, renewable energies and Smart-Grids.