FUNDAMENTAL FREQUENCY CONTOUR IN HEBREW SYNTHESIS BY RULE – AN IMITATION OF A SPEECH COMMUNICATION MODEL

Asher Laufer*, Yehuda Sonnenblick† and Aryeh Teitelbaum‡

**The Hebrew University, Mount Scopus, Jerusalem, Israel, †The Jerusalem College of Technology, Jerusalem, Israel; ‡The Jerusalem College of Technology, Jerusalem, Israel*

ABSTRACT

Our Hebrew speech synthesis (although based entirely on rules) attempts to resemble natural speech production. Our speech communication model assumes that speech (both in production and perception) is constructed from sequential segments, which form a continuum. Communication occurs simultaneously on various levels, which can be differentiated by the length of their segments. The basic segment length is that of the phoneme. Higher levels, consisting of longer segments (syllables, stress-units, sense-groups, breath-groups, and intonation-groups), conform to associated rules that affect the lower levels.

The F0 contour is chosen to demonstrate synthesis rules derived from our model. It is shown how F0 can be created step by step, taking into account the different layers of speech production or perception. We demonstrate how F0 is molded as a result of the influence of a few levels and how mutual effects among levels lead to the construction of the composite fundamental frequency contour.

1. INTRODUCTION

The authors of this paper are involved in a major project to create synthesized speech from regular Hebrew text. The method employed for synthesis in this system is based entirely on rules, yet it attempts to resemble natural speech production. Therefore, the rules of synthesis are designed in such a way that they imitate a model of natural speech communication.

The first part of this paper is devoted to a description of our model of speech communication. The second part will concentrate on one specific parameter of speech synthesis, namely the fundamental frequency. Some of the F0 synthesis rules as derived from our model will be demonstrated. It will be shown how the F0 contour can be created step by step, taking into account different levels of speech production.

2. THE SPEECH COMMUNICATION MODEL

The essence of our model of speech communication is very simple. Speech, both from the production and from the perception points of view, can be considered a linear sequence of segments. On the other hand, speech is complex because it is produced (and perceived) on several parallel levels simultaneously.

The various levels of speech differ mainly in their segment length. The segment length of the basic level is that of the phoneme. Segments of higher levels are (see table 1): the syllable, the stress-unit (laymen may call it a foot or "word"), the sense-group (which may be called a clause or "phrase"), the breath-group (often called a "sentence"), and the intonation-group (paragraph). According to our model, speech is a linear concatenation of segments transmitted simultaneously on various levels. Since the segment length of the basic level is that of the phoneme, speech can be looked upon as a chain of phonemes (consonants and vowels). At the same time we can say that speech is a chain of syllables, or stress units, or sense-groups, or breath-groups, or intonation-groups (Table 1, lines 1-6).

Level	Type of Segment	Concatenation of Segments
1	**Phoneme** (consonants & vowels)	P+P+P+P+P+P+P+P+P+P+P+P+P+P+P+P+P+P+P+P+P+P+P+P+P+P+P+P+P+P+P (/CVCCCVCVCVCVCCCVCVCCVCVCVCVCCVCVCCCVCCV/)
2	**Syllable**	Syll.+Syll.+ Syll.+ Syll.+ Syll.+ Syll.+ Syll.+Syll.+ Syll.+ Syll.+Syll.+ Syll.+ Syll.
3	**Stress Unit** (foot or "word")	Stress-unit +Stress-unit + Stress-unit +Stress-unit + Stress-unit +Stress-unit + Stress-unit +Stress-unit + Stress-unit +Stress-unit
4	**Sense-group** (clause or "phrase")	Sense-group +Sense-group +Sense-group +Sense-group +Sense-group +Sense-group + Sense-group
5	**Breath-group** ("sentence")	Breath-group + Breath-group + Breath-group + Breath-group + Breath-group
6	**Intonation-group** ("paragraph")	Intonation-group + Intonation-group

Table 1. Levels of speech communication and their respective segment concatenations

In addition to this basic common characteristic (that each level is composed of a concatenation of its segments), there are two additional characteristics common to all levels: (1) A segment of a specific level is built of elements which are themselves the segments of the next lower level. (2) There is at least one element in each segment that serves as a nucleus, which binds together all the elements in the segment. As a result of this, neighboring elements within one segment are bound to each other more closely than neighboring elements that belong to different segments.

For example, consider the syllable level. (1) The syllable segment is composed of elements, namely phonemes, which are themselves the segments of the next lower level. (2) In each syllable there is one element, usually a vowel that serves as a nucleus, which holds the elements of this syllable together.

The principles of this model can be illustrated by employing the model to construct the framework of a representative paragraph step by step, as follows: The basic level is a concatenation of phonemes (some of which are consonants and some of which are vowels; Table 2, line 1). The phonemes are then grouped into a concatenation of syllables (Table 2, line 2). Note that borders of syllables are marked with periods.

Climbing up one level, to the stress-unit level, we again find the same principles: (1) The stress-unit segment is composed of the elements of its lower level, i.e. of syllables. (2) In each stress-unit we find one syllable which serves as a nucleus, the stressed syllable that holds all the syllables of this stress-unit together. The result of this is that the syllables within one stress-unit are tied to each other more closely than neighboring syllables of different stress-units. Note that the stress-unit nuclei have been marked with apostrophes and each stress-unit is underlined (Table 2, line 3).

Going up another level to the "sense-group" level, we find at least one prominent word that serves as its nucleus [7, 8]. Once more, the elements of this unit are the segments of its lower level, the stress-units. In Table 2 each asterisk marks a "sense-group" nucleus, each sense-group is double-underlined, and two neighboring sense-groups are separated by a vertical line (Table 2, line 4).

<table>
<tr><th>Level</th><th>Type of Segment</th><th>Concatenation of Segments</th></tr>
<tr><td>1</td><td>Phoneme</td><td>/CVCVCVCVCVCCCVCVCVCVCVCVCVCVCCVCCVC/</td></tr>
<tr><td>2</td><td>Syllable</td><td>/CV.CV.CV.CV.CVC.CCV.CV.CV.CV.CV.CV.CV.CVC.CVC.CVC/</td></tr>
<tr><td>3</td><td>Stress-unit</td><td>/CV.'CV. CV.CV.'CVC. CCV.'CV.CV.'CV.'CV.CV.CV.'CVC.CVC.'CVC/</td></tr>
<tr><td>4</td><td>Sense-group</td><td>/CV.'CV.CV.CV.*CVC| CCV.*CV| CV.*CV| 'CV.CV.CV.'CVC.CVC*CVC/</td></tr>
<tr><td>5</td><td>Breath-group</td><td>/CV.'CV.CV.CV.*CVC| CCV.*CV || CV.*CV| 'CV.CV.CV.'CVC.CVC*CVC/</td></tr>
<tr><td>6</td><td>Intonation-group</td><td>{ /CV.'CV.CV.CV.*CVC| CCV.*CV || CV.*CV| 'CV.CV.CV.'CVC.CVC*CVC/ }</td></tr>
</table>

Table 2. Levels of speech communication and representative segment concatenations

As noted, "sense-groups" are clustered into a higher unit we call a "breath-group". Again, the sense-groups are the elements of the breath-group. In table 2, line 5, a double vertical line marks the border between two breath-groups. "Breath-groups" are assembled into an even higher unit we call an "intonation-group" (enclosed in curly brackets in table 2, line 6).

All in all, according to our model, speech can be described as a concatenation of phonemes, that are grouped into a concatenation of syllables, that are grouped into a concatenation of stress-units, that are grouped into a concatenation of sense-groups, that are grouped into a concatenation of breath-groups, that are grouped into an intonation-group.

This can be illustrated by a concrete example of one short Hebrew intonation-group ("paragraph") [SOUND 0210_01.WAV]. The speech is described in Table 3 according to the aforementioned levels.

There is an even lower, more fundamental level than the phoneme level. This is the allophone level, which can be designated as level Zero (Table 3, line 0).

<table>
<tr><th>Level</th><th>Characteristic</th><th>Segment Series</th></tr>
<tr><td></td><td>Hebrew Transcription</td><td>• • • • • • • • • • • • ; • • • • • • • • • • • • • • .</td></tr>
<tr><td>0</td><td>Allophones</td><td>[ανι μΙϖακεΣ σλιΞΑ ανι λοΕ μΙδαβεΡ ιϖριτ]</td></tr>
<tr><td>1</td><td>Division into phonemes</td><td>/ʔANIMEVAKESHSLIXAʔANILOMEDABERʔIVRIT/</td></tr>
<tr><td>2</td><td>Division into syllables</td><td>/ʔA.NI.ME.VA.KESH.SLI.XA.ʔA.NI.LO.ME.DA.BER.ʔIV.RIT/</td></tr>
<tr><td>3</td><td>Division into stress-units</td><td>/ʔA'NI.ME.VA'KESH.SLI'XA. ʔA'NI.'LO.ME.DA'BER.ʔIV'RIT/</td></tr>
<tr><td>4</td><td>Division into sense-groups</td><td>/ʔA'NI.ME.VA*KESH | SLI*XA | ʔA*NI | *LO.ME.DA'BER.ʔIV'RIT/</td></tr>
<tr><td>5</td><td>Division into breath-groups</td><td>/ʔA'NI.ME.VA*KESH SLI*XA || ʔA*NI *LO .ME.DA'BER.ʔIV'RIT/</td></tr>
<tr><td>6</td><td>Intonation-group</td><td>{ /ʔA'NI.ME.VA*KESH SLI*XA || ʔA*NI *LO .ME.DA'BER.ʔIV'RIT/ }</td></tr>
</table>

Table 3. An example of a Hebrew paragraph segmented according to various levels

As mentioned, the essence of our model is that each higher level segment is composed of segments of the level below it. As a result, a higher level possesses associated appropriate rules that influence the segments of its lower levels. Such rules were applied in our speech synthesis program. This influence will be illustrated in the case of one specific synthesis parameter, namely the fundamental frequency.

3. THE FUNDAMENTAL FREQUENCY CONTOUR

Our synthesis program deals with changes in the F0 contour on various levels. First, the changes associated with the highest levels are applied. Next, the changes due to the successively lower levels are introduced. Finally, the changes due to the basic levels, those of the phoneme and allophone, are added. The interaction of the changes associated with the various levels leads to the construction of the composite fundamental frequency contour.

3.1. Width of pitch ranges in a section, in an intonation-group and in a breath-group

Observations and measurements on natural Hebrew speech led us to introduce into our synthesis calculations various rules dictating the hierarchy of "pitch ranges" in the high levels. (A "pitch range" is defined as the range of F0 values in an interval, measured from the lowest possible F0 to the highest possible F0.)

We have observed that consecutive breath-groups, which form one paragraph (intonation-group), have their "pitch ranges" arranged in a descending order, as shown by the pairs of horizontal lines in figure 1. The differences in the width are mainly in the top end. The first breath-group in the paragraph is wide, the last – is narrow, and there is a general narrowing of each group in between. Such behavior helps to join the breath-groups together into one integral intonation-group. This feature has, therefore, been incorporated into our software calculations.

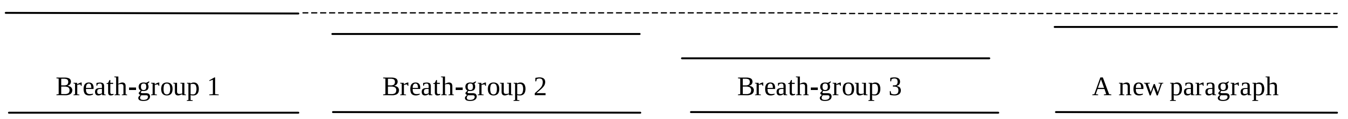

Figure 1. Descending pitch ranges of consecutive breath-groups, which form one paragraph

At a higher level we observed a similar phenomenon. Consider a section of speech that is composed of two paragraphs. After the completion of a full intonation-group a subsequent new series of descending pitch ranges begins for the following intonation-group. The pitch range at the beginning of the second paragraph is slightly narrower than the pitch range at the beginning of the first paragraph.

A similar phenomenon was observed also at a lower level, the breath-group level. Therefore, we applied rules that cause consecutive sense-groups, which comprise a specific breath-group, to have their pitch ranges arranged in a gradually descending order. This phenomenon helps to combine these sense-groups together in our perception into one integral breath-group. At the beginning of the next breath-group the pitch range sequence will be "reset", except that the pitch range of the first sense-group of this next breath-group will become a little narrower than that of the first sense-group of the previous breath-group (Figure 2).

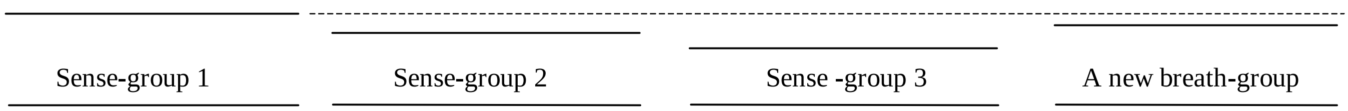

Figure 2. Descending pitch ranges of consecutive sense-groups, which form one breath-group

3.2. Fundamental frequency contour at the sense-group ("phrase") level

A sense-group was chosen as the smallest speech unit to be treated by our synthesizer. Our major calculations of F0 are made at the sense-group level. A sense-group can be divided into four parts: the pre-head, the head, the nucleus and the tail. The possible sense-groups are summarized in the formula:

Sense-group = (pre-head +) (head +) nucleus (+tail)

Note that the parts in parenthesis are optional in Hebrew. We treat a nucleus and its (optional) tail as one unit and call its F0 contour a "nuclear tone". We employ six types of "nuclear tones" in our synthesis; this increases the possible number of "intonation patterns". (These terms are defined in [7] and resemble those of [8]. Other researches use various names for similar units, such as: "Tone-groups", "Tone-units", "Phonological phrases", "Intonational phrases", "pitch contours"… see, for example: [3, 4].)

We can listen to our six "nuclear tones" embedded in one particular sense-group, creating six different intonation patterns [SOUND_0210_02.WAV].

3.3. Sense-group component

The place of the nucleus plays an important role in the F0 contour. We can hear this effect if we choose one particular sense-group with a fixed "intonation pattern" and change only the location of the nucleus [SOUND_0210_03.WAV].

3.4. Fundamental frequency contour at the stress-unit level

On a still lower level, that of the stress-unit, we again employ changes in this same physical component, the F0 contour, to cause the distinctive stress in Hebrew [6]; compare this to the findings in [1, 5]. A sound sample [SOUND_0210_04.WAV] illustrates a stress difference in three minimal pairs in Hebrew.

Note that fluctuations of F0 on a lower level than the sense-group level are referred to as "microintonation" [9].

3.5. Fundamental frequency contour at the phoneme level

In Hebrew, as in other languages, it has been observed that F0 rises after voiceless consonants [2, 6]. Accordingly, an appropriate rule was implemented in our synthesis. Thus, the F0 contour is changed also at the phoneme level.

4. CONCLUSIONS

We have presented a simple model of natural speech communication. We have demonstrated the model by describing how the F0 contour is formed through our synthesis program. It should be noted, however, that we have hardly scratched the surface of the rules needed to build just the fundamental frequency contour, which is only one of the parameters needed to synthesize speech. We have shown that our rules are based upon a model of natural speech, in which higher levels influence the most basic level.

ACKNOWLEDGMENTS

We would like to thank our benefactors Mr. and Mrs. M. Wohl and an anonymous foundation for their support of this project.

REFERENCES

[1] Bolinger D.L. 1958. A theory of speech accent. *Word*, 14, 109-149.

[2] Catford J.C. 1977. *Fundamental Problems in Phonetics*. Bloomington, Indiana University Press.

[3] Cruttenden A. 1986, *Intonation*. London, Cambridge University Press,.

[4] Crystal D. 1969. *Prosodic Systems and Intonation in English*. Cambridge, Cambridge University Press.

[5] Fry D.B. 1958. Experiments in the perception of stress. *Launguage and Speech*, 1, 126-152.

[6] Laufer, A. 1972. *Synthesis by Rule of a Hebrew Idiolect*. Ph. D. thesis, University of London, London.

[7] Laufer, A. 1987. *Intonation*. Jerusalem, Publications of the Institute of Jewish Studies of the Hebrew University (in Hebrew).

[8] O'Connor, J.D. and G.F. Arnold. 1973. *Intonation of Colloquial English*. second edition. Great Britain, Longman.

[9] t' Hart J., Collier R. & Cohen A. 1990. *A Perceptual Study of Intonation: An Experimental-Phonetic Approach to Speech Melody*. Cambridge, Cambridge Univ. Press.