Journal of Computer Science 5 (2): 86-89, 2009
ISSN 1549-3636

Extraction of Arabic Standard Micromelody

[1,2]A. Chentir, [2]M. Guerti and [3]D.J. Hirst
[1]Department of Electronic Saâd Dahlab University, Blida, Algeria
[2]Department of Electronic National Polytechnic College, Algiers, Algeria
[3]Laboratoire Parole et Langage, Aix En Provence, France

Abstract: Problem statement: In the early days of speech synthesis research the obvious focus of attention was intelligibility. But many researchers agree that the major remaining obstacle to fully acceptable synthetic speech is that it continues to be insufficiently natural. **Approach:** In this study, we exploited microvariations of fundamental frequency (F_0) of speech (intrinsic and co-intrinsic effects) to extract micromelody effect in Standard Arabic in view to improve synthesis speech systems. **Results:** We found that Arabic voiced consonants micromelody effect exists and seems to be possible to be included in a prosodic generating unit by a simple model. **Conclusion:** This preliminary result need to be tested on larger corpora and must be following by incorporating a microprosodic model of duration and intensity.

Key words: Arabic standard, prosody, micromelody, fundamental frequency

INTRODUCTION

The analysis of prosody is important[1] in speech synthesis because it gives us the basis for making prosodic effects around our utterance plans (phonological prosodic processing) and later to arrive at suitable rendering strategies for the marked prosody (phonetic prosodic processing).

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 which, at the physiological level of the production of the speech, corresponds to the frequency of vibration of the vocal cords. Intensity is essentially connected to the energy of the sound while the acoustic duration corresponds to its time of emission[2]. These three parameters harmonize in uneven proportions to give to every language its particular prosodic characteristics.

Prosody is often defined on two different levels[3]:

- An abstract, phonological level (phrase, accent and tone structure
- A physical phonetic level (fundamental frequency, intensity and duration)

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.

So, microprosody is defined as the intrinsic and co-intrinsic influences of fundamental frequency (F_0), duration and intensity, due to segment identity and to phonetic context[3].

Although for several year synthetic speech has been fully intelligible from a segmental perspective, there are areas of naturalness which still await satisfactory implementation[4]. 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.

In order to reinforce the existing systems of synthesis and recognition of Standard Arabic (SA), we made us in this study a new approach to determine the micromelody effect in Standard Arabic using our scripts in Praat[5], the program for speech analysis and synthesis and results obtained by the Praat script for MOMEL (MELodic Modelisation)[6]. Results shows that Arabic micromelody can be extract and its effect can be simply included in prosodic bloc generation.

Momel algorithm: MOMEL[6] is developed to allow automatic modelling of a raw fundamental frequency curve as a sequence of target points defining by a quadratic spline function.

Corresponding Author: Amina Chentir, Electronic Department, Saâd Dahlab University, B.P. 270, Blida, Algeria

Momel modelling: MOMEL consist of 4 stages to automatically modelling the curves[7]. There are:

- Pre-processing of F_0
- Estimation of target candidates
- Partition of candidates
- Reduction of candidates

In short, pre-processing of F_0 consist of reassign value for values which are more than a given ratio (reassign 0 for out of pitch value). Estimation of target candidates involves the process of eliminating out-of range value and label target value. Partitioning of target candidates is done by partitioning the window of 200 ms into two and comparing the value for the mean of the two. Reduction will eliminate outlaying candidate value and the remaining target let is then recalculated as final target of the segment. At the end of the MOMEL modelization, a quadratic spline function is used to give a close fit to the original curve[8].

MATERIALS AND METHODS

Corpus: One native Arabic-speaker pronounced 16 sentences including all Arabic phonemes. The recording was made in an anechoic recording chamber in the Laboratory Parole et Langage (LPL) in Aix-en-Provence. The Praat computer program [5] 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.

Method of Extraction of Micromelody Effect (EME): In our approach, we followed the following stages:

Stage 1: We execute MOMEL algorithm to our corpus with intention to extract the corresponding melodic curves from.

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

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

Stage 4: Once the melodic data representing the microprosodic effect calculated, we then execute our 2nd PRAAT script, which we developed to model microprosodic profile for every used consonant.

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

1st script algorithm: creation of microprosodic profile: In order to extract the microprosodic 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 microprosodic effect thanks to the ratio: microprosody = F_0 / Momel_F_0
- Recording results in a file with extension *.mpp

2nd script algorithm: Modelling a microprosodic profile of a consonant: In order to model with precision the microprosodic evolution's effect of a consonant, we adopted the following approach: Knowing that each consonant lies between two vowels, we carry out to extract measurements of microprosodic effect of F_0 at the following points (Fig. 1).

With: C = Consonant, V = Vowel, Duration = Consonant duration, S = Start, M = Middle, E = End1 = 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

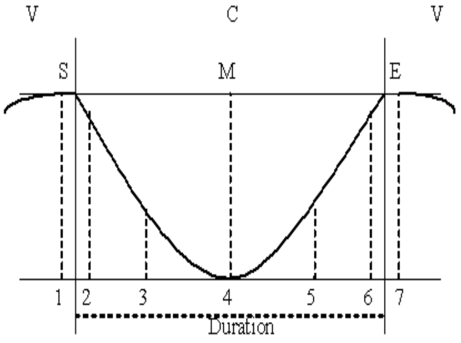

Fig. 1: Simplistic Representation of points retained to estimate the microprosodic variations of F_0

. 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 microprosodic 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. 1
- Deduct the microprosodic values corresponding to the 7 calculated points
- Record obtained results in a file (*. dat) in order to exploit them later by Excel

RESULTS

Knowing that each consonant can appear several times throughout the used corpus, we have then opted for calculation of the median value of each of the 7 points corresponding in microprosodic 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).

Figure 2 shows the microprosodic profile evolution of phoneme [b] (case of 6 possible values of results) where x-axis represents the corresponding selected points and y-axis represents the mpp: the ratio microprosody = F_0 / Momel_$F_{0.}$ We note that all curves generally, follow the same trajectory.

Figure 3 shows median values obtained from the microprosody evolution of phoneme [b]:

- We noted that although there is a well variation of the micromelodic curve, this last one is very weak (about 0.045)
- We also noticed that this variation is always (for the most voiced studied consonants) maximum 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. 4) where variations are between 1.003 and 1.011.

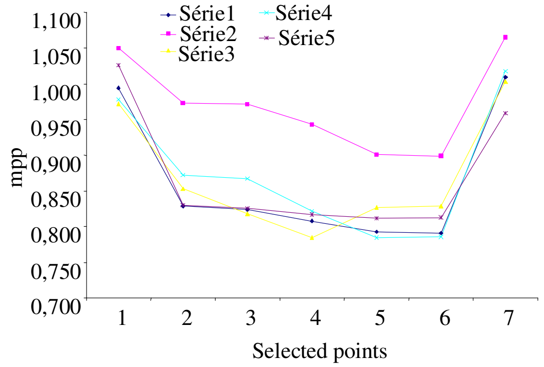

Fig. 2: Evolution of microprosodic profile of phoneme [b]

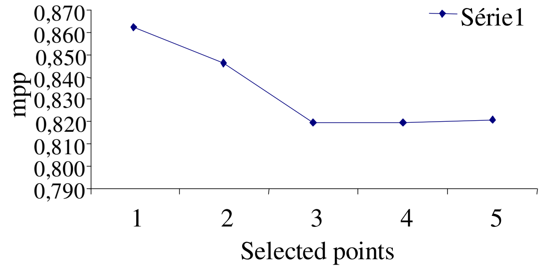

Fig. 3: Median values of microprosodic profile for phoneme [b]

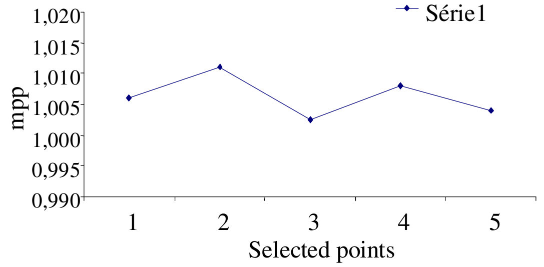

Fig. 4: Median values of microprosodic profile for phoneme [n]

DISCUSSION

By calculating the global median value for each studied voiced Arabic consonants, we then obtained the Table 1.

From Table 1, several observations were made:

- 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 F_0

Table 1: median Values of Arabic voiced phonemes

Phonème	Médian	Phonème	Médian
ب /b/	0.85	م /m/	1.00
د /d̪/	0.83	ن /n/	1.00
ض /ð̞̣/	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 /ẓ/	0.95		

- The microprosodic effect is almost non-existent for the case of nasals, semivowels and the liquid consonants.
- The microprosodic effect is more evident for the fricatives than for the occlusives.
- The microprosodic effect exists certainly but it is very weak, which does not require a complex mathematical expression to model its variation.
- 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

CONCLUSION

In this study, we present a new method which makes it possible to extract most automatically possible, the micromelodic information from speech signal using the original curve of fundamental frequency and of its macromelodic curve obtained using algorithm MOMEL.

Results obtained come to reinforce the idea that the microprosodic effect exists in fact. But the variance analysis pushes us to propose quite simply, only one additional relative lowering of the macromelodic curve, if we wish to improve the most simply possible naturalness of the synthesized voice.

However, complementary studies concerning microprosodic effect of duration as of energy are to be envisaged to complete our analysis.

REFERENCES

1. Monaghan, A., 2002. State-of-the-art summary of European synthetic prosody R & D. In Keller, Bailly, Terken & Huckvale (eds) Improvements in Speech Synthesis. Chichester: John Wiley & Sons, Ltd. pp. 93-103. ISBN: 9780471499855 Online ISBN: 9780470845943.
2. Monaghan, A., 2002. Prosody in Synthetic Speech: problems, solutions and Challenges. In Keller, Bailly, Terken & Huckvale (eds) Improvements in Speech Synthesis, Part II, Issues in Prosody. Chichester: John Wiley & Sons, Ltd. Chapter 8, pp. 87-92 ISBN: 9780471499855 Online ISBN: 9780470845943.
3. Tatham, M. and K. Morton, 2005. Developments In Speech Synthesis. Chichester. John Wiley and Sons, Ltd. ISBN: 047085538X (HB), DOI: 10.1002/0470012609.
4. Keller, E., 2002. Toward greater naturalness: Future directions of research in speech synthesis. In: Improvements in Speech Synthesis. Keller, Bailly, Monaghan, Terken and Huckvale (Eds.). Chichester. John Wiley and Sons, Ltd., pp: 3-17. ISBN: 9780470845943, Online ISBN: 9780470845943.
5. Boersma, P. and D. Weenink, 2008. Praat: Doing phonetics by computer (Version 5.0.32) [Computer program]. http://www.praat.org
6. Hirst, D.J. and R. Espesser, 1993. Automatic modelling of fundamental frequency using a quadratic spline function. Travaux de l'Institut de Phonétique d'Aix, 15 : 71-85. http://aune.lpl.univ-aix.fr/~hirst/articles/1993%20Hirst&Espesser.pdf
7. Hirst, D.J., A. Di Cristo, and R. Espesser, 2000. Levels of Representation and Levels of Analysis for Intonation. In Prosody: Theory and Experiment. M. Horne (Ed.). Dordrecht: Kluwer Academic Press. ISBN: 978-0-7923-6579-2
8. Samsudin, N.H. and T.E. Kong, 2005. Adjacency analysis for unit selection speech model using MOMEL /INTSINT. Proceedings for MMU International Symposium on Information and Communications Technologies, Nov. 24-25, Petaling Jaya, Malaysia, pp: 1-4. http://www.cs.bham.ac.uk/~nhs/MyPapers/Adjacency%20Analysis%20for%20Unit%20Selection%20Speech%20Model%20Using%20MOMEL%20INTSINT%20(M2USIC).pdf