Proc. 13th NCC, Riyadh. Nov., 1992.

Assessment of Standard Arabic Acoustic Phonetic Decoder

M. Djoudi[1]

Laboratoire d'Informatique de l'Université de Poitiers
Poitiers cedex France

ABSTRACT

Within the framework of the automatic recognition of continuous speech, we have developed SAPHA system : an acoustic phonetic decoder of standard Arabic. In This article, we present a first assessment of the system.

SAPHA makes possible the analytic recognition of phonemes in continuous speech for multispeakers. Combining other linguistic information (lexicon, syntax, semantic and pragmatic), the system can be considered as an important step towards a system for oral dialog between man and machine. It also acts as a module of a dictation machine.

The recognition module is realized as form of an expert system based on production rules. Knowledge is acquired after a phonetic study of Arabic carried out on DJOUMA corpus, which is composed of 50 sentences pronounced by 11 speakers (7 males and 4 Females) . This study also allows us to adopt a strategy for signal segmentation into large phonetic classes and determine the distinctive values of parameters used during recognition. In the same way, the manual labelling of the corpus sentences allows us to test the performances of the system [6].

Keywords : Assessment, Standard Arabic, Phonetic Decoding, Expert System

[1] This work has been done in the CRIN-CNRS / INRIA Lorraine research center (Nancy, France)

1 INTRODUCTION

Arabic language has been studied several times in the past. These studies relate to either the phonetic aspect or the linguistic component. However, up to now, the question of automatic recognition has hardly been talked. The recognition of Arabic poses several problems due to the particular phonetic characteristics of the language. Arabic contains 28 consonants, to each of them corresponds a particular phoneme. The original feature of the system is based on the glottal, pharyngeal and emphatic consonants.

The glottal and pharyngeal consonants are distinguished from the rest of the consonants by having distinct vertical places of articulation. A vertical place of articulation is defined as a set of anatomical locations from palate to the glottis inclusive. In contrast, a horizontal place of articulation is from the lips to the uvula, inclusive. The emphatic consonants are described as having a second place of articulation at the pharynx level [4]. The Arabic vowel system is generally described as being comprised of six vowel sounds. To each of the short vowel /a/, /i/, and /u/ corresponds a vowel of a longer duration /aa/, /ii/ and /uu/. The temporal opposition short vs. long is fundamental at the grammatical and semantic levels [1].

2 SAPHA SYSTEM

SAPHA system is designed for acoustic phonetic decoding of Arabic. Its structure consists of modules, it receives as a way of access the speech signal previously digitalized and as a result sends back a phonetic lattice. Around recognition modules, procedures for phonetic analysis and graphical display and also evaluation modules of system performances have been developed. The evaluation requires a corpus of balanced sentences pronounced by several speakers and manually labelled. The main stages of the system are segmentation, phonetic features extraction and segment identification [3].

2.1 Segmentation

This module performs the segmentation of the speech signal into broad phonetic classes by using non-contextual algorithms based on simple criteria. The main purpose of the segmentation is to reduce the combinatorial explosion during the recognition and to allow a centering for an automatic labelling. This segmentation is carried out by a set of three modules and consists of the segmentation of the speech signal into three large phonetic classes i.e. :

- VOY : vowels (/a/, /i/, /u/, /aa/, /ii/, /uu/),
- PLO : plosives (/t/, /k/, /?/, /b/, /d/, /q/, /t/)
- FRI : fricatives (/z/, /f/, /□/, /s/, /ʃ/, /□/, /h/, /z/, /s/)

The vowel segmentation process uses the total energy maxima and the energy of the frequency band [250-2500 Hz]. The segmentation algorithm of plosives uses the absence of energy beyond 600 Hz. A zero-crossing number of the signal and the center of gravity are used for segmentation of the fricatives.

Each of the three segmentation procedures produces a list of detected segments. Concerning the problem of inclusions and intersections the following remarks are observed :

If two segments are included in one other, according to the position of the maxima, either one segment having two characteristics (example fricative and a vowel for /i/ : class FRIVOY, plosive and fricative for /f/ : class PLOFRI) or two segments (for example a plosive then a fricative for /z/ which reveals as being a plosive followed by a fricative) are produced. In the latter case, limits of the segments are calculated through spectral difference. In the case of two disconnected segments, the segment existing between them will be labelled as sonnant if its duration is long, otherwise it will be attached to its neighbors. The class of sonnants consist therefore of the trill /r/, the lateral /l/, the nasals /m/ and /n/, the semi-vowels /w/ and /y/, as well as the fricatives having a formantic structures.

2.2 Extraction of features

The extraction of pertinent phonetic features is a very important stage in the process of phonetic decoding. To each phonetic parameter a procedure is associated to perform it. The values of these parameters will be used at the moment of the activation of a rule for the labelling module. These parameters are :

- segment duration,
- degree of voicing,
- position of the burst,
- characteristics of the burst,
- formant tracking,
- formant transitions,
- lower limit of noise,
- center of gravity.

2.3 Labelling

The SAPHA labelling module is an expert system based on production rules. From the segments provided by the segmentation module, this module tries to find out the pronounced phonemes by using the parameters extracted at the previous stage and the rules of the knowledge base. It is made up of a knowledge base of phonemes identification and an inference engine. This system has already been used for the phonetic decoding of French in the framework of the APHODEX project [2].

Knowledge base

A rule is made up of several parts, which can be optional :

- rule number
- commentary,
- left context (list of phonemes)
- right context (list of phonemes)
- premiss (condition on measures from the precedent, the actual or the following segment),
- conclusion (list of phonemes with scores).

Here is an example of rule :

R223
C Rule for /q/ context /a/ and /aa/ **C**
CONTEXT_RIGHT [a aa]
IF
burst-present_ACT &
^(burst-freq_ACT 2200 2600)
THEN [q 70]

The result scale varies between -100 (absolutely false) and +100 (absolutely true), the zero corresponding to the state of absolute uncertainty. The Knowledge base is presently made up about 250 rules.

Inference engine

The inference engine allocates a list of one or several phonemes to each segment detected by the segmentation module. The activation of a rule depends on condition expressed in left and right contexts, on the conclusion and on segmentation. A plausibility is assigned to each hypothesized phoneme. The conclusion of a rule may also be an action carried out or a list of phonemes. To process the fuzziness that can be found in the rules, the engine uses a reasoning mechanism based on fuzzy logic [5].

3 SYSTEM ASSESSMENT

3.1 Segmentation results

We have tested the segmentation algorithms on DJOUMA corpus, manually segmented. The results given on table 1 are obtained by comparison with the manual labelling.

Table 1 : Segmentation results

Phon	Nb	Plo	Voy	Fri	Aut	FriVoy	PloFri	Omis	Taux
a	623	1	552	2	14	15	0	39	91%
i	339	3	197	2	5	106	0	26	89%
u	149	1	127	0	4	1	0	16	85%
aa	181	0	163	0	1	9	0	8	95%
ii	58	0	27	2	0	24	0	5	87%
uu	42	1	32	0	2	1	0	6	78%
t	170	165	1	1	1	0	0	2	97%
k	51	49	0	0	0	0	0	2	96%
?	93	66	1	0	23	0	0	3	70%
b	77	73	0	0	3	0	0	1	94%
d	80	75	0	1	4	0	0	0	93%
q	61	60	0	0	1	0	0	0	98%
ṭ	36	33	0	0	1	0	0	2	91%
ḍ	10	2	2	0	5	0	0	1	50%
z	32	2	0	28	1	1	0	0	96%
f	60	4	0	55	0	0	1	0	93%
θ	13	0	0	12	1	0	0	0	92%
s	48	0	0	47	0	0	0	1	97%
ʃ	26	0	0	25	0	1	0	0	100%
χ	10	0	0	10	0	0	0	0	100%
ħ	58	1	0	41	12	1	0	3	72%
∂	8	1	0	0	5	1	0	1	62%
z	12	0	0	12	0	0	0	0	100%
γ	19	2	0	3	12	1	0	1	63%
ε	59	1	2	0	43	2	0	11	72%
h	29	0	2	0	17	0	0	10	58%
ṣ	22	0	0	22	0	0	0	0	100%
∂.	18	6	0	1	10	0	0	1	55%

Phon	Nb	Plo	Voy	Fri	Aut	FriVoy	PloFri	Omis	Taux
m	117	3	5	0	84	0	0	25	71%
n	161	10	2	0	108	1	0	40	67%
l	205	1	4	26	119	2	0	53	58%
r	112	1	4	2	78	1	0	26	69%
w	49	1	1	0	41	0	0	6	83%
j	65	1	1	19	17	2	0	25	26%

The segmentation results by class are summed up on table 2.

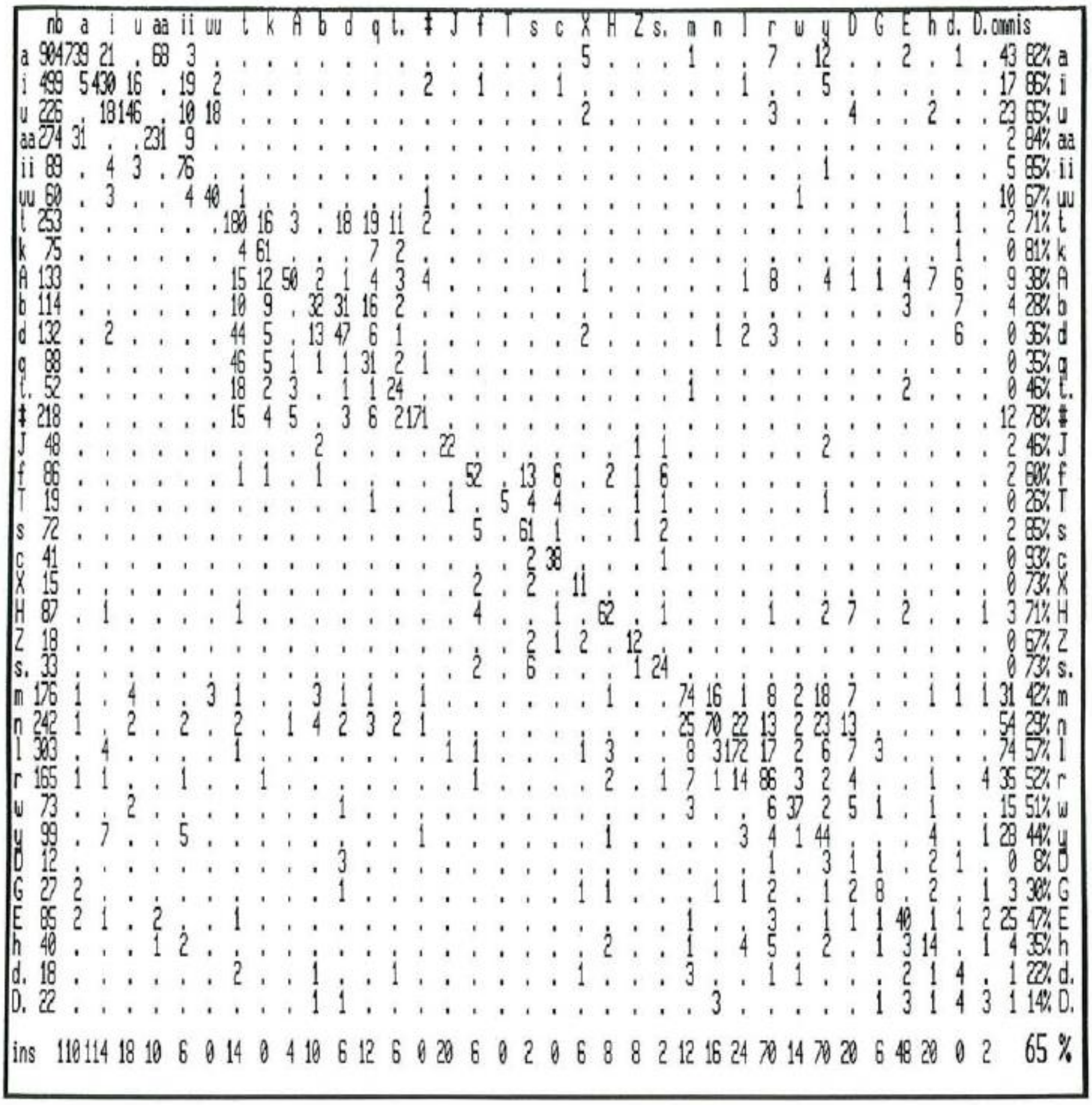
Table 2 : Segmentation results by class

Class	Present	Found	Inserted
Vowels	1392	1254 (90%)	124 (9%)
Plosives	720	672 (93%)	63 (9%)
Fricatives	281	256 (91%)	99 (35%)
Sonnants	852	539 (63%)	193 (21%)

From these results, the following remarks can be made :

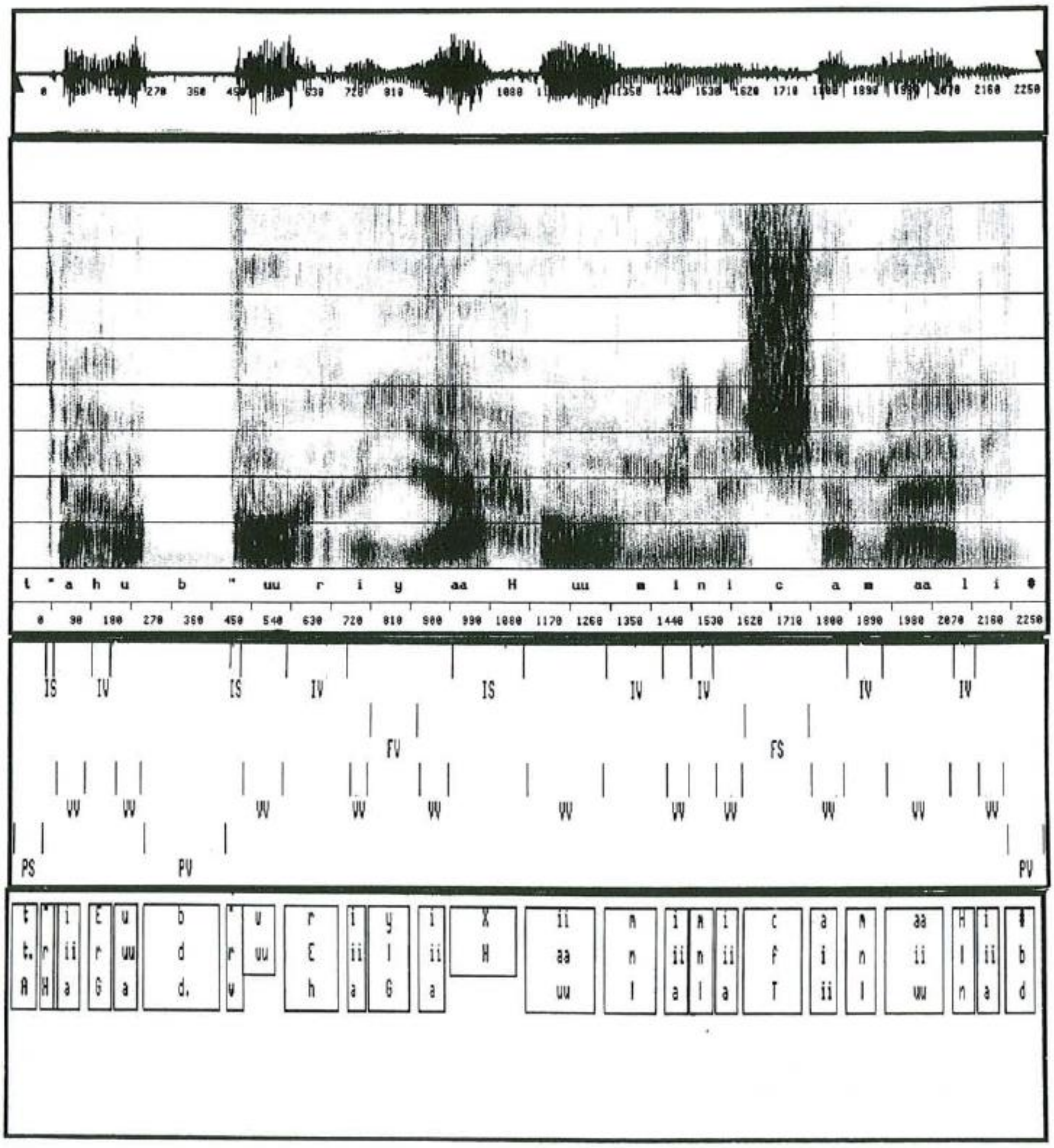
The corpus being balanced, some phonemes are scarcely present, their result not being very representative. An evaluation with other corpus is therefore need to draw conclusions about them.

On the whole, the algorithms provide a good score in the segmentation (over 90%). The vowels are often omitted at the end of the sentence. The most frequent omissions affect the class of sonnants : when two sonnants are adjacent, the system sends back only one segment, the other segment being automatically omitted. The /n/ is often omitted at the end of the sentence.



	nb	a	i	u	aa	ii	uu	t	k	A	b	d	q	t.	‡	J	f	T	s	c	X	H	Z	s.	m	n	l	r	w	y	D	G	E	h	d.	D.	ommis		
a	904	739	21	.	68	3	.	.	.	.	.	.	.	.	.	.	.	.	.	.	5	.	.	.	1	.	.	7	.	12	.	.	2	.	1	.	43	82%	a
i	499	5	430	16	.	19	2	.	.	.	.	.	.	.	2	.	1	.	.	1	.	.	.	.	.	.	1	.	.	5	.	.	.	.	.	.	17	86%	i
u	226	.	18	146	.	10	18	.	.	.	.	.	.	.	.	.	.	.	.	.	2	.	.	.	.	.	.	3	.	.	4	.	.	2	.	.	23	65%	u
aa	274	31	.	.	231	9	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	2	84%	aa
ii	89	.	4	3	.	76	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	5	85%	ii
uu	60	.	3	.	.	4	40	1	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	10	67%	uu
t	253	.	.	.	.	.	.	180	16	3	.	18	19	11	2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	1	.	2	71%	t
k	75	.	.	.	.	.	.	4	61	.	.	.	7	2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	0	81%	k
A	133	.	.	.	.	.	.	15	12	50	2	1	4	3	4	.	.	.	.	.	1	.	.	.	.	.	1	8	.	4	1	1	4	7	6	.	9	38%	A
b	114	.	.	.	.	.	.	10	9	.	32	31	16	2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	3	.	7	.	4	28%	b
d	132	.	2	.	.	.	.	44	5	.	13	47	6	1	.	.	.	.	.	.	2	.	.	.	.	1	2	3	.	.	.	.	.	.	6	.	0	36%	d
q	88	.	.	.	.	.	.	46	5	1	1	1	31	2	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	0	35%	q
t.	52	.	.	.	.	.	.	18	2	3	.	1	1	24	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	2	.	.	.	0	46%	t.
‡	218	.	.	.	.	.	.	15	4	5	.	3	6	2	171	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	12	78%	‡
J	48	.	.	.	.	.	.	.	.	.	2	.	.	.	.	22	.	.	.	.	.	.	1	1	.	.	.	.	.	2	.	.	.	.	.	.	2	46%	J
f	86	.	.	.	.	.	.	1	1	.	1	.	.	.	.	.	52	.	13	6	.	2	1	6	.	.	.	.	.	.	.	.	.	.	.	.	2	60%	f
T	19	.	.	.	.	.	.	.	.	.	.	.	1	.	.	1	.	5	4	4	.	.	1	1	.	.	.	.	.	1	.	.	.	.	.	.	0	26%	T
s	72	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	5	.	61	1	.	.	1	2	.	.	.	.	.	.	.	.	.	.	.	.	2	85%	s
c	41	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	2	38	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	0	93%	c
X	15	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	2	.	2	.	11	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	0	73%	X
H	87	.	1	.	.	.	.	1	.	.	.	.	.	.	.	.	4	.	.	1	.	62	.	1	.	.	.	1	.	2	7	.	2	.	.	1	3	71%	H
Z	18	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	2	1	2	.	12	.	.	.	.	.	.	.	.	.	.	.	.	.	0	67%	Z
s.	33	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	2	.	6	.	.	.	1	24	.	.	.	.	.	.	.	.	.	.	.	.	0	73%	s.
m	176	1	.	4	.	.	3	1	.	.	3	1	1	.	1	.	.	.	.	.	.	1	.	.	74	16	1	8	2	18	7	.	.	1	1	1	31	42%	m
n	242	1	.	2	.	2	.	2	.	1	4	2	3	2	1	.	.	.	.	.	.	.	.	.	25	70	22	13	2	23	13	.	.	.	.	.	54	29%	n
l	303	.	4	.	.	.	.	1	.	.	.	.	.	.	.	1	1	.	.	.	1	3	.	.	8	3	172	17	2	6	7	3	.	.	.	.	74	57%	l
r	165	1	1	.	.	1	.	.	1	.	.	.	.	.	.	.	1	.	.	.	.	2	.	1	7	1	14	86	3	2	4	.	.	1	.	4	35	52%	r
w	73	.	.	2	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	3	.	.	6	37	2	5	1	.	1	.	.	15	51%	w
y	99	.	7	.	.	5	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	1	.	.	.	.	3	4	1	44	.	.	.	4	.	1	28	44%	y
D	12	.	.	.	.	.	.	.	.	.	.	3	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	3	1	1	.	2	1	.	0	8%	D
G	27	2	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	1	1	.	.	.	1	1	2	.	1	2	8	.	2	.	1	3	30%	G
E	85	2	1	.	2	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	3	.	1	1	1	40	1	1	2	25	47%	E
h	40	.	.	.	1	2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	2	.	.	1	.	4	5	.	2	.	1	3	14	.	1	4	35%	h
d.	18	.	.	.	.	.	.	2	.	.	1	.	.	1	.	.	.	.	.	.	1	.	.	.	3	.	.	1	1	.	.	.	2	1	4	.	1	22%	d.
D.	22	.	.	.	.	.	.	.	.	.	1	1	.	.	.	.	.	.	.	.	.	.	.	.	.	3	.	.	.	.	.	1	3	1	4	3	1	14%	D.
ins		110	114	18	10	6	0	14	0	4	10	6	12	6	0	20	6	0	2	0	6	8	8	2	12	16	24	70	14	70	20	6	48	20	0	2		65 %	

Table 3 : Decoding result



Fi 1 : Example of decoding

Sentence : translation of "The winds blow from the North"

The insertions are relatively few in the classes of vowels and plosives. On the other hand, they are much more important in the fricatives and the sonnants. Phonemes responsible for the insertions in the class of fricatives are /y/ and to lesser degree the /l/.

The /y/ is present at more than 50% in the class of fricatives. The insertions in the class of vowels are due to the consonants having formantic structure. The /n/ may present an occlusion, it is thus segmented as a plosive. The /?/ in intervocalic surroundings presents weak formants, it is, therefore, segmented as a sonnant.

3.2 Recognition results

The result of the phonetic labelling is given by the confusion matrix of table 3. The evaluation is made on DJOUMA corpus manually labelled for 3 male speakers. For each segment, we kept the best three labels. Besides, on figure 1, we present an example of a sentence which has been automatically labelled by the system

4 CONCLUSION

In this paper, we have presented a first assessment of an acoustic phonetic decoding of standard Arabic. Our objective in this work is to use as well as the approach based on knowledge new methods of acoustic decoding in order to improve the global rate of recognition and to integrate the phonetic SAPHA decoder in a recognition system of spoken Arabic language.

REFERENCES

[1] S. H. Al Ani Arabic Phonology, an Acoustical and Physiological investigation. Mouton the Hague, 1970.

[2] N. Carbonnel, J.P. Haton, D. Fohr, F. Lonchamp and J.M. Pierrel. APHODEX, Design and Implementation of an Acoustic Phonetic Decoding Expert System. Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing, Tokyo, 1986.

[3] M. Djoudi, J.P. Haton. The SAPHA Acoustic Phonetic Decoder System for Standard Arabic. Proceedings of 1990 International Conference on Spoken Language Processing, Kobe (Japan), 1990.

[4] M. Djoudi, H. Aouizerat, J.P. Haton. Phonetic Study and Recognition of Standard Emphatic Consonants. Proceedings of 1990 International Conference on Spoken Language Processing, Kobe (Japan), 1990.

[5] M. Djoudi. Utilisation des techniques d'intelligence artificielle pour le décodage acoustico-phonétique de l'Arabe Standard. Proceedings of First Maghrebin Symposium on Programming ans Systems, Algiers 1991.

[6] M. Djoudi Contribution à l’étude et à la reconnaissance automatique de la parole en Arabe standard. Doctorat de l’Université de Nancy 1, November 1991.