THE LABELLING OF PROMINENCE IN SWEDISH BY PHONETICALLY EXPERIENCED TRANSCRIBERS

Eva Strangert and Mattias Heldner
Department of Phonetics, Umeå University, Sweden

ABSTRACT

An IPA-based system has been agreed upon for labelling Swedish prosody. In the present study this system is evaluated by assessing the inter-transcriber reliability in prominence labelling of nine expert subjects. The study also explores the acoustic (F0) basis for observed variability in the assignment of focus accent, the highest prominence label.

INTRODUCTION

Recently, as large corpora of prosodically labelled speech are needed for quantitative computational modelling of speech, great efforts are being taken to develop transcription systems meeting high standards on reliability. Thus, before extensive use of a system is initiated, it must be evaluated. The TOBI (TOnes and Break Indices) system developed for transcribing English prosody has been evaluated in a number of studies eg. [1,2]. Reyelt [3] evaluated a number of variants of prosodic transcription for German within the VERBMOBIL project. For Swedish, an IPA-based system has been agreed upon for labelling prosody (prominence and boundary phenomena), the details of which have been described in [4]. We have used this system in two studies [5,6] comparing the labelling of boundaries and prominences in spoken Swedish made by phonetically experienced and non-experienced transcribers.

In the present study, the scope has been widened. One purpose, which it shares with the former studies [5,6], is to evaluate the transcription system used for labelling. In particular we want to estimate the extent to which experienced phoneticians and speech researchers vary in their labelling of prominences when presented with samples of read and spontaneous Swedish. In addition, the study aims at exploring the acoustic basis, specifically F0-characteristics, for the variability in labelling that we predict will occur. In particular, we want to establish the extent to which the variability associated with the assignment of focus accent is explainable in terms of F0-cues.

Beckman [7] reviews the research on acoustic correlates to perceived stress in English. Referring to study [8], Beckman [7, p 60-62] makes clear that the dependence of perceived stress on F0-cues is complex, and varies with the position of the word in the sentence. Further, Wells [9] concludes that F0-cues play an important role for perceived prominence in English, although various other cues contribute, too. Although F0 is not assumed to be the only cue to prominence in Swedish – Bruce [10] also mentions temporal correlates, and there are also data reported in [11] indicating temporal correlates – it is believed to be an important determiner of focus accent. Thus, relating perceived focus accent to F0-events seems reasonable in the light of previous research [12] according to which focus accent is intimately tied to a F0-rise following a word accent F0-fall timed differently for words with acute and grave accent, respectively.

EVALUATION OF THE TRANSCRIPTION SYSTEM

Method

The 9 subjects participating in the study are all phoneticians or speech researchers with wide experience in prosody from different sites in Sweden. All are native-born Swedes.

The subjects transcribed two kinds of recorded speech material. One was an excerpt, 233 words long, from an authentic news cable read aloud. The other was a 252-word-long excerpt of spontaneous speech, a retelling of the story read aloud. Both recordings were made in a soundproof room and rendered by the same male Swedish speaker.

Each expert was sent the recorded material and instructions for labelling prominence according to the IPA-based Swedish system. Following this, four levels of prominence were distinguished and labelled accordingly for each word in the material: no stress (unmarked), secondary stress (,), primary stress/accented (') and focus acccent (").

Subsequent analyses included coding the data (no stress=0; secondary stress=1; primary stress=2; focus accent=3) and statistical analyses to estimate reliability.

Labelling data

Table 1 shows the labelling of prominences by the nine experts in a sample of the read material. The words in the text are ordered vertically in the first column. The following nine columns contain the individual labellings of the transcribers and the tenth column the means of these labellings for each word. The data presented give a rough indication of the reliability of labelling.

Table 1. Labelling by nine transcribers. 0=no stress, 1=secondary stress, 2=primary stress, 3=focus accent.

Word	Transcribers 1 — 9								$\bar{X}$
enligt	0	0	1	0	0	0	0	0	0	0.1
libyska	3	2	3	3	3	3	2	3	3	2.8
uppgi…	2	2	2	2	2	2	2	2	2	2
föll	0	0	2	0	0	1	0	0	0	0.3
åtta	2	3	3	2	3	3	2	3	2	2.6
450-k…	2	2	3	3	3	2	2	2	2	2.3
över	0	0	1	0	0	0	0	0	0	0.1
Tripoli	2	2	3	3	3	3	2	3	3	2.7
och	0	0	0	0	0	0	0	0	0	0
Bengazi	2	3	3	2	3	2	2	3	2	2.4
när	0	0	0	0	0	0	0	0	0	0
de	0	0	0	0	0	0	0	0	0	0

Inter-transcriber reliability

Generally, reliability concerns the extent to which measurements are repeatable in a variety of conditions. Within this framework we will consider two aspects, the one concerning the extent to which the transcribers covary, that is, give relative labelling values that are correlated, and the other concerning the extent to which the transcribers give identical labels. We will henceforth refer to the first as 'reliability' and the second as 'agreement'. All computations are made with acute and grave accent words pooled.

The inter-transcriber reliability (Cronbach's alpha) for prominence is .98 for read and .97 for spontaneous speech (difference not significant). That is, the transcriptions are highly reliable in the sense of relative labelling consistency irrespective of the material.

To determine the reliability in the more strict sense of agreement, that is identical matching, we used the same test as Silverman et al. [1] and Pitrelli et al. [2]. They calculated the agreement across all possible pairs of transcribers for each word of each utterance labelled. The index was calculated as the average percentage of agreeing pairs and, according to the criterion set in [1], the agreement should be at least 80%. Calculated on our data, this index is 78% and 71%, for the read and spontaneous speech respectively, thus indicating a somewhat higher agreement on the read speech. There are several differences between TOBI and our system which make comparisons complicated. For the TOBI transcribers, the task was to decide whether a word had a pitch accent or not, and if so, what kind of pitch accent. The indices reported for these tasks were 86% and 64% respectively for the 4 most experienced of their 20 transcribers.

We also calculated an index estimating the extent to which *all* the transcribers made *exactly* the same judgements on each word. A detailed account of these calculations and other evaluation data presented here are given in [6].

F0 IN RELATION TO PROMINENCE LABELS

Method

The subsequent analysis was made on 60 acute and 55 grave accent words judged to be focussed (that is, having a prominence degree of 3, according to our coding) by two or more of the nine transcribers. For each of these words a prominence mean score based on the labelling of all nine transcribers was calculated.The words were digitized at 44.1 kHz. Measurements were made in both the read and spontaneous speech of the size of the word accent fall and the focus accent rise.

To calculate the falls and rises four measuring points were defined, primarily on the basis of the F0 tracings, see the illustrations in Figure 1: (1) The beginning of the word accent fall; the highest point in the word accent fall. (2) The end of the word accent fall; the lowest point in the word accent fall. (3) The beginning of the focus accent rise; the lowest point in the focus accent rise. For acute accent words this point coincides with (2). For grave accent words it either coincides with (2) or, in the case of longer words, may be located at som distance from (2).

(4) The end of the focus accent rise; the highest point in the focus accent rise. In a few cases in which the critical F0-events were not easily located, additional criteria were used, determined on the basis of the patterns observed in the unequivocal cases. We also used [13] as a reference when deciding on these additional criteria.

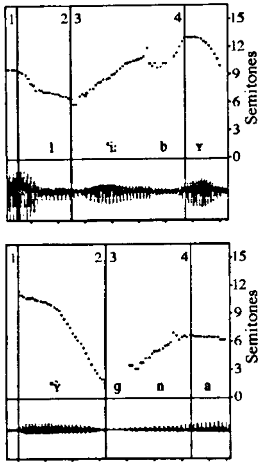

Figure 1. Measurement points. Above : underlined portion of libyska (acute accent); below: underlined portion of byggnad (grave accent).

The word accent fall is defined as the difference between points (1) and (2) and the focus accent rise is defined as the difference between points (3) and (4) measured in semitones. In addition we tested two other F0-parameters, differences (focus accent rise–word accent fall) and ratios (focus accent rise/word accent fall).

Results

The majority of the prominence mean scores for all acute and grave accent words included in the analysis fell in the range between 2 and 3. (It should be recalled that a word judged to be focussed is coded as 3 in our analysis. Therefore, mean scores close to 3 indicate a general agreement on the word as being focussed.) The prominence mean scores were then used in multiple regression analyses to determine if, and to what extent, the measured F0 movements (with word accent fall and focus accent rise as the independent variables) could explain the variability in the prominence scores.

The results demonstrate insignificant effects of the word accent fall in the read as well as the spontaneous speech and for words with acute and grave accent alike. The focal accent rise, on the other hand, is significantly correlated with perceived scores ($p<.05$) both in the read and spontaneous speech and for acute and grave accent words (Figure 2 a-d). That is, the greater the size of the rise, the stronger the agreement on focus accent. Both kind of data therefore corroborate previous results demonstrating greater effects on perceived prominence of the rise than the fall [11,12]. However, the R-square values, correlations in terms of explained variance, are quite low for all four regression models, .14 and .26 for the acute accent and .41 and .38 for the grave, indicating other influences than F0 on perceived prominence cf. [11].

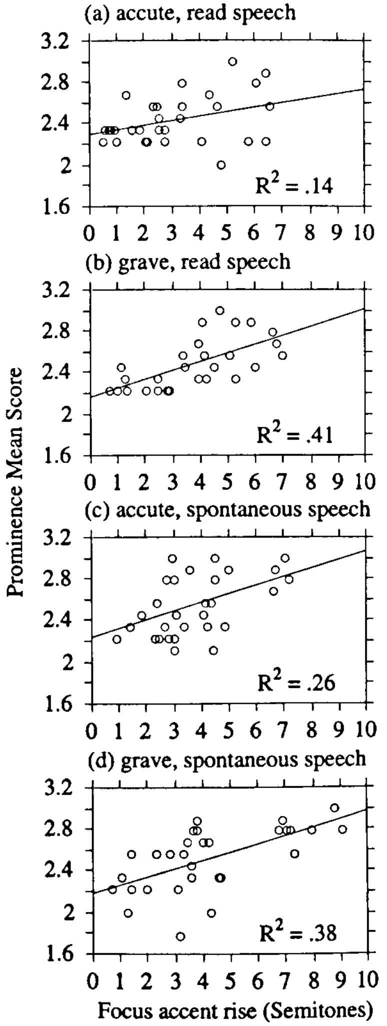

Figure 2. Regression analyses of size of focus accent rise and prominence mean score for 60 acute and 55 grave accented words in read and spontaneous speech.

We also did regression tests with differences as well as ratios between the focus accent rise and the word accent fall as independent variables, but neither of them reached significance.

CONCLUSIONS

In this prosodic transcription evaluation we have demonstrated the capacity of the system as used by expert transcribers. The reliability is high as well as the intertranscriber agreement. Exploring the acoustic basis for observed variability associated with the assignment of focus accent, we found that the greater the F0-rise, the stronger the agreement on focus accent. That is, the size of the focus accent cues the degree of prominence. Yet it explains only part of the variation. In conclusion then, there are other important cues to perceived prominence (focus accent) than those investigated here. We are in the process of conducting a study including temporal as well as other cues to perceived focus accent.

ACKNOWLEDGEMENTS

We gratefully acknowledge the contributions of Gösta Bruce, Rolf Carlson, Claes-Christian Elert, Anders Eriksson, Gunnar Fant, Eva Gårding, Olle Kjellin, Anita Kruckenberg, Per Lindblad, Ulla Sundberg, without whose help this study would not have been possible.

This research was supported by grants from the Swedish HSFR/NUTEK Language Technology Programme.

REFERENCES

[1] Silverman, K., Beckman, M., Pitrelli, J., Ostendorf, M., Wightman, C., Price, P., Pierrehumbert, J. & Hirschberg, J. (1992), "TOBI: A standard for labeling English prosody". In *ICSLP 92 Proc.* pp. 867-870, Banff, Alberta, Canada.

[2] Pitrelli, J., Beckman, M. & Hirschberg, J. (1994), "Evaluation of prosodic transcription labeling reliability in the ToBI framework". In *ICSLP 94 Proc.* pp. 123-126. Yokohama, Japan.

[3] Reyelt, M. (1993), "Experimental investigation on the perceptual consistency and the automatic recognition of prosodic units in spoken German", Proc. ESCA Workshop on Prosody, *Working Papers*, 41, pp. 238-241, Dept. of Linguistics and Phonetics, Lund Univ.

[4] Bruce, G. (1994), "Prosodisk strukturering i dialog". In *Svenskans beskrivning* 20. pp. 9-23. Lund: Lund University Press.

[5] Strangert, E. & Heldner, M. (1994), "Prosodic labelling and acoustic data", *Working Papers*, 43, pp. 120-123, Dept. of Linguistics and Phonetics, Lund Univ.

[6] Strangert, E. & Heldner, M. (1995), "Labelling of boundaries and prominences by phonetically experienced and non-experienced transcribers", *PHONUM 3*, Dept. of Phonetics, Umeå Univ.

[7] Beckman, M.E. (1986), *Stress and Non-Stress Accent*, Dordrecht: Foris Publications.

[8] Nakatani, L. & Aston, C. (1978), "Acoustic and linguistic factors in stress perception", Unpubl. ms, Bell Lab.

[9] Wells, W.H.G. (1986), "An experimental approach to the interpretation of focus in spoken English", in C. Johns-Lewis (Ed.), *Intonation in Discourse*, 53-75, London: Croom Helm.

[10] Bruce, G. (1983), "Accentuation and timing in Swedish", *Folia Linguistica*, vol. 17, pp. 221-238.

[11] Sundberg, U. (1994), "Tonal and temporal aspects of child directed speech", Working Papers 43, pp. 128-131 Dept. of Linguistics and Phonetics, Lund Univ.

[12] Bruce, G. (1977), *Swedish Word Accents in Sentence Perspective*, Lund: CWK Gleerup.

[13] Engstrand, O. (1989), "Phonetic features of the acute and grave word accents: data from spontaneous speech", *PERILUS X*, pp. 13-37, Inst. of Linguistics, Univ. of Stockholm.