Proceedings of Fonetik '97, PHONUM,
Reports from the Department of Phonetics Umeå University
No. 4, May 1997, pp. 121–124.

Some Observations on the Influence of F_0 and Duration to the Perception of Prominence by Swedish Listeners

Per Sautermeister and Robert Eklund
Per.G.Sautermeister@telia.se, Robert.H.Eklund@telia.se
Telia Research AB, Spoken Language Processing, SE–136 80 Haninge, Sweden

Abstract

Experiments have been conducted that deal with prosodic prominence in reiterant speech in order to determine the relative contribution of F_0 and duration to the perception of prosodic prominence by Swedish listeners. F_0 and duration were manipulated independently on different syllables in the stimuli. The results show that F_0 is considered primary cue by most subjects. Furthermore, duration only does not seem to be a sufficient cue to the perception of prominence to many of the subjects.

Introduction

Pitch, duration, and to some extent intensity, are known to contribute to the perception of prosodic prominence. Previous studies have shown that F_0 is the most important factor at play in prosodic prominence for languages such as Scottish English (Currie 1980; Currie 1981), American English (Maeda 1976), British English (Halliday 1970; de Pijper 1983) and Dutch (van Katwijk 1974). In Gussenhoven and Blom (1978) the relative contribution of F_0, duration and intensity to the perception of prominence was examined for Dutch listeners. In a similar study by Lehiste and Fox (1993), Swedish listeners were used, but their study dealt with duration and intensity only. In this study, F_0 and duration are studied independently in order to determine their relative contribution to the perception of prosodic prominence in Swedish.

Method

An utterance consisting of three nonsense syllables /mamama/ was recorded in neutral tone. Out of this utterance, the middle /ma/ was extracted in order to avoid effects typical of initial and final syllables. The extracted syllable was duplicated into a five-syllable utterance /mamamamama/ with an equal syllable duration of 379 ms and a total duration of 1.895 s. The utterance was then resynthesized with pitch and duration modifications, resulting in high quality speech stimuli.

In order to study the relative contributions of F_0 and duration, two experiments were created where these parameters were independent variables. F_0 was varied without any durational changes on the same syllable. Likewise, duration was varied without any F_0 changes on the same syllable. Six different syllable durations were used: 379 ms (100% of original length), 455 ms (120%), 531 ms (140%), 606 ms (160%), 682 ms (180%) and 758 ms (200%). F_0 variations were realized with a hat shaped accent on six different excursion levels, equidistant on an ERB-rate scale (see Hermes and Van Gestel 1991): 0.0 E, 0.3 E, 0.6 E, 0.9 E, 1.2 E and 1.5 E. In Terken (1991) it is shown that the declination of an utterance affects the perception of F_0 peaks. The stimuli in this study were resynthesized with a declination, with an initial value of 120Hz and a final value of 100Hz.

In the first experiment, an intra-utterance comparison, F_0 and duration were varied independently on the second and the fourth syllable of the same utterance. F_0 and durations

were manipulated in both positions and randomized. This gave a total of 72 stimuli. The subjects were asked to decide which syllable they found to be the linguistically most prominent. Subjects thus were given six alternatives: any of the five syllables o*none*.

In order to avoid the possibility of effects caused by the position of the manipulated syllable, a second experiment, an inter-utterance comparison, was created, where F_0 and duration were varied in two different utterances. In both utterances the second syllable was manipulated. All stimuli were presented in all pairwise combinations, which resulted in a total of 72 stimulus pairs. The subjects were asked to decide in which utterance they found the second syllable linguistically most prominent. Subjects thus were presented with three alternatives: any of the two utterances or *none*.

There were 22 subjects, all of whom were native Swedes. The subjects had no formal education in phonetics or linguistics. Each subject carried out the experiments using a graphical user interface on a computer. The subjects could listen to the stimuli as often as desired. Written and oral instructions were given to the subjects prior to the experiments.

Results

Experiment 1

The results do not appear to show any differences due to positional effects, i.e., whether the F_0 excursion was located on the second or the forth syllable, and vice versa, for the duration lengthening. The results show that the majority of the responses fall on the syllable on which F_0 was manipulated (hereafter referred to as F_0 *syllable*). The total amount of responses are distributed as follows: F_0 *syllable*: 67.7%, *duration syllable*: 13.6%, *none*: 16.5%, *syllable 1*: 0.9%, *syllable 3*: 1.1% and *syllable 5*: 0.3%. Only 2.3% of the total number of responses fall into the categories *syllable 1*, *syllable 2* and *syllable 5*, which can be considered unexpected responses. These responses are considered to be unsignificant and are not dealt with any further. The *none* responses derive mainly from stimuli where there were no or small modifications to F_0 and duration. Figure 1 and Figure 2 show the percentages of responses that fall in the F_0 and duration syllable categories, respectively, for different values of F_0 excursion size and duration lengthening. The results show that F_0 is by far the more important prosodic parameter to signal prominence. However, the results also seem to indicate that different strategies can be used. Three of the subjects seem to rely primarily on duration when duration gets long enough.

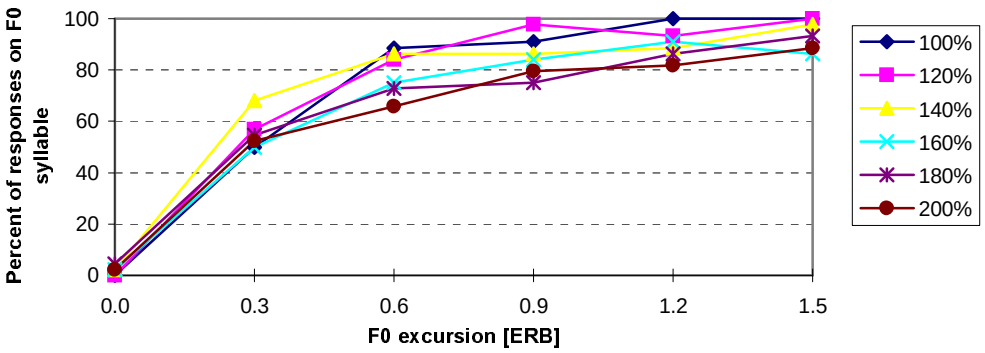

Figure 1. The percentage of responses to the F_0 manipulated syllable being the most prominent as a function of F_0 excursion size. The different lines indicate different durational lengthenings on another syllable.

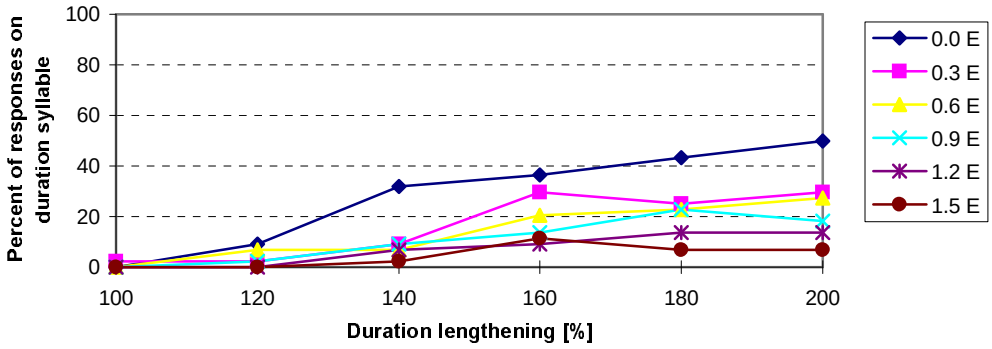

Figure 2. The percentage of responses to the duration manipulated syllable being the most prominent as a function of durational lengthening. The different lines indicate different F_0 excursion values on another syllable.

Experiment 2

There are no major differences between experiment 2, where the subjects knew which syllable was manipulated, and experiment 1, where the subjects were not given that information. Some subjects commented that experiment 2 was an easier task than experiment 1. The distribution of the total amount of responses is as follows: F_0 *syllable*: 71.4%, *duration syllable*: 12.9% and *none*: 15.7%.

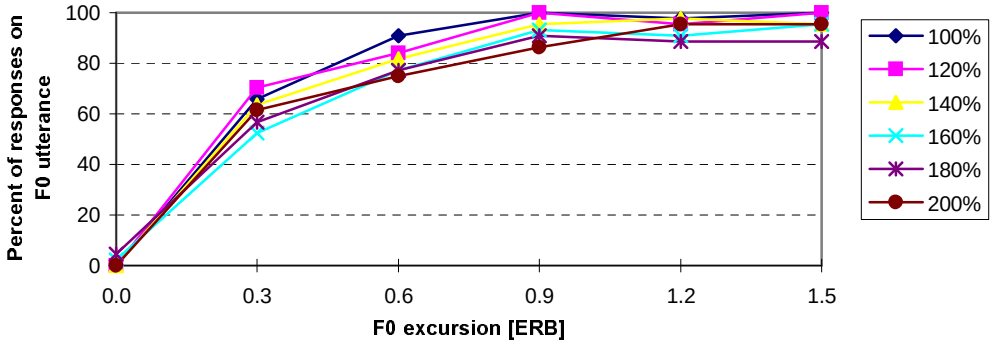

Figure 3. The percentage of responses to the second syllable of the F_0 manipulated utterance being the most prominent as a function of F_0 excursion size. The different lines indicate different durational lengthenings in the other utterance.

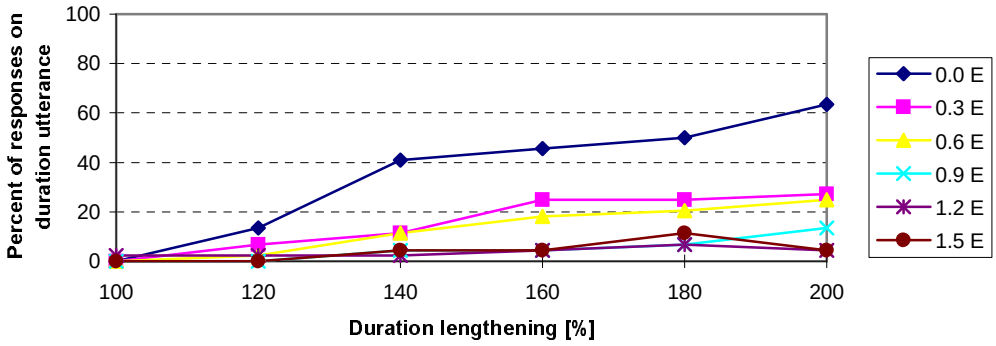

Figure 4. The percentage of responses to the second syllable of the duration manipulated utterance being the most prominent as a function of duration lengthening. The different lines indicate different F_0 excursion values in the other utterance.

Discussion

The results seem to strenghten the general notion that there are "primary" and "secondary" cues in the perception of prosodic prominence. F_0 was considered the primary cue by an overwhelming majority of the subjects.

Heldner (1996) argues that F_0 is not a necessary cue to the perception of prosodic prominence, and that listeners, lacking F_0 cues, can make use of durational or other information to perceive syllables as prominent. In this study, it is shown that if both F_0 and duration are present, and "competing", F_0 is considered the more important cue by most listeners. When no F_0 excursions were present, there was a tendency among the subjects to perceive the syllable with duration lengthening as prominent. However, far from all subjects exhibited this behaviour. Thus, in utterances with unchanged F_0 and manipulated durations, many subjects responded that no syllable was linguistically prominent, even for long durations. This would seem to indicate that, to these subjects, duration does not take on the signalling role, and that F_0 *is* indeed a necessary cue to the perception of prosodic prominence.

References

Currie, K. L. 1980. An initial search for tonics.*Language and Speech* 23: 329–350.

Currie, K. L. 1981. Further experiments in the 'search for tonics'. *Language and Speech* 24: 1–28.

Gussenhoven, C. and Blom, J. G. 1978. Perception of Prominence by Dutch Listeners. *Phonetica*, Volume 35: 216–230.

Halliday, M. A. K. 1970. *A Course in Spoken English: Intonation*. Oxford University Press, Oxford.

Heldner, M. 1996. Is an F_0-rise a necessary cue to perceived focus in Swedish*? Nordic Prosody VII*. (Forthcoming.)

Hermes, D. J. and Van Gestel, J. C. 1991. The frequency scale of speech intonation. *Journal of the Acoustical Society of America*, Volume 90: 97–102.

Van Katwijk, A. 1974. *Accentuation in Dutch*. Van Gorkum, Assen.

Lehiste, I. and Fox, R. A. 1993. Influence of duration and amplitude on the perception of prominence by Swedish listeners. *Speech Communication*, Volume 13: 149–154.

Maeda, S. 1976. *A Characterization of American English Intonation*. MIT Press, Cambridge.

de Pijper, J. R. 1983. *Modelling British English Intonation*. Foris Publications, Dordrecht.

Terken, J. 1991. Fundamental frequency and perceived prominence of accented syllables. *Journal of the Acoustical Society of America*, Volume 89: 1768–1776.