New Approach to Modeling Micromelodic Effects

Amina Chentir[1], Daniel J. Hirst[2], Mhania Guerti[3]

[1]*Department of Electronics, LATSI Laboratory*
Saâd Dahlab University, Blida, Algeria
chentiramina@yahoo.fr

[2]*Laboratoire Parole et Langage*
Université de Provence , Aix En Provence, France
Daniel.hirst@lpl-aix.fr

[3]*Department of Electronics*
National Polytechnic College (ENP), Algiers, Algeria
mhaniag@yahoo.fr

Abstract- **In this paper, we describe a new approach, allowing as much as possible, an automatic extraction of the micromelodic information of the speech signal using the original curved of the fundamental frequency and its macromelodic curve obtained by MOMEL, the algorithm of MOdelisation MELodic. The results reinforce the idea that microprosodic effect exists and seems to be possible to be included in a prosodic generating unit by a simple model to provide improved naturalness in speech synthesis.**

Keywords- **MOMEL, Micromelody, Fundamental Frequency, Arabic Standard.**

I. INTRODUCTION

In order to extract the micromelodic effect for possible improvement of naturalness of synthesized speech, we present in this paper, a new method for modeling the micromelodic effect in Standard Arabic (SA), using the software speech analysis PRAAT [1] and MOMEL algorithm [2].

From the acoustic point of view, prosody refers to the phenomena linked to the variation in the time of the parameters of pitch, intensity and duration. The perception of pitch is essentially related to fundamental frequency F_0 which, at the physiological level of the production of the speech, corresponds to the frequency of vibration of the vocal cords [3].

The algorithm of melodic modelisation gives a representation of the melodic curve, which characterises the temporal variations of the laryngeal frequency, by the way of a quadratic spline function.

F_0 variations can be considered as the superposition of two phenomena: the macroprosodic effects which can be considered as the elocution intonative choice and microprosodic effects, which are linked to the phonetic constituents of the sentence. The macroprosody allows to apply a global approach of the melodic curve when the microprosody gives local variations. It's often the case with consonants like [f], [s], [ch] that are most of the time unvoiced. For that reason, the F_0 curve will reach discontinuities with such consonants.

Although for several years, synthetics speech has been fully intelligible from a segmental perspective, there are areas of naturalness which still await satisfactory implementation. High quality text-to-speech synthesis systems require accurate prosody labels to generate natural-sounding speech. In these systems, prosody is assigned based on information extracted from text [4].

Developed in 1993, MOMEL algorithm transforms the dashed curve of the designed fundamental frequency as a continuous curve. The purpose of the algorithm is to separate the macroprosodic component from the microprosodic component, rejected as linguistically irrelevant.

II. MELODIC MODELISATION: MOMEL

The MOMEL Algorithm lies on the acceptation that the melodic curve can be, by pieces, approximated with a second degree polynomial. A moving window of length A (Typically 300ms) covers the acoustic signal. In each window, the F_0 curve is calculated, and an approximation can be given by such a polynomial, with the only purpose to minimise the quadratic error between the initial curve and the polynomial. Initial points that are more than 5% below the polynomial are set to zero for the approximation process (That is the microprosoy filter). A polynomial is then recalculated with the remaining points, and so on until no new points are set to zero [5].

Then, the resulting polynomial can be considered as the best second degree representation, so its vertex is calculated and saved as a candidate. The following step consists of the extraction of the target points from those candidates.

To reach that goal, the time domain must be sliced (divided) to separate the candidates from each other (the R value must be chosen in order to be adapted to the speaker's speech rate). On each partition, candidate averages are calculated, and also the typical deviation (in time and in frequency). Values that are outside the acceptation area (centered on the average and delimited by the typical deviation) are deleted, and the average is recalculated with the kept points. This average gives a target point.

The process is done for each partition in order to obtain a sufficient number of target points to well approximate the F0 curve with a quadratic function. That is the last step of the algorithm: to fit the target points with a quadratic spline function [5].

The following parameters can be changed when the program is called:

H_{zmin}: minimal accepted F_0 value
H_{zmax}: maximal accepted F_0 value
A: size of the initial analysis window (default 300ms)
Delta: maximal accepted error percentage for the polynomial approximation
R: size of the second window for the partitions choice (default 200ms)

The following figure represents an F_0 curve on which candidates have been put (black points), as well as target points (green circle) and the MOMEL quadratic approximation, in red.

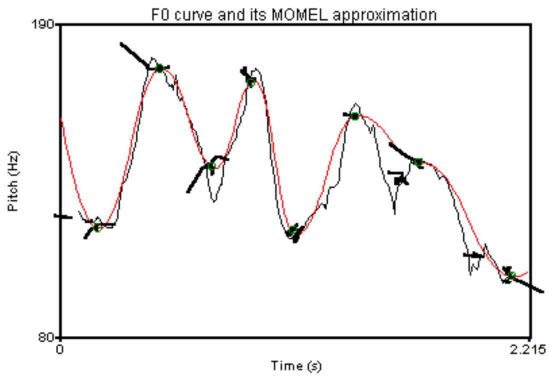

Fig.1 F_0 curve and its MOMEL melodic stylisation

III. Method of Extraction of Micromelody Effect (EME)

From the moment the researchers agree to recognize the existence of intrinsic variation and co-intrinsic, it seems unwise to dismiss them. We will seek to assess their significance in Standard Arabic, then in order to judge the appropriateness of retaining them or not.

Granted that microprosodic effects resulting from constraints production are relatively stable, it is possible to separate the macroprosodic component (which reflects changes in F_0 intentionally motivated) through an appropriate method of calculus. Thus the use of MOMEL [2] allows extracting from the F_0 raw curve, the suprasegmental profile resulting from macroprosodic component.

The method that we present below will allows us to retrieve the micromelodic information that has been filtered when running MOMEL algorithm on real melodic curve, for the case of voiced consonants in SA.

A. Corpus and Materials

One native Arabic-speaker pronounced 16 sentences including all Arabic phonemes (Table 1). The recording was made in an anechoic recording chamber in the Laboratory Parole et Langage (LPL) in Aix-en-Provence. The Praat computer program [1] was then used to analyse and manipulate the speech data. Sentences are then, segmented and aligned semi-automatically in phonemes and at the end, a Phonetic Transcription is made.

TABLE1
SENTENCES USED IN THE CORPUS

Sentences	
ذَهَبَتْ جَارَةُ أَحْمَدِ إِلَى الجَزَائِرِ.	الطِفْلُ الذَكِيُّ شَرِبَ الحَلِيبَ اللَّذِيذَ.
وَجَدَ شَعْبَانُ خَرُوفًا.	أَضَاعَتْ صَابِرَةُ نَظَارَةً فَوْقَ المِنْضَدَةِ.
أُمُّ البَطَلِ عُمَرْ رَافَقَتْهُ فِي سَفَرِهِ مِنْ فَاسْ إِلَى تُونِسْ.	قَرَأَتْ زَيْنَبُ عِشْرِينَ كِتَابًا خِلاَلَ العُطْلَةِ.
سُكَانُ المَغْرِبِ أَكْثَرُ عَدَدٍ مِنْ سُكَانِ الجَزَائِرِ بِكَثِيرٍ.	أَحْمَدُ يَأْكُلُ دَائِمًا فِي نَفْسِ المَطْعَمِ.
وَزَّعَ سَاعِي البَرِيدِ ظَرْفًا وَاحِدًا اليَوْمَ.	يَعْتَبِرُ جَمَالُ أُسْتَاذَ اللُّغَةِ شَخْصًا غَيْرَ مَرْغُوبٍ فِيهِ.
وَضَعَ الوَلَدُ الصَّغِيرُ مِحْفَظَتَهُ فَوْقَ الطَاوِلَةِ	شَاهَدَ خَالِدٌ السَيَارَةَ المَسْرُوقَةَ أَمَامَ مَكْتَبِ البَرِيدِ.
تَتَسَبَبُ الغَازَاتُ المُنْبَعِثَةُ مِنَ المَصَانِعِ فِي تَلَوُثِ الهَوَاءِ.	قَابَلَ مُصْطَفَى المُرَاسِلَ نَفْسَهُ.
شَارَكَتْ خَوْلَةُ وَ خَدِيجَةُ فِي الحَفْلِ.	وَضَعَ ثَابِرُ العِقْدَ حَوْلَ جِسْمِ خَرُوفِهِ.

B. EME Methodology

To achieve the automatic Extraction of the Micromelodic Effect of different voiced consonants present in our corpus, we followed the following approach [6]:

Step 1: We extract the melodic curves of our corpus by running MOMEL algorithm.

Step 2: We then proceed to a manual correction of the melodic curves modelled by MOMEL.

Step 3: We then execute our 1st PRAAT script, which allows us to determine the micromelodic profile looked for.

Step 4: We then execute our 2nd PRAAT script, which we developed to model micromelodic profile for every used consonant.

Step 5: We then transfer calculated values towards the Excel software in order to calculate the median values and to plot the various corresponding curves.

1) 1st Script: Creation of Micromelodic Profile

In order to extract the micromelodic effect, we carried out, for each file, already treated by MOMEL, the various following operations:

- Reading maximum and minimal values of F_0, already recorded in their respective files
- Reading real values of F_0 recorded in the file with extension *.Hz
- Reading the target values calculated by Momel from the file with extension *.PitcTier
- Realize a quadratic interpolation of the reading target values
- Deduction of the corresponding values of F_0 (Momel_F0)
- Calculation of micromelodic effect thanks to the ratio: micromelody = F_0 / MomelF_0
- Recording results in a file with extension *.mpp

2) 2nd Script: Modeling a Micromelodic Profile of a Consonant

In order to model with precision the micromelodic evolution's effect of a consonant, we adopted the following approach [7]:

Using only the structure [VCV], where each consonant lies between two vowels, we carry out to extract measurements of microprosodic effect of F_0 at the following points (Fig. 2). With: C = Consonant, V = Vowel,

Duration = Consonant duration,
S = Start, M = Middle, E = End,
1= S - Δ, 2 = S + Δ,
3 = Duration *25%,
4 = Duration*50%,
5 = Duration *75 %,
6 = E - Δ, 7 = E+Δ

Where, Δ is a positive parameter defined by the user.

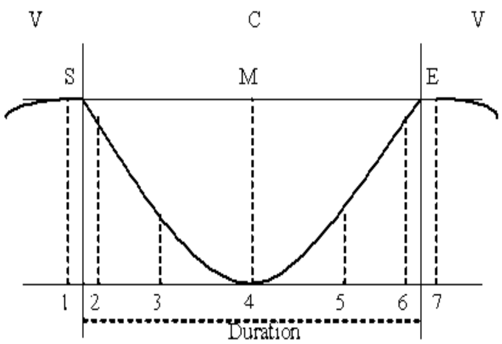

Fig. 2 Simplistic Representation of points retained to estimate the micromelodic variations of F_0

We thus proceeded, for each file already treated by the 1st script, to the execution of the 2nd script, according to the following steps:

- Reading sound file (*.wav)
- Reading the corresponding phonetic transcription file (*.Text Grid)
- Reading the corresponding micromelodic profile file (*.mpp)
- Detection of the required consonant
- Detection of corresponding points S and E
- Calculation of the detected consonant's duration
- Calculation of times corresponding to the 7 points defined on Fig. 2
- Deduct the micromelodic values corresponding to the 7 calculated points
- Record obtained results in a file (*. dat) in order to exploit them later by Excel

IV. Results

We have opted for calculation of the median value of each of the 7 corresponding points in micromelodic profile. This choice is justified by the fact that contrary to the arithmetic mean which is considered as an average of size, the median is rather considered as an average of position and it is not influenced by the extreme values (possibly very large or very small).

Once the calculation of the median made, we proceed to the layout of corresponding profile (we don't take into accounts both extreme values, S - Δ et E + Δ).

Fig. 3 shows the micromelodic profile evolution of phoneme [b] (case of 5 possible values of results) where x-axis represents the corresponding selected points and y-axis represents the mpp: the ratio micromelody = F_0 / Momel_$F_{0.}$ We note that all curves generally, follow the same trajectory.

Median values obtained from the micromelody evolution of phoneme [b] are presented on Fig. 4:

- We distinguished that there is a variation of the micromelodic curve, but this last one is very weak (about 0.045)
- We also noticed that maximum variation (for the most voiced studied consonants) occurred at the level of the M point
- This variation becomes almost nil for certain consonants. For the liquid consonant [n], we obtain the following results (Fig. 5) where variations are between 1.003 and 1.011 (an area of about 0.008). This also allows us to affirm that in the case of this nasal, the micromelodic effect is almost absent (mpp ≈ 1).

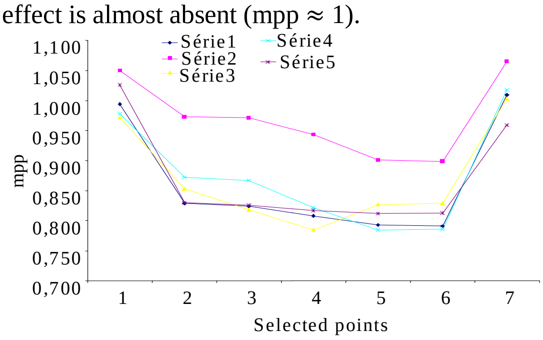

Fig. 3 Evolution of Micromelodic Profile of Phoneme [b]

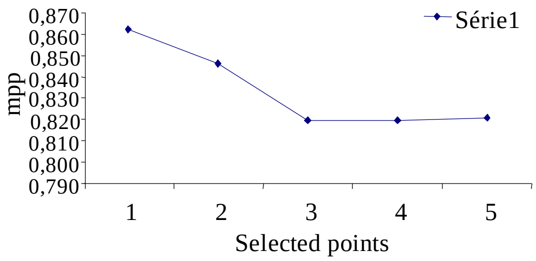

Fig. 4 Median Values of Micromelodic Profile for Phoneme [b]

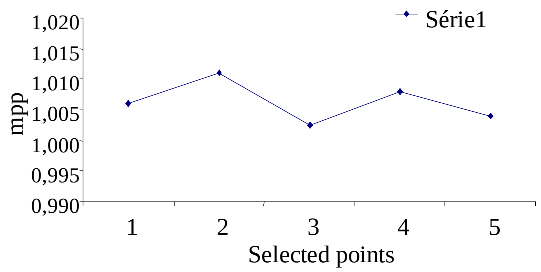

Fig. 5 Median Values of Micromelodic Profile for Phoneme [n]

V. Discussion

Table 2 gives the global median value for each studied voiced Arabic consonants.

From Table 2, we deduce several observations:

- Calculated Median values lie between 0.81 and 1.00. It implies that the MOMEL's modeled frequency approaches very strongly the real value of fundamental frequency F_0
- The micromelodic effect is almost non-existent for the case of nasals, semivowels and the liquid consonants.
- The micromelodic effect is more evident for the fricative than for the occlusive consonants.
- The micromelodic effect exists certainly but is very weak, which does not require a complex mathematical expression to model its variation.

Table 2

Median Values of Arabic voiced phonemes

Phoneme	Median	Phoneme	Median
ب /b/	0.85	م /m/	1.00
د /d̪/	0.83	ن /n/	1.00
ض /ð̣/	0.81	ل /l/	0.99
ز /z/	0.95	ج /ʤ	0.87
ذ /ð/	0.86	ر /r/	0.96
غ /ɣ/	0.95	و /w/	0.99
ع /ʕ/	0.93	ي /j/	0.96
ظ /z̪/	0.95		

The obtained results come to strengthen the theory of several researchers who maintain the idea that the micromelodic effect can be very well neglected, what affects not at all the good quality of the corresponding speech synthesis.

VI. Conclusion

In this paper, we present a new method which consists to extract as automatic as possible, the micromelodic information.

In order to achieve the extraction of this micromelodic effect, we exploited the real curve of the fundamental frequency and the macormelodic curve obtained by MOMEL algorithm

The relationship between these two calculated values represents the micromelodic effect that we look for and tried to model appearance by taking a few and very specific points.

It is clear that micromelody effect exists but is very low down for voiced consonants of Standard Arabic. If we wish to improve the most simply possible naturalness of the synthesized voice, the analysis pushes us to propose simply, only one additional relative lowering of the macromelodic curve,.

The obtained results are to be tested on more important corpuses of Arabic and complementary studies concerning microprosodic effect of duration and energy are to be envisaged to complete our analysis.

References

[1] P. Boersma and D. Weenink, *Praat: Doing phonetics by computer* (Version 5.2.06) [Computer program], December 2010. http://praat.org

[2] D.J. Hirst and R. Espesser, "*Automatic modelling of fundamental frequency using a quadratic spline function*". Travaux de l'Institut de Phonétique d'Aix, 1993, vol. 15, pp. 71-85.

[3] A. Monaghan, "*Prosody in Synthesis Speech: problems, solutions and challenges*". In Keller, Vailly, Terken & Huckvale (eds) Improvements in Speech Synthesis, Part II. Issues in Prosody. Chichester: John Wiley & Sons, Ltd. Chapter 8, pp. 87-92, 2002.

[4] E. Keller, "*Toward greater naturalness: Future directions of research in speech synthesis*". In: Improvements in Speech Synthesis, Keller, Bailly, Monagham, Terken and Huckvale (Eds). Chichester John Wiley and sons, Ltd, pp. 3-17, 2002.

[5] G. Rolland, "*Automatic stylisation of the fundamental frequency F0 using MOMEL*". Documentation and translation into a Praat script , ICP , November 2000.

[6] A. Chentir, M. Guerti and D.J. Hirst, "*Extraction of Arabic Standard Micromelody*". Journal of Computer Science 5 (2): 86-89, 2009. ISSN 1549-3636.

[7] A. Chentir, D.J. Hirst and M. Guerti, "*Modeling Micromelodic Effects in Standard Arabic Using MOMEL*". The 17th IEEE International Conference on Electronics, Circuits and Systems (ICECS 2010), December 12-15, 2010, in Athens, Greece.