PROSODY AS A MARKER OF INFORMATION FLOW IN SPOKEN DISCOURSE*

MARC SWERTS and RONALD GELUYKENS
Institute for Perception Research (IPO), Eindhoven

This study concerns the role of prosody in the structuring of information in monologue discourse, from the point of view of production as well as perception. Two prosodic variables were investigated: speech melody and pauses. Melodically, it was found that local intonation features (falling *vs.* rising tones) are employed to indicate discourse boundaries. On a more global level, speakers appear to use relative height of pitch peaks and of average pitch values as markers of information units. Furthermore, speakers manipulate both the distribution of pauses and their relative length to mark information flow. A perception experiment was carried out to evaluate the perceptual impact of both speech melody and pauses. It was found that, in the absence of semantic cues, both melodic and pausal information is used by listeners to process the incoming signal in terms of discourse structure.

Key Words: discourse, prosody, boundary cues, pauses

INTRODUCTION

Discourse is more than just a sequence of sentences. It is generally accepted that discourse can be segmented into macro-structures which have some internal coherence and which exceed the level of the sentence. Such larger-scale information units (which we will label topics) may well have some specific linguistic correlates. Indeed, it seems reasonable to assume that a speaker will try to mark these units as clearly as possible to facilitate comprehension. The current paper investigates in some detail what role prosody plays in this respect.

Language users have at their disposal various non-prosodic means with which they can indicate which stretches of speech constitute topical units, such as particular discourse

---

* The authors are also affiliated with the University of Antwerp (UIA and UFSIA, respectively), and with the Belgian National Science Foundation (NFWO). A pilot version of this research was reported in Swerts and Geluykens (1993). Thanks are due to Jacques Terken for allowing us to make use of the speech materials employed in Terken (1984), and to René Collier, Nina Grønnum, Bruno Repp, Margreet Sanders, Jacques Terken, and an anonymous reviewer for their detailed comments on an earlier version of this paper.

Address for correspondence: Institute for Perception Research (IPO), P.O. Box 513, NL-5600 MB Eindhoven, The Netherlands.

particles (Schiffrin, 1987), morpho-syntactic devices (e.g., left-dislocation; see Geluykens, 1992a), and the lay-out of a written text. The functional use of such cues for structuring information flow is an area which has received a lot of attention in the recent discourse-analytic literature, and this has led to a better understanding of the underlying principles (e.g., Givón, 1983, and many others).

Not much is known with the same degree of certainty, however, about the role of prosody in structuring topic flow. This is especially true for spontaneous speech, which is, after all, the most common, least restricted form of language use.

There is a tradition of experimental research into intonational and durational correlates of macro-units, but it has been largely limited to read-aloud speech. For instance, Lehiste (1975, 1980), Bruce (1982), and Thorsen (1985, 1986; see also Grønnum, 1992) have reported that spoken paragraphs possess a characteristic melodic supra-structure. Lehiste (1975) also observed that a speaker can signal by temporal means whether a sentence is paragraph-final or not. Moreover, strong indications were found that these prosodic properties are perceptually relevant: Listeners appeared to use such suprasegmental information to decide correctly where in a paragraph an individually presented sentence had to be situated. (Relations between discourse structure and prosodic features in read aloud speech have also been investigated by Brubaker, 1972; Grosz and Hirschberg, 1992, and Sluijter and Terken, 1992). So far, however, there has been no comparable experimental research into spontaneous speech. Spontaneous language use generally involves only little pre-planning (see Levelt, 1989, for a more thorough discussion), and this may have strong repercussions on the prosodic properties of macro-units .

There are some studies on the contribution of prosodic features to topical organization in spontaneous types of discourse (e.g., Brown, Currie, and Kenworthy, 1980; Yule, 1980; Menn and Boyce, 1982; Johns-Lewis, 1986; Esser, 1988), but these are mostly corpus-based rather than experimental. Also, some attention has been paid to prosody in conversation analysis (e.g. Sacks, Schegloff, and Jefferson, 1974); however, in this research tradition, prosody is dealt with in a very limited way, based on rather loose transcription methods, and restricted to one or two prosodic variables (usually final pitch movements) on a local level. More fundamentally, however, analyzing spontaneous types of discourse with respect to prosody poses some considerable methodological problems which cannot be easily solved through corpus-based research, and which have led us to opt for an experimental approach to prosodic characteristics of topics in spontaneous discourse. The two major methodological drawbacks are: (i) The danger of circularity, and (ii) the need for lexico-syntactically comparable speech fragments (to allow for perceptual evaluation).

METHOD

Problems

The danger of circularity. Since topic structure often cannot be identified unambiguously, especially in uncontrolled spontaneous speech materials, it is tempting to use prosodic criteria for determining discourse structure. Such an approach ultimately leads to circularity, since it begs the question of the role of prosody in demarcating topic structure.

Brown et al. (1980) are conscious of this when they write: "We are uncomfortably aware that, until an independent theory of topic-structure is formulated, much of our argument in this area is in danger of circularity" (Brown et al., 1980, p. 27).

We will propose a method which tries to avoid this particular pitfall by employing data of a very specific type, i.e. instruction monologue. The production task is such that the resulting discourse is, by its very nature, easily segmentable into coherent units, viz. the individual instructions. This uncontroversial segmentation, based on the non-linguistic task structure (see also Terken, 1984), will be taken as the basic criterion for identifying the topical units. This allows us to determine the impact of prosody without any *a priori* assumptions about its role in structuring discourse.

The need for perceptual evaluation. Another important point we want to make is that the general hypothesis about the role of prosody should in fact be reformulated into a speaker and a listener aspect. In other words, it needs to be investigated whether prosodic correlates of topics are relevant for both participants in verbal communication. From a production point of view, we claim that speakers demarcate larger-scale units by means of a clear prosodic structure; in particular, we will investigate two parameters – melodic variation (pitch height and pitch movement) and temporal variation (pauses). From a perceptual point of view, we expect that these prosodic variables function as cues for the listener, i.e. that they are used to facilitate the decoding of discourse structure. Indeed, it is not self-evident that all prosodic features generated in production are perceptually relevant, or relevant to the same degree.

So far, as mentioned above, research combining acoustic measurements and perceptual evaluation of prosodic macro-structures has mainly been limited to analyses of read-aloud speech (but see Lehiste, 1979; Kreiman, 1982; Schaffer, 1984). As there are no constraints on the flow of the discourse, the topics it covers, and the way there is progression from one topic to the next, it is not easy to collect comparable materials – which are indispensable for perceptual evaluation – in totally unrestricted discourse. The production task we used (see below), instruction monologue, is designed to elicit comparable topic structures, which allows for inter-speaker comparison and yields convenient data for perception tests.

We will concentrate on spoken monologue rather than dialogue, as this should give us information flow in its purest form, undistorted by interactional considerations such as the turn-taking system (which are of course equally interesting; cf. Schaffer, 1983; Cutler and Pearson, 1986; Geluykens and Swerts, 1992, 1993).

The data

In order to avoid the methodological problems mentioned above, the data used in the production part of this paper consist of three different, Dutch instruction monologues; they were originally collected by Terken (1984) for research into accent distribution, but we are employing them here for different purposes. (For the details of the total experimental design, see Terken, 1984.) These monologues consist of a series of instructions from a speaker to a listener whose task is to assemble the front view of a house from a set of ready-made pieces of cardboard (e.g., a roof, a front door, etc.). From Terken's original eleven recorded speakers, three (one male (HZ) and two females (SK, NE)) were selected for further analysis, as they produced the most fluent stretches of speech (i.e. produced without metacomments on the task itself and without elaborate repair sequences). Speakers SK, HZ, and NE produced 13, 12, and 10 instructions overall, respectively.

Instructions in the monologues are generally of the following form (slashes are used to facilitate reading and are not meant to indicate utterance boundaries; English glosses are approximations rather than literal translations):

(1) A 1. dan hebben we het zwarte vierkant
then we have the black square

B 2. daar gaan we nu een dak opzetten / dat is het groene driehoek
now let's put a roof on it / that is the green triangle

B 3. de grote groene driehoek / die zetten we er boven op
the large green triangle / we place that on top of it

C 4. dan pakken we het woonkamerraam / dat draaien we met de
then we take the living room window / we turn its

5. kleurzijde om / en leggen het links onderin / met wat ruimte
coloured side up / and put it left at the bottom / with some space

6. eronder / zodat de lange kant evenwijdig ligt aan de onderkant
underneath it / so that the long side is parallel to the bottom

7. van het huis
of the house

D 8. dan pakken we de voordeur / en die zetten we een eindje
then we take the front door / and we put that a little to the

9. rechts van het raam / met de korte zijde naar onder
right of the window / with the short side down

First of all, a referent (e.g., het zwarte vierkant ('the black square')) is introduced, which constitutes the core of the instruction, and on which some action has to be performed. As an (optional) second component, the speaker may then give some additional information to facilitate referent-identification for the listener; this will be labelled elaborative material (e.g. het grote groene driehoek ('the large green triangle') in line 3 of (1), which is an elaboration on een dak ('a roof') mentioned in line 2). Next comes the actual instruction, which may be spread out over several clauses (e.g. C in (1)), but which can also be quite short (e.g. B in (1)), depending of course on the level of difficulty of the task which has to be performed. Finally, the speaker sometimes provided a kind of summing-up statement to make it clear that an instruction has now been completed. In other words, each instruction can be regarded as a coherent group of sentences dealing with one particular topic, viz. the building block at hand.

The great advantages of this type of speech are that it is (i) spontaneous, i.e. produced without any prior planning, and without any specific limitations on the structure being imposed, and (ii) controlled, in that the semantic content of the produced speech is determined by the actual instructions (see also Swerts and Collier, 1992). The three speakers thus produced speech which shows a lot of variety as to syntax, etc., but which has enough similarities to allow for inter-speaker comparisons. What is more, the speech is recorded under sufficiently good acoustic conditions to enable us to employ it for testing the perceptual impact of the prosodic variables involved.

Additional evidence for discourse structure

The previous section has shown that the monologues have some internal organization reflecting the different instructions, which are easy to identify. It would be preferable, however, to have some additional textual motivation for our structural analysis. This is desirable for two reasons. First of all, it provides us with supplementary evidence for the macrostructure we have assigned to the text. Secondly, it should allow us to generalize over different types of texts, and thus expand our analysis to spoken materials other than the instruction monologues analyzed here.

For our linguistic verification of macrostructure, we can make use of the concepts developed by Geluykens (1988, 1992a, 1992b, in press). We can determine salient points in the discourse based on the 'newness' (see also Halliday, 1967; Chafe, 1976, 1987; Prince, 1981) of the referent in the overall flow of the information. In Geluykens (1992a, 1992b), for instance, new discourse topics are defined as non-recoverable pieces of information which have some degree of persistence. Non-recoverable information (see Geluykens, 1988) is information which 'cannot be retrieved from the preceding context', i.e. which does not actually occur in the previous discourse record. Persistent elements are those which recur, in some surface form, in the subsequent discourse. A distinction is made between direct persistence (recurrence of a co-referential form of the same entity) and indirect persistence (through a semantically closely related referent, e.g. one which has a part-whole relationship with the referent).

If we apply these notions to the excerpt presented in (1), for instance, we get the following entities qualifying as non-recoverable information: het zwarte vierkant ('the black square'; line 1), een dak ('a roof'; line 2), het woonkamerraam ('the living room window'; line 4); de voordeur ('the front door'; line 8). Furthermore, all referents mentioned above are persistent. If we examine one of them in detail, for instance het woonkamerraam (instruction C), we find the following evidence of direct persistence: dat ('that'; line 4), het ('it'; line 5), and eronder ('underneath it'; line 6). Instances of indirect persistence are de kleurzijde ('the coloured side'; line 5); de lange kant ('the long side'; line 6). This linguistically motivated topical structure thus confirms the non-linguistic task structure.

Occasionally, there are also some interactional clues present which give further support to our analysis, in the form of particles which function as 'discourse markers' (Schiffrin, 1987) indicating the start of a new topical unit. Concentrating on excerpt (1) again, we find the marker dan ('then') in several locations (lines 1, 4, and 8), and the marker nu ('now') in another (line 2), all of them corresponding to the start of a new topical unit.

Other speakers employ other interactional markers (e.g., speaker SK very consistently uses the voiced pause uhm before each new topical unit). Once again, we get independent confirmation of the topical structure.

Summing up: Our definition of instructions as topical units can also be motivated linguistically. It remains to be seen to what extent prosodic information reinforces the discourse structure, but at least this can be investigated without any recourse to circular arguments based on prosodic organization.

PRODUCTION DATA

In this study, as pointed out above, both production and perception were taken into account. As regards speech production, two prosodic variables were investigated with regard to the supra-structure of discourse: (i) the intonation or speech melody (melodic boundary markers, scaling of pitch maxima, and mean pitch of subsequent clauses) and (ii) the temporal structure, more specifically the distribution and relative duration of pauses. Other prosodic features, such as laryngealization and loudness, were not investigated, though we do not exclude their potential relevance for highlighting discourse structure. Final lengthening was not studied because it has been claimed (Lehiste, 1979, p.195) that there is no obvious difference in length between sentence-final and paragraph-final syllables. Moreover, the determination of lengthening would require comparisons of identical syllables in different discourse contexts, which were not available in our data.

The monologues were digitized with a 10 kHz sampling frequency at 12 bits, with low-pass filtering (cut-off frequency: 4.8 kHz; slope: 90 dB/octave); the fundamental frequency (F_0), as the acoustic correlate of perceived pitch ('t Hart, Collier, and Cohen, 1990), was then determined by means of the method of subharmonic summation (Hermes, 1986).

Speech melody: boundary tones

Hypothesis. Brown et al. (1980), among others, have argued that intonation is often exploited to signal topic continuity or finality through the use of different melodic boundary tones (see also Bolinger, 1989; Cruttenden, 1981, 1986). Their claim is that so-called low terminals are regularly associated with the end of a topic, whereas non-low terminals would serve to indicate that there is more to come on the same topic. We hypothesized that this tendency would also occur in our monologues.

Measurements. At the end of each clause in the monologues, the course of F_0 was examined from the last accent till the beginning of the next clause, the latter being operationally defined as a syntactic entity containing a finite verb (sometimes implicitly). Clauses rather than intonational phrases were taken as the basic units of analysis to ensure that melodic boundary tones could be assessed in a non-circular manner. A classification was made into low-ending contours and high-ending contours on the basis of two melodic transcriptions carried out by two experienced intonation researchers other than the authors. Transcriptions were made using the IPO-grammar of Dutch intonation (see 't Hart et al. 1990). Contours of type /1Ø/ (= accent-lending rise, followed by high continuation), /1E/ (= same rise, followed by half-fall), /12/ (= same rise, followed by continuation rise) and /A2/

TABLE 1

Distribution of low and non-low boundary tones as a function of discourse position in the monologues of speakers SK, HZ, and NE

	Speaker SK		Speaker HZ		Speaker NE	
	End of instruction		End of instruction		End of instruction	
	yes	no	yes	no	yes	no
Boundary tones						
Low	12	5	9	7	9	3
Non-low	5	38	3	18	3	12

(= accent-lending fall, followed by continuation rise) were interpreted as non-low tones (H%); contours of type /AO/ (= accent-lending fall, followed by low continuation) were interpreted as low tones (L%). The distribution of these is shown in Table 1.

Results and discussion. Table 1 shows that there is a significant association of low-ending contours with instruction-finality on the one hand, and of high-ending contours with non-finality on the other hand (SK: χ^2=38.5, p<0.001; HZ: χ^2=13.1, p<0.001; NE: χ^2=18.8, p<0.001). These findings appear to confirm the earlier claims by Brown et al. (1980). There is thus indeed a correspondence between the topical structure of the discourse and the use of boundary tones. The majority of the low-ending contours are located at the ends of the final clauses of the various instructions: their function seems to be to signal that an informational unit has been completed. Most of the non-low contours occur within instructions; they signal that there is still more to come on the same topic. A typical example of this general tendency is instruction (2) below by speaker SK :

(2)

1. dan pakken we de plant (H%)
   then we take the plant

2. en die plant die zetten we naast de voordeur (H%)
   and that plant we put next to the front door

3. links naast de voordeur op de grond (H%)
   to the left of the front door on the ground

4. dus gelijk met de onderkant van de voorgevel (H%)
   so it is parallel to the bottom side of the facade

5. midden tussen de voordeur en het woonkamerraam (L%)
midway between the front door and the living room window

In this instance, all instruction-internal clauses end in a high contour, while the instruction itself ends in a low one.

It should also be pointed out that there is some variation among speakers. Speaker HZ, for instance (cf. Table 1), appears to be slightly less consistent in his use of low contours than the other two speakers, as only 56% of his low contours are instruction-final, compared to 71% and 75% for SK and NE, respectively.

Note that, for all three speakers, the exceptions occur mostly in one direction: There are only 11 (out of 79 instances; 13.9%) high-ending contours in final position, whereas there are 15 (out of 45 instances; 33.3%) low-ending ones in non-final position. In other words, while a high-ending contour almost invariably signals that an instruction has not ended, it is not the case that a low-ending one automatically implies instruction-finality.

Though this is difficult to determine on independent grounds, it appears that some instances of non-final low contours turn out not to be real counter-evidence; looking at specific discourse contexts sometimes reveals that some of the apparent exceptions might be functionally motivated. For instance, some non-final low boundary tones occur at points where the information conveyed can be considered 'complete' from an informational point of view, i.e. sufficient to enable a listener to successfully execute the instruction of the speaker. However, as a sort of afterthought, there is a subsequent clause providing some details that are redundant from a purely informational point of view or that are so obvious or deducible from the previous discourse that they are not really necessary to be communicated. In (3), which is instruction 3 of speaker SK, an example is presented as an illustration of what is meant:

(3)

1. dan pakken we de voordeur (H%)
then we take the front door

2. en die zetten we rechts in het zwarte vierkant (H%)
and we put that right in the black square

3. rechtsonder (H%)
at the bottom right

4. zodat de smalle kant van de voordeur
so that the small side of the front door

5. tegen de onderkant van het zwarte vierkant aanzit (H%)
sits on the bottom side of the black square

6. en een klein stukje een centimeter of twee
and a little bit, about two centimeters,

7. vanaf de rechterzijkant van het zwarte vierkant (L%)
from the right side of the black square

8. dus de onderkant van de voordeur loopt gelijk met
so the bottom side of the front door runs parallel with

9. de onderkant van het zwarte vierkant van de voorgevel (L%)
the bottom side of the black square of the front view

In (3), it can be observed that the first occurring low boundary (line 7) is located at a position where the instruction is informationally complete. The subsequent utterance only paraphrases the semantic content of the previous part of the instruction. Of course, this non-essential information may facilitate the communication between speaker and listener, as it confirms what has been said in earlier utterances.

In the three monologues, five instances were found of low boundary tones that occurred at the end of an informationally complete unit, i.e. one in which all the relevant information appears to have been conveyed already, but nevertheless does not coincide with the end of the total instruction. Of course, it is not always straightforward to decide when an instruction can be considered 'informationally complete', and one risks running into circularity; we have mainly used our own intuitions as a criterion here, and have only considered very clear-cut cases of redundant information as afterthoughts.

Summing up this section, we can conclude that local pitch cues are indeed used by the speaker to signal discourse structure. As expected, low-ending contours are generally associated with finality, while high-ending contours are associated with continuation.

Speech melody: F_0 maxima

Hypothesis. Another melodic variable is the height of the F_0 maxima in accent-lending pitch movements. Terken (1984) has already observed in these speech materials that noun phrases introducing a new topic into the discourse were always accented. This finding can be interpreted to mean that by accentuation the speaker gives an indication of the non-recoverability of the information conveyed. As the referents introduced are always non-recoverable, they always get an accent. Moreover, these new items constitute the topics of the subsequent discourse and are therefore made prominent in order to signal that the referent must be given preferential status in the listener's discourse model. As the new topics reflect major changes in the information flow, and as they are very salient points in the instruction task, it would make sense to make them even more prominent than other accented words by giving them a higher than average pitch peak.

Measurements. As the basis for our measurements, we took the two independent intonation transcriptions mentioned earlier and considered the positions where a pitch accent had been heard. F_0 maxima were defined as the end of an accent-lending rise (type 1) or the beginning of an accent-lending fall (type A). There were some difficult cases, namely the abrupt pitch rises that occurred relatively late in one-syllable words and that seemed to

TABLE 2

Distribution of F_0 peaks (= or $\neq$ highest within instruction) on topic NPs as a function of position within instruction in the monologues of speakers SK, HZ, and NE

	Topic NP = highest F_0 peak	Topic NP $\neq$ highest F_0 peak
Speaker SK		
first accent	5	2
non-first accent	4	2
N =	9	4
Speaker HZ		
first accent	5	2
non-first accent	3	2
N =	8	4
Speaker NE		
first accent	3	3
non-first accent	2	2
N =	5	5

consist of an accent-lending and a non-accent-lending part. It was difficult to locate the exact transition point between these two parts, as they present themselves as one fluent, complete F_0 movement. Therefore, the F_0 maximum of the entire movement was taken as a measurement.

In order to compare topic peaks with non-topic peaks, we have consistently regarded the highest F_0 peak on the introduction proper (i.e. the clause containing the first mention of the new topic NP) as the relevant 'topic peak' (see our discussion of example (1) above).

Results and discussion. Table 2 gives the data on the correlation between highest F_1 maxima and topic NPs. These data support our hypothesis that topic peaks are more prominent than non-topic peaks, as the accents on the topic-introducing noun phrases indeed appear to differ qualitatively from other accents in the discourse: They are located higher in the speaker's register. Such very prominent accents may function as 'warning signals' from the speaker to the listener that a new topical unit has been started. However, there is some degree of speaker variation in that NE does not provide the topic-introducing noun phrase with the highest F_0 maximum quite as consistently as HZ and SK do (50% versus 67% and 69%, respectively).

When we look at the average height of F_0 peaks in different discourse locations (see

TABLE 3

Mean peak height (and standard deviations) of accents (in Hz) in different discourse positions in monologues of speakers SK, HZ, and NE

	first accent		not first accent		last accent
	topic	non-topic	topic	non-topic	non-topic
Speaker					
SK	332 (49)	269 (26)	334 (43)	252 (23)	210 (37)
HZ	190 (17)	153 (17)	181 (27)	154 (21)	116 (11)
NE	314 (35)	290 (28)	308 (45)	297 (29)	268 (27)

Table 3), we find that F_0 peaks on the topic-introduction NP are the highest, irrespective of whether this NP has the first accent within an instruction or not; t-tests reveal that differences between peak height of topic-NP accents in first position and those not in first position are not significant. Conversely, accents which are not on the topic-introducing NP do not get the highest peak, even if they are in first position in the instruction. The difference in peak height between topic NPs on the one hand and non-topic NPs on the other hand is significant for speakers SK (t=3.04, p<0.05) and HZ (t=3.81, p<0.05); this is not the case, however, for NE (t=1.16, n.s.), though there is clear trend in her data, similar to the effect in the data of the other two. This means that the highest peak is on the topic-NP, irrespective of whether this is the first F_0 maximum or not. It is precisely these latter cases which show clearly that the occurrence of the highest F_0 maximum is not just a matter of linear order, but is indeed exploited by the speaker to co-occur with the new topical referent, which thus gets the most salient F_0 peak in the discourse stretch.

It should also be stressed here (as is also shown by Table 3) that the first peak of the topical unit is always far higher than the final peak of the preceding unit. They differ significantly if the first peak is the topic-NP (SK: t= 6.27, p<0.001; HZ: t= 5.06, p<0.001; NE: t=2.93, p<0.01); if the peak is not the topic-NP, the same trend can be observed, though significantly so only for SK (t= 3.04, p<0.01). In other words, there is always an upward shift in peak height at the boundary between two topical units, especially if the topic-NP peak is at the very beginning of the instruction. It would appear, therefore, that the relative heights of F_0 peaks serve a dual function: (i) demarcating the topical units themselves, by an upgrade in peak height; (ii) highlighting the new topic, by giving it the highest F_0 peak within a topical unit. This points to a very sophisticated use of global F_0 features by the speaker, and shows that we should also look beyond the local level when studying the discourse functions of F_0 variation. This is even more true for the F_0 variation studied in the following section.

Speech melody: global use of F_0 values

Hypothesis. Our analysis of the relative heights of F_0 peaks in the preceding section

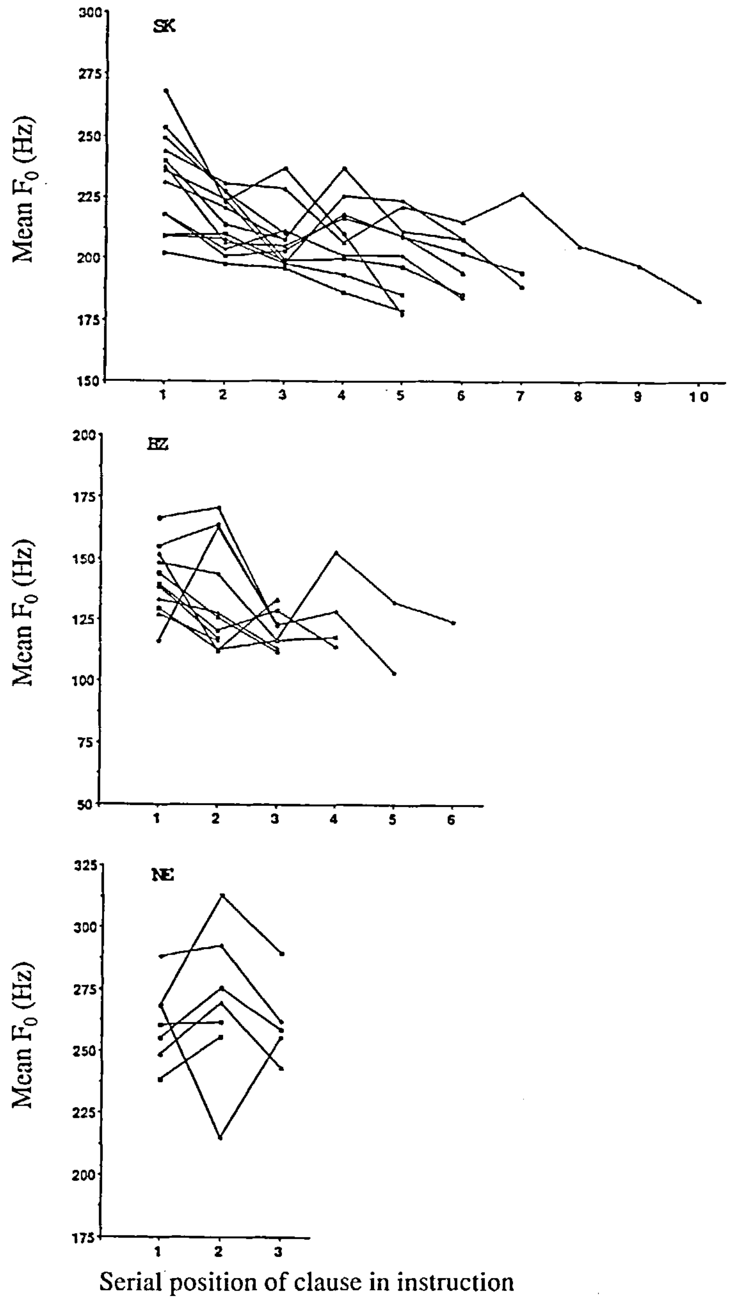

Fig. 1 Mean F_0 (in Hz) of successive clauses in the instructions of speakers SK, HZ,and NE; identical symbols connected with lines represent means within the same instruction.

has already shown that speakers can make subtle use of F_0 features over longer stretches of speech. It would therefore be reasonable to assume that speakers also use F_0 on a more global level. One might ask whether there could be a kind of overall declination of F_0 over an instruction, followed by a reset on the following instruction.

Measurements. To test this hypothesis, some more global calculations were performed on the fundamental frequency data. As a way of determining the supra-clausal declination just mentioned, the mean F_0 was determined over the range of one clause. If there was a large pause (i.e. longer than 1000 msec), the clause was split up into two separate units. The measured means of F_0 of subsequent clauses for the three speakers are shown in Figure 1.

Results and discussion. It can be seen that the larger-scale informational units of speakers SK and HZ exhibit a global phonetic characteristic: F_0 is, on average, relatively high at the beginning of a unit, and it then slowly decreases over the course of the instruction; at the beginning of a new topical entity, F_0 is again shifted up. However, the macro-units of speaker NE, which were on average much shorter, do not have this general prosodic feature.

At this point, it is not yet clear how the global decrease in the mean F_0 (that is, in the data of HZ and SK) should be interpreted. It is our impression that it is the composite result of two mechanisms: a general lowering of F_0 register and a global decrease in the excursion size of the movements (see also Swerts, Bouwhuis, Collier, 1993). In any case, it suggests that relatively global correlates can be observed in spontaneous discourse (a claim which has been questioned a few times in the literature). Of course, this does not mean that the same finding also holds for conversational, non-monitored speaking style.

Pauses

Hypothesis. A second major prosodic dimension which speakers may manipulate to structure their information flow is the temporal one, more particularly the use of pauses in discourse. First of all, we would expect that the occurrence of pauses is to some extent dependent on discourse structure, in that topic shifts are accompanied by pausal boundaries. Secondly, such pausal boundaries can be expected to be more pronounced – i.e. longer– than topic-internal pauses.

Measurements. Pauses were operationally defined as periods of silence, equal to or longer than 100 msec; they were measured directly on the digitized speech waveform.

Results and discussion. The likelihood of occurrence of a clause-final pause appears to be very high at two important discourse locations. Firstly, pauses are present at all transitions between instructions, i.e. between all topical units, as expected. Secondly, pauses also consistently occur right after the topic-introducing phrase or clause; sometimes the pause is after the elaborative material rather than on the introduction proper, but always before the beginning of the actual instruction part. These two locations constitute crucial information flow positions, as will be shown below.

This already indicates that it would be a mistake to interpret pausal structure purely in syntactic terms. Indeed, some of the pauses are present at shallow structural breaks, e.g. in between preposition and noun, as in (4):

(4) 1. dan brengen we één rood gordijn aan met (0.39 sec)
then we apply one red curtain with

2. afgeknipte scherpe puntjes
sharply cut-off edges

In other words, not all pauses are at clause boundaries (as was to be expected). Secondly, it is also the case that not all clause boundaries are marked by a period of silence. This is clearly exemplified by (5) below, where a major clause boundary does not coincide with the presence of a pause, after dan heb ik nog een groen frotje over:

(5)	1.	dan heb ik nog een klein groen frotje over *then I have another small green thingie left*	(no pause)
	2.	dat zal wel een bloempotje zijn *that will probably be a flowerpot*	(0.43 sec)
	3.	dat zetten we bij het voorraam onder *that we put by the front window below*	(7.18 sec)

We have also looked at the respective lengths of the pauses in various discourse locations, i.e. (i) between instructions, (ii) after the clause or phrase introducing a new topic, and (iii) in other positions. Results can be found in Table 4.

This table shows that, in these three monologues, pause duration is dependent on the topical structure of the discourse: The longest silent intervals are found between instructions (differing significantly from pauses within instructions for all three speakers: SK: t=5.93, p<0.001; HZ: t=4.08, p<0.001; NE: t=10.54, p<0.001); within a topical unit, the pauses after the topic introduction are consistently longer than in other locations, though this difference is only significant for speakers SK (t=4.96, p<0.001) and NE (t=3.23, p<0.01). Though the three speakers differ considerably in their absolute pause lengths, they all share this same pattern of varying pause duration as a function of discourse location.

Having established that there is a strong relationship between pause structure and the topical organization of the discourse, it remains to be explained what cognitive or communicative factors might account for this regularity. In the following, a few tentative hypotheses will be presented that need further experimental verification. First of all, one could argue that pausal structure is a result of cognitive processing by the speaker. In this view, the silent intervals between instructions reflect the planning carried out by speakers before they embark on the next instruction. Similarly, the pause after the topic-introduction could be caused by the speaker's planning of how to develop the newly introduced topic in the subsequent discourse. The subsequent utterances within the same topical unit would then require less processing, and pauses are consequently shorter. However, since the experimental task is quite simple and does not need considerable mental effort, the cognitive explanation does not appear to give a full account of factors governing pause length.

An alternative hypothesis (which is not mutually exclusive with the previous one) concerns the communicative goals of the speaker, and his/her accommodation to the needs

TABLE 4

Mean durations of actually occurring pauses (in sec) and standard deviations (s.d.) at various discourse locations in the monologues of speakers SK, HZ, and NE

	between instructions			within instructions		
	mean	n	s.d	mean	n	s.d.
SK	2.89	13	1.21	1.11	90	1.00
HZ	1.82	13	0.88	0.84	50	0.81
NE	7.60	10	1.64	1.38	26	1.55

	after topic-introduction			elsewhere		
	mean	n	s.d.	mean	n	s.d.
SK	2.21	13	0.98	0.92	77	0.88
HZ	1.04	12	0.98	0.78	38	0.76
NE	2.65	8	1.80	0.81	18	1.05

of the listener. By manipulating pause length, one could argue, speakers are trying to make it easier for their interlocutor to process discourse structure. In such a way, they are not only marking which major chunks of the discourse (the instructions) belong together, but are also drawing attention to the newly introduced topics by making them prosodically more salient, and hence more easily identifiable as new discourse topics. This outcome is very compatible with the finding that two out of three speakers generally provided the topic-introducing NPs with the highest F_0 maxima.

This latter view (and, especially, the discussion about the pauses in post-topic-introduction position) is also compatible with a third, more interactional explanation. Although the speech materials under investigation consist of monologues, the test setting really was a communicative one: Speakers had to give instructions to listeners who were physically present. There is thus an interactive dimension to this discourse, despite the fact that no verbal or visual feedback was possible, unlike real conversation (see Terken, 1984, for the experimental design). In Geluykens (1992a, 1992b) it is argued that referents in conversation are introduced in a collaborative manner, through a three-stage interaction between a speaker and a listener. These stages are, respectively, introduction by the speaker of a new referent, acknowledgement by the listener of this new referent, and establishment of the new referent by the speaker, by developing it as a discourse topic (see (6), simplified from Geluykens, 1992a):

(6) C: Professor Worth asked me to get some books for him

B: oh yes yes

C: I've just arranged for those to be sent over by taxi

Acknowledgment, it is claimed, can be either verbal, as in (6), (usually through a short acceptance signal such as yeah, mhm, and the like), or implicit, without an overt linguistic signal. In the latter case, the speaker pauses to give the listener the opportunity to take in the new referent cognitively, but also to enable him/her to reject the new referent if s/he should feel that way inclined. Given the normal politeness principles operative in conversation (see Brown and Levinson, 1987), such rejection is not very likely. In the vast majority of the corpus-data analyzed by Geluykens (1992a, 1992b) it was found that referent-introductions were followed by an overt acknowledgment signal, but also often by a pause.

The long post-topic-introduction pauses in our data could therefore be argued to reflect this interactive dimension, giving the listener the chance to process the new referent, but also, theoretically, giving him/her the opportunity to intervene if necessary, either to request more information or to short-circuit the referent. In other words, some of these pauses would thus be quasi-conversational and essentially interactive. As it is impossible to verify this unequivocally in the data, such a statement needs further experimental support.

PROSODIC FEATURES AS PERCEPTUAL CUES

In the preceding section, we have shown that speakers do structure the information flow of a spontaneously produced monologue by means of particular melodic variables and by pausal structure; it remains to be examined to what extent this prosodic information is relevant to a listener. We want to evaluate the communicative import of both speech melody and pausal structure for the clarification of the discourse structure, as it is not self-evident that all prosodic aspects of speech production are necessarily vital for speech understanding. But, potentially, they may help the listener in his/her decoding of an incoming speech signal.

Hypothesis. More specifically, we wanted to tackle two questions in our perceptual approach. First of all, we intended to shed some light on the pure contribution of the prosodic layer separate from the lexical and syntactic information conveyed. To address this question, we performed a listening experiment with band-pass filtered speech, i.e. speech that is unintelligible but which has an identifiable prosodic shape (see also Kreiman, 1982; Schaffer, 1984; Swerts et al., 1992).

Secondly, a perceptual approach also created an opportunity to evaluate the relative perceptual relevance of the variables studied and to ascertain whether either of the two is more crucial from the listener's point of view. As our speakers appeared to make use of different prosodic strategies to structure their discourse (with different emphases on different prosodic variables), we had the chance to investigate these relative contributions by systematically eliminating prosodic information from the speech presented to the listeners.

Materials. For each of the three speakers investigated (SK, HZ, and NE), we took an excerpt of their speech consisting of five successive instructions of varying length (starting with instruction 2 of each speaker), plus the beginning of a sixth. This gave us five actual instruction boundaries. Following a method already applied by Lehiste (1979), Kreiman (1982) and Schaffer (1983, 1984), these excerpts were then band-pass filtered with a low-pass cut-off value of 310 Hz and a high-pass cut-off value of 260 Hz, resulting in unintelligible speech, without any lexical or syntactic cues remaining. This 50-Hz band, based on Schaffer (1983,1984), ensured minimal intelligibility while at the same time being the least disruptive to prosodic structure.

We then created four different prosodic conditions for all three excerpts (thus resulting in 12 chunks of speech in all). In Condition 1, all the prosodic information, including melody and pause durations, was kept intact. In condition 2, all pauses were made equally long, using the average 'unmarked' duration (i.e. not between instructions, and not after a topic-introduction) for that particular speaker (0.92 sec for SK, 0.78 sec for HZ, and 0.81 sec for NE). Naturally, we only manipulated pauses which existed in the original version. Speech melody was left unchanged. In Condition 3, we kept the original pause durations, but, using a waveform-manipulation technique (Charpentier and Moulines, 1989), pitch variation was eliminated by making F_0 monotonous at a level of 200 Hz for SK, 150 for HZ, and 250 Hz for NE, these values approximating the average pitch levels of the respective speakers. Condition 4, finally, involved both manipulation of pause lengths (as in condition 2) and monotone pitch (as in condition 3).

Subjects and procedure. Eight subjects (all students and staff at IPO) were presented with all 12 versions of the manipulated speech. They were seated in a sound-proof studio, hearing the filtered speech through loudspeakers. Order of presentation of the different versions was randomized, to compensate for potential learning effects. Subjects were instructed to say 'yes' ('ja' in Dutch) as quickly as possible, whenever they thought that a major discourse unit had ended (the targets being, of course, the five instruction endings). Subject responses were recorded together with the filtered speech, to determine the degree of correspondence between actual boundaries and response locations.

Results and discussion. In order to estimate the effects of melody and pause, the perceptual results were analyzed in the following way. We computed two numbers for each listener: (a) all end-of-instruction locations that received "ja" responses, expressed as a percentage of the total number of instruction endings; (b) all "ja" responses that occurred at end-of-instruction locations, expressed as a percentage of the total number of "ja" responses given for each speech fragment. In other words, measure (a) expresses how many instruction endings were correctly identified, while measure (b) expresses the frequency of responses at correct locations compared to the overall number of responses. Table 5 gives the mean percentages of both these computations for the three speakers.

Two-way ANOVAs (F_0 variation by pause duration variation) on the percentages of end-of-instruction locations receiving a "ja" response (measure (a)) showed that, for all three speakers, there was a significant effect of melody (SK: $F(1, 28)=24.72, p<0.001$; HZ: $F(1, 28)=17.11, p<0.001$; NE: $F(1, 28)=7.83, p<0.01$) and of duration (SK: $F(1, 28)=20.92$, $p<0.001$; HZ: $F(1, 28)=20.36$, $p<0.001$; NE: $F(1, 28)=50.48$, $p<0.001$); while the interaction between these factors was never significant. From this point of view, both melody and pause duration contribute significantly to the perception of discourse structure in an additive way.

The picture is somewhat different, however, when two-way ANOVAs (F_0 variation by pause duration variation) are computed on the number of "ja" responses occurring at

TABLE 5

Results of perception test for the four conditions (see text) of the monologue of speakers SK, HZ, and NE

Conditions		C1	C2	C3	C4
Variables:	Melody	Original	Original	Constant	Constant
	Pause	Original	Constant	Original	Constant
Speaker SK					
% end of instr. receiving "ja"		97.5	77.5	75.0	37.5
% "ja" occurring at end of instr.		65.6	58.3	42.3	24.5
Speaker HZ					
% end of instr. receiving "ja"		82.5	62.5	65.0	25.0
% "ja" occurring at end of instr.		74.0	74.0	73.6	54.1
Speaker NE					
% end of instr. receiving "ja"		85.0	50.0	75.0	27.5
% "ja" occurring at end of instr.		82.9	72.5	92.8	34.4

end-of-instruction locations (measure (b)). The effect of melody is significant for SK ($F(1, 28)=61.47$, $p<0.001$), marginally significant for NE ($F(1, 28)=3.75$, $p=0.063$), and non-significant for HZ ($F(1, 28)=1.19$, n.s.). The effect of duration is significant both for SK ($F(1, 28)=11.89$, $p<0.01$) and NE ($F(1, 28)=22.20$, $p<0.001$), but non-significant for HZ ($F(1, 28)=1.11$, n.s.). Moreover, for speaker NE, there was a significant interaction between the two factors ($F(1, 28)=10.82$, $p<0.01$): Melody only contributed when pause variation was absent. From this perspective, the analysis shows that there is some speaker-dependent variation.

We can thus conclude unequivocally that listeners are able to deduce discourse structure from prosody. Both pause duration and pitch variation appear to be important perceptual cues.

Interestingly, however, these results do not apply to the same degree for each speaker. This is understandable if one takes the speech production of the three speakers into account. In the production analysis, we observed that speaker NE gave the clearest pause duration signals, in the sense of making the largest variations in pause duration; with regard to speech melody, however, she was less clear , especially as far as global use of F_0 values was concerned. Given these production factors, the perception results can be interpreted more easily: It turns out that, if one removes the clearer production signal for this speaker, i.e. pause duration (condition 2), perception suffers more than if one removes the less clear signal, i.e. pitch variation (condition 3). Speaker SK, on the other hand, gave the clearest speech melody signals, in the sense of being the most consistent with regard to its local and global melodic supra-structure, while at the same time being less clear about pausal signals, in the sense of not differentiating very clearly in pause length. Not surprisingly, then, given these production characteristics, perception suffers more than it does for NE if one removes pitch variation (condition 3). Speaker HZ lies somewhat in between these two extremes, in the sense that the results, for both measures (a) and (b), are similar for condition 2 and condition 3.

What all this boils down to is the following. While it is true that both speech melody and pause duration are shown here as important perceptual cues, this result should be amended in the following way. Both cues work well when combined, though perception does not always suffer to the same extent if one removes either of the two cues. This is a result of the speaker variation with respect to the use of prosodic variables and its consequences for perception: The likelihood of correct perception appears to depend on how the prosodic features are realized in speech production. This shows that it does indeed make a lot of sense to investigate not just production, but also to evaluate perception.

To round off this section, we should say a few words about condition 4, in which both pitch melody and pause duration were kept constant (Table 5). Not very surprisingly, listeners performed worse in this condition than they did in all of the other ones. This was to be expected, given the very minimal nature of the potential cues offered. Of course, this does not necessarily imply that there are no relevant cues present in some of the remaining supra-segmental information, such as speech tempo, relative loudness, etc.

GENERAL DISCUSSION AND CONCLUSION

Having dealt with the analysis of spontaneous discourse prosody both from the speaker's and the listener's point of view, we will now consider our specific results from a wider perspective and try to relate them to findings of comparable studies on this subject.

With regard to the speaker, it was found that larger-scale information units are indeed phonetically encoded in pausal and melodic properties. On a local level, it appears that these supra-segmental features are used for demarcative purposes. A speaker may separate two successive units by the use of a long pause. Melodically, s/he can provide the terminal edges of the units with conspicuous melodic features. First of all, the ends of the instructions are generally provided with a low boundary tone, differing from the non-low tones within instructions. Secondly, at the beginning of a unit, one often finds a high-pitched accent on a topic NP, presumably serving as a warning signal to the listener that a new information unit has started. It would appear that finality cues and the subsequent long pauses are more important than initial cues, which are somewhat redundant and merely serve as confirmations that a new unit has started, since in the perception test listeners were usually able to respond before a new topical unit had actually started. At any rate, we can conclude that speakers use local prosodic devices to signal the boundaries of a discourse unit.

These findings are compatible with the results of Brown et al. (1980) and Yule (1980). The latter, more specifically, introduces the concept of a 'paratone', which is "frequently identifiable in spontaneous speech, not by its internal structure, but by its boundaries" (p.47). This statement seems to imply that the internal body of the topical unit is thought to have no particular prosodic correlates (at least not in spontaneous discourse), apart from the fact that 'topic-continuation' is signalled throughout the unit by non-low boundary tones.

However, our acoustic observations brought to light that the internal structure of a topical unit can also be reflected in melodic features, as speakers can provide a coherent stretch of discourse with a gradually varying, global melodic supra-structure (though there seems to be some speaker-dependency here). Comparable global trends have previously been reported to occur in spoken paragraphs in read-aloud speech (Bruce, 1982; Thorsen, 1985, 1986; Sluijter and Terken, 1992). In this speech mode, the observation that such larger-scale textual units can get some superordinate prosodic structures is not surprising, given the fact that speakers are able –quite literally– to have an overview of the sentences constituting a paragraph. The finding that speakers are also able to supply global prosodic structure in spontaneously produced speech is interesting, since it shows that, even when there is no predetermined structure, speakers are able to pre-plan their ongoing speech in such a way that its hierarchical global structure is reflected in its prosody.

Of course, the instructions we employed are of a relatively simple nature, and speakers may be able to provide these with a gradually varying melodic feature because they can foresee to some extent the amount of speech they will devote to a particular building block. The situation may be very different in daily, non-monitored small-talk (Umeda, 1982; Ayers, 1992). Moreover, unlike in our monologues, in which the transmission of information is predominant, the prosodic structure in informal conversations is likely to be influenced by many other features; these may be interactional, attitudinal and/or emotional in nature.

From the listener's point of view, global melodic features, having a whole discourse unit as their domain, are likely to be very useful, as they may enable him/her to predict an upcoming ending. Indeed, Menn and Boyce (1982) reported that the bottom of a speaker's pitch range is fairly constant, so that a listener would be able to anticipate an approaching end if the speaker is reaching the bottom of his/her range. The predictive value of prosody in this respect is treated in more detail in Grosjean (1983), Geluykens and Swerts (1993) and Swerts (1993).

In general, our perceptual tests revealed that listeners are quite sensitive to the types of prosodic indications of information structure discussed above. In the literature, one also finds attempts to investigate the communicative impact of prosody in a more direct way, by studying whether text comprehension is influenced or facilitated by adequate macro-prosodic structure. These questions have been addressed in tests with synthesized speech using the implementation of read-aloud paragraph prosody in a text-to-speech algorithm. Silverman (1987) found that ambiguous texts could be disambiguated by means of paragraph prosody, while Sluijter and Terken (1992) found that texts with adequate paragraph prosody were judged to be more 'natural' than those lacking such a component. It will be a challenging task to find an adequate experimental paradigm to study, along similar lines (e.g. by means of rule-based implementations), the communicative impact of macro-prosody of more spontaneous types of speech.

The results obtained in this investigation show that the methodology we employed is a useful one for analyzing prosodic aspects of discourse. First of all, the way the spoken material was elicited gave us some control over the kind of discourse produced by the speakers, while still resulting in relatively unconstrained, spontaneous speech. The framework developed for analyzing discourse structure gave us an independent criterion for evaluating the topical structure of the monologues investigated. Secondly, the perception experiment enabled us to evaluate the perceptual impact of the prosodic features produced by the speakers in a sufficiently controlled manner. This combination of speech production analysis under controlled circumstances, and experimental evaluation of perceptual significance has thus proved to be a good strategy for analyzing functional aspects of prosody in larger-scale discourse units. This emphasis on the listener's as well as on the speaker's point of view is something which is, of necessity, missing from the corpus-based research in the field, and requires a rigorous experimental approach.

Naturally, this study has its limitations. Perhaps the most important one is that, the speech produced here, although it was spontaneous, was also a restricted kind of monologue discourse. It remains to be seen to what extent the results obtained here can be extrapolated to more interactive types of discourse such as dialogue. In dialogue, other functional dimensions, such as the turn-taking system, come into play. Since prosody is also employed for signalling this interactive dimension, it would be interesting to see in what way these two dimensions, i.e. information flow and interaction, are realized through prosody. We have already started some experimental research dealing with this problem (Geluykens and Swerts, 1992), but much work remains to be done.

(Received May 7, 1993; accepted October 26, 1993)

REFERENCES

AYERS, G.M. (1992). Discourse functions of pitch range in spontaneous and read speech. Presented at the Linguistic Society of America Meeting, Philadelphia, PA, January 1992.

BOLINGER, D.L. (1989). *Intonation and Its Uses: Melody in Grammar and Discourse*. Stanford: Stanford University Press.

BROWN, G., CURRIE, K., and KENWORTHY, J. (1980). *Questions of Intonation*. London: Croom Helm.

BROWN, P., and LEVINSON, S.C. (1987). *Politeness: Some Universals in Language Usage*. Cambridge, U.K.: Cambridge University Press.

BRUBAKER, R.S. (1972). Rate and pause characteristics of oral reading. *Journal of Psycholinguistic Research*, **1**, 141—147.

BRUCE, G. (1982). Textual aspects of prosody in Swedish. *Phonetica*, **39**, 274—287.

CHAFE, W.L. (1976). Givenness, contrastiveness, definiteness, subjects, topics, and points of view. In C.N. Li (ed.), *Subject and Topic* (pp. 21—55). New York: Academic Press.

CHAFE, W.L. (1987). Cognitive constraints on information flow. In R. Tomlin (ed.), *Coherence and Grounding in Discourse* (pp. 21—55). Amsterdam: Benjamins.

CHARPENTIER, F., and MOULINES, E. (1989). Pitch-synchronous waveform processing techniques for text-to-speech synthesis using diphones. *Proceedings Eurospeech '89* (pp. 13—19), Paris, 1989.

CRUTTENDEN, A. (1981). Falls and rises: Meanings and universals. *Journal of Linguistics*, **17**, 221—231.

CRUTTENDEN, A. (1986). *Intonation*. Cambridge, U.K.: Cambridge University Press.

CUTLER, A., and PEARSON, M. (1986). On the analysis of prosodic turn-taking cues. In C. Johns-Lewis (ed.), *Intonation in Discourse* (pp. 139—155). London: Croom Helm.

ESSER, J. (1988). *Comparing Reading and Speaking Intonation*. Amsterdam: Rodopi.

GELUYKENS, R. (1988). Five types of clefting in English discourse. *Linguistics*, **26**, 823—841.

GELUYKENS, R. (1992a). From Discourse Process to Grammatical Construction: On Left-Dislocation in English. Amsterdam/Philadelphia: Benjamins.

GELUYKENS, R. (1992b). Topics in English conversation: On topic-introduction in conversational discourse. Ms.

GELUYKENS, R. (in press). *The Pragmatics of Discourse Anaphora in English: Evidence from Conversational Repair*. Berlin: Mouton de Gruyter.

GELUYKENS, R., and SWERTS, M. (1992). Prosodic topic- and turn-finality cues. *Proceedings of the Workshop on Prosody in Natural Speech* (pp. 63—70), University of Pennsylvania, August 1992.

GELUYKENS, R., and SWERTS, M. (1993). Local and global prosodic cues to discourse organization in dialogues. *Proceedings of the ESCA Workshop on Prosody* (pp. 108 — 111), Lund University, September 1993.

GIVÓN, T. (ed.) (1983). *Topic Continuity in Discourse: A Quantitative Cross-Language Study*. Amsterdam/Philadelphia: Benjamins.

GRØNNUM, N. (1992). *The Groundworks of Danish Intonation*. Copenhagen: Museum Tusculanum Press.

GROSJEAN, F. (1983). How long is the sentence? Prediction and prosody in the on-line processing of language. *Linguistics*, **21**, 501—529.

GROSZ, B., and HIRSCHBERG, J. (1992). Some intonational characteristics of discourse structure. *Proceedings of the International Conference on Spoken Language Processing* (pp. 429—432). Banff, Canada, October 1992.

HALLIDAY, M.A.K. (1967). Notes on transitivity and theme in English, part 2. *Journal of Linguistics*, **3**, 199—244.

'T HART, J., COLLIER R., and COHEN, A. (1990). *A Perceptual Study of Intonation*. Cambridge, U.K.: Cambridge University Press.

HERMES, D.J. (1986). Measurement of pitch by subharmonic summation. *Journal of the Acoustical Society of America*, **83**, 257—264.

JOHNS-LEWIS, C. (ed.) (1986). *Intonation in Discourse*. London: Croom Helm.

KREIMAN, J. (1982). Perception of sentence and paragraph boundaries in natural conversation. *Journal of Phonetics*, **10**, 163—175.

LEHISTE, I. (1975). The phonetic structure of paragraphs. In A. Cohen and S. Nooteboom (eds.), *Structure and Process in Speech Perception* (pp. 195—206). Berlin: Springer-Verlag.

LEHISTE, I. (1979). Perception of sentence and paragraph boundaries. In B. Lindblom and S. Öhman (eds.), *Frontiers of Speech Communication Research* (pp. 191—201). London: Academic Press.

LEHISTE, I. (1980). Phonetic characteristics of discourse. *Acoustical Society of Japan, Transactions of the Committee on Speech, April 1980*, 25—38.

LEVELT, W. J.M. (1989). *Speaking: From Intention to Articulation*.Cambridge, MA: MIT Press.

MENN, L., and BOYCE, S. (1982). Fundamental frequency and discourse structure. *Language and Speech*, **25**, 341—383.

PRINCE, E.F. (1981). Toward a taxonomy of given-new information. In P. Cole (ed.), *Radical Pragmatics* (pp. 223—255). New York: Academic Press.

SACKS, H, SCHEGLOFF, E.A., and JEFFERSON, G. (1974). A simplest systematics for the organization of turn-taking in conversation. *Language*, **50**, 696—735.

SCHAFFER, D. (1983). The role of intonation as a cue to turn taking in conversation. *Journal of*

Phonetics , **11**, 243—257.

SCHAFFER, D. (1984). The role of intonation as a cue to topic management in conversation. *Journal of Phonetics*, **12**, 327—344.

SCHIFFRIN, D. (1987). *Discourse Markers*. Cambridge, U.K.: Cambridge University Press.

SILVERMAN, K.E.A. (1987). The structure of fundamental frequency contours. Unpublished doctoral dissertation, University of Cambridge.

SLUIJTER, A., and TERKEN, J.M.B. (1992). The development and perceptive evaluation of a model for paragraph intonation in Dutch. *Proceedings of the International Conference on Spoken Language Processing* (pp. 353—356). Banff, Canada, October 1992.

SWERTS, M. (1993). On the prosodic prediction of discourse finality. *Proceedings of the ESCA Workshop on Prosody* (pp. 96—99), Lund University, September 1993.

SWERTS, M., BOUWHUIS, D.G., and COLLIER, R. (1993). A perceptual study of melodic cues to 'finality'. *IPO Annual Progress Report* , **27**, 19—28.

SWERTS, M., and COLLIER, R. (1992). On the controlled elicitation of spontaneous speech. *Speech Communication* , **11**, 463—468.

SWERTS, M., and GELUYKENS, R. (1993). The prosody of information units in spontaneous monologue. *Phonetica*, **50**, 189—196.

SWERTS, M., GELUYKENS, R., and TERKEN, J.M.B. (1992). Prosodic correlates of discourse units in spontaneous speech. *Proceedings of the International Conference on Spoken Language Processing* (pp. 421—424). Banff, Canada, October 1992.

TERKEN, J.M.B. (1984). The distribution of pitch accents in instructions as a function of discourse structure. *Language and Speech*, **27**, 269—289.

THORSEN, N.G. (1985). Intonation and text in Standard Danish. *Journal of the Acoustical Society of America*, **77**, 1205—1216.

THORSEN, N.G. (1986). Sentence intonation in textual context – Supplementary data. *Journal of the Acoustical Society of America*, **80**, 1041—1047.

UMEDA, N. (1982). "F_0 declination" is situation dependent. *Journal of Phonetics*, **10**, 297—290.

YULE, G. (1980). Speaker's topics and major paratones. *Lingua* , **52**, 33—47.