PROSODIC CHARACTERISTICS OF INFORMATION STRUCTURE IN SPONTANEOUS DISCOURSE IN DUTCH

Monique E. van Donzel
Institute of Phonetic Sciences/IFOTT, University of Amsterdam

ABSTRACT

This paper presents an overview of the results of our study on the prosodic aspects of information structure in spontaneous discourse. We recorded speech material of eight speakers of standard Dutch. They read aloud a short story, which they subsequently retold in their own words. The verbatim transcriptions of these retold versions were analyzed for discourse structure (boundaries and information status), using a purely text-based framework. These analyses are taken as a reference point to which acoustic realization and perceived structure are related.

The aims of the study are to find out what acoustic means are used by speakers to signal the structure of spontaneously spoken discourse, and how these cues are used by listeners to detect the structure of the message.

Results show that discourse boundaries are marked with high boundary tones, also at locations where a low tone was expected. Heavier boundaries are marked with longer pauses. Listeners use acoustic pauses more than boundary tones as a cue for discerning phrasing. Furthermore, there appeared to be an ordering in the percentage 'pitch accented' and 'perceived as prominent' for speakers and listeners relative to information status: new > inferrable > modifier > discourse marker > verb.

1. INTRODUCTION

Spoken discourse, as produced in everyday communicative situations, basically involves three different aspects: the speaker, the listener, and the message itself. *Speakers* may use various prosodic means to signal the structure of the message they are producing. They will mark certain words as more important than others, for instance by pitch accenting the important information. They will also divide the whole message into smaller parts, such as paragraphs and sentences. To indicate final or non-final boundaries, they may use boundary marking pitch movements and/or pauses. When listening to such spoken discourse, *listeners* have certain ideas about the structure of the incoming text. They perceive certain words or word groups as more important than others, while they are also able to detect different types of boundaries, such as sentence boundaries and paragraph boundaries [1]. The *message* itself, i.e. the text as produced by the speaker, also has a structure. Assuming that this message is more or less coherent, it can be divided into paragraphs, sentences, clauses, phrases, etc. Apart from prosodic means, the speaker also has a variety of linguistic means available to indicate the structure of the message.

The present paper presents an overview of the results from our study on the prosodic aspects of information structure in discourse [2]. The two main research questions are: I) What acoustic means are used by speakers to signal the structure of spontaneously spoken discourse?, and II) How are these cues used by listeners to detect the structure of the message?

In this study, we have concentrated on intonational and temporal aspects, more specifically boundary tones, pitch accents, and pauses in relation to information structure in spontaneous discourse.

2. METHODS

2.1 Speakers, discourse analysis, and listeners

Four male and four female speakers of Dutch were selected as speakers. They were all students or staff members of the Institute of Phonetic Sciences. The speakers were asked to read aloud a short story in Dutch [3]. After a short break they were asked to retell the same story in their own words, with as many details as possible. During the retelling a listener was present to create a more natural story-telling situation. This procedure resulted in eight spontaneously retold versions of the same story (hereafter 'retold version'). All recordings were made in an anechoic room on DAT-tape. The retold versions were stored as digitized audio files (sample rate 48 kHz, 16-bit precision).

Verbatim transcriptions of each of the eight retold versions were made by the author of the present paper. These transcriptions were subsequently analyzed for discourse structure, using a purely text-based framework 'Information Structure In Discourse (see [2] for further details and an elaborate discussion of the framework). On a global level, a distinction was made in paragraphs, sentences, and clauses. On a more local level, word groups were labeled according to their information status in new, inferrable, or evoked information. Furthermore, discourse markers and modifiers were labeled, as well as verbs.

In order to obtain perceptual judgements, twelve listeners, all students, participated in a listening test. They were asked to mark the structure of each of the retold versions, on the basis of the speech signal rather than on the basis of the text. The listeners had to indicate non-final, sentence final, and paragraph final boundaries, using conventional punctuation (',' for non-final, '.' for sentence final, '//' for paragraph final). They also had to underline which words or word groups they perceived as being emphasized by the speaker. The verbatim transcriptions without any punctuation were used as an answer sheet.

2.2 Measuring procedures

The location of boundary-marking and accent-lending pitch movements was obtained by presenting the spoken versions of

the retold stories to eight Dutch intonation experts, and having them indicate F0-movements in the matching verbatim transcriptions. They were asked to determine where in the discourse the speaker had realized pitch accents and boundary marking pitch movements. They furthermore had to indicate whether the boundary marking pitch movement was a high tone ('continuity') or a low tone ('finality'). The eight retold versions were randomly ordered and distributed over the experts in such a way that each speaker was evaluated by three different experts. The results were processed in the following way: two out of three experts had to agree on the location of a pitch accent in order for a word to count as accented, or for a boundary to be marked with a pitch movement. The type of boundary (high or low) was determined by the majority of judgements; and in case only two experts marked the boundary, and one marked it as high and one as low, it was labeled 'ambiguous'.

Pauses were measured directly in the speech signal. The minimum duration for a pause was 150 ms, to insure that closure time of stop consonants were excluded. In case of filled pauses, the filler as well as the preceding and following silence was included in the pause.

For each perceived discourse boundary, the perceptual boundary strength (PBS) was determined. The judgements given by the naive listeners are taken as a reference point. PBS is computed relative to the number of listeners and the type of perceived finality (1 point for each non-final judgement, 2 for each sentence final, and 3 for paragraph final judgement). PBS values are clustered into three groups: Weak, Strong, and Extra Strong perceived boundaries. Our method of determining PBS differs from the one used by [7], but reflects the same idea.

2.3 Hypotheses

On the basis of results from similar experiments for Dutch [5,6] in relation to our framework, we formulated hypotheses about the relation between information structure and prosodic realization. Table 1 at the end of this paper gives a schematic overview.

3 DISCOURSE BOUNDARIES

3.1 Structural discourse boundaries

For each discourse boundary (clause, sentence, or paragraph; as determined by the textual discourse analysis) we checked whether it was accompanied by a boundary tone (either high of low) and/or a pause (either silent or filled). Four strategies are thus possible to mark boundaries: no tone & no pause, no tone & pause, tone & no pause, and tone & pause. Figure 1 presents the distribution of temporal and intonational strategies used by the speakers. Due to space limitations, only the data are presented across speakers. For individual data see [2].

The heavier the boundary, the more both a boundary tone and a pause are used. *Clause* boundaries are frequently not marked with any of the two prosodic events (38%; no tone, no pause). If a boundary is marked, it is mostly done by a pause (24%) or a tone (24%). Relatively few cases are marked with both a tone and a pause (15%). *Sentence* boundaries are in only few cases not marked by any prosodic event (5%). A majority of the cases is marked with both a tone and a pause (38%). If only one cue is used, this is predominantly a tone (35%) rather than a pause (22%). *Paragraph* boundaries are always marked prosodically, with either a pause (38%), a tone (11%), or both (51%). In most cases they are marked with both cues. If only one cue is used, it is a pause rather than a tone, contrary to what we saw for sentence boundaries.

Figure 1 does not give information about the type of boundary tone or pause used by the individual speakers. This was investigated separately. Here it is of importance to know that predominantly high tones were used by the speakers, also to mark sentence and paragraph boundaries. This is contrary to what we had expected. Furthermore, mainly silent pauses were used to mark boundaries (rather than filled pauses).

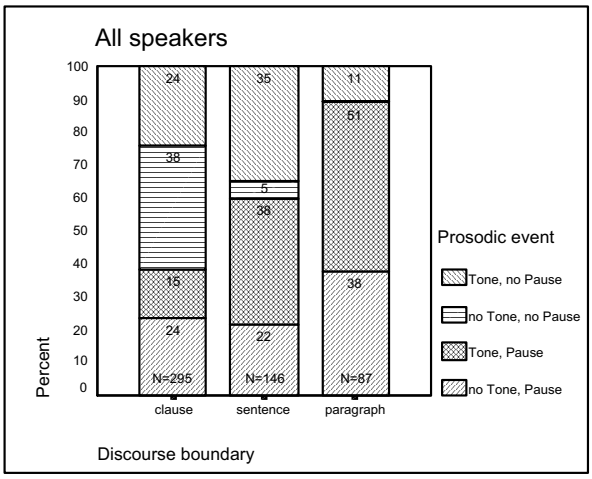

Figure 1. Distribution of prosodic events realized at various types of *structural* discourse boundaries, across speakers.

3.2 Perceived discourse boundaries

In this section we will look at how the boundaries as perceived by the listeners match with the realized prosodic cues by the speakers. For each perceived discourse boundary by at least one of the twelve naive listeners, we checked whether it was accompanied by a boundary tone (either high or low) and/or a pause (either silent or filled). The same four strategies are possible as in the previous section (no tone & no pause, no tone & pause, tone & no pause, tone & pause). Figure 2 presents the distribution of temporal and intonational strategies used by the speakers at perceived discourse boundaries. Data are presented across speakers.

The heavier the perceived boundary, the more cues seem to be used by the speakers, thus the more a boundary tone as well as a pause is realized. Boundaries perceived as *Weak* are in a majority of the cases not marked by any prosodic event at all (no tone, no pause; 30%). If Weak boundaries are prosodically marked, the main prosodic strategy used by the speakers at seems to be the realization of only a pause (27%) or a tone (28%). The use of both boundary tones and pauses is relatively small (15%). Boundaries perceived as *Strong* are in the vast majority realized

with both a tone and a pause (77%), and virtually never without any prosodic event (1%). They are also realized with only a pause (12%) or a tone (10%), but this strategy is used little. Boundaries perceived as *Extra Strong* are realized with at least a pause (19% with only a pause). In the majority of the cases, also a tone is realized (81%). No Extra Strong boundaries are perceived where no prosodic event is realized, or where only a tone is realized.

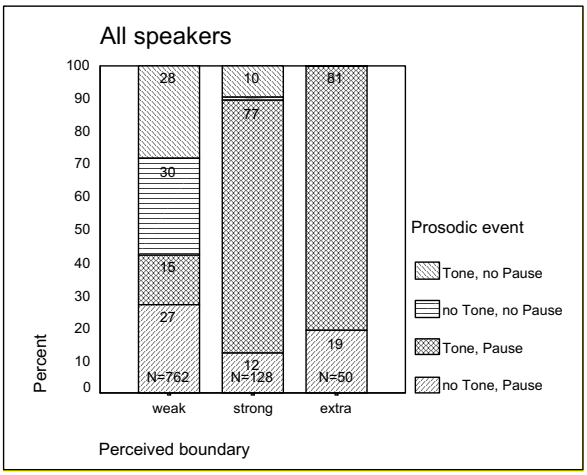

Figure 2. Distribution of prosodic events realized at various types of *perceived* discourse boundaries, across speakers.

4 INFORMATION STATUS

4.1 Perceived prominence and realized pitch accents in relation to information structure

For each of the perceptually prominent words as well as for the pitch accented words in the discourse we checked the information status. We also checked the number of prominent pitch accents (i.e. both perceived as prominent and realized with a pitch accent). The absolute number of perceptually prominent and pitch accented words differed across speakers. Since the discourses produced by the speakers are not equal in length, the data can best be interpreted as percentage relative to the total number of words per information category. This is shown in Table 2, together with the mean number of prominent and pitch accented words per clause.

The category *new* is interesting, since the percentage exceeds 100%, both for prominence and accentuation. In the textual discourse analysis, which determines the total number of elements per category, concepts (comparable to NPs) are labeled rather than individual words. Our data suggest that within each concept, which may consist of more than one word, at least one is perceived as prominent and is realized with a pitch accent. Furthermore, they indicate that more words within one concept may be perceived as prominent or accented by the speaker. The percentages for *inferrable* information (96% prominent, 86% accented) are as predicted by the hypotheses. *Modifiers* are in general perceived as prominent in only half the cases, and are accented in only 41%. This is less than we had expected. *Discourse markers* are perceived as prominent in only 28%, whereas we had expected a larger percentage, since these elements represent the major turning points in the discourse. One explanation could be that the clear linguistic form and function of these elements is taken as a sufficient cue by the speakers in the production of the discourse, and is therefore not marked acoustically prominent, and are thus not perceived as such. This could mean that the speaker assumes that the listener does not need the extra prosodic information to recognize and/or process discourse markers.

Table 2. Mean percentage perceptually prominent and pitch accented words per information category, and mean number of prominent and pitch accented words per clause, across speakers.

	Prominent	Accented	Prom & Acc
New	121	104	98
Inferrable	96	86	89
Modifier	50	41	82
Disc.marker	28	18	67
Verb	53	45	82
per clause:	2.3	1.9	1.6

The data for prominent pitch accents clearly show that in general the majority of pitch accents are also perceived as prominent. This is of course as expected. However, we see a hierarchy relative to information type. Pitch accents realized on new information are most often also perceptually prominent, those on inferrable information somewhat less, and those on modifiers and verbs even less, and those on verbs still less. This hierarchy is nearly identical to the ones found for perceived prominence and pitch accented: new information is accented and perceived as prominent more often than inferrable information, followed by modifiers and discourse markers, and finally verbs. These results are in accordance with [8], who has found that in instruction monologues, new and inferrable information is prominent (prominence defined as pitch), whereas evoked information is not. She used a similar taxonomy to determine information status.

On average, there are 2.3 prominent words per clause, while there are only 1.9 pitch accented words. This means that the presence of a pitch accent was apparently not the only cue for listeners to perceive prominence (otherwise the numbers would have been equal). In other words, the speakes may have used other means to highlight important information. Furthermore, not all pitch accents are perceptually prominent, otherwise the mean number of prominent pitch accents would not be lower than the mean for accented.

5 CONCLUSIONS AND DISCUSSION

In how far do our results meet the hypotheses, as formulated in Table 1? Table 3 presents an overview of the realization and perception of discourse boundaries and focal structure, according to the same setup as in Table 1.

Speakers mark information structure in discourse in terms of phrasing (boundaries) and in terms of focal structure (informative words). Structural discourse boundaries, i.e. determined on textual information, are prosodically marked, but the specific

means used to do this are dependent on the type of boundary: the heavier the boundary the more a boundary is realized prosodically, and the more prosodic cues are used. These cues are, in order of importance, silent pauses and high boundary tones.

Information expressing the lexical content of the discourse, such as new or inferrable information or modifiers, is marked predominantly by pitch accent. There is an ordering in the accentability of information status in discourse: new information is more often accented than inferrable information, which is more often accented than verbs, followed by modifiers and finally discourse markers.

Listeners make use of the prosodic information provided by the speakers to detect the structure of spoken discourse. For the perception of discourse boundaries, pausing is more important than pitch movements, especially if it concerns heavier boundaries. Discourse boundaries are mainly perceived as non-final. The perception of prominence (i.e. the important parts of the discourse) is mainly triggered by pitch accents, and is also dependent on the type of information. The same ordering applies as for accentability: new > inferrable > verbs > modifiers > discourse markers.

ACKNOWLEDGEMENTS

I would like to thank Florien Koopmans-van Beinum and Louis Pols for careful reading and useful suggestions on earlier versions of this paper.

REFERENCES

[1] Lehiste, I. (1979). Perception of sentence and paragraph boundaries. In Lindblom, B. and Öhman, S. (eds), *Frontiers of speech communication research*, Academic Press, London, 191-201.

[2] Van Donzel, M.E. (1999). *Prosodic aspects of information structure in discourse*. PhD Dissertation, University of Amsterdam, LOT Series 23.

[3] Carmiggelt, S. (1966). Een triomf. In *Fluiten in het donker*. ABC Boeken, Amsterdam.

[4] Boersma, P. (1997). *Praat: doing phonetics by computer*. Manual can be found at http://fonsg3.hum.uva.nl/praat/praat.html.

[5] Swerts, M. (1994). *Prosodic features of discourse units*. PhD Dissertation, Eindhoven University.

[6] Blaauw, E. (1995). *On the perceptual classification of spontaneous and read aloud speech*. PhD Dissertation, Utrecht University.

[7] Sanderman, A.A. (1996). *Prosodic phrasing*. PhD Dissertation, Eindhoven University.

[8] Brown, G. (1983). Prosodic structure and the given/new distinction. In Cutler, A. and Ladd, D.R. (eds), *Prosody: Models and Measurements*, Springer Series in Language and Comm. 14, 67-77.

Table 1. Hypothesized realization of discourse structure by means of boundary tones, pauses, and pitch accents, and the hypothesized use of prosodic cues by the listeners in the perception of discourse boundaries and prominence.

	Realized		Perceived		
BOUNDARIES	Boundary tone	Pause	Boundary	Tone	Pause
Clause	High	Yes, shorter	Weak	High	Yes, shorter
Sentence	Low	Yes, longer	Strong	Low	Yes, longer
Paragraph	Low	Yes, still longer	Extra Strong	Low	Yes, still longer
INFORMATION	Pitch accent		Prominence		
New	Always		Always		
Inferrable	Often		Often		
Modifier	Always		Always		
Disc. Marker	Always		Always		
Verb	Never		Never		

Table 3. Actual realization of information structure by means of boundary tones, pitch accents, and pauses, and of the prosodic cues used by the listeners in the perception of discourse boundaries and prominence.

	Realized	Perceived
BOUNDARIES	Tone, pause, both, none	Tone, pause, both, none
Clause	None, high tone or (short) pause	None, tone or pause
Sentence	High tone and (longer) pause	Tone and pause
Paragraph	High tone and (longest) pause	Always at least pause, often tone
INFORMATION	Pitch accent	Prominence
New	Always	Always
Inferrable	Often	Very often
Modifier	Always	Often
Disc. Marker	Always	Less often
Verb	Never	Often