Umeå University, Department of Philosophy and Linguistics
PHONUM 9 (2003), *181-184*
Available online at http://www.ling.umu.se/fonetik2003/

Relating perceptual judgments of upcoming prosodic breaks to F0 features

Rolf Carlson (1) and Marc Swerts (2)*

*(1) CTT, KTH, Sweden (2)University of Tilburg, The Netherlands and Universitaire Instelling Antwerpen, Belgium *Names in alphabetic order*

The paper reports on a study of perceptually based predictions of upcoming prosodic breaks in spontaneous Swedish speech materials. The question tackled here is to what extent listeners are able, on the basis of prosodic features, to predict the occurrence of upcoming boundaries, and if so, whether they are able to differentiate different degrees of boundary strength. To answer these questions, an experiment is conducted in which spontaneous utterance fragments (both long and short versions) are presented to listeners, who are instructed to guess whether or not the fragments are followed by a prosodic break, and if so, what the strength of the break is. Results reveal that listeners are indeed able to predict whether or not a boundary (of a particular strength) is following the fragment.

1. Introduction

In a recent paper, Carlson and Swerts (2003) a listener-oriented approach to prosodic boundaries is described. The specific hypothesis tested in that study is that speakers not only encode prosodic breaks locally at the places where they occur (e.g. in the form of silent pauses), but that they also pre-signal these breaks in advance. In the current paper we will summarize this report on perceptually based predictions of upcoming prosodic breaks and present some additional analysis of possible acoustic correlates. The work is carried out within the Swedish project "Boundaries and groupings - the structuring of speech in different communicative situations" (GROG). The objective of this project is to model the structuring of Swedish speech in terms of prosodic breaks and groupings (Carlson et al., 2002).

Earlier work has shown that listeners are not only sensitive to the absence or presence of a boundary, but that it also matters how "strong" the boundary is. For instance, a few phonetic studies that focused on the exact nature of the prosodic cues that lead to the perception of a break, consisted of experiments in which listeners were asked to rate the prosodic boundary strength on a given scale (e.g. Dutch: Sanderman, 1996; Swedish: Strangert, and Heldner, 1995; Fant et al. 2000; Hansson, 2003). The results of these studies reveal that perceived boundary strength is heavily dependent on the occurrence of a silent pause, even to the extent that it may overrule the contribution of other parameters. In addition, we know from previous work on prosody modeling that there are indeed (phonetic) features which presignal upcoming breaks (e.g. Swerts et al., 1994; Ferrer et al. 2002).

We have conducted a variant of the gating paradigm, basically an experiment in which spontaneous Swedish utterance fragments are presented to listeners, who are instructed to guess whether or not the fragments are followed by a break, and if so what its strength is.

2. Experiment

The speech corpus, was selected from one interview of a female politician (GS) that was originally broadcast on public Swedish Radio. After the entire interview was prosodically labeled by three independent researchers in the project (Heldner and Megyesi, 2003), 60 utterance fragments (each about 2 seconds long) by GS were selected. The exact initial cutting point was moved to the nearest word boundary, whereas the final cutting point was fixed. The fragments all preceded the word "och" (and) in their original context, and the fragments were cut right before the silent interval (if any) before that word. The choice to use the word "och" was partly syntactically motivated, given that the fragments then all occurred in comparable syntactic positions before an identical conjunction. The fragments differed regarding the presence or absence of a break in between the end of the fragment and the word "och", i.e., as annotated by our independent labelers by a majority voting procedure: in about one third of the cases, there was a strong intervening break, one third of the fragments preceded a weak break and one third was followed by no break at all. From these longer fragments, we then constructed short versions consisting of only the final word of the fragment. The 120 different stimuli (long and short versions, preceding a strong, weak or no boundary) were mixed and presented sequentially to our listeners (13 students in logopedics from Umeå university) via a specifically designed interface, which allows to run perception experiments through the internet using a standard web browser with audio facilities. To minimize possible learning effects, each subject was presented with a differently randomized list of stimuli. Their task was to rate, for each stimulus, on a 5-point scale whether they felt that the fragment preceded no boundary (1), a strong boundary (5), or a boundary having a strength in between these two extremes (2-4). The actual test was preceded by a short introduction which briefly explained a few concepts (such as prosodic boundary) and the actual task. No feedback was given on the "correctness" of their responses, and there was no interaction with the experimentors. During the test, subjects could listen as many times as needed to a given stimulus before giving an answer, but they could not return to a previous stimulus after a response had been entered.

3. Results

In Figure 1a the results from the perceptually based prediction experiment are presented. Since each word stimulus also can be found as part of a 2 seconds fragment it is possible to correlate the perceptually based prediction of upcoming prosodic breaks based on different sized context. Figure 1b shows that there is a significant correlation (r = 0,89) between the two fragment sizes.

According to Hansson (2003) the intonation can be regarded to be a secondary perceptual feature for boundary strength prediction compared to pause duration. Since the pause duration feature by default is missing in our case it is interesting to study the relation between some intonation cues and the judged boundary strength. The word sized fragments were acoustically analysed in terms of presence/absence of final creak, using spectrographic analysis, Figure 2a, and the median f0 value of the last voiced 50 ms of the word, Figure 2b. A small but significant correlation between the final f0 value and the boundary strength was found, (r=0,62.) Several other tested f0 cues turned out to have less predictive power.

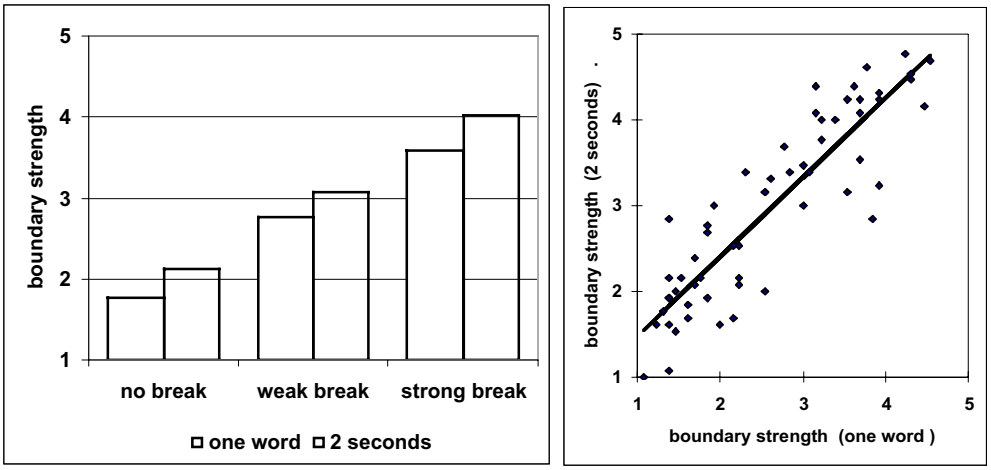

Figure 1. a) Perceived upcoming boundary strength. Data grouped according to boundary strength and fragment size. b) Correlation between perceived upcoming boundary strength for each word in isolation and in a 2 seconds fragment. Regression coefficient r = 0,89.

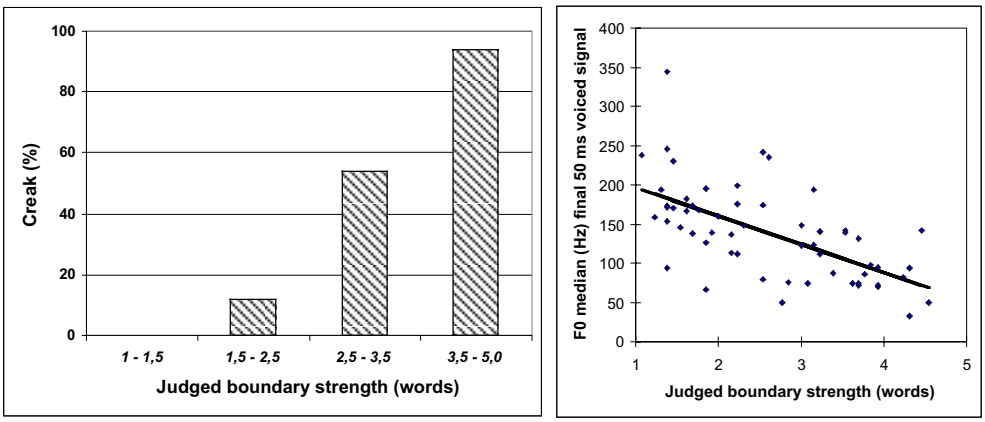

Figure 2. a) Number of stimuli with creaky voice (in %) for different judged boundary strength intervals. b) F0 median of last voiced 50 ms and judged boundary strength.

4. Discussion and Conclusion

The results of our current study show that a listener is able to predict a possible upcoming break, based on properties of the preceding context alone. One of the interesting findings is that the responses for the two types of stimuli, namely 2-sec fragments and 1-word stimuli, are not fundamentally different, as is clear from the high correlation between the two sets of responses. Yet, there is an overall difference between the responses in that the longer context has significantly higher values for all three classes (no boundary, weak boundary, strong boundary). The finding that the overall pattern for the two sets of stimuli is essentially the same implies that we cannot conclude that longer context leads to a higher amount of "correct" responses. At first sight, this may seem a surprising outcome, as one might have

expected that the task of guessing an upcoming boundary would be easier for 2-sec stimuli, given that for these stimuli, subjects literally have more speech materials at their disposal for making a decision. Our contrary finding suggests that the final word contains important prosodic and syntactic features that facilitate the prediction of upcoming breaks. It is clear that some of the important boundary predictors may indeed be located in the final word, including features like type of boundary tone preceding the break, final lengthening, loudness patterns and possible effects of voice quality (e.g. the amount of creakiness).

Similarly, the one word stimuli have some linguistic information in terms of parts-of-speech information which can be of value for the prediction. This leaves us with the question as to what the strength relationship is in cue value between the prosodic and syntactic features for predicting upcoming boundaries. We conjecture that the ability to predict prosodic boundaries is primarily based on acoustic cues and can not be over-ruled by a break prediction based only on syntactic features. On the other hand the syntactic structure probably has a predictive power on where a break is placed and acoustically realized.

5. Acknowledgments

Marc Swerts is also affiliated with the Fund for Scientific Research – Flanders (FWO - Flanders). We would like to thank Theo Veenker for help with setting up the experimental environment and the members of the GROG team for prosodically labeling, useful discussions and cooperation.

6. References

Baron, D., Shriberg, E., Stolcke, A. (2002) Automatic Punctuation And Disfluency Detection In Multi-Party Meetings Using Prosodic And Lexical Cues, *ICSLP 2002*, Denver, USA.

Carlson R, Granström B, Heldner M, House D, Megyesi B, Strangert E, Swerts M (2002). Boundaries and groupings - the structuring of speech in different communicative situations: a description of the GROG project. *Proc of Fonetik 2002*, TMH-QPSR, 44.

Carlson R, Swerts M (2003) Perceptually based prediction of upcoming prosodic breaks in spontaneous Swedish speech materials, *Proc. ICPhS 03*.

Fant G, Kruckenberg A, Liljencrants J (2000) Acoustic-phonetic Analysis of Prominence in Swedish. In A Botinis (ed.), *Intonation, Analysis, Modelling and Technology* (Kluwer)

Ferrer L, Shriberg E, Stolcke A (2002), Is the speaker done yet? Faster and more accurate end-of-utterance detection using prosody, *ICSLP – 2002*, Denver, USA

Hansson P (2003) *Prosodic Phrasing in Spontaneous Swedish*. Travaux de l'institut de linguistique de Lund 43, Dept. of Linguistics and Phonetics, Lund University, Sweden.

Heldner M, Megyesi B (2003) Exploring the prosody-syntax interface in conversations, *Proc. ICPhS 03*.

Sanderman, A. (1996). *Prosodic phrasing. Production, perception, acceptability and comprehension*. PhD thesis, Eindhoven University of Technology

Strangert E, Heldner M (1995) Labelling of boundaries and prominences by phonetically experienced and non-experienced transcribers. In *PHONUM 3*, pp. 85-109. Umeå: Department of Phonetics, Umeå University.

Swerts M, Collier R, Terken J (1994). Prosodic predictors of discourse finality in spontaneous monologues. *Speech Communication* 15, 79-90.