Study and quantification of the declination for the Arabic speech synthesis system PARADIS

A. Benabbou*, N. Chenfour[+], A. Mouradi[†]

*FST Fès-Saïss,
abenabbou@yahoo.fr
[+]Faculté des Sciences Fès
chenfour@yahoo.fr
[†]ENSIAS Rabat
mouradi@ensias.ma

Abstract

The modeling of the melody in a Text-To-Speech System is indispensable to have a good quality of synthesis and to approach the naturalness. The study of the melody generally includes the analysis of the local melody events relating to the accent and the declination of the global melody contour of an utterance.

In this paper, we will present an experimental study of the declination phenomenon concerning the Arabic language. Our observations and results are of a great contribution for the quality of synthesis in our Text-To-Speech system PARADIS (Psola ARAbic DIsyllable Synthesizer). The melodic model that we have integrated in PARADIS is based on the prediction of a declination line on which local melodic events related to stressed syllables would be superimposed.

Our study will include the description and the classification of the declination line in a context of isolated sentences. The classification will be established mainly according to the modality and the number of syllables in the sentence. We will also study the phenomenon related to the resetting of F0 value which often affect the declination.

1. Introduction

Our speech synthesis system PARADIS is able to synthesize any completely vowelled Arabic text. It is based on concatenation of disyllables using TD-PSOLA synthesizer (Chenfour et al., 2000). In order to improve the quality of synthesis, it is imperative to incorporate the melodic processing in our system. We then studied the two principal melodic phenomena : the declination that affect the whole sentence and the melodic events generally related to the presence of stressed syllables.

In this paper we are interested only by the declination witch is a universal phenomenon in all languages (Beaugendre, 1994). Modeling the declination in a TTS system makes it possible to increase the naturalness of the synthesis which would be practically absent if no declination is made.

To perform a melodic analysis of the corpus, we realized an analysis system WinF0 based on a short-term temporal algorithm inspired from the AMDF algorithm (Doval, 1994). This system visualize the curves of the temporal signal, energy and F0. It produces the semi-automatic syllables labeling, and also extract all declination information from the signal.

The quantification of the declination will be established by computing the slopes for the top line and the base line of the fundamental frequency. These two lines represent the dynamic of the F0 values from the beginning to the end of the sentence.

The last section in this paper is related to the analysis of another event which often accompanies the declination, namely the F0 reset. We then compute the partial top and base lines on each group of the sentence delimited by an F0 reset occurrence. Moreover, we will study F0 reset for sentences which end with a descending melodic contour and those with an ascending melodic contour.

2. Melodic analysis system

To establish a melodic model it is necessary to have reliable tools for calculation and treatment of the F0 curve. To this end, we realized an analysis system WinF0 (figure 1).

The system gives possibilities to select a portion or the totality of the speech signal for listening and to perform a semi-automatic labeling. To refine the process of labeling, the knowledge of the signal energy curve is generally very useful. This functionality is also integrated in WinF0.

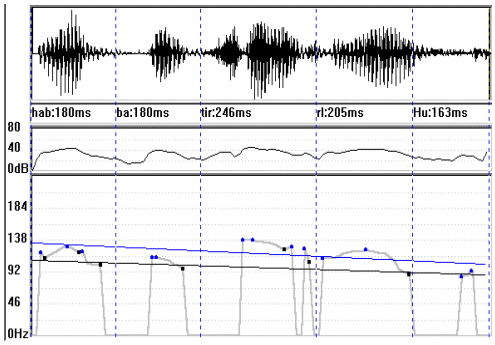

Figure 1: Analyze by the WinF0 of the declarative sentence "هبت الريح" /habbati rrIHu/ : temporal signal curve, labeling in syllables with duration (in ms), energy curve (in dB), F0 curve (in Hz), top and base lines.

2.1. F0 computing method

The F0 algorithm integrated in our analysis system is an alternative of the algorithm AMDF (Average Magnitude Function Difference). This algorithm consists in minimizing a function of dissimilarity F(t) which is based on the difference between an extracted frame and a same size frame shifted with t samples on the signal :

$$F(t) = \frac{\sum_{i=1}^{n} |s_i - s_{i+t}|}{\sum_{i=1}^{n} |s_i|}$$

where N represents the size of the frame to be analyzed, $(S_i)_{i=1, N}$ represent its samples and $(S_{i+t})_{i=1, N}$ the samples of a same size frame shifted by t samples number, $t \in [1, N]$.

The denominator term normalize the F(t) function in order to make comparisons with an absolute threshold. This threshold is calculated experimentally and allows us to decide for the periodicity (voiced/unvoiced) of the analyzed frame.

2.2. Declination computing method

There is certainly no method making it possible to quantify in a faithful way the declination line. The majority of the authors define it by a linear factor over the length of the sentence : the declination slope. Consequently, we established in the WinF0 system a method of computing slopes relating to the top and base lines, based on the local F0 extremums. This method is based on the linear regression, which permits to represent the global tendency of a chronological data set by a least-squares line. The principal informations generated by the system are the equations of the two lines of declination, their slope, the initial and final F0 frequencies (table 1 and figure 2).

Sentence : « اشتريت التذكرة » /ei^ ta ray tut tav ki rah /		
Number of syllables : 7 **Total duration :** 1426 ms **Initial Frequency :** 129 Hz **Final Frequency :** 83 Hz		
	Top line	**Base line**
Equation	-30*x+139.245	-20*x+111.867
Slope	-0.907 ST/syl	-0.727 ST/syl

Table 1: Slopes of the top and base lines estimated for the declarative sentence " اشتريت التذكرة "/ ei^taraytu ttavkirah /.

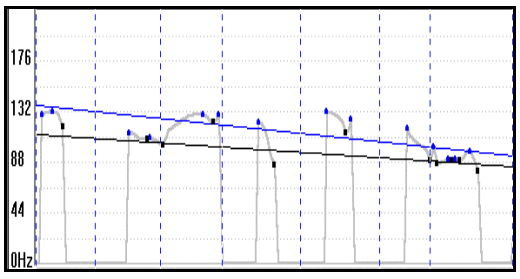

Figure 2: The top and base lines for the declarative sentence: " اشتريت التذكرة " / ei^taraytu ttavkirah /.

The values of the slopes are calculated in semitone per syllable (ST/Syl), knowing that the variation D in semitones (ST) , between two values f1 and f2 of the fundamental frequency is :

$$D = 12 \times \log_2\left(\frac{f1}{f2}\right)$$

The value in ST/Syl simply consists in dividing this variation by the number of syllables of the analyzed sentence.

3. Study of the declination

The study of the declination is made according to the phrase modality. Therefore, we recorded by three male speakers a corpus of 90 isolated sentences, representing the various phrase modality in Arabic language. The declarative and exclamatory sentences were recorded in a narrative style. On the other hand, imperative, interrogative and call sentences were recorded in a dialogued style.

We will determine the slopes of the top and base lines for each sentence of the corpus. Thus, for each speaker and each phrase modality, we will statistically compute the tendency of the slopes of these lines depending on the number of syllables, then we determine the average tendencies on the three speakers.

3.1. Declarative modality

For each declarative sentence of the corpus we calculated the values of the slopes for the top and base lines. All the slopes obtained are negative i.e. the top and base lines are decreasing from the start to the end of the sentence. This phenomenon was already observed in many studies for other languages (Beaugendre, 1994). Moreover, the slopes of the long sentences are weaker than those obtained for the short ones.

As an example, we present in the figure 3 a graph of the declination slopes values computed for one speaker. We have represented the values of the slopes in –ST/Syl, so that their decreasing will be quite illustrated. Thus, we can notice that the degree of the slopes of the top and base lines decrease when the number of syllables increases.

This decrease has a tendency curve represented by an equation obtained from the various values of slopes in function of the number of syllables. We found the best approximation for these tendency curves using power functions of general equation : $a.X^b$ (determination coefficient R^2 approaching 1). A tendency curve represent then a function of declination which the parameter is the number of syllables of the sentence.

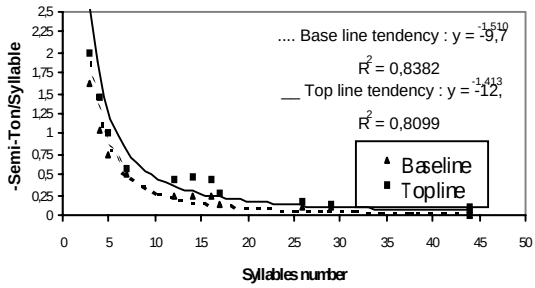

Figure 3: Curves of tendency of the slopes for the various declarative sentences of the corpus. The base and the top lines are respectively represented by a dotted and a solid lines

In table 2, we summarize the results of tendency curves for the three speakers of the corpus. Finally, we determined the equation of an average tendency (geometric mean of power functions) from the equations of the tendencies obtained for each speaker. This average tendency can be considered independent of the speaker.

Speaker	Base line tendency	Top line tendency
Speaker 1	$Y=-9.73\ X^{-1.51}$	$Y=-12.08\ X^{-1.41}$
Speaker 2	$Y=-17.35\ X^{-1.49}$	$Y=-15.43\ X^{-1.31}$
Speaker 3	$Y=-10.34\ X^{-1.51}$	$Y=-13.66\ X^{-1.38}$
Average	$\mathbf{Y=-12.04\ X^{-1.50}}$	$\mathbf{Y=-13.65\ X^{-1.36}}$

Table 2: Equations of the tendencies of the top and base lines for each speaker and their average. Parameter X represents the number of syllables of the sentence and Y represents the value of the slope in ST/Syl.

3.2. Exclamatory modality

Using the same methodology, we summarize in table 3 the equations found for the slope tendencies of the top and base lines for each speaker as well as the average tendencies.

Speaker	Base line tendency	Top line tendency
Speaker 1	$Y=-0.37\ X^{+0.54}$	$Y=-6.67\ X^{-0.84}$
Speaker 2	$Y=-0.15\ X^{+0.90}$	$Y=-5.39\ X^{-0.80}$
Speaker 3	$Y=-0.48\ X^{+0.18}$	$Y=-5.13\ X^{-0.81}$
Average	$\mathbf{Y=-0.30\ X^{+0.54}}$	$\mathbf{Y=-5.70\ X^{-0.82}}$

Table 3: Equations of the slope tendencies of the top and base lines for each speaker and their average in the case of the exclamatory modality.

The equation of the average tendency of the base line slopes shows a negative coefficient -0.30. This means that the base line slope is decreasing from the start to the end of a sentence.

However, compared with the result in a declarative modality the base line slope is steeper for a long sentence than for a short one. This is due to the positive exponent in the tendency equation (+0.54).

3.3. Imperative modality

In the Arabic language, imperative sentences are generally similar to the declarative sentences in term of grammatical components, but generally start with a verb with the imperative mode preceded in certain cases by a negation particle. The analysis of the slope tendencies for each speaker as well as the average tendencies are given in table 4.

Speaker	Base line tendency	Top line tendency
Speaker 1	$Y=-8.57\ X^{-1.38}$	$Y=-16.74\ X^{-1.39}$
Speaker 2	$Y=-14.49\ X^{-1.59}$	$Y=-21.26\ X^{-1.48}$
Speaker 3	$Y=-39.57\ X^{-2.23}$	$Y=-18.29\ X^{-1.40}$
Average	$\mathbf{Y=-17.00\ X^{-1.73}}$	$\mathbf{Y=-18.66\ X^{-1.42}}$

Table 4: Equations of the slope tendencies of the top and base lines for each speaker and their average in the case of the imperative modality.

We can notice that the imperative modality indicates a top line slope more steeper than in declarative modality.

3.4. Call modality

The call sentences in the Arabic language, generally consist of a very reduced number of words, often one or two words preceded by a call particle, sometimes only one isolated word (a name or a proper name). In the table 5, we have computed the tendencies of the declination line slopes for the three speakers and their average.

Speaker	Base line tendency	Top line tendency
Speaker 1	$Y=-16.10x^{-1.77}$	$Y=-10.19x^{-1.03}$
Speaker 2	$Y=-11.37\ x^{-1.88}$	$Y=-17.06x^{-1.55}$
Speaker 3	$Y=-30.75x^{-2.28}$	$Y=-25.30x^{-1.80}$
Average	$\mathbf{Y=-17.78\ X^{-2.02}}$	$\mathbf{Y=-16.38\ X^{-1.47}}$

Table 5: Equations of the slope tendencies of the top and base lines for each speaker and their average in the case of the call modality.

We note that the base line slope is weaker in the call sentences comparatively to the declarative sentences, while the top line slope is approximately the same. This implies that the melodic dynamics, from the start to the end of the sentence, decrease faster in the call modality than in the declarative modality.

3.5. Interrogative modality

We will separate the interrogative sentences into two different categories according to the generated melodic contour which can be final descending or final ascending (Benabbou, 1997). The first category is characterized by the interrogation particles : "من" / man / (who), "ما" / mA / (what is) "ماذا" / mAvA / (what), "لماذا" / limAvA / (why), "كيف" / kayfa / (how), "كم" / kam / (how much), "أين" / eayna / (where) and "متى" / matA / (when) The second category is represented by the two equivalent particles "هل" / hal / and "أ" / ea / (is) or by the interrogative sentences with no particles.

3.5.1. Final descending melodic contour

In table 6, we report the tendencies equation of the declination lines obtained for this category. We can note that the declination lines are decreasing throughout the

sentence but the two lines have slopes steeper than those in the declarative modality.

Speaker	Base line tendency	Top line tendency
Speaker 1	$Y=-7.29\ X^{-1.12}$	$Y=-23.37\ X^{-1.51}$
Speaker 2	$Y=-6.30X^{-1.07}$	$Y=-11.83X^{-1.16}$
Speaker 3	$Y=-18.22X^{-1.48}$	$Y=-29.78X^{-1.57}$
Average	$\mathbf{Y=-9.42\ X^{-1.22}}$	$\mathbf{Y=-20.18\ X^{-1.41}}$

Table 6: Equations of the slope tendencies of the top and base lines for each speaker and their average in the case of the interrogative modality with a final descending melodic contour.

3.5.2. Final ascending melodic contour

In the same way, we represent the equations found for the tendencies of the declination lines in table 7.

Speaker	Base line tendency	Top line tendency
Speaker 1	$Y=-10.39\ X^{-1.85}$	$Y=+185.39\ X^{-2.63}$
Speaker 2	$Y=-4.61\ X^{-0.85}$	$Y=+228.35\ X^{-2.88}$
Speaker 3	$Y=-1.30\ X^{-0.24}$	$Y=+184.88\ X^{-2.78}$
Average	$\mathbf{Y=-3.43\ X^{-0.98}}$	$\mathbf{Y=+198.44\ X^{-2.76}}$

Table 7 : Equations of the slope tendencies of the top and base lines for each speaker and their average in the case of the interrogative modality with a final ascending melodic contour.

Comparing with the first category, we can notice that the top line presents a typical melodic rise in this category of interrogation (the coefficient is positive +198.44), while the base line always shows a descending melodic contour. However, the negative exponent in the equation (-2.76) indicates that the degree of the rise decreases depending on the number of syllables.

4. Declination reset

The declination for long sentences (without punctuation), generally declarative, is often accompanied by the phenomenon of F0 reset. The melodic curve then presents resetting which describe an irregular increase or reduction in the fundamental frequency, then giving a new reference for the declination.

F0 reset is generally marked by a brief pause that denotes a breathing group. But f0 reset is not systematic, it does not appear obligatorily with each pause. Thus, the localization of this phenomenon is not an easy task. However, as the synchronous F0 reset with the pause is most frequent and easiest to locate on the signal, we have focused our analysis only on this context.

Also, we studied F0 reset for two particular classes of sentences according to their final melodic contour : descending or ascending.

In figure 4 we give the analysis results for a sentence with a final descending melodic contour and its declination line slopes in table 8.

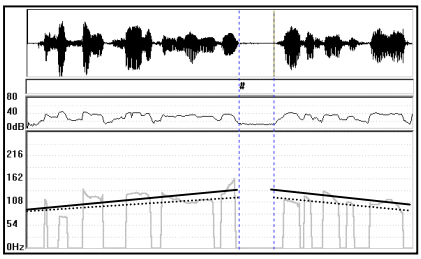

Figure 4: Presence of a F0 reset with pause (#) for the declarative sentence " تحسنت حالة المريض # بعد شرب الدواء " / taHassanat HAlatu lmarIDi # bacda ^urbiddawAe /.

Groups of breathing	1st group	2 nd group
Number of syllables	10	6
Slope of the base line	+0.640 ST/Syl	-0.456 ST/Syl
Slope of the top line	+1.046 ST/Syl	-0.779 ST/Syl

Table 8 : Results of analysis of F0 reset for the sentence in figure 4.

Also, we present in figure 5 the analysis results for a sentence with a final ascending melodic contour and its declination line slopes in table 9.

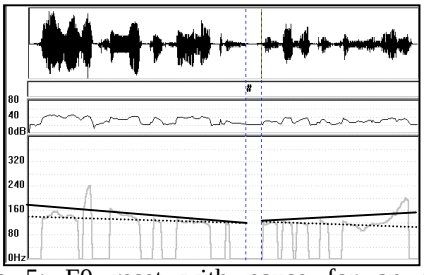

Figure 5: F0 reset with pause for an ascending interrogative sentence : " هل وجدتم ما قلته لكم عندما كنتم هنالك ؟ " / hal wajadtum mAqultuhu lakum # cindamA kuntum hunAlik? /.

Groups of breathing	1st group	2 nd group
Number of syllables	10	8
Slope of the base line	-0.313 ST/Syl	-0.145 ST/Syl
Slope of the top line	-0.790 ST/Syl	+0.320 ST/Syl

Table 9: Results of analysis of F0 reset for the sentence in figure 5.

In spite of the reduced number of studied sentences, our preliminary results are in conformity with those for other languages such as French (Martin, 1987) :

- the division of a group ending with a final descending melodic contour implies the presence of rising contours at the end of each group created.
- the division of a group ending in a rising contour involves the presence of contours going down at the end from each group created.

5. Conclusion

In this paper, we have confirmed the existence of the phenomenon of the declination for the Arabic language, by analyzing a multispeaker corpus of isolated sentences recorded in a reading style. We have presented the WinF0 analysis system which permits us to study the slopes of the top and base lines in function of each type of Arabic phrase modality. Results were then represented in tendency functions. Those functions allow to determine the declination line slopes of a sentence according to its modality and number of syllables.

We also studied the F0 reset phenomenon that generally accompanies the declination in the case of long sentences. We have noted that the F0 reset features are perceived differently according to the final melodic contour (descending or ascending).

6. References

Aubergé, V. 1991. La synthèse de la parole : des règles aux lexiques. thèse de Doctorat d'informatique, Université Pierre Mendès France, Grenoble.

Beaugendre, F. 1994. Une étude perceptive de l'intonation du français, développement d'un modèle et application à la génération automatique de l'intonation pour un système de synthèse à partir du texte . Thèse de doctorat en sciences de l'Université de Paris XI.

Benabbou, A. 1997. Implémentation sur PC d'un système à formants pour la synthèse par règles de la parole arabe. Thèse de Doctorat de 3ème cycle, Univ. Mohamed V, Rabat, Juillet 1997.

Chenfour, N.; Benabbou, A.; Mouradi, A. 2000. Optimisation du synthétiseur TD-PSOLA pour le Sytème de Synthèse de la Parole Arabe PARADIS. MCSEAI'2000, Fès 2000, p.369-378.

Doval, B. 1994. Estimation de la fréquence fondamentale des signaux sonores. Thèse de Doctorat de l'Université Paris VI en informatique.

Martin, P. 1987. Prosodic and rythmic structures in French. Linguistics 25, p.925-949, © Mouton de Gruyter, Amsterdam.