International Journal of Emerging Technology and Advanced Engineering
Website: www.ijetae.com (ISSN 2250-2459, ISO 9001:2008 Certified Journal, Volume 6, Issue 10, October 2016)

Acoustic Phonetic Decoding in Arabic Language

Karim Tahiry[1], Laila Elmaazouzi[2], Elmustapha Louragli[3], Yassir Menssor[4], Abdelmajid Farchi[5]
[1,3,4,5]*Faculty of Sciences & Technics / IMII Laboratory, Settat, Morocco*
[2]*Graduate School of Technology / LAPSSII Laboratory, Safi, Morocco*

***Abstract*— The present study investigates acoustic phonetic decoding in Arabic language, our goal is to segment Arabic continuous speech into words. Our system to acoustic phonetic decoding is based on two basic features of speech signal which are the time-domain features and spectral centroid. After extracting these features, a simple dynamic thresholding is applied to detect the word boundaries. We recorded an Arabic corpus containing sentences in order to apply and test our acoustic phonetic decoding system.**

***Keywords*— Arabic continuous speech, speech segmentation, short Time Energy, Spectral centroid.**

I. Introduction

Arabic language is the second most spoken language in terms of number of speakers (more than a billion people worldwide). However, it is much less researched compared to other languages [1]. The automatic speech recognition is a human-machine communication, when we listen to a familiar language; we perceive and recognize a succession of words. So to recognize continuous speech, we segmented it into words. Acoustic phonetic decoding is used to decode the acoustic signal into linguistic units (phonemes, syllables, words ...) its main functions are extraction of relevant parameters, segmentation of continuous speech into words, identification of those units or labels. Segmentation is then a main function of the acoustic phonetic decoding [2].

Many studies have investigated the segmentation of continuous speech in different languages. klara vicsi and gyorgy szaszak (2005) [3] have studied segmentation of continuous speech in word and phrasal level in Hungarian and Finnish languages by examination of supra-segmental parameters using two approaches. They used rule-based approach for the detection of emphasized syllables and data-driven approach based on HMM method. V. Kamakshi Prasad et al. (2004) [4] have presented a new algorithm to automatically segment a continuous speech signal into syllable-like segments. Their algorithm for segmentation is based on processing the short-term energy function of the continuous speech signal. Amit Juneja and Carol Espy-Wilson (2002) [5] have examined a methodology for combining acoustic-phonetic knowledge with statistical learning for automatic segmentation and classification of continuous speech.

Matthew Harold Davis (2000) [6] has investigated the role of different sources of information. He has divided up the speech stream using acoustic cues in the speech signal and statistical regularities in the structure of the language. Regine andre-obrecht (1998) [7] has discussed a statistical approach for the segmentation of the continuous speech signal.

Other studies have investigated the segmentation of Arabic continuous speech. Tolba et al. (2005) [8] have developed a new algorithm for Arabic speech Consonant and Vowel segmentation based on wavelet transform and spectral analysis without linguistic information. Al-Manie et al. (2010) [9] have reported a method dividing the Arabic speech into regions of low and high energy corresponding to voiced and unvoiced phonemes. Mohamed S. Abdo et al. (2014) [1] have developed an algorithm for segmenting a subset of emphatic and non-emphatic sounds automatically from continuously spoken Arabic speech. Their method is based on peaks detection from delta function of Mel Frequency Cepstral Coefficients “MFCC”.

The principal motivation of this work is to contribute to the experimental literature on segmentation of continuous speech in Arabic language. The findings presented here are the implementation of an algorithm based on acoustic features in order to study the Arabic continuous speech segmentation. The primary goal is the words detection in continuous speech according to factors such as short Time Energy and Spectral centroid.

This paper is organized as follows. The first section outlines the methods and tools employed and the experiments carried out. The second section comprises presentation and discussion of the results. In the last section, there is a summary of the findings and presentation of conclusions.

II. Methods

A. General processing

For our experiments, we used the speech digitized with a 22050 Hz sampling rate to compute the short Time Energy and Spectral centroid. The speech utterances are subsequently divided into time segments of 11.6 ms with an overlap of 9.6 ms.

International Journal of Emerging Technology and Advanced Engineering
Website: www.ijetae.com (ISSN 2250-2459, ISO 9001:2008 Certified Journal, Volume 6, Issue 10, October 2016)

Each segment is first hamming windowed, then a 512-point fast Fourier transform (FFT) is computed and the short Time Energy and Spectral centroid are calculated.

B. Short Time Energy

During the production of speech, the excitation signal is considered voiced or unvoiced according to, respectively, the presence or absence of the vocal folds vibration. The Short-time energy of voiced sounds is higher than unvoiced ones or silence. It is then often used in speech processing to distinguish between voiced and unvoiced speech segments [10]. It is generally calculated on a short time basis by windowing the signal at a particular time, squaring the samples and taking the average. The short-time energy function of a speech frame with length N is defined as:

$$E = \frac{1}{N}\sum_{m=1}^{N}[x(m)\mathrm{w}(\mathrm{n}-\mathrm{m})]^2$$

x(m) : the discrete-time audio signal.
w(n) : window function.

C. Spectral centroid

Spectral centroid is used in digital signal processing to characterize a spectrum. It indicates the location of the "centre of gravity" of the spectrum. This feature is a measure of the spectral position, with high values corresponding to “brighter” sounds [11]. It is given by the following equation:

$$\mathrm{Ct} = \frac{\sum_{n=1}^{N} n.|Mt[n]|^2}{\sum_{n=1}^{N}|Mt[n]|^2}$$

Mt[n] : the magnitude of the Fourier transform at frame t.

n : the frequency sample.

D. Algorithm for speech decoding

First we apply a windowing to the acoustic signal for applying short-term speech characteristics, the following steps are:

- ✓ Calculate an energetic threshold (Te), by characterizing the noise part.
- ✓ Calculate short time energy (STE) of acoustic signal.
- ✓ Calculate a spectral centroid threshold (Tsc), by characterizing the noise part.
- ✓ Calculate spectral centroid (SC) of acoustic signal.
- ✓ Test if STE & SC are respectively higher than Te & Tsc
    - If yes the flag is marked 1
    - If not the flag is marked 0
- ✓ Map these flags on the signal to find the boundaries.

III. RESALTS AND DISCUSSIONS

Participants were twenty adults Moroccan speakers of Arabic Modern Standard. The group included 10 female and 10 male subjects ranged in age from 19 to 27 years. They were asked to produce sentences containing words giving in Table.1.

TABLE I
ARABIC SENTENCES USED FOR CREATION OF DATABASE

Sentences	**Arabic sentences**
Sentence 1	إرتفع سعر البنزين في السوق الدولية
Sentence 2	ذهبنا إلى الغابة يوم السبت للترفيه
Sentence 3	يقام مهرجان الموسيقى الروحية في مدينة فاس
Sentence 4	إنتقل إلى عفو ربه
Sentence 5	إنتقل فوق جمله

Arabic has a very rich consonantal system (stop, sonorant, fricative) and a relatively impoverished vocalic system. Arabic standard had three short vowel phonemes divided into two close vowels (palatal /i/ and labio-velar /u/) and one open vowel (guttural /a/) and three long vowels (/aa/,/uu/and/ii/) [12].

TABLE 2
SEGMENTATION RESULTS

Sentences	Segmentation Rate (%)
Sentence 1	93.3%
Sentence 2	97.1%
Sentence 3	90.5%
Sentence 4	92.4%
Sentence 5	95.8%

Table 2 shows the experimental results in terms of Segmentation Rate over various utterances. We can see that this system is performant and easy to perform (it is based on only two parameters). We can also see that the segmentation rate is lower for:

- ✓ The utterances containing consonants with low energy which were considered as a silence.
- ✓ The wrong co-articulation of words by speakers.

To enhance the performances of this system, we can put little pause between words when we record sentences to have a best co-articulation of words.

IJETAE
EXPLORING RESEARCH AND INNOVATIONS

International Journal of Emerging Technology and Advanced Engineering
Website: www.ijetae.com (ISSN 2250-2459, ISO 9001:2008 Certified Journal, Volume 6, Issue 10, October 2016)

IV. Conclusion And Future Work

In this study, we have presented an algorithm based on short time energy and spectral centroid in order to segment Arabic continuous speech into words. The choice of short-term characteristics allowed us to distinguish voiced from unvoiced speech components. The energy of the voiced segments is higher than that of unvoiced ones or silence. The spectral centroid helps to distinguish the noise located in the low frequencies. The interest of this algorithm is due to his simplicity and his high performances (accuracy about 90 %). In future researches, we aim to apply phonetic decoding to solve the problems caused by week consonants and to develop techniques to articulate words in a sentence without silence between words.

REFERENCES

[1] Mohamed, S. A., Kandil, A. H., Fawzy, S. A. 2014. MFC Peak Based Segmentation for Continuous Arabic Audio Signal. In Proceedings of the Middle East Conference on Biomedical Engineering (MECBME).

[2] Calliope. 1989. La parole et son traitement automatique. Mason.

[3] Vicsi, K., Szaszak, G. 2005. Automatic Segmentation of Continuous Speech on Word Level Based on Supra-segmental Features. International journal of speech technology 8(4). 363–370.

[4] Prasad, V. K., Nagarajan, T., Murthy, H. A. 2004. Automatic segmentation of continuous speech using minimum phase group delay functions. Speech Communication 42. 429–446.

[5] Juneja, A., Wilson, C. E. 2002. Segmentation of continuous speech using acoustic-phonetic parameters and statistical learning. In Proceedings of the 9th International Conference on Neural Information Processing (ICONIP '02).

[6] Davis, M. H. 2000. Lexical Segmentation in Spoken Word Recognition. Doctoral Thesis. Birkbeck College. University of London.

[7] Regine, A. 1998. A New Statistical Approach for the Automatic Segmentation of Continuous Speech Signals. Ieee transactions on acoustics, speech, and signal processing. vol. 36. No. 1.

[8] Tolba, M. F., Nazmy, T., Abdelhamid, A. A., Gadallah M. E. 2005. A novel method for arabic consonant/vowel segmentation using wavelet transform. International journal on intelligent cooperative information systems. IJICIS Vol. 5. No. 1. pp. 353-364.

[9] Mohammed, A. A, Alkanhal, M. I., Mansour, M. A. 2010. Arabic speech segmentation: Automatic verses manual method and zero crossing measurements. In processing of Indian journal of science and Technology. Vol. 3 No. 12.

[10] Rabiner, L. R., and Sambur, M. R. 1975. An Algorithm for Determining the Endpoints of Isolated Utterances. The Bell System Technical Journal. 54(2):297-315.

[11] Giannakopoulos, T. 2009. Study and application of acoustic information for the detection of harmful content and fusion with visual information. Ph.D. dissertation, Dept. of Informatics and Telecommunications. University of Athens. Greece.

[12] .Watson, J. C. E. The Phonology and Morphology of Arabic Watson. The phonology of the world's languages. Oxford University press.