

AUDITORY ORGANIZATION OF PROMINENCE AND CHUNKING IN SPOKEN SWEDISH

Robert Bannert

Department of Phonetics, Umeå University, Sweden

ABSTRACT

In an investigation of the macro-prosodic organization of spoken Swedish, different aspects of the listeners' variation were studied. Two listener groups, students at the beginner's level and trained phoneticians, had to mark the most prominent words and the chunks they could hear in speech samples of spontaneous Standard Swedish. Variations concerning each prominent word and chunk, the number of scores per item and the number of scores per listener are presented.

INTRODUCTION

In the past, prosodic organization of speech has been studied mostly in texts read-aloud. In order to arrive at a theory of speech, it is imperative to investigate the macro-prosodic structure of spontaneous speech from a perceptual point of view. A research programme with this goal was launched some time ago [1, 2], continuing previous research [3, 4]. Two significant prosodic features were selected, the highest degree of prominence (focus accent) and chunking (phrasing).

Table 1. General distribution of the listeners' scores: prominence, chunking; students, experts; three categories. The first line gives the number of scores, the second line the percentage.

	Prominence					
	female speaker			male speaker		
scores	0	≤ 50%	> 75%	0	≤ 50%	> 75%
26 students	107	68	9	87	83	3
	55	35	6	47	45	2
5 experts	143	34	13	142	23	15
	73	17	7	76	12	8

	Chunking					
	female speaker			male speaker		
scores	0	≤ 50%	> 75%	0	≤ 50%	< 75%
29 students	135	46	6	125	44	8
	69	24	3	68	24	4
3 experts	155	20	13	151	17	13
	79	10	7	82	9	7

For this pilot investigation, aiming at the development of methodological insights into the study of perceptual modelling of the macro-prosodic organization of spoken Swedish, two samples of spontaneous (monologue) speech were used. They were produced by a female and a male speaker of Standard Swedish, about 40 years old, both with academic backgrounds. The speech sample of the female speaker contained 196 words, thus, there is a possible 196 possible votes for prominence and 195 votes for chunking. The speech sample of the male speaker contained 186 words and 185 possible chunk boundaries. Each speech sample had the duration of approximately one minute. The listeners had to mark the most prominent words and the chunking on a sheet of paper where the speech samples were given in orthographic representation. However, no punctuation marks were used. Four listener groups participated. A group of 26 students and of 5 trained phoneticians scored separately for prominence, another group of 29 students and 3 experts marked for chunking. The speech samples were presented from a loudspeaker four times.

LISTENERS' SCORES

First the general distribution of the scores will be given, followed by the marks for prominence and chunking.

General distribution

As a first rough measure of the distribution of votes (markings, scores) of the listeners, a simplified account of the data is given in Table 1. The complete data is to be found in [5].

From Table 1 it can be seen clearly that experts vote in a much more consistent way across all the categories and speakers than the students. This difference could be expected, although the instructions for both groups were formulated in an identical, though general way of expression. The experts, it has to be assumed, reformulated and defined the instructions in phonetic-prosodic terms which the students had not learned yet.

Prominence and chunking

The group scores for each word in the first part of the speech sample of the female speaker are shown, prominence in Figure 1, chunking in Figure 2. The results are representative for the rest of this speech sample and also for the male speaker's speech sample.

The histograms of Figure 1 show a rather large variation in the scoring of the listeners and listener groups. In some words, students and experts agree rather well and to a high degree, in other words they score quite differently.

Figure 2 shows the percentage of the chunking for the students and the experts. Even in this case, listeners vary considerably in hearing chunks.

VOTES PER WORD AND CHUNK

Instances of numbers of scores per word for prominence for the students are given in Figure 3 for the female and male speech samples. No word received the highest number of possible votes per word, namely 26. Contrary to the experts, the students show a rather even and low distribution over the whole range, except for the lowest part, 1 and 2 votes per word.

Even this observation can be interpreted in the same way as above, namely that experts are more consistent in their scores due to their knowledge of prosody and the acoustic correlates of

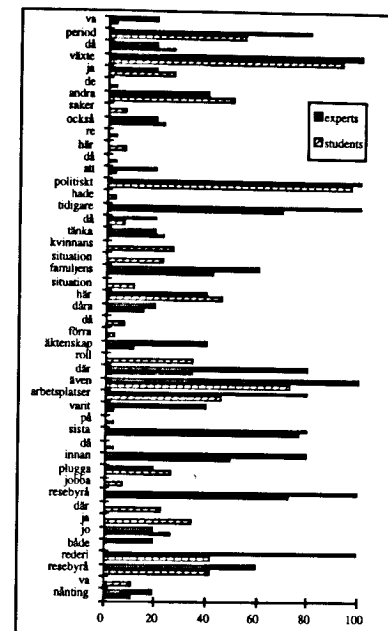


Figure 1. Percentage of total scores for the most prominent words by 26 students and 5 experts. Female speaker, first part of speech sample.

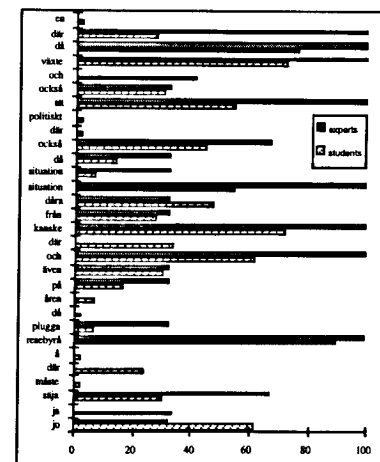


Figure 2. Percentage of the segmentation of the speech sample into chunks by 29 students and 3 experts. Female speaker, first part of speech sample.

prominence. However, it should be noted in any case, that experts, too, show

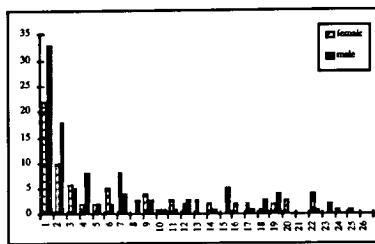


Figure 3. Instances of number of scores per word (prominence). Female and male speaker.

a relatively high degree of uncertainty, expressed in the rather high number of only one vote per word, especially with respect to the female speaker. For this bias, there does not seem to be an easy explanation because three of the experts were woman.

The scores for chunking appear to be similar for students and experts, i.e. both groups show a high degree of uncertainty. The student group gives the highest numbers of votes, 28 and 29 respectively, only to 5 chunks for the male speaker and only to 1 chunk for the female speaker. There were 40 chunks that received 1-28 votes (cf. Table 1). The experts give three votes, the highest number for this group, to 15 chunks out of 40 for the female speaker, and to 13 chunks out of 34 for the male speaker. Only one vote per chunk is given to 16 chunks for the female and to 14 chunks for the male speaker.

In comparison, the data suggest that students have great difficulties in recognizing the highest degree of prominence (focus accent) and chunking in spontaneous Swedish. Experts do better in recognizing focus accent. They, too, are rather bad at assigning chunk boundaries unanimously.

LISTENERS' VOTES

As an illustration of the individual variation among listeners, Figure 4 shows the scores for prominence for both speech samples and both groups. It is striking to observe how great the difference between raters can be. The lowest number of votes for prominent words in the texts, namely 6 votes, is given by listener no. 20 for the speech sample of the male speaker. This listener and listener no. 11 (7 votes) are very

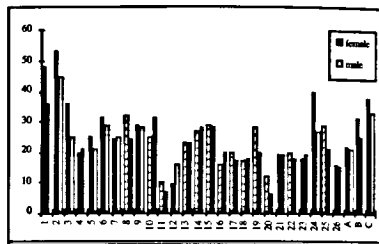


Figure 4. Individual distribution of listener scores for prominence. Students and experts, female and male speaker.

thrifty when they spend their votes. At the opposite end of the range we find listener no. 2 who gives 53 votes for the most prominent words of the female speaker. This means that one word out of four is heard as a most prominent word by listener no. 2, while listeners nos. 20 and 11 only hear one word out of 30 as most prominent.

The most striking aspect, even for chunking, is the great variation between the individuals. The average score for the students is 19.4 votes for both speakers. The experts' average score is almost 25 votes for both speakers.

AGREEMENT BETWEEN STUDENTS AND EXPERTS

In spite of all the inter- and intra-group variation, accounted for in this paper, there is one clear difference to be noted between the groups. In Table 1, it could be seen that experts, trained phoneticians specializing in prosody, score more consistently across varying conditions compared to students who have no training in prosodic theory and labelling, nor experience in carrying out such a listening test. An interesting question arises: How much agreement is to be found between the students and the experts in identifying prominence and chunking in spontaneous speech?

In order to give a quantitative answer to this question a statistical analysis, a simple regression analysis, was conducted, given the assumption of a linear relationship between the variables and their interdependency.

Only in one case, the scores for chunking of the male speaker, is the regression coefficient high, 0.97. In the other three cases, it is about 0.7. It can be interpreted such that there is not an

excellent, although a rather good agreement between the scoring of the students and the experts. However, when the coefficient of determination is taken into consideration, this interpretation has perhaps to be restricted. Although about two thirds of the variation is explained by the independent variable (0.647 - 0.688), in one case not even half of it is explained (0.453). Therefore, the two listener groups do not agree well in scoring prominence and chunking in spontaneous Swedish.

CONCLUSIONS

When the pioneering, and now classical, work by Gårding [7] was published, prosodic research at then time, it could be said, was in its infancy. Since then significant contributions to the understanding of prosody and its role in speech communication have been made. One great insight in the dimension of stress or prominence in Swedish was achieved by Bruce [6] where he demonstrated convincingly that the famous and puzzling Swedish word accents have to be isolated from focus or phrase accent. Focus accent that signalled the most prominent words depending on context is mainly characterized by a tonal rise following the word accent fall in Standard Swedish. This separate tonal rise is a very marked cue and is easy to be heard. Therefore it was expected that listeners would easily hear focussed word but would have difficulties to decide upon chunking.

At a first glance, the results of this study appear to point to the interpretation that focus accent and phrase boundaries are non-existent in spontaneous speech or that listeners organize spontaneous speech in quite different ways using maybe divergent strategies. However, there are strong reasons to believe in the opposite interpretation. Listeners process the speech flow by applying rather general macro-prosodic strategies. This does not mean of course that listeners would identify prominent words or chunks categorically. On the contrary, these prosodic features, opposed to segmental features like nasal or rounded, do not function in a binary fashion. It seems at this stage of research that hypothesis (1) obviously was not justified.

When the tonal manifestation of the focus accent is concerned, we know that the size of the rise may vary considerably in speech. However, a survey of the use of focus accent, its distribution and the variations in manifestation, in spontaneous speech of different varieties of Swedish is badly needed. This applies also to chunking. We know that silent interval, low F_0 and final lengthening, single or combined, are strong cues to phrase boundaries. Unfortunately, we still know very little about the rôle that voice quality, intensity and perhaps other features play for chunking.

References

- [1] Bannert, R. (1993), Macro-prosodic organisation of spoken Swedish: Some preliminary observations on focus accent and chunking. *Reports from the Department of Phonetics, Umeå University, PHONUM 2*, pp. 21-39.
- [2] Bannert, R. (1994), Listeners' Identification of Prominence and Chunking in Spoken Swedish: Variation and Consistency. *Working Papers, Department of Linguistics, Lund University 43*, pp. 10-13.
- [3] Bannert, R. (1987), From Prominent Syllables to a Skeleton of Meaning: a Model of Prosodically Guided Speech Recognition. *Proceedings of the 11th International Congress of Phonetic Sciences, Tallinn, USSR. 2*, pp. 73-76.
- [4] Bannert, R. (1991), Automatic Recognition of Focus Accent in German. *Journal of Semantics 6/7*. (Special Issue: Views of Focus. Proceedings of a Workshop on Focus, Intonation, and Semantics).
- [5] Bannert, R. (1995), Variations in the perceptual modelling of macro-prosodic organization of spoken Swedish: prominence and chunking. *Reports from the Department of Phonetics, Umeå University PHONUM 3*, pp. 7-30.
- [6] Bruce, G. (1977), Swedish Word Accents in Sentence Perspective. *Travaux de l'Institut de Linguistique de Lund XII*. Lund: Gleerup.
- [7] Gårding, E. (1967), Prosodiska drag i spontant och utpläst tal. In: Holm, G. (ed) *Svenskt talspråk*, pp. 40-85. Stockholm: Almqvist och Wiksell.