Listeners' Identification of Prominence and Chunking in Spoken Swedish: Variation and Consistency

Robert Bannert
Department of Phonetics, Umeå University
S - 901 87 Umeå, Sweden

ABSTRACT
In a pilot study, two groups of listeners, phonetically experienced and non-experienced, scored the most prominent words and the speech chunks in a short passage of spontaneous Swedish spoken by a male and a female Stockholm speaker. In addition to the previously reported results, different aspects of variation of the listeners' scores are investigated in order to highlight the special difficulties connected with naturally spoken Swedish.

INTRODUCTION
In spite of the vast amount of prosodic research conducted during the last few decades where data and knowledge about prosodic features and phenomena have been accumulated in different languages, including Swedish (Gårding 1967), there seems to be no acceptable **theory** of the large prosodic units of language. However, it is generally known that spontaneous as well as read speech is organized by the speaker into pieces or chunks of varying lengths that can contain several syllables, words, and groups of words. Natural speech is characterized by the fact that these chunks are grouped and kept together by prosodic means resulting in **grouping (phrasing, chunking)** and **binding (coherence)**. They can be termed the basic principles of macro-prosodic organization of speech. It is strongly believed that a very high degree of prominence, signalled as focus accent and pertaining to a higher level unit of prosodic organization of speech, contributes essentially to the coherence of discourse.

A number of phonetic-acoustic cues are used to express these two principles of macro-prosodic organization of speech: intonation (tonal features at boundaries and within the groups), final lengthening and speech tempo (temporal features), intensity, voice quality, and the spectrum of segments. Even these cues are subject to much variation, and such conditioning factors may be speech tempo, speech style, involvement, kind of text, individual characteristics, emotions, attitudes, etc.

As to the organization of speech into larger units than foot (stress group), i.e. prosodic phrases and still larger units, attempts have been made to postulate larger prosodic units departing from the syntactic structure of sentences as the unit of analysis. Today it is generally agreed that a sentence has an autonomous prosodic structure besides its syntactic, semantic, and pragmatic structure. It is also clear that there are links between these different structures of a sentence. During the eighties an increasing interest in semantics and pragmatics can be discerned within linguistics, which is true of sentence and discourse levels as well. This new interest was concentrated especially around focus accent and the information structure of discourse.

PROSODIC ORGANIZATION OF SPEECH
When surveying the literature, it becomes clear that phonetic research up to now has not devoted major efforts to describing the organization of spontaneous speech, i.e. the substance-driven analysis of its phonetic organization, its division into chunks, its macro-prosodic structure and the interplay between different levels of speech and language starting from the speech signal. Research was above all concerned with rules

in order to predict boundaries departing from the syntactic structure of sentences, and the distribution of focus accent as a function of different linguistic variables.

With some exceptions, phonetic research both in Sweden and abroad has not yet reached a satisfactory level of scientific sophisitication as the study of the organization of speech in chunks of different size and how these chunks are processed by the listener in speech perception (cf. Svensson 1974, Selkirk 1981) is concerned.

Principles of macro-prosodic organization
It is widely accepted that the macro-prosodic organization of speech in general, spontaneous and read aloud, is characterized by two main principles, known as **grouping** (phrasing, chunking) and **binding** (coherence). The signalling to the listener of the prosodically structured parts of speech is achieved by different acoustic means of expression. From a phonetic and communicative point of view, the prosodic features most investigated are pauses (silent intervals), boundary tones (the acoustic correlate of which is an Fo movement or an Fo level), and final lengthening. Even if most such research is done on English, a fair amount of knowledge has been - and is being - gathered on Swedish, too. However, it should be pointed out that the object of these investigations has not been spontaneous speech but texts read aloud (frame sentences and texts, often referred to as lab speech).

When listening phonetically and carefully to spontaneous speech, it becomes quite obvious that the macro-prosodic organization of this speech style comprises other features as well. Spontaneous speech includes not only acoustic features expressing linguistically, phonologically, and phonetically relevant elements, it also contains non-linguistic parts, e.g. various vocalization (sounds of hesitation, filled pauses, laughter, hawking, whistling, smacks, clicks, breathing, etc), false starts, and repairs.

Dealing with macro-prosodic structuring of spontaneous speech, especially in a dialogue or discourse situation, one is struck by the vast amount of events, linguistic and non-linguistic ones, that are to be found in the flow of the signal produced by the speaker. Moreover, one is compelled to look briefly at certain visual cues that the speaker uses and that accompany his prosodic gestures like stressing (accentuation, lending prominence to words and syllables) and phrasing. Therefore, in a somewhat widened perspective of viewing the macro-prosodic organization of spontaneous speech, the following tentative list should be considered an attempt to localize the different elements or cues that listeners (and speakers) may use in order to process (or to structure, respectively) sponanteous speech. In Table 1, the cues are listed according to the two basic principles of macro-prosodic organization, namely grouping and binding. Question marks appearing in several instances express uncertainty.

IDENTIFICATION OF PROMINENCE AND CHUNKING
A pilot study was carried out, focussing on the perceptual aspects of macro-prosodic organization of spontaneous speech samples of two Stockholm speakers (Bannert 1993). A non-experienced group of students and an expert group participated.

Listeners' scores
Considering the tonal rise of the focus accent as a stable cue to focus accent and thus a high degree of prominence, it could be expected that listeners have no cifficulty in recognizing a focussed word. In contrast to this, one would expect a great amount of uncertainty when listeners have to decide on speech chunks. This is a consequence of hesitations, corrections, and other features, even non-linguistic ones, which are characteristic of naturally executed speech.

The listeners' scores for focus and chunking, surmounting a threshold of 50%, are reported in Bannert (1993). For the sake of convenience, the identification scores for the most prominent words of the female speaker are repeated in Fig. 1. It can be seen that listeners, students and experts alike, are far from any consistent agreement as to

which word they hear as most prominent. It is even the case that students mark words more than 50% while experts do not (e.g. helskotta) and vice-versa (e.g. där, äktenskap). Thus it is quite obvious that listeners do not recognize the most prominent word or the focussed word unanimously. Instead, listeners display a great variability. This finding is contrary to expectation. So is the result for chunking.

Table 1. *Tentative list of phonetic cues or gestures involved in the two basic principles of macro-prosodic organization of spontaneous speech.*

GROUPING (local events)	BINDING (global events)
1. Pauses (silent intervals)	-- (?)
2. Boundary tones (initial/final)	Direction of Fo course, declination
3. Final lengthening	-- (?)
4. Word accents	Word accents
5. Rhythmic alternation (syllable weight)	Rhythmic alternation (syllable weight)
6. Fo of focus accent	Fo of focus accent
7. Tonal re-setting	Tonal range (width)
8. Phrase final decrease of intensity (loudness)	Variation of intenstity (loudness)
9. Change of speech tempo	Speech tempo (speed of articulation)
10. Voice quality	-- (?)
11. Non-linguistic events: (a) acoustic/auditory (b) visual (head nodding, hand strokes)	-- (?) visual (sweeping hand and body movements)

Looking at the listeners' scores for chunking, it seems that they show a higher degree of consensus. Here the experts score more unanimously than do the students. The high degree of agreement between listeners in marking chunking, in spite of all the interfering factors, is also to be found in another kind of speech material, namely texts read aloud (Strangert and Heldner, this volume). However, even as to the chunking scores, there is variability in the listeners' behaviour.

VARIATION AND CONSISTENCY

In order to prepare the research envisaged, it is considered a strong requirement to learn the details about the variation of the listeners' behaviour in their listening tasks. One cannot be totally satisfied just counting the total groups' scores and setting some appropriate threshold. On the contrary, it would be very interesting to study the variation in the litseners' behaviour. If some regularity in the scores of the groups and the individual listeners could be discerned, one could perhaps draw some significant conclusions as far as the processing of linguistic and non-lingusitic cues in the spontaneous speech material is concerned. At a later stage, the listeners' scores have to be correlated to the acoustic cues in the signal.

Thus I will study, in various respects, the amount of variation in the listeners' scores for prominence and chunking. Data on the inter- and intra-listeners' variability of the votes will be presented. First I will look at the **whole range of the variation** of the non-experienced student group and the experts (earlier only the scores above 50% were considered, as the focus of interest was on the listeners' agreement). Second, the **relationships** between the scores of the students and the experts on corresponding

words or chunks will be analysed. Finally, the **individual variation** of each listener, students and experts as well, will be investigated. It is hoped that this knowledge of the variation of the listeners' behaviour will provide us with new insights in the processing of spontaneous speech by human listeners.
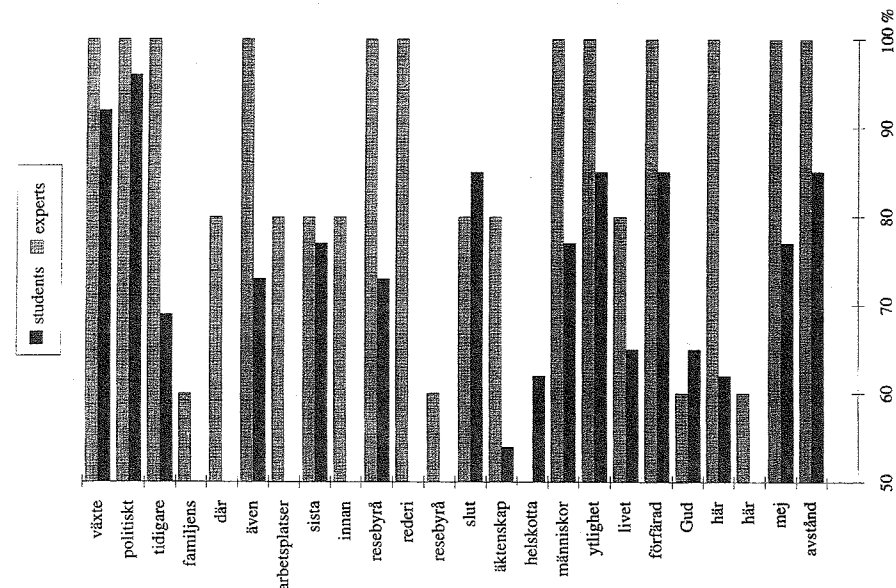

Figure 1. *Identification of the most prominent words by 29 students and five experts. Students scores of more than 50% are given as a histogramme. Female speaker.*

REFERENCES

Bannert, R. 1987. 'From Prominent Syllables to a Skeleton of Meaning: a Model of Prosodically Guided Speech Recognition. *Proceedings of the 11th International Congres of Phonetic Sciences, Tallin.* Vol. 2, 73-76. Expanded version in : Lund University, Department of Linguistics and Phonetics. Working Papers 29, 1-30.

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, Stuttgart.

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

Gårding, E. 1967. 'Prosodiska drag i spontant och uppläst tal'. In G. Holm (ed.), *Svenskt talspråk,* 40-85. Stockholm: Almqvist & Wiksell.

Strangert, E., and M. Heldner. 1994. Prosodic Labelling and Acoustic Data. This volume.