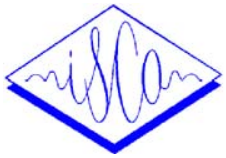
ISCA Archive
http://www.isca-speech.org/archive



Speech Prosody 2004
Nara, Japan
March 23-26, 2004

# Perceiving Prominence and Emotion in Speech – a Cross Lingual Study

*Noam Amir, Bat-Chen Almogi, Ronit Gal*

Department of Communications Disorders, Sackler School of Medicine,
Tel-Aviv University
noama@post.tau.ac.il

## Abstract

Suprasegmentals in general, and the pitch contour in particular, contain a large amount of information pertaining to gestural intentions and the emotional state of the speaker. In this study we compare perceptual identification tasks of prominence and inquiry on one hand, and anger, on the other hand, as performed by two separate groups: a group of native Hebrew speakers, and a group of native Arabic speakers, who have acquired Hebrew as a second language. All of the perceptual tests were carried out on Hebrew speech. Analysis of the results revealed near categorical perception of prominence for native speakers only. Overall, native speakers identified prominence more readily than non-native speakers. Concerning anger, both groups identified nearly the same subset of utterances as expressing anger, though the Arabic speakers consistently rated them as having a lower *degree* of anger.

## 1. Introduction

Suprasegmental features of the speech signal convey a large amount of non-textual information, such as gestural information, prominence, and emotion. Though these are expressed through a complex interplay of pitch, duration and intensity, it can be stated that pitch is an important component, if not the primary one, in this interaction.

Prominence, which causes one word to stand out as a focus of a sentence, is usually achieved by heightening of pitch and by lengthening the duration of stressed syllables [1,2]. This is true of many languages, and has been shown to be true for Hebrew also [3]. Terken [2] has shown that raising of pitch is the most important such indicator, though recent findings of the authors [4] show that it is not always present. On the other hand, there is no agreed-upon value of pitch increase that is necessary to create the perception of prominence. Grant [5] sets this threshold between 10 and 18 Hz, whereas Rosen & Fourcin [6] quote 7 to 12 Hz, depending on slopes in the pitch contour and the average pitch of the speaker.

It is widely agreed upon in the literature that emotional information is also conveyed by suprasegmental features, mainly pitch. Murray and Arnot [7] gave a qualititative overview of the influence of various emotions on pitch, though there does not yet exist a definitive study on the subject. The manifestations of emotions in speech have been discussed across many languages [8, 9, 10, 11 and many more], using various methods of elicitation [8, 10, 12, 13], various classification systems of emotions, and different methods of verifying the emotional content [8, 14]. Cross lingual issues, on the other hand, have barely been touched upon as of yet.

In the present study we examine the correlation between the perception of 1) prominence and 2) inquiry, which are relatively simple and straightforward to quantify, and 3) anger, which is far more difficult to quantify acoustically. Our primary goal was to compare the perception of the above three types of non-verbal information in Hebrew speech, between two groups of listeners: native Hebrew speakers, and native Arabic speakers, speaking Hebrew as second language. We wished to observe whether a difference between these two groups in the perception of a measurable characteristic such as prominence, will carry over to a more complex phenomena such as anger.

## 2. Objectives, assumptions and hypotheses

Several objectives guided the approach taken in this study. On the most basic level, we set out to determine the threshold of pitch rise in a certain word that would be necessary to create the impression of narrow focus. Our assumption was that this threshold would depend on the familiarity of the speaker with the language, and therefore that this threshold would be higher for non-native speakers.

Concerning inquiry, or question-type intonation, we wished to find the threshold for rising terminal pitch that would create this impression. It was not clear at the outset whether the universality of this type of expression would cause the threshold to be similar for native and non-native speakers.

Next, assuming that expression of anger is in some ways similar to the expression of prominence – i.e. it is often found to be accompanied by large pitch range and sharp peaks in the pitch contour – we wished to examine whether there would be a correlation between performance on the simpler intonation tasks and the identification of angry utterances. Comparison here can be carried out on an individual level, and also on a group level – native vs. non-native speakers.

## 3. Method

The perceptual experiment consisted of two sub-experiments, described below, carried out on two groups of subjects.

### 3.1. Subjects

The subjects were 22 female students in the dept. of communication disorders, Tel-Aviv University. 12 were native Hebrew speakers, 10 were native Arabic speakers who had acquired Hebrew on an academic level. The latter had been exposed to Hebrew during at least 11 years, and did not require a preparatory course in Hebrew before the commencement of their studies. Ages ranged between 20 and 28, and all subjects had normal hearing.

10.21437/SpeechProsody.2004-86

### 3.2. Intonation experiment

The subjects were presented with multiple repetitions of two three-word sentences, which translate to: "the boy bought bread" (*hayeled kana lexem*), and "I'll sit in the kitchen" (*ani eshev bamitbax*). Subjects heard twenty variants of the first sentence, in random order, and then 20 variants of the second sentence, in random order. Each sentence was recorded once by a male speaker, with broad focus. Using PRAAT software, the intonation contour was then manipulated to obtain 20 variants:

- Narrow focus: to create narrow focus on each word, the pitch of the stressed syllable was raised in 10 Hz increments, from 10 to 50 Hz.
- Inquiry: inquiring intonation was created through a final rise, in increments of 20 Hz, from 20 to 100 Hz.

After each presentation, the subjects were asked if they noticed whether any additional meaning was present in the utterance, according to 5 possible responses (for the first phrase):

1) No additional meaning
2) The boy, not the girl, bought bread
3) The boy bought, not ate, bread
4) The boy bought bread, not cheese
5) The boy bought bread?

Similar responses were constructed for the second phrase.

### 3.3. Emotion identification experiment

24 utterances taken from televised political talk shows were presented to the speakers in random order. These utterances were used in a previous study [15] on anger also. The utterances were judged by the authors to contain various degrees of angry speech, and neutral speech. Subjects were requested to mark whether the utterances expressed anger, no emotion, or "other". If they judged them to contain anger, they were asked to rate its degree on a scale from 1 (mild anger) to 4 (intense anger).

One difficulty in this experiment is deciding on a "correct" identification. Since the utterances were not taken from acted speech, some form of decision had to be made as to the definition of a consensual agreement. In a previous study it was decided that any utterance whose sum of scores was above half the maximum would be considered angry. This type of decision is in fact biased against utterances that would be judged mildly angry, even by all listeners. Instead, we chose as angry, all utterances that were judged angry by more than 75% of the native Hebrew speaking listeners.

## 4. Results

### 4.1. Identification of prominence

Several statistical analyses were performed on the results of the perceptual study. For the intonation experiment, a comparison of percent-identified vs. degree of pitch change was carried out for each group of listeners separately. The results are presented in figure 1. The native Hebrew speakers exhibited categorical perception, with a significant change in perception level occurring between a pitch rise of 20 Hz and 30 Hz. The non-native speakers did not exhibit this behavior, having the percent-identified rising gradually from 22% at 10 Hz to 72% at 50 Hz. On the other hand, at all pitch rises the native speakers judged a significantly larger number of utterances as having narrow focus. As a group, for all pitch rises, the native speakers judged significantly more utterances as having narrow focus – 64%, vs. 52% for the non-native speakers.

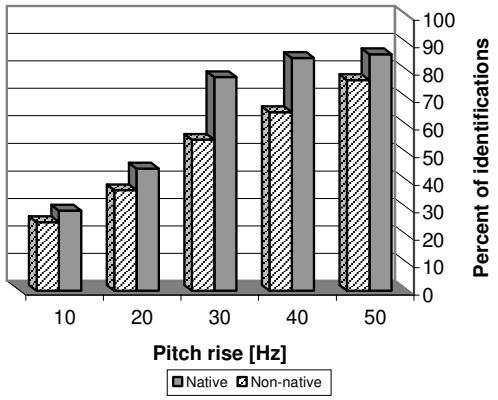



Figure 1: *Percent-identified of prominence*

### 4.2. Identification of questions

Figure 2 presents the results for identification of inquiry. We observe that even for pitch rises as large as 100Hz, not all subjects perceived the resultant intonation as representing a question. Though the graph suggests categorical perception, statistical analysis did not reveal this to be significant. Also, identification by non-native speakers was higher than that of native speakers. Overall the non-native speakers rated 51% of the utterances as questions, vs. 40% for the native speakers, yet this difference was also not found to be statistically significant. This was due to the large variability in the results of this experiment.

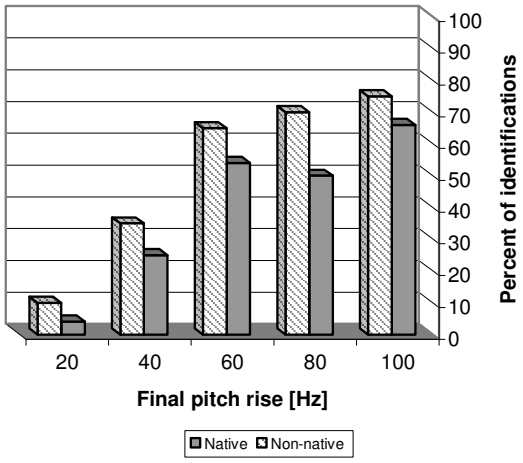



Figure 2: *Percent-identified of inquiry*

### 4.3. Correlation between prominence and question phrases

One of our initial conjectures was that subjects which identified prominence more readily, would also identify questions more readily. A Pearson correlation was calculated between the results of these two identification tasks, for each subject and for each group. No significant correlation was found.

### 4.4. Identification of emotions

The judgment of the native Hebrew speaking group was taken as a baseline judgment. Applying the criterion discussed above – i.e. agreement between 75 percent of the listeners - 8 out of the 24 phrases were judged as angry. Interestingly, this was in very good agreement with the judgment of the non-native speakers: The latter judged 7 out of these 8 phrases as angry.

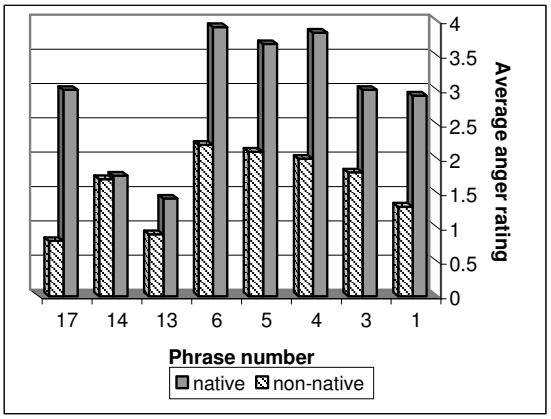



Figure 3: *Average ratings of angry phrases*

A separate comparison was performed on the average rating of *degree* of anger, appearing in figure 3. Though the two groups judged nearly the same group of phrases as angry, the native speakers rated nearly all of these phrases with a significantly higher degree of anger.

## 5. Discussion

It is interesting to note that the native Hebrew speakers exhibited a behavior very similar to categorical perception in the judgment of prominence. The significant jump in judgment of prominence between 20 and 30 Hz of pitch rise is slightly above that found by Grant [5], who set this at 10-18Hz. Evidently, this threshold, and even the manner in which prominence is perceived is language dependant. The non-native speakers did not exhibit categorical perception, and the percentages for this group were consistently below those for the native speakers.

It should be stressed that the threshold suggested above is not the threshold for perception of changes in pitch. A pilot study revealed that listeners could detect changes in pitch of less than 10 Hz, but needed a larger change in order for it to be construed as lending prominence.

It is possible that over a larger sample, the non-native speakers would exhibit categorical perception also. It is clear though that the non-native speakers needed a larger pitch rise to perceive prominence, and this can be attributed either to their unfamiliarity with the language, or to differences in Hebrew and Arabic prosody. This is an issue which certainly merits further attention, especially since there is a dearth of studies on Arabic prosody.

The perception of phrases as inquiries, vs. final pitch rise, did not exhibit categorical behavior for either group. As a rule, larger pitch rises were necessary here than for prominence, and interestingly – there was not found to be a significant difference between the groups. It seems that the template for inquiring intonation is similar enough between the two languages so that identification was equally easy for native and non-native speakers.

The difference between results for prominence and inquiry is strengthened by the correlation test carried out for individuals. No significant correlation was found between the identification of prominence and the identification of inquiry, carried out on an individual or group basis. In other words, subjects who identified prominence more readily, did not necessarily identify questions more readily, and vice versa.

The results for identification of anger were particularly interesting. Though both groups chose nearly identical groups of phrases for *presence* of anger, the native Hebrew speakers rated the *degree* of anger significantly higher. This suggests that there is a strong degree of universality in the manner in which anger is expressed. The differences in the average ratings could be explained in several ways: it could be that nuances conveying the degree of anger are easier for native speakers to perceive, or simply that expression of the degree of anger is different to a certain extent in Arabic prosody.

It should be noted that although Hebrew and Arabic are both Semitic languages, having similar syntax and morphology, there are many differences on phonological and prosodic levels between the two languages. Arabic has been spoken continuously in the Middle East for at least a millennium, whereas Hebrew is a revived language. Being revived mainly by European immigrants, the phonological and prosodic aspects demonstrate a strong European influence.

## 6. Conclusions

The results of this study demonstrate that there is a certain correlation between the difficulties encountered by non-native speakers in different prosodic identification tasks. The various ways in which these difficulties are manifested can be surprising, on the other hand. To summarize the main conclusions: Identification of prominence was more difficult for non-native speakers, whereas identification of questions was equally easy for native and non-native speakers. Since the expression of anger has no relation to inquiring prosody, and is more similar to the expression of prominence, it was expected that non-native speakers would also have more difficulties in identifying anger. In fact, they turned out to be equally proficient in identifying the presence of anger, but consistently judged it to be weaker anger than the native speakers.

The results of this study have interesting implications. Evidently, a very good familiarity with a second language, even in a minority population exposed to this second language on a daily basis, does not ensure the ability to interpret the nuances of emotion (or even prominence) conveyed by native speakers of this language. Clearly, this can be a source of misunderstanding and discomfort in interaction between the two types of population. Further study could further clarify the issue, to determine whether the differences between the two groups were due to differences in the characteristic prosody of Hebrew and Arabic, or to the fact that the non-native speakers were simply less accustomed to Hebrew. One possibility is to carry out a "symmetric" study, using Arabic stimuli and having the following two groups of listeners: native Arabic speakers, and native Hebrew speakers with Arab as a second language.

## 7. References


[1] Cooper, W.E., Eady, S.J., Mueller , P.J., 1985, Acoustical aspects of contrastive stress in question-answer contexts, *J. Acoust. Soc. Am.* 77(6), 2142-2154.

[2] Terken, J., 1991. Fundamental frequency and perceived prominence of accented syllables. *J. Acoust. Soc. Am.* 89(4), 1768-1776

[3] Mixdorff H., Amir N. 2002, The prosody of Modern Hebrew – a quantitative study, *Proceedings of Prosody 2002, Aix en Provence*

[4] Amir N., Silber-Varod, V., Izre'el S., Characteristics of intonation unit boundaries in spontaneous spoken Hebrew – perception and acoustic correlates, *forthcoming*

[5] Grant, K.W., 1987, Identification of intonation contours by normally hearing and profoundly hearing –impaired listeners, *J. Acoust. Soc. Am.* 82(4), 1172-1178

[6] Rosen, S., Fourcin, J. 1986, in: *Frequency selectivity in hearing* Moore, B.C.J.(ed), Orlando: Harcourt Brace Jovanovic, 392-475.

[7] Murray I.R., Arnott, J.L. 1993. Toward the simulation of emotion in synthetic speech: A review of the literature on human emotion. *J Acous Soc Am* 93:1097-1108

[8] Amir N., Ron S., Laor N. 2000, Analysis of an emotional speech corpus in Hebrew based on objective criteria, *Proceedings of ISCA workshop on speech and emotion, Belfast*. 65-69

[9] Engberg I, Hansen A, Andersen O, Dalsgard P 1996. Design, recording and verification of a Danish emotional speech database.Proceedings *Proceedings of ICSLP 1996* pp 1695-1698

[10] Kehrein, R. 2002. Prosodie und Emotionen, Tuebingen: Niemeyer

[11] Scherer, K.R. 1986. Vocal affect expression: A review and a model for future research, *Psychological Bulletin*, 99 (2) 143-165

[12] Douglas-Cowie, E., Cowie, R., Schroder, M., 2000, A new emotion database: Considerations, sources and scope, *Proceedings of the International ISCA Workshop on Speech and Emotion, Belfast*

[13] Scherer, K.R., Ceshi, G., 2000, Studying affective communication in the airport: The case of lost baggage claims, *Personality and Social Psychology Bulletin*, 26(3) pp. 327-339

[14] Cowie, R., Douglas-Cowie, E., Savvidou, S., McMahon, E., Sawey, M., Schroder, M. 2000, Feeltrace: an instrument for recording perceived emotion in real time, *Proceedings of the International ISCA Workshop on Speech and Emotion, Belfast*

[15] N. Amir, S. Ziv, R. Cohen 2003, Characteristics of authentic anger in Hebrew Speech, *Proceedings of Eurospeech 2003, Geneva, Switzerland*