Prosodic Labelling and Acoustic Data

Eva Strangert and Mattias Heldner
Department of Phonetics, Umeå University
S-901 87 Umeå, Sweden

ABSTRACT

Data on the labelling of boundaries and prominences in read and spontaneous speech have been collected from ten non-expert and one expert transcribers and analyzed for their inter-subjective variability. The labellings are matched with acoustic data to explore the relevant cues used by the transcribers.

INTRODUCTION

Most work on prosody relies on some kind of labelling of the prosodic features investigated. In the ideal case the labelling is reproducible and independent of the specific transcriber. Everyday experiences as well as more systematic studies, however, indicate that this is very often not the case. For example, Lieberman (1965) and Brown, Currie and Kenworthy (1980) have shown that even phonetically experienced transcribers vary considerably in prominence labelling tasks on English speech material. Bannert (1993) reports similar findings for Swedish in a study based on spontaneous speech that included both phonetically experienced and inexperienced subjects.

These observations form the basis for the work presented here, the aim of which is to learn about the variability in labellings of boundaries and prominences. To that end analyses of labellings made on both read and spontaneous speech by a number of phonetically inexperienced subjects were undertaken. The investigation will also include comparisons with phonetically trained transcribers to highlight questions like: To what extent and how do expert and non-expert transcribers differ? Is the variation among expert transcribers similar to non-expert variation? To date we have labellings made by one highly qualified expert transcriber. These data were compared to the non-expert labellings.

The study also aims at taking the first steps to explore the acoustic/percepual basis for the variability. Relating prosodic labelling to acoustic data could give hints as to the cues used by the transcriber. We are, for example, interested in the answers to such questions as: How are boundaries and prominences signalled to the listener? Are there alternative cues? Which are the most robust cues, that is, how should boundaries and prominences be signalled to be perceived - and labelled - consistently by several transcribers?

METHODS

10 non-experts (phonetically untrained university students) and 1 expert (highly qualified in prosodic transcription) listened to and made labellings of two recordings rendered by the same male Swedish speaker. One was an excerpt (233 words long) from an authentic news cable read aloud and the other an excerpt (252 words long) of spontaneous speech, a retelling of the story in the read-aloud speech. All tests were run individually.

The non-experts marked the boundaries they perceived as either strongly (=2) or weakly (=1) signalled. Cases without a boundary were left unmarked (=0). Three prominence levels of words were distinguished, extra prominent (=2), prominent (=1), and not prominent (left unmarked, =0). The figure within parentheses is the respective coding used for each category. The expert made similar markings for boundaries. For prominences four categories were distinguished and coded accordingly to enable direct comparisons with the non-experts: focussed (=2), primary stressed (=1), secondary stressed (=0,5) and unstressed (=0).

Acoustic analyses were conducted using ESPS Waves+ and SoundScope software. The analyses included oscillograms, fundamental frequency tracings and spectrograms.

LABELLING BY NON-EXPERTS

Table 1 shows the labelling of boundaries and prominences respectively, by 10 non-expert transcribers in a sample of the read speech material.

Table 1. *Labelling of a sample of read speech by 10 non-experts (T1-10). The individual labellings are given with the mean score and standard deviation for each word. Above: labelling of boundaries: 2=strongly marked, 1=weakly marked, 0=no boundary. Below: labelling of prominences: 2=extra prominent, 1=prominent, 0=not prominent.*

Word	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	mean	sd
enligt	0	0	0	0	0	0	0	0	0	0	0	0
libyska	0	0	0	0	0	0	0	0	0	0	0	0
uppgifter	1	1	0	0	1	0	0	1	1	0	.5	.527
föll	0	0	0	0	0	0	0	0	0	0	0	0
åtta	0	0	0	0	0	0	0	0	0	0	0	0
450-kilosbom..	1	2	1	1	1	1	1	1	1	0	1.0	.471
enligt	0	0	0	0	0	0	0	0	0	0	0	0
libyska	2	0	0	0	2	2	2	2	1	1	1.2	.919
uppgifter	1	0	0	0	0	1	1	1	0	0	.4	.516
föll	0	0	0	0	0	0	0	0	0	0	0	0
åtta	2	1	2	1	2	0	2	2	2	2	1.6	.699
450-kilosbom..	2	0	2	0	2	2	1	1	1	1	1.2	.789

The samples seem to indicate a greater consistency, i.e. agreement between transcribers, in the labelling of boundaries as compared to prominences. This is confirmed by an analysis of the entire read speech material and the same is true of the spontaneous speech. Similar observations for spontaneous speech have been made by Bannert (1993). Moreover, comparing the read speech and the spontaneous, the labelling of boundaries on the one hand is more consistent in the read speech, while prominence labelling is similarly inconsistent in the two speaking styles. These conclusions may be inferred from the dispersion data. The means of the standard deviations for the four conditions studied are: .03 (bound, read), .18 (bound, spont), .37 (prom, read) and .36 (prom, spont).

COMPARING NON-EXPERT AND EXPERT LABELLING

We correlated the mean scores of the non-experts' word labellings (that is, means as in Table 1) with the corresponding expert scores for each word. A regression analysis revealed a positive, approximately linear relationship for boundaries and prominences in both the read and spontaneous speech. However, the correlations vary among the different conditions in terms of explained variance; the squared correlation coefficients (r^2) were .95 (bound, read), .61 (bound, spont), .58 (prom, read) and .54 (prom, spont), respectively. Thus the best agreement between the expert and non-expert labelling was found for boundaries in read speech, while the agreement was the least for prominences in spontaneous speech. The analyses for read speech, boundaries and prominences, are shown in Figure 1.

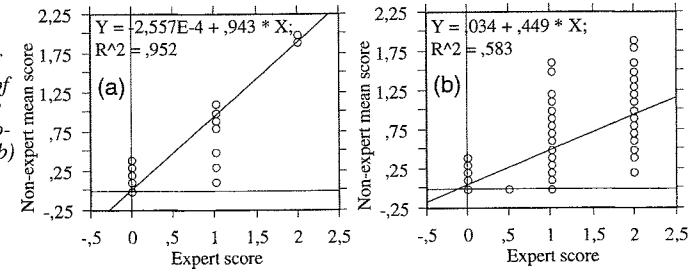

Figure 1. *Regression analyses of labellings of boundaries (a) and prominences (b) in the read speech.*

The reason for the disagreement, the unexplained variance, may be inferred from Table 1. The variance includes variation among individuals, among words and interaction effects.

Concluding, it is evident that with increasing boundary/prominence scores by the expert, the non-experts generally increase their scores accordingly, though the non-experts and the expert to some extent use different scales . The regression coefficients vary between .94 (bound, read) and .44 (prom, spont) indicating that in the latter case the non-experts use higher scores less frequently than the expert. That is, the expert is less restrictive in judging words to be prominent than the non-experts.

ACOUSTIC DATA

Here we present examples of data with bearing on the relation between the acoustics and strong prominence labelling, focussing on the expert versus (average) non-expert labelling. Figure 2 shows oscillograms and Fo curves for five words in their respective contexts. They have been chosen because they are representative of groups of words that have been given a high score by both the expert and the non-experts, the expert alone, or the non-experts alone. In analyzing the data we are particularly interested in finding out if, and to what extent, they agree with the characteristic pattern for strong prominence (sentence accent) hypothesized in the intonation model for Swedish presented by Bruce (1977). This pattern is an Fo rise following a word accent fall, timed differently in words with the grave and acute accent respectively.

Considering the five words presented in Figure 2 (a-c grave accented, d-e acute accented), there are apparent reflections of the model, though not in all words. (a) is an example of a word with high expert and non-expert scores combined with the acoustic characteristics predicted by the model. (b) scored high by a majority of the non-experts, but was not recognized as strongly prominent by the expert. This word does not conform to the model, as there is almost no sentence accent rise. (c) contains a sentence accent rise. It was scored 2 by the expert, but was given a low score by the group of non-experts. One would hypothesize that the small rise passed unnoticed by the non-experts. Similarly, the moderate rise in (d) might be the basis for the low non-expert score. (e) finally, just like (a), obtained high scores from both the expert and the non-expert. In this word the sentence accent rise is very evident, although the word accent fall is missing.

Conclusions based on the data presented here will have to be postponed till more detailed analyses have been undertaken. In particular, the match between the prominence scores and the Fo data will be analyzed systematically rather than in the qualitative fashion reported above. Possible contributions of other cues than Fo will also be considered as well as the possibility that different individuals may rely on different cues to prominence. The same might be true also for the match among the acoustics and the perception and labelling of boundaries, which will be analyzed in a similar fashion.

ACKNOWLEDGEMENTS

We are grateful to Gösta Bruce for prosodic labelling and discussions about principles for transcribing prosody, and to Martin Englund for carrying out acoustic analyses for this study. The work has been carried out within the Language Technology Program under a grant from the Swedish Council for Research in the Humanities and Social Sciences.

REFERENCES

Bannert, R. 1993. 'Macro-prosodic organization of spoken Swedish: some preliminary observations on focus accent and chunking'. *PHONUM 2, Reports from the Department of Phonetics, University of Umeå*, 21-39.

Brown, G., Currie, K. L. & Kenworthy, J. 1980. *Questions of intonation*. London: Croom Helm.

Bruce, G. 1977. *Swedish word accents in sentence perspective*. Lund: Gleerup.

Lieberman, P. 1965. 'On the acoustic basis of the perception of intonation by linguists'. *Word* 21, 40-54.

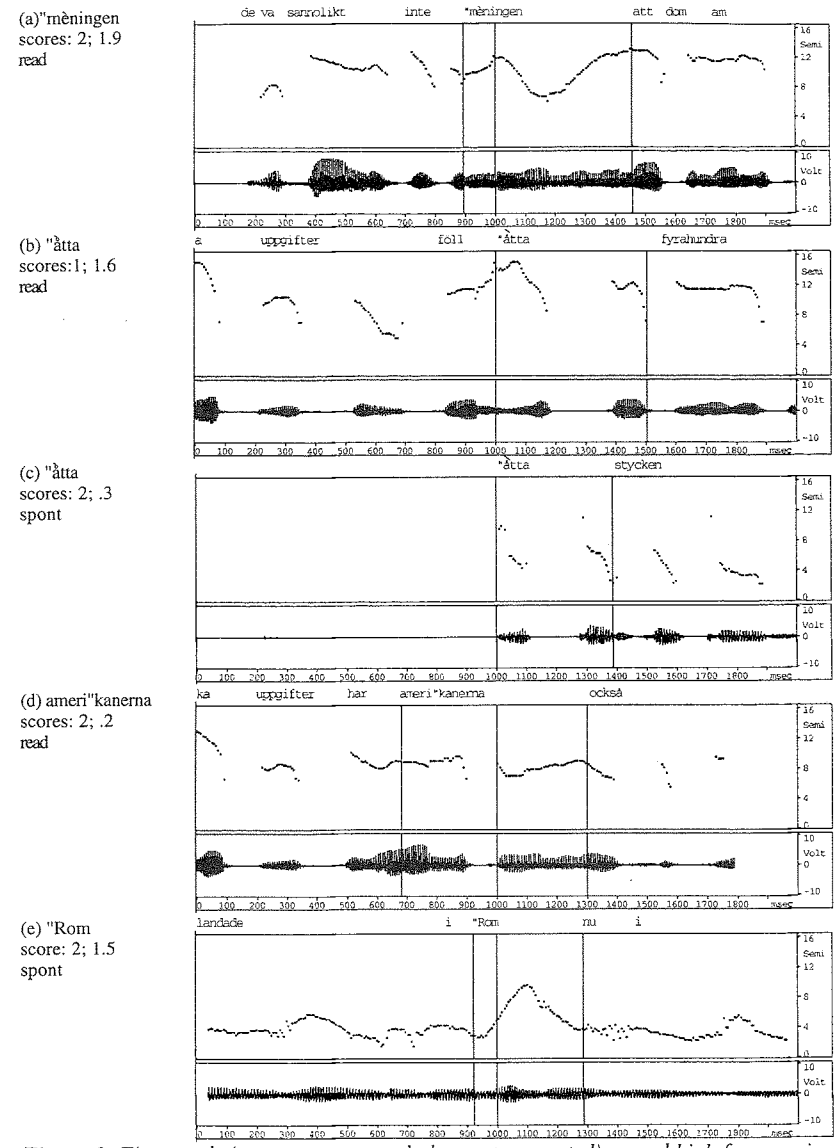

Figure 2. *Five words (a-c grave accented, d-e acute accented) scored high for prominence by both the expert and the non-experts (a, e), the non-experts alone (b) or the expert alone (c, d). Expert (first) and mean non-expert (second) scores and speaking style (read/spont) are indicated for each word. Time alignment: onset of stressed vowel. Additional vertical bars are positioned at the beginning and end of the word.*