

Acoustical study of tachdid and idgham in Arabic Standard. Application for speech synthesis.

Kamel FERRAT

*Scientific and technical research center for the
development of the Arabic language (C.R.S.T.D.L.A)
1, rue Djamel Eddine al-Afghani – Bouzareah
ALGIERS ALGERIA
Phone/fax : (213) (021) 94-12-38*

E-mail: kferrat@wissal.dz

Abstract : This work contributes to the modelisation of the "tachdid" and "Idgham" in Standard Arabic language, in order to synthesize speech using the Mbrola method and to reduce the VC₁V (Vowels, Consonant, Vowels) units of the dictionary in a speech synthesis. The number of pre-stored units can be diminished to the half thanks to the modelisation of the V C_gV, where C_g represents the geminate consonant C. This may be possible by a modification of some acoustical parameters of the corresponding wave's form of the VCV units.

We have judged useful to study the modifications undergone by the curve of the energy, the temporal durations and the formantic values (features) of the subsequent vowels using the Computing Speech Laboratory CSL of Kay Elemetrics and the Praat and Matlab Windows Programs. For the application, we have utilized the Mbrola method (Multi Band Re-synthesis Overlap Add) which is one of the most efficient to have the natural of speech.

Keywords: Arabic speech synthesis, Mbrola, gemination, formants, energy, tachdid, idgham.

1 Introduction

The " tachdid ", indicated in Standard Arabic Language by the sign (ˆ) called *cheddah* , is often confused wrongly or rightly with the French term "*gemination*". This characteristic is one of the most significant phenomenon characterising the Standard Arabic language. All the consonants of this language can be geminated except for the glottal consonant [*Hamza*].

Various works have been realized to show the variations that the organs of the vocal tract undergo at the time of the gemination and the resulting acoustic effects. It remains that the majority of this works doesn't converge all toward the same results. Some controversies blemished the study, notably those relating to the modifications of the organ movements and the resulting acoustic effects undergo.

This phenomenon which presents only one simple means of expression in much of languages like French, is very pertinent in Arabic. For example, the pronunciation of the sentence [h.ad.ara eddarsa] (he attended the course) defers the sense of the pronunciation of the sentence [h.ad.d.ara eddarsa] (he prepared the course). (t. and d. explain that it concern

an emphatic consonants).

This modest work aims to explain some acoustical effects, resulting of the pronunciation of a geminate consonant by opposition to its non geminate correspondent in a same context and its possible application in Arabic Speech. We studied the modifications that the temporal durations and the formantic values of the subsequent vowels undergo. The shape (allure) of the energy curve during the pronunciation of phoneme has been compared to the shape (allure) of its geminate correspondent in the same context.

We applied the found models to synthesize the Standard Arabic speech using the Mbrola method. Many researchers specified the interest to develop systems of synthesis capable to produce multilingual speech. The integration of the lingual variations within a same synthesiser appears to us essential, so that the systems of synthesis are of an universal range. The MBROLA project, initiated in 1992 by the Polytechnic Faculty of Mons (Belgium) answers to this preoccupation. It proposes an algorithm of synthesis using the concatenation of diphones.

2 The tachdid in Arabic language

Various definitions have been given to the notion of tachdid in Arabic Standard. We can mention the definition of Sibawayh, which is very old but always of actuality, quoted by A. Roman in his Arabian phonology study. For Sibawayh, "It's heavy to employ its language with leaving a place of articulation for making there at once return. Also, because of this tiredness which the realisation of two identical articulations brings, this realisation is rejected in favour of the gemination of two identical [harf] so it has there that only one elevation in the language" (Roman, 1983).

According to a study made by Dkhissi-Boff "the geminate consonant doesn't present two distinct articulatory movements, but only one single movement, which differs from that of the simple consonant, by its great articulatory stability and its very significant duration" (Dkhissi-Boff, 1983).

For P. Delattre, the articulation of geminate achieves itself in two phases and presents two summits of activity. On the other way, for Rousselot, J Cantineau, J.F Bonnot and M.C. Dkhissi-Boff, the geminate consonants are achieved in only one phase of great articulatory stability in which, the duration is definitely affected. Other works explain that one duration of the preceding vowel inversely proportional to the articulatory strength of the subsequent consonant (Attaoui, 1993).

Generally, the *tachdid* is defined as phenomenon of reinforcement of the consonant articulation which prolongs the duration of it approximately half and increases the intensity. This phenomenon is sometimes called *dedoublement*, although there is not truly repetition of the consonant.

A consonant can also be assimilated in the pronunciation by a different consonant that follows it, provoking a gemination (in Arabic, it concerns the notion of "idgham"). For example: because of articulatory heaviness, [man yakun] is pronounced [mayyakun].

According to Ozza Obaid Deass, three sorts of idgham exist in Arabic (Ozza Obaid, 1991). When the subsequent phonemes:

- have the same mode and point of articulatory. Example: "qad daxalu" (اقد دخلو) that becomes in the pronunciation "qaddahalu", and also "bal la yaxafun" (بل لا يخافون) pronounced "balla yaxafun";
- have the same point of articulatory, but differs in the mode. Example: "qad tabayyana" that is pronounced "qattabayyana";
- have a nearly mode and point of articulatory. Example: « bal rafaEahu » (بل رفعه) pronounced «barrafaEahu ».

3 Acoustic study

We used for the acoustic study, the module CSL

(Computer Speech Laboratory) for Windows of Kay Elemetrics (figure 1 and 2), and the software of analysis and treatment Matlab and Praat.

The study was carried out starting from seven (7) different records from two masculine and feminine speakers, researchers in the CRSTDLA (Algiers). The feminine speaker is aged of 32 years and work in the laboratory of didactics. She has a licence in language and Arabian literature. The masculine speaker is aged of 36 years, member of the speech communication laboratory. We have recorded the values of the formants of the following vowels (V_f) during the transitions, the temporal durations (in milliseconds) of the preceding vowel (V_p), the following vowel (V_f), the geminate consonant (C_g) and its opposite not geminate (C).

3.1 The duration Study

Some average durations of the subsequent vowels and the geminate / not geminate consonants are illustrated in the table 1.

duration (ms) phonemes	V_p	V_f	d
[d.]	70.5	97.4	61.0
[d.d.]	48.4	119.2	166.4
[b]	68.2	71.5	74.8
[bb]	53.8	86.5	211.7
[d]	74.8	67.0	96.1
[dd]	70.4	75.1	204.4
[x]	80.0	91.7	109.0
[xx]	68.1	103.2	197.4
[k]	75.7	114.2	108.4
[kk]	59.2	125.4	203.1
[t]	82.0	95.1	111.4
[tt]	68.7	110.4	231.7
[y]	67.8	75.5	116.7
[yy]	55.0	79.7	213.1

Table 1. The average durations of the geminate /not geminate consonants and the subsequent vowels.

While putting respectively :

$D_1 = V_{p2} / V_{p1}$, where V_{p2} and V_{p1} are respectively the durations of vowels which precedes C_g and C.

$D_2 = V_{f2} / V_{f1}$, where V_{p2} and V_{p1} are respectively the durations of vowels which follows C_g and C.

$D_3 = D_2 / D_1$, where D_2 and D_1 are respectively the durations of the consonants C_g and C.

We can obtain the table 2 of coefficients of the durations ratios :

coeffs phonemes	D ₁	D ₂	D ₃
[d.]	0.68	1.22	2.72
[b]	0.79	1.21	2.83
[d]	0.94	1.12	2.12
[x]	0.85	1.12	1.81
[k]	0.78	1.10	1.87
[t]	0.84	1.16	2.08
assimilation	0.81	1.05	1.83

Table 2. Coefficients of the durations ratios

By analysing these results, we have noted the following remarks:

- The durations of the geminate consonant are important compared to its opposite not geminate;
- A larger duration of the following vowel on the presence of gemination ;
- A fall duration of the preceding vowel on the presence of gemination.

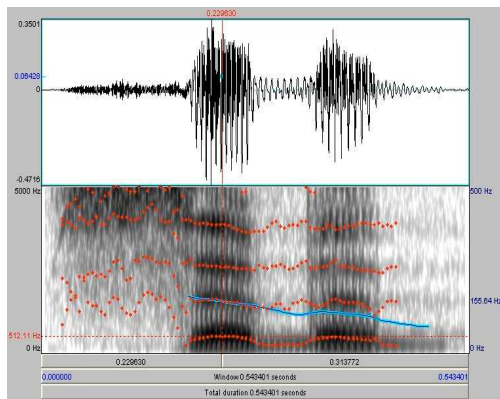


Figure 1. Representation spectrographic of the Arabic word [sabab]

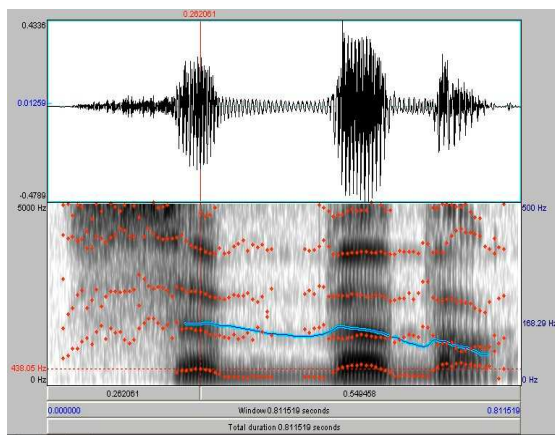


Figure 2. Representation spectrographic of the Arabic word [sabbaba]

3.2 The energy Study

In the figure 3, we presents example aspects of the energy curve during the occlusion phase (silence) of the stop phonemes compared to its geminates correspondents

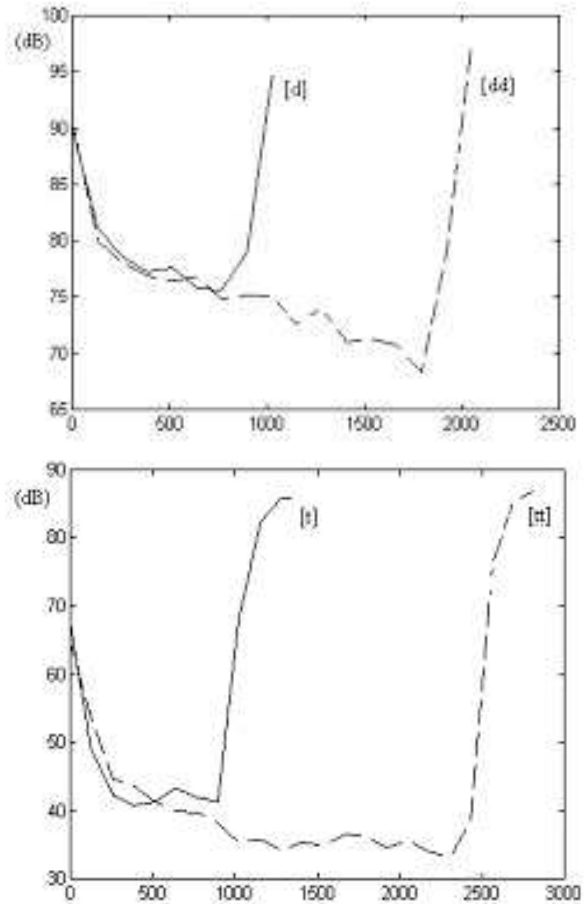
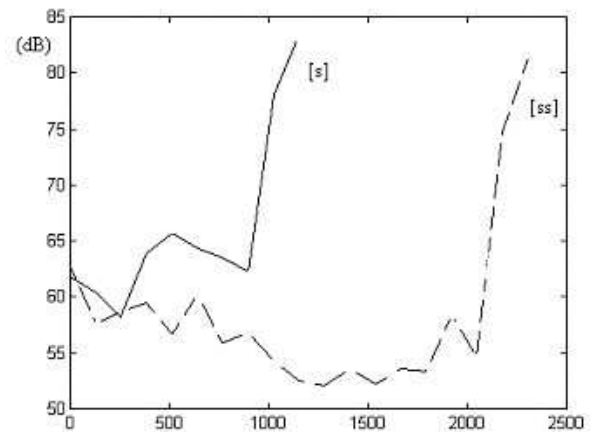


Figure 3. Representation of the energy during the occlusion of the stop phonemes

In the figure 4, we presents example aspects of the energy curve during the pronunciation of the fricative phonemes compared to its geminates correspondents



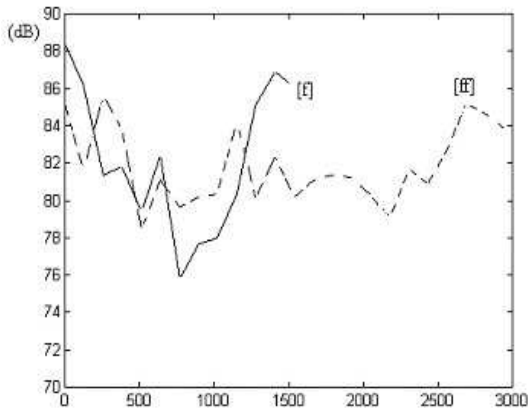


Figure 4. Representation of the energy during the pronunciation of the fricative phonemes

The study of the various curves of energy shows us that:

- The duration of occlusion of the geminate consonant is much more important than its opposite not geminate;
- The curve of amplitude of energy continues to go down because of lengthening of the duration of the occlusion (silence) of the occlusive consonant.

We can nevertheless note that the spectrum of the geminate consonant presented, in practically all the studied cases, energy uniformly repartee and without notable discontinuity. In this direction, it is difficult to separate the end from the first consonant and the beginning of the second consonant.

It converges towards the direction of the thesis of the realization of the geminate consonant in only one phase with a great articulatory stability.

3.3 Study of the formantic values

- The values of the F1 and F2 in the level formantic transitions of the following vowel slightly falling, contrary to the value of F3 witch is stable or slightly rising. For the preceding vowel, we also notice a falling of F1, in presence of the gemination. In a general way, we note that in contact of a geminate consonant, the transitions and the stable parts of the formants F1 et F2 of the subsequent vowel are lower, notably for the formant F1 (figure 5 and 6)..

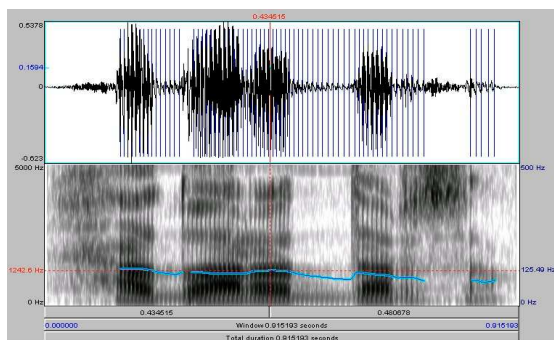


Figure 5. Spectrographic representations of the sentence

(*h.ad.ara eddarsa*)

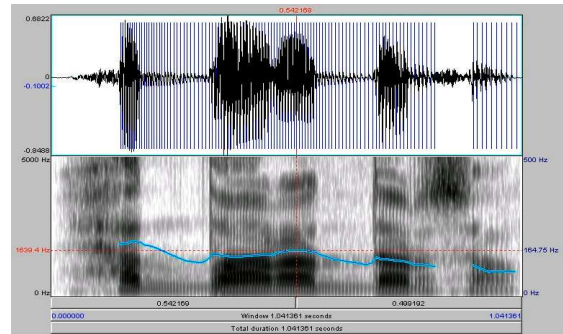


Figure 6. Spectrographic representations of the sentence (*h.ad.d.ara eddarsa*)

For the assimilation or “Idgham”, we present the following values :

[man yaku:n] : the time of the emission of the consonant [y] : 0.065s approximately, values of the formants of the subsequent vowels of [y] : F1=520Hz, F2=2240Hz and F3=2640Hz (figure 7).

[mayyaku:n] : the time of the emission of the geminate consonant [yy] = 0.150s, values of the formants of the vowel which follows the geminate [yy]: F1=440Hz, F2=2200Hz and F3=2680Hz (figure 8).

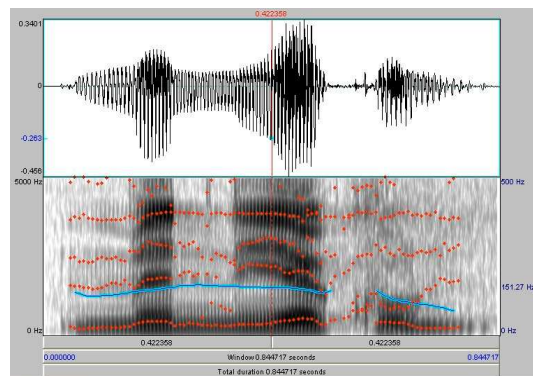


Figure 7. Representation spectrographic of the word [man yaku:n]

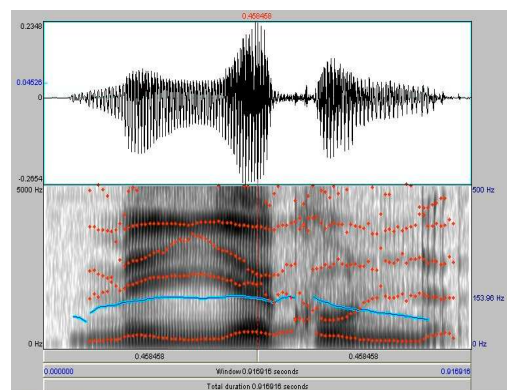


Figure 8. Representation spectrographic of the word [mayyaku:n]

5 Application of the rules of the tachdid and the idgham to synthesize speech in Standard Arabic, using the Mbrola method

5.1 The Mbrola method

MBROLA (Multi Band Resynthesis Overlap Add) is a speech synthesis method using the concatenation of the sounds (synthesis from pre-recorded units) (Dutoit & al., 1998). The main advantage of a speech synthesis with Mbrola is that it permits:

- The smoothing of the spectral discontinuities appearing of part and others of the points of concatenation (problem that we meet in the automatic reading of texts from the Arphon system of the speech laboratory of the CRSTDLA) ;
- A very fluid speech synthesis;
- A capacity memory of the dictionary units appreciably reduced.

The Mbrola module receives the informations above in entry, as a file format "file.pho". It generates then in exit, an audio file "file.wav". The principle of Mbrola is to write a sentence in phonetic symbols and to provide the following informations, using the Mbrolign program :

- The pauses (comma, point, semicolon,...) ;
- The duration of every phoneme;
- The movement of the pitch F_0 (beginning of the pitch in relation to the duration of the phoneme and its value).

5.2 application for Arphon system (CRSTDLA)

The ARPHON system, finalized in the speech laboratory of the CRSTDLA, uses a dictionary of 3760 variable size units (polysounds) (Benbellil & all, 2002). The size of the dictionary is 18Mo. What is very important, notably for a possible demonstration reader's on the Web site of the center. It is therefore more that necessary to compress this data base for its exploitation from the server of the application. The modelling of the tachdid and the idgham answers this need while decreasing this data base considerably and contribute to a better security of the data especially since the rules will be integrated directly in the reader's program and therefore its executable.

These rules once clarifications in the different contexts are applied to the Arphon automatic reader of the CRSTDLA. The different stages to generate our speech synthesis are :

- The generation of the words from the Arphon automatic reader;
- Their resynthesize (passage through the Mbrola module) permitting to modify the words by a better improvement of the natural, modifying the applicable parameters that are the duration and the pitch (while respecting the gotten models of the acoustic analysis).

The quality of the gotten speech is good (figure 9

and 10). The audible discontinuities through the Arphon automatic reader are eliminated thanks to the contribution of these parameters. Some following examples show the gotten results after resynthesis of the speech by Mbrola.

```
; .pho file created with MBROLIGN v1.0
; Ferrat Kamel, kferrat@wissal.dz,
; resynthesis of the word "h.ad.ara"
; generated by the Arphon reader
_46
X 60
a 71 13 135 51 140
d. 61 14 140 70 135 98 130
a. 97 57 130 77 128 96 125
r 33 14 121
a 93 14 102 55 98 95 85
_44
```

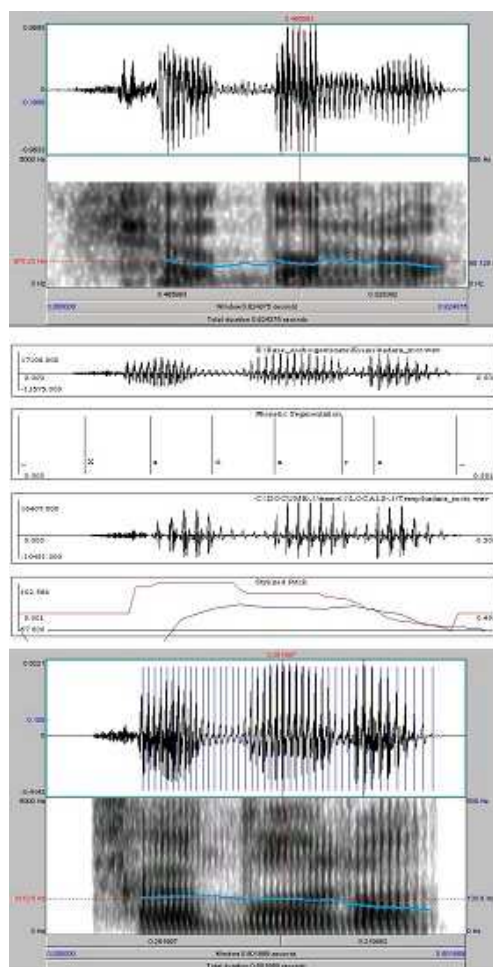


Figure 9. Resynthesis by Mbrolign of the word [h.ad.ara] generated by Arphon

```
; .pho file created with MBROLIGN v1.0
; Ferrat Kamel, kferrat@wissal.dz
; resynthesis of the word "h.ad.d.ara"
; generated by the Arphon reader
_46
X 60
a 48 13 135 51 140
d. 166 14 140 70 135 98 130
```

a. 119 57 130 77 128 96 125
 r 33 14 121
 a 93 14 102 55 98 95 85
 _ 44

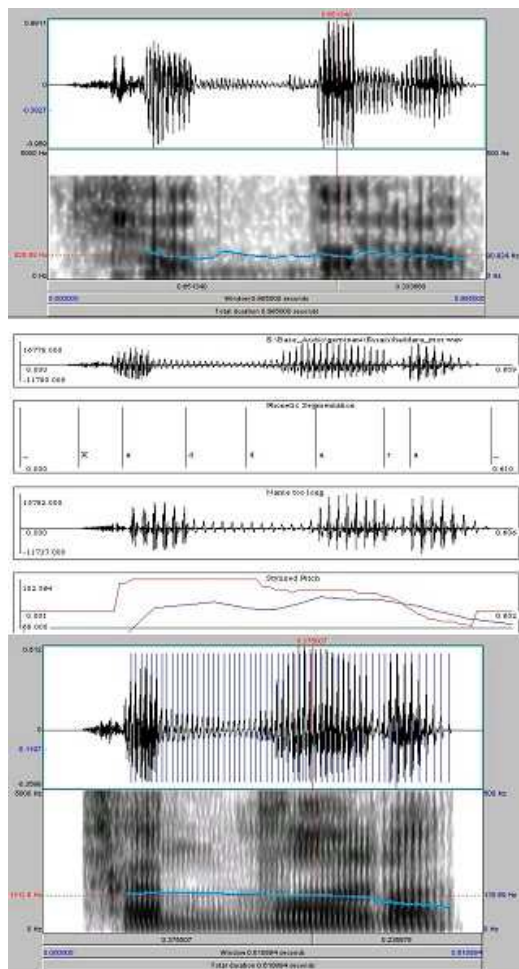


Figure 10. Resynthesis by Mbrologn of the word [h.ad.d.ara] generated by Arphon.

Conclusion

This experience showed us that the contribution of the MBROLA method in the Standard Arabic speech synthesis is very interesting, especially when we re-synthesis the word. The insertion of the Mbrologn module, in the program of the Arphon automatic reader, proves to be very interesting seen the quality of the gotten speech resynthesis. Besides, this work showed us that by an adequate modelling of the geminate phonemes in the different contexts, we can get a speech synthesis without having to resort to the conception of the VC_gV units in the sound units dictionary of the Arphon automatic reader. What will permit us to reduce the size of this dictionary considerably. This reduction is much important that we project to put a demonstration of the Arphon reader on the Web site of the research center CRSTDLA (Algiers).

References

- M. Attaoui, 1993, *Force articulatoire et gémation en arabe marocain de Fès*. Travaux de l'Institut de Phonétique de Strasbourg, T.I.P.S. 23, 1993, Page 22 ;
- K. Benbellil & all., 2002, *Conception et réalisation d'un lecteur automatique de textes en Arabe Standard. Le système Arphon*, In séminaire international sur l'automatisation de la langue arabe, organisé par l'Académie de langue arabe de Aman (Jordanie).
- M.C. Dkhissi-Boff, 1983, *Contribution à l'étude expérimentale des consonnes de l'arabe*. Travaux de l'Institut de Phonétique de Strasbourg, N° 15 ;
- T. Dutoit & al., 1998, *The Mbrologn Project : Towards a set of high quality speech synthesizers free of use for non commercial purposes*. Faculté Polytechnique de Mons, TCTS Lab, 31 Bvd Dolez, B-7000 Mons, Belgium, Email: mbrologn@tcts.fpms.ac.be ;D. Ozza Obaid, 1991, *Fan attadjouid*. Dar Ibn Hazm, Beyrouth, Liban ;
- A. Roman, 1983, *Etude de la Phonologie et de la Morphologie de la Koinè Arabe*. Tome I, page 351, Université de Provence, France ;