ISCA Archive
http://www.isca-speech.org/archive

First International Conference on
Spoken Language Processing (ICSLP 90)
Kobe, Japan
November 18-22, 1990

Phonetic Study and Recognition of Standard Arabic Emphatic Consonants

M. Djoudi[1] H. Aouizerat[1,2] J.P. Haton[1]

[1] CRIN — INRIA Lorraine France [2] Université d'Alger

Abstract

In this paper we propose a phonetic study of emphatic consonants in standard Arabic, which is essentially based on the spectrographic examination of these consonants in various production contexts. This study allowed us to determine the pertinent parameters to the recognition of emphatic consonants in continuous speech. The recognition stage uses the SAPHA system, developed for the acoustic-phonetic decoding of standard Arabic and deals with the following points:

- segmentation of the speech signal into large phonetic classes;
- extraction of parameters necessary for the recognition of emphatic consonants;
- phonetic identification of these consonants by using an expert system based on production rules.

The results of the segmentation and of the recognition for three male speakers are given commented.

1 Introduction

The recognition of Arabic poses several problems, due to the particular phonetic characteristics of the Arabic language. The original feature of the phonetic system of Arabic is mainly based on the presence of emphatic consonants. Grammarians and phoneticians have granted a special importance to the emphatic phenomena and moreover, several physiological and acoustic studies have been carried out on emphatic consonants. Our study is simply acoustic, it relies on an analysis of the spectrograms of the consonants in several contexts and focuses on the phonetic recognition of emphatic consonants in continuous speech.

2 Experimental protocol

This study consists of the spectrographical analysis of a corpus of words pronounced by four male Algerian speakers, students in Nancy research computer center[CRIN]. The speech acquisition and analysis are carried out with the SNORRI software, developed at CRIN.

The corpus

In order to collect representative data for emphasis [8], we have proceeded as follows: we have put together as many words as possible for each emphatic phoneme and its non-emphatic peer in their three possible states: beginning, medial and final position. Each speaker pronounce the words of the corpus two by two, the word containing the emphatic phoneme and the word containing the non-emphatic one separated by a time interval of about 2 seconds, so that at the time of the analysis we obtain on the graphic console the 2 words together. The spectrographic analysis was focused on the segments in question and also on the left and/or right vocalic contexts in which they are. Moreover, as our work aims at continuous speech recognition, isolated words alone are not sufficient, since they do not answer the purpose of the criterion which really takes in account the production context of the phonemes and the coarticulation. It was therefore necessary to use the DJOUMA corpus, composed of 50 phonetically balanced sentences [4].

The analysis tool

SNORRI is an interactive signal editor designed for both speech specialists and non-specialists. It has the classical tools to acquire and reproduce speech signals and also to compute spectrograms and many other functions, which allow a finer analysis of the speech and the manipulation of the corpora. SNORRI runs on a CONCURENT 5600 computer. It uses a 12 or 16 bits acquisition card. The display takes place on a colour bitmap console. SNORRI is programmed in C and needs the window system and graphic primitives of CONCURRENT [7].

3 Phonetic study

The consonantal system of Arabic comprises 28 consonants [9]. The special feature of the system is based on the presence of pharyngal, glottal and emphatic consonants. Emphatic consonants are described as having a second point of articulation at the level of the pharynx [6], [1]. These consonants depend on a buccal phenomenon, which consists of the action of the root of the tongue being put backwards and of the lowering of the back of the tongue [2]. We include in emphatic consonants, the phonemes [/ṭ/, /ṣ/, /ḍ/, /ð̣/] defined by the first Arab grammariens [10]. Nevertheless, the real pronunciation of /ḍ/ has disappeared since a long time [1], it is confused nowadays with /ð̣/. Thus we deal only with only [/ṭ/, /ṣ/, /ð̣/] and their non-emphatic counterparts [/t/, /s/, /ð/]. The results of the acoustic investigation appears are given in the following subsections.

/t/ vs /ṭ/

The spectra of these two phonemes are relatively similar. It is made up of a silence interval ranging between 80 and 100 ms, followed by a burst with a center of gravity comprised between 4000 and 4600 hz. The /t/ may be affricated, generally presenting a friction noise when it is followed by an /i/. The important difference between /t/ and /ṭ/ is to be found in the values of the beginning of the formant-2 of the vowels before and after the two phonemes. Thus, the value of the beginning of F2 of /a/ close to /t/ ranges between 1400 and 1500 hz, the one of F2 of /i/ close to /t/ ranges between 1900 and 2100 hz and the one of F2 of /u/ close to /t/ ranges between 1100 and 1200 hz. On the other hand , the value of F2 of /a/ close to /ṭ/ ranges between 1000 and 1300 hz, the one of F2 of /i/ close to /ṭ/ between 1700 and 1900 hz and the one of F2 of /u/ close to /ṭ/ between 900 and 1050 hz.

/s/ vs /ṣ/

The spectra of fricatives /s/ and /ṣ/ present the same features, ie a strong noise of friction in high frequency from 3 khz. The value of the beginning of F2 of /a/ close to /s/ ranges between 1400 and 1600 hz, the one of F2 of /i/ close to /s/ between 2000 and 2200 hz and the one of F2 of /u/ close to /s/ between 1100 and 1300 hz. On the contrary, the value of the beginning of F2 of /a/ close to /ṣ/ ranges between 1000 and 1200 hz, the one of F2 of /i/ close to /ṣ/ from 1800 to 2000 hz and finally the one of F2 of /u/ close to /ṣ/ between 700 and 900 hz.

/ð/ vs /ð̣/

Defined as fricatives, these two consonants appear rather like plosives and sonorants with weak formants: F1 ranging between 250 and 350 hz, F2 between 1100 and 1300 hz and F3 between 2200 and 2300 hz. The value of the beginning of F2 of /a/ close to /ð/ ranges between 1300 and 1600 hz, the one of F2 of /i/ close to /ð/ between 1800 and 2000 hz and the one of F2 of /u/ close to /ð/ between 1150 and 1250 hz. On the contrary, the value of the beginning of F2 of /a/ close to /ð̣/ ranges between 1000 and 1100 hz, the one of F2 of /i/ close to /ð̣/ between 1600 and 1800 hz and the one of of F2 of /u/ close to /ð̣/ between 800 and 900 hz.

4 The recognition

The SAPHA system

Sapha is an acoustic-phonetic decoder system for standard Arabic, developed for the purpose of the analytic recognition of phonemes in continuous speech in multi-speaker context [5]. The system is made up of a set of modules and includes the acoustic, phonetic and phonologic that are themselves contained in the lower level of a speech recognition system [3]. It aims at :

- segmenting the speech signal into broad phonetic classes by using non contextual algorithms. These classes are vowels, plosives and fricatives. The rest of consonants is put into the class of unknown.
- automatically extracting indices pertinent to the phonetic recognition. These indices are :
  - segment duration.
  - degree of voicing.
  - position and features of the burst.
  - values of formants.
  - formant transitions.
  - lower limit of friction noise.
- identifying the phonemes by using an expert system. It consists of a knowledge base of phonemes identification under the form of production rules and an inference engine, which works in forward and backward chaining while carrying out a left to right analysis segment after segment.

Recognition of emphatic consonants and of their non-emphatic peers

Our goal consists in recognizing the consonants /s/, /ṣ/, /t/, /ṭ/, /ð/, /ð̣/ by using the knowledge we have acquired from our experience in the reading of speech spectrograms. This knowledge is described under the form of production rules. Rules are represented by structured objects composed of the following fields:

- a rule number for identifying the rule;
- a comment: the expert may describe here the semantic content of the rule, using natural language;
- left and right contexts: each context consists of a list of phones which describe all the phonetic contexts in which the rule applies;
- premisses: for any given speech segment the premisses explicit the conditions that the acoustic phonetic properties of the segment must verify in order that the rule may be activated;
- a conclusion: it may be either a list of scored phones or an action to be executed by the system.

The right and/or the left context part may not exist in a rule. Nevertheless, in our study, to descriminate between an emphatic segment and its non-emphatic peer, it was essential to take into consideration not only the right context but also the left one. Regarding the premisses, each one corresponds to a visual indice, that we have deduced from the reading of spectrograms and characterizing a given segment. Thus, for the recognition of the plosives /t/ and /ṭ/, we have relied our study on:

- the presence of the burst,
- the center of gravity of the burst,
- the voicing
- the formant transitions of the following and preceding vowels.

As far as the fricatives /s/ and /ṣ/ are concerned, we have used

- the degree of voicing,
- the lower limit of the friction noise
- the formant transitions of the adjacent vowels.

As for /ð/ and /ð̣/, we have used

- the values of the formants F1 and F2,
- the degree of voicing
- the formant transitions of the preceding and following vowels.

For the six phonemes we have implemented 60 rules.

5 Results and Discussions

We have tested the segmentation algorithms on the DJOUMA corpus, manually segmented. The given results are calculated in comparison with this manual labelling for 3 male speakers.

phonemes	Plosives	Fricatives	Unknown
t	100%		
ṭ	100%		
s		100%	
ṣ		100%	
ð	41%	17%	41%
ð̣	45%	17%	38%

Although defined as fricatives, /ð/ and /ð̣/ appeared rather like plosives and sonorants(generally at the beginning position and between vowels) than fricatives(in final position). Among the causes of the variety of nature of these two phonemes is the tendency nowadays in pronouncing the plosive /d/ instead of /ð/ and /ð̣/, also, when geminated they are automatically plosives. It is the reason, in our opinion, why the percentage of the plosives is higher in relation to sonorants and fricatives, concerning /ð/ and /ð̣/.
For the identification, the three phonemes that totalize the best scores are held as being labels of the segment. The percentage is given in the next table.

phonemes	recognized	inserted
t	85%	8%
ṭ	84%	7%
s	82%	9%
ṣ	80%	9%
ð	59%	11%
ð̣	55%	10%

Results for /t/, /ṭ/, /s/ and /ṣ/ are encouraging with regards to /ð/ and /ð̣/ which are highly ambiguous. Neverthless, to reduce the percentage of insertions for all phonemes, other rules have to be implemented.

6 Conclusion

We have presented a phonetic study of emphatic consonants in standard Arabic, which allowed us to determine the pertinent parameters for continuous speech recognition. Results already obtained have to be improved, especially for /ð/ and /ð̣/. Thus, we will implement other rules and in the near future we will introduce the notion of phonetic triplet.

References

[1] Salman H. Al. Ani. *Arabic Phonology. An Acoustical and Physiological Investigation.* Mouton & Co N.V., 1970.

[2] J. F. Bonnot. *Recherche expérimentale sur la nature des consonnes emphatiques de l'Arabe classique.* Technical Report 9, Travaux de l'institut de phonétique de Strasbourg, 1977.

[3] N. Carbonell, J. P. Haton, D. Fohr, F. Lonchamp, and J. M. Pierrel. APHODEX, design and implementation of an acoustic-phonetic decoding expert system. *IEEE International Conference on Acoustics, Speech and Signal Processing*, 1986.

[4] M. Djoudi, D. Fohr, and J. P. Haton. Phonetic Study for Automatic Recognition of Arabic. *Proceedings of European Conference on Speech and Technology*, 2:268–271, September 1989.

[5] M. Djoudi, D. Fohr, and J. P. Haton. SAPHA: un système expert pour le décodage acoustico-phonétique de l'arabe standard. *Première Conférence Maghrébine sur le Génie logiciel et l'Intelligence Artificielle*, Septembre 1989.

[6] A. Giannini and M. Pettorino. *The Emphatic Consonants in Arabic.* Giardini editori e stampatori, 1982.

[7] Y. Laprie. Un système d'étude interactif de la parole. *Actes des 17ème Journées d'Etudes sur la Parole*, 71–76, Sep 1988.

[8] K. Norlin. *A Phonetic Study of Emphasis and Vowels in Egytian Arabic.* Lund University Departement of Linguistics, 1987.

[9] A. Hadj Salah. Arabic Linguistics and Phonetics. In *Applied Arabic Linguistics and Signal & Information Processing*, pages 3–22, Hemisphere publishing corporation, 1987.

[10] Sibawayh. *EL KITAB, traité de grammaire arabe.* H. Derembourg, 1889.

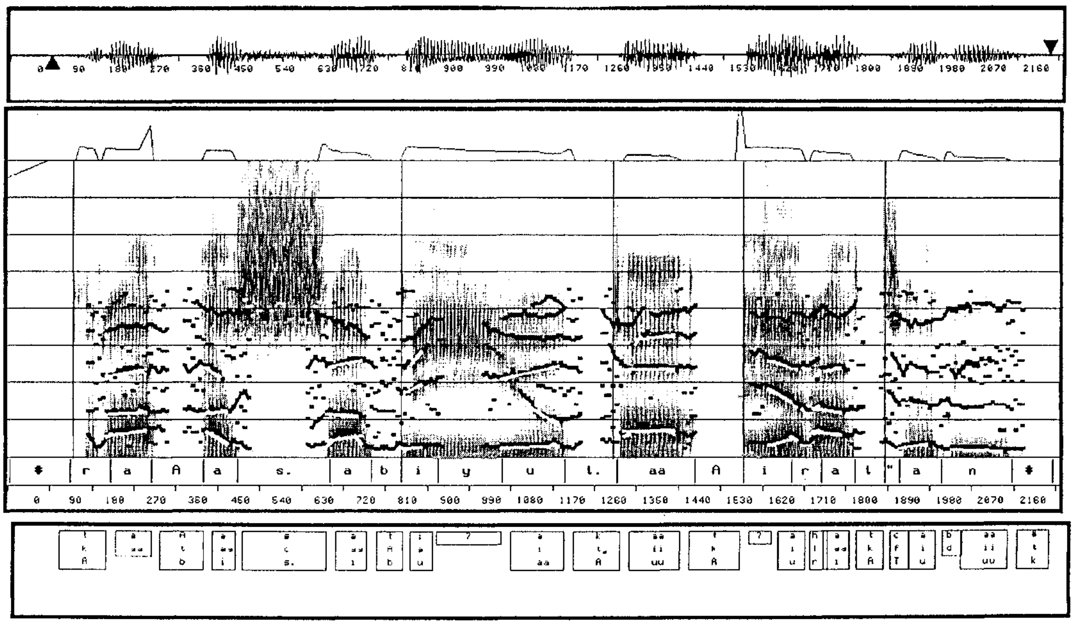

Figure 1: Temporal signal and spectrogram

Sentence : 'The kid have seen a plane'