A Model of F0 Contour for Arabic
Affirmative and Interrogative Sentences

Omar A. G. Ibrahim\textsuperscript{1}, S.H. El-Ramly\textsuperscript{1} & N.S. Abdel-Kader\textsuperscript{2}

\textsuperscript{1} Electronics and Comm Eng. Dept, Ain Shams University, Cairo, Egypt.
\textsuperscript{2} Electronics and Comm Eng. Dept, Cairo University, Cairo, Egypt.

Abstract

This Paper presents the results of analyzing the global contour of the fundamental frequency (F0) for Arabic sentences and developing a model that represents it. The work concentrated on analyzing only affirmative and interrogative isolated ‘read-loud’ sentences. The work is divided into two parts:

1) Extracting the common characteristics and differences between the F0 plots of affirmative and interrogative Arabic sentences, and 2) Analyzing the effect of change in sentences length on the characteristics and differences extracted in the first part of the study. The model obtained can be easily implemented in speech synthesizers to improve its intonation.

1. Introduction

After having obtained the highest possible graphem-to-phoneme translation fidelity leading to minimum ambiguity in the synthesized speech, researchers are now working on improving the suprasegmental level (prosody) [1]. The fundamental frequency is known to be an important element in prosody and is subject to researches aiming at analyzing its contour and linking it to the underlying speech. Results of these researches find their way to be applied to speech synthesizers in order to improve their prosody, i.e., making synthesized speech sound more natural.

Lot of models were developed to represent the F0 contour. Some are based on a set of straight lines (stylized contours) [2,3,4,5], others depend on target points joined by transition equations [4], while Fujisaki et al. used a series of critically damped impulses and step functions [6,7,8]. All these models try to simulate the local changes in the F0 contour as a result of stressed syllables or other factors. As a different approach, Aubergé et al. [9,10] thought of intonation in terms of global contours resulting from high-level controlled gestures.

This paper presents a model obtained by analyzing the global F0 contour for Egyptian Arabic sentences. Only affirmative and interrogative sentences were examined due to the type of corpus used in this research which is ‘isolated read-loud’ sentences.

For the whole study, a set of five native speakers participated by pronouncing sentences from the above stated types. The speakers were asked to read the sentences in a very natural way without knowing the goal of the study. For the sake of diversity, the speakers had different ages and genders as follows (name/gender/age: AM/M/19; GI/M/65; HM/F/23; OM/M/29 and SL/F/56).

For the first part of the study a set of three affirmative and three interrogative sentences were selected. Each speaker pronounced five times each of the three affirmative and the three interrogative sentences. For the second part a set of seven affirmative and seven interrogative sentences were selected. Only four speakers pronounced only once each of the seven affirmative and the seven interrogative sentences. This makes a total of 206 recorded and analyzed Arabic sentences totaling 231 sec of speech.

The first part of the study is presented in sections 2, 3 and 4; where the choice of sentences is given in section 2, the analysis method is described in section 3 and the results and interpretation are presented in section 4. The results for the second part and its interpretation are presented in section 5. Finally the paper is concluded in section 6.

2. Choice of sentences for the first part

In this research three affirmative and three interrogative sentences were used. The sentences are:

Affirmative sentences:
1. Sabah elkher ya Ahmad. (E1) (Good morning Ahmad.)
2. Eshams talaa ennaharda. (E2) (The sun is shining today.)
3. Enta roht elmadrasa. (E3) (You went to school.)

Interrogative sentences:
1. Amalt aih embareh ? (Q1) (What did you do yesterday?)
2. Yatara mzaker kwayes ? (Q2) (Are you studying well?)
3. Enta roht elmadrasa ? (Q3) (You went to school?)

In this paper E1, E2, E3, Q1, Q2 & Q3 will be used to refer to each of the above sentences.

A particular attention is made in the choice of the interrogative sentences to cover three different types of question:

1. The first type (Q1) is known as a ‘Wh’ question. It always contains a ‘question tool’ like (what, when, where, etc…) and the subject is hidden (as per the Arabic grammar); i.e. it has the syntax of a question.
2. The second type (Q2) is known as a ‘Yes/No’ question. The answer to it usually starts with ‘yes’ or ‘no’. It may contain a ‘question tool’ and the subject is hidden (as per the Arabic grammar); i.e., it has a partial syntax of a question.
3. The third type (Q3) is known as a declarative question. It does not contain any ‘question tools’ and its syntax is similar to a declarative or affirmative sentence but pronounced as a question. (compare with E3)

3. Analysis method

F0 plot is generated for each repetition of the above sentences for each speaker. Linear trendlines are obtained for each F0 plot. These linear trendlines should indicate the global behavior of the F0 plots for the recorded utterances. Local F0 behavior does not contribute into the calculation of these.
global trendlines, i.e., the speakers individualities are removed.

The global trendlines method is described as follows[11,12]:

1. A straight line equation that best represents all the points of the F0 plot is obtained using the least error squares method. This general trendline splits the plot into upper and lower regions (i.e., points on the F0 plot having values above the value of the trendline and other points having values below the trendline).

2. The points of the F0 plot having values above the general trendline are considered as a stand-alone plot and a linear trendline that best represents all these upper points is generated (Upper Trendline).

3. A lower trendline is obtained by considering all the points of the F0 plot having values below the general trendline as a stand-alone plot. A linear trendline that best represents all these lower points is then generated.

4. Each of these trendlines is represented by a straight line equation \( y=mx+b \) where \( m \) is the slope of the line and \( b \) is the intercept.

5. Values of the upper and lower trendlines slope (U.Slp. and L.Slp. respectively) are recorded for each utterance. These upper and lower trendlines are considered to be the guidelines of the global behavior of the F0 plots of the recorded utterances used for the study.

Figure 1 illustrates this method. The points above the general trendline are represented with circles whereas those below the general trendline are represented by continuous curve.

Other values of importance were also calculated:

1. The avg. laryngeal frequency (ALF) of each speaker.

2. The attack F0 (the beginning frequency) (Att) of each utterance and the ratio (Att/ALF) for each speaker.

3. The intercepts of the upper and lower trendlines with the beginning time of each utterance (UI & LI respectively) and the ratios (UI/ALF and LI/ALF) for each speaker.

4. Results and interpretation of the first part

4.1. Results

Table 1 gives the ALF values for each speaker. Results are reported for each of the five repetitions of each of the six sentences by each of the five speakers and the summary for each sentence across all the speakers is presented in tables 2 (affirmative sentences) and 3 (interrogative sentences). It is noted that the ALF of female speakers is higher than that of the male speakers (which is expected). It is also noted that the results of all affirmative sentences are very close, therefore, an average result for all three affirmative sentences is calculated and used to present this type of sentences. Figures 2, 3, 4, and 5 give a schematic representation of the results. It is noted that the affirmative sentences have negative U.Slp and L.Slp. Q1 has a negative U.Slp while L.Slp is positive. Q2 and Q3 have both positive U.Slp and L.Slp. But in Q2 both trendlines are almost parallel while U.Slp in Q3 has a relatively much higher value than L.Slp. These results are discussed in the next sub-section.

Table 1: Avg. Laryngeal Freq. (ALF) of participating speakers

<table>
<thead>
<tr>
<th>Speaker</th>
<th>AM</th>
<th>GI</th>
<th>HM</th>
<th>OM</th>
<th>SL</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALF in Hz</td>
<td>113.8</td>
<td>115.2</td>
<td>265.4</td>
<td>114.6</td>
<td>205.3</td>
</tr>
</tbody>
</table>

Table 2: Summary results of all affirmative sentences

<table>
<thead>
<tr>
<th>Sentence</th>
<th>U. Slp</th>
<th>L. Slp</th>
<th>UI/ALF</th>
<th>LI/ALF</th>
<th>Att/ALF</th>
</tr>
</thead>
<tbody>
<tr>
<td>E1</td>
<td>-29.03</td>
<td>-18.27</td>
<td>1.07</td>
<td>0.88</td>
<td>0.91</td>
</tr>
<tr>
<td>E2</td>
<td>-39.64</td>
<td>-31.84</td>
<td>1.12</td>
<td>1.00</td>
<td>1.01</td>
</tr>
<tr>
<td>E3</td>
<td>-31.14</td>
<td>-24.42</td>
<td>1.03</td>
<td>0.92</td>
<td>0.94</td>
</tr>
<tr>
<td>Average</td>
<td>-33.27</td>
<td>-24.84</td>
<td>1.07</td>
<td>0.93</td>
<td>0.95</td>
</tr>
</tbody>
</table>

Table 3: Summary results of all interrogative sentences

<table>
<thead>
<tr>
<th>Sentence</th>
<th>U. Slp</th>
<th>L. Slp</th>
<th>UI/ALF</th>
<th>LI/ALF</th>
<th>Att/ALF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q1</td>
<td>-18.57</td>
<td>23.55</td>
<td>1.15</td>
<td>0.86</td>
<td>0.80</td>
</tr>
<tr>
<td>Q2</td>
<td>123.15</td>
<td>108.18</td>
<td>0.84</td>
<td>0.69</td>
<td>0.86</td>
</tr>
<tr>
<td>Q3</td>
<td>160.14</td>
<td>101.09</td>
<td>0.91</td>
<td>0.75</td>
<td>0.92</td>
</tr>
</tbody>
</table>

Figure 1: Illustration of the Trendline Method (utterance E1 by speaker GI)
4.2. Interpretation of results

The results of the affirmative sentences is comparable to the results obtained by other researchers for other languages and is known in the literature by ‘declination’ attitude [4,13].

We can comment on the results obtained from the interrogative sentences by saying that they generally have a rising lower trendline opposed to a declining lower trendline in declarative sentence. i.e. the question has a general rising attitude opposing the general declining attitude in the affirmative sentences. Other researches on other languages reached similar conclusions ( rising contour, rise in last syllable, etc...) [4,13].

But why are there different upper trendline slopes for the three different interrogation types. This can be explained by the fact that the speech elements over all levels (syntax, pragmatic, prosodic, etc...) contribute to the delivery of information ( or message ) from the speaker to the listeners. If one of these elements contributes strongly into the delivery of the message, the other elements will have less contribution and vice-versa. [11,12]

Q1 has a very strong question formation from the syntactic point of view (a question tool and a hidden subject), therefore the prosodic contribution into the delivery of the message ( a question ) is relatively weak and tend to be the closest to the affirmative attitude. Q2 has less strong syntactic contribution into the delivery of the message (only a hidden subject and the question tool is not a must), the prosodic contribution is then stronger and demonstrated by a rising upper trendline. Q3 has very week syntactic contribution into the delivery of the message (it is a question) therefore the prosodic contribution is even stronger compared with Q1 & Q2. This is demonstrated by a more rising upper trendline.

These results demonstrate that the F0 plot in Egyptian Arabic contains prosodic information that is linked to the type of utterance pronounced which if used in Egyptian Arabic synthesizers will improve the naturalness of their output synthesized speech.

5. Results and Interpretation of the Second Part

For the second part of the study which focuses on analyzing the effect of change in sentences length on the characteristics and differences extracted in the first part of the study we chose seven affirmative and seven interrogative sentences that vary in length. The interrogative sentences were all from the (Wh question) type.

Each sentence is repeated once and a comparison is made between the seven sentences of each set. The same technique of analysis (Global Trendlines) is used in this part of the study. An average for all speakers is taken for each sentence. Figure 6 shows the effect of sentence length on the U.Slp. and L.Slp. values for affirmative sentences, while figure 7 shows the effect of sentence length on UI/ALF, LI/ALF and Att/ALF values for affirmative sentences. It is noted that both U.Slp and L.Slp remain negative (i.e. the declining attitude is preserved) regardless of the sentence length but their values decrease with the increase in the sentence length. This is matching with results obtained for other languages and other F0 modeling techniques take it
Att/ALF ratios. The length has relatively no effect on UI/ALF, LI/ALF and Att/ALF ratios. It is also noted that the sentence length is keeping its signs regardless of the sentence length but the other two ratios are relatively not affected by sentence length. This paper presented a well defined model that can describe the F0 attitude for affirmative and interrogative Arabic sentences. This model is gender independent because its parameters are linked to the ALF value of the speaker, therefore it can slide up and down the frequency scale depending on the required speaker ALF. This model is not affected by the local F0 movements and does not represent them, therefore it can be used to synthesize the intonation of any sentence regardless of its stress patterns.

6. Conclusion
This paper presented a well defined model that can describe the F0 attitude for affirmative and interrogative Arabic sentences. This model is gender independent because its parameters are linked to the ALF value of the speaker, therefore it can slide up and down the frequency scale depending on the required speaker ALF. This model is not affected by the local F0 movements and does not represent them, therefore it can be used to synthesize the intonation of any sentence regardless of its stress patterns.

7. References