Proceedings of the IRCS Workshop on Prosody in Natural Speech, August 5-12 1992

PROSODIC TOPIC- AND TURN-FINALITY CUES

Ronald Geluykens and Marc Swerts *

Institute for Perception Research (IPO)
P.O. Box 513, 5600 MB Eindhoven, The Netherlands

ABSTRACT

This paper describes an acoustic analysis of prosody in a variety of experimental types of dialogue. Subjects cooperatively had to perform a speaking task (i.e. describe simple rows of differently colored figures and signal their structure) and a listening task (i.e. respond to discourse boundaries in the speech produced by the interlocutor and take over as soon as the other had finished). It was found that the demarcation of discourse units by means of various intonation contours and accent shifts is largely dependent on the kind of discourse setting, in that speakers clearly take into account whether a conversational partner is likely to interrupt or not. Moreover, subjects appear not just to exploit local cues to signal the boundaries of larger-scale units. Our study reveals that they also have at their disposal: (1) a specific type of intonation contour (a level tone), occurring well before the actual end, that pre-signals that a unit will soon be rounded off; (2) topline-declination over the course of a topical unit that is different in final position than in non-final position; (3) a gradual shift in prominence in a NP from the adjective to the noun position over the course of a discourse unit.

1. INTRODUCTION

1.1. The problems

Conversation can be taken to be a kind of talk in which two or more participants alternate in speaking about particular topics (after Levinson 1983). Such a definition implies that a dialogue has both informational and interactive aspects: not only do interlocutors exchange ideas (topic dimension), they must also do so in an organized manner by regulating that no two speakers talk simultaneously (turn-taking dimension). Various conversation and discourse analysts have tried to find out what the linguistic devices are for signalling topical coherence in verbal com-munication on the one hand, and for governing the turn-taking mechanism on the other hand,

According to many researchers (e.g. Brown, Currie & Kenworthy 1980, Johns-Lewis 1986), prosody has a predominant role in structuring the flow of interactive spoken discourse. In particular, it is generally taken for granted that such suprasegmental features as speech melody, tempo, pause, etc. are used to demarcate both topical units and turns. Research has mainly concentrated on how the end of such units is signalled, as marking of finality is considered most relevant.

This paper also deals with the prosodic demarcation of discourse units in a conversation context and primarily focusses on finality cues. It addresses some problems largely neglected in the literature. Firstly, the notions 'turn' and 'topic' clearly are non-overlapping. For instance, it is quite conceivable that a speaker changes a topic within his turn or that a topic is continued over the turns of the participants. Therefore, we feel that it needs to be investigated how the prosodic structuring of topic flow interferes with that of the turn-taking, and vice versa (see section 2). Secondly, from experimental studies we know that a discourse unit may be rounded off on a global level too (e.g. Leroy 1984; Swerts, Bouwhuis & Collier (in prep.)). Hence, we will also try to find out whether there are some more indications that finality may be signalled prosodically well before the end of a unit (see section 3). The study reported upon here is experimental in nature, concentrates on Dutch and is mainly production-oriented. It includes both instrumental and auditory analyses.

1.2. Notes on methodology

Any experimental investigation on the role of prosody in governing both information flow and interaction in natural speech data is likely to run into methodological problems. On the one hand, in daily conversations, there are many potentially relevant variables involved that may all contribute to prosodic structure and that are difficult to vary independently. On the other hand, an experimental setting might be that far removed from ordinary dialogue that it gives us little insight in how the latter really works. In another paper (Swerts & Geluykens 1992), we show how one could get rid of some of these limitations for a monologue situation; in this paper, we develop a methodology which tries to find a compromise between, on the one hand, data which are spontaneous but too 'open-

ended' and uncontrollable, and, on the other hand, an experimental situation in which spontaneous interaction is still to some extent possible. In our dialogue set-up, we will attempt to vary the two levels mentioned above — topic flow and turn taking— independently.

1.3. The data

From a total of ten test subjects, two participants at a time were seated in front of each other in a sound-treated studio. In between them, a screen was placed, so that they could only hear, but not see each other. Together they had to perform five slightly different experiments that consisted of a speaking and a listening instruction: while one of the two was speaking, the other had to execute a listening task. Also, in all but the first experiment, there was a constant switching of roles, in that, during a test, speakers at particular times became hearers, and vice versa. (A schematic representation of the different discourse settings is given in Table 1; see appendix for all tables).

	instruction to speaker	instruction to listener
exp 1 (monologue)	- signal breaks between series	- indicate perceived breaks on sheet
exp 2 (dialogue)	- signal end of row	- take over at end of turn
exp 3 (dialogue)	- signal breaks between series - signal end of row (end of row always equals end of series)	- indicate perceived breaks on sheet - take over at end of row
exp 4 (dialogue)	- signal breaks between series - signal end of row (end of row never equals end of series)	- indicate perceived breaks on sheet - take over at end of row
exp 5 (dialogue)	- signal breaks between series - signal end of row (end of row sometimes does, sometimes does not equal end of series)	- indicate perceived breaks on sheet - take over at end of row - indicate whether end of row is equal to end of series or not

Table 1: Schematic representation of the different discourse settings used (further explanations in the text)

The general speaking task in the different experiments was to convey relatively simple chunks of information, viz. describe from left to right rows of differently colored geometrical figures (see also Swerts & Collier, in press; Levelt 1989). Care was taken that no figure or color occurred in two successive positions, to avoid effects of given-new information on production. Within each row, some successive figures were visually presented as belonging together by drawing connecting lines in between them; subsequent series (which we will label 'topics') were being presented as being unconnected there being no visual link (see figures 1A-B for examples). In this way, series of two to seven figures were created randomly, and, except in experiment 2, each row consisted of at least two of such series (subjects were not told how many series to expect). When describing the rows with series of figures, a speaker was not allowed to use lexical or syntactic cues to clarify the structure of the rows, and therefore could only exploit prosody to, for instance, signal the breaks between series or indicate the end of a row.

In experiment 1, a monologue setting designed to test information flow structuring independently from turn-taking considerations, the speaker was instructed to describe rows with series of geometrical figures to his partner in such a way that the major breaks between successive series became apparent. His partner, the hearer, had to try and detect these breaks, indicating this on an aswer sheet in rows with numbers (see figure 1C).

In experiment 2 a kind of enforced turn-taking was introduced. Both of the participants were given an instruction sheet, that consisted of five rows with geometrical figures and five rows with numbers, and each row of figures alternated with a row of numbers. Rows with figures were used for the description task, rows with numbers for the listening task. Subject A received a sheet with a first row of figures, subject B had a sheet with a first row of numbers. There was no subdivision in series present in the rows of figures, so speakers only had to signal when the other had to take over. The hearer just had to count the figures described, indicate this in his row with numbers and start describing his row with figures as soon as he thought that the other had stopped describing his.

In experiment 3, the instruction sheets given to the subjects were identical to those of experiment 2, except that the rows of figures were subdived into smaller series of geometrical figures. The task assigned to the speaker was now twofold: he had to make clear to the hearer when his row was completed, so that the other could start describing his row with figures; he also had to indicate where in his row the breaks occurred between successive series. The task to the hearer was also twofold: he had to take over, as soon as he thought that the other had rounded off a row, and he had to indicate on his answer sheet where he heard the major breaks. Note that in this experiment, the end of a row always coincided with the end of a series.

Experiment 4 was the same as experiment 3, except that a row of figures ended in an incomplete series (indicated visually as in figure 1B), so that the interlocutor had to, as it were, finish this series after taking over the floor. In this experiment , all the rows ended with such an incomplete series.

Experiment 5, finally, was a combination of experiment 3 and 4. The speaking and listening tasks again were the same as in experiment 3, only this time rows either could end in an incomplete or a complete series.

Speakers were asked to make this difference clear to the hearer, who had to try and indicate this on an answer sheet, by using either '>' for continuation or 'II' for finality (figures 1 C-D). As this task was more complex, subjects were asked to do this experiment twice (data from both sessions were analyzed).

Data from experiments 1 to 5 were auditorily analyzed to investigate the interference of topic and turn demarcation (see section 2). Measurements of global indicators of finality (see section 3), are based on an auditory analysis of experiments 3 to 5, and on an instrumental investigation of the speech materials from experiment 3.

2. INTERFERENCE OF PROSODIC DEMARCATION OF TOPICS AND TURNS

In this section, it is discussed how the prosodic demarcation of topical units may interfere with cues that signal turn-taking, and vice versa. In 2.1., we treat the distribution of various intonation contours as a function of discourse position. In 2.2., we look at accent structure in relation to the topic- and turn-dimensions.

2.1. Final contours

Since the distinction between falling versus rising contours is often claimed to be a powerful marker of finality versus continuation in discourse, we have auditorily determined the shape of the contour at several crucial locations in the patterns produced by our speakers. Contour shapes are depicted in figure 2, which has to be consulted together with Table 2 to get the full picture. For our purposes, it appears to be sufficient if we classify contours according to two parameters: type of final movement, which determines the contour label, and end-point of that movement in the pitch range. The 'normal' pitch range is divided into a low, mid and high part, to which two marked values are added, viz. very high and very low. Table 2 only depicts the major trends for each speaker.

Three rising contours ending in mid-position can be distinguished which are used topic-internally in all settings. As there appears to be no systematic difference in the distribution of these three rises, and since they do not differ as to direction of movement or end-point, all are referred to as rise-to-mid (RM) in Table 1. (Note that some series-internal contours are systematically different, however; these will be discussed in section 3.1., and are left out of the picture for now.) For some turn-internal finality markings, a rise-to-high (RH) contour is employed, especially in the more complex settings. For most outspoken finality, speakers mostly use a fall-to-low (FL).

When only one kind of finality has to be marked, as is the case in experiments 1 and 2, this is rather consistently done through a FL. Note, however, that some speakers try to signal the additional hierarchical organization in experiment 1 by using a RH for 'minor' topic finality; this indicates that, even in such a monologue setting, there is no simple correspondence between falling tones on the one hand, and finality on the other hand. In experiment 3, FLs are reserved for the [+topic][+turn]-final positions, whereas turn-internal topics are marked by a RH (except for one speaker). To conclude from this that falling intonation is primarily reserved for marking turn finality, however, would be a mistake, as can be deduced from the results of experiment 4, in which both RHs and FLs are used to signal within-turn topic finality. Even in those cases where RHs are employed, however, speakers very rarely use a fall to signal turn-finality, presumably because in this setting, it is in conflict with topic-continuation.

couple:	1		2		3		4		5	
speaker:	1	2	3	4	5	6	7	8	9	10
exp 1										
[-topic]	RM	RM	RM	RM	RM	RM	RM	RM	RM	RM
[+topic]	RH	RH	FL	FL	FL	RH	FL	FL	RH	RH
[+topic series]	FL	FL	FL	FL	FL	FL	FL	FL	FL	FL
exp 2										
[-turn]	RM	RM	RM	RM	RM	RM	RM	RM	RM	RM
[+turn	RH	RH	FL	FL	FL	FL	FL	FL	FR	FL
exp 3										
[-turn] [-topic]	RM	RM	RM	RM	RM	RM	RM	RM	RM	RM
[-turn] [+topic]	RH	RH	RH	RH	RH	RH	FL	RH	RH	RH
[+turn] [+topic]	FL	FL	FL	FL	FL	FL	FL	FL	FL	FL
exp 4										
[-turn] [-topic]	RM	RM	RM	RM	RM	RM	RM	RM	RM	RM
[-turn] [+topic]	RH	RH	RH	RH	FL	RH	FL	FL	RH	RH
[+turn] [-topic]	L	FL	RF	RF	RH	RH	RH	RH	L	L
exp 5										
[-turn] [-topic]	RM	RM	RM	RM	RM	RM	RM	RM	RM	RM
[-turn] [+topic]	RH	RH	RH	RH	RH	RH	FL	FL	RH	RH
[+turn] [-topic]	L	RV	RF	RF	L	L	RH	RV	RV	RH
[+turn] [+topic]	FL	FL	FL	FL	FV	FL	FV	FL	FL	FL

Table 2: pitch movements in topic- and turn-final positions (see also figure 2)

Speakers use various alternatives in the most complex setting 5. On the whole, there seem to be two major strategies: (i) speakers either create another rise or fall level by going beyond the commonly employed pitch range, resulting in both falls-to-very-low (FV) and rises-to-very-high (RV); (ii) or they create another major tone by varying pitch movement, mostly resulting in a level tone (L), with a pitch which stays at mid-level towards the end, or a rise-fall (RF), which can be defined as a falling movement preceded by a rise in pitch. In these more complicated tasks, speakers appear to have no problems signalling these different categories (they also score well perceptually, cf. section 5).

Generalizing, one can say that speakers reserve a FL for the cases where the two types of finality occur together. In cases where there is only topic-finality ([+top][-turn]), a RH is most often used (as in exp 3); in cases where there is only turn-finality ([-top][+turn]), a variety of patterns occurs, but most speakers are very well able to keep these three levels distinct.

It can be concluded, then, that prosodic demarcation by means of various intonation contours is largely dependent on the type of discourse setting: speakers clearly take into account whether a conversational partner is likely to interrupt him or not, which manifests itself in the intonational characteristics of his utterances. Summarizing, final falls are regularly used to signal both topic- and turn-finality when they are not in conflict; otherwise, low falls are reserved for the 'deepest' finality level, whereas high rises and/or other tones serve to signal other finality dimensions, both informationally and interactionally. It would thus be a mistake simply to equate 'falling' prosody with 'finality' without being more specific.

2.2. Accent structure

Having established that speakers use different types of pitch contours to structure their speech and signal various kinds of finality, we will now investigate a second prosodic dimension which appears to be relevant for the demarcation of discourse units, viz. accent placement. We have pointed out in the data description (see section 1.3.) that, in order to avoid interference from the given-new structure of the discourse, each string of figures was constructed in such a way that in each figure, both adjective and noun could be considered non-recoverable (Geluykens 1988, 1991, 1992), or 'new' information, in the sense of not being predictable from the preceding context. In such a situation, given the rules for 'neutral' accenting, one would expect the noun to carry the strongest accent, with perhaps a secondary accent on the adjective. An auditory analysis of the data, however, shows a different picture. For each of the experiments, the description of each figure was auditorily evaluated (independently by both authors, with a third session in cases of doubt), and put into one of two categories: adjective-noun compounds with strongest accent on the adjective (A), and compounds with the strongest accent on the noun (N). The percentage of 'untypical' A-accents was then calculated for each figure, relative to its position in a topical string. Results can be found in Table 3.

Table 3 shows several things. First of all, it was clear from listening to the data that not all speakers treated accentuation the same way: although the majority of them appeared to vary accent placement, and put the main accent sometimes on the adjective rather than on the noun, there were three speakers ('N-N') who rather consistently accented the noun (as we expected all of them to do). We thus distinguished two categories of speakers. Secondly, the 7 'variant' speakers ('A-N') all exhibited a clear pattern, in that there was a clear shift from a high percentage of accents on the adjective in initial position in a series, through a slightly lower percentage in mid-positions, to a very low percentage of accents on the adjective in final position. In Table 3, the averages for initial positions, final positions, and all mid-positions (i.e. positions 2-3-4 for a string of five figures, etc.) are depicted.

	initial position	mid-positions	final position
exp 1			
7 A-N speakers	32.4 %	28.6 %	2.7 %
3 N-N speakers	6.5 %	11.7 %	3.2 %
exp 2			
7 A-N speakers	45.1 %	37.2 %	0.0 %
3 N-N speakers	6.7 %	10.6 %	0.0 %
exp 3			
7 A-N speakers	70.3 %	58.1 %	7.2 %
3 N-N speakers	10.0 %	28.0 %	6.0 %
exp 4			
7 A-N speakers	48.6 %	34.8 %	2.9 %
3 N-N speakers	33.3 %	39.5 %	6.7 %
exp 5			
7 A-N speakers	66.9 %	58.2 %	3.3 %
3 N-N speakers	16.3 %	11.2 %	2.0 %
exp 3 + 4 + 5 (mean)			
7 A-N speakers	61.9 %	50.4 %	4.5 %
3 N-N speakers	19.9 %	26.2 %	4.9 %

Table 3: percentage of A-accents in three major positions (auditory analysis)

The table shows clearly that the behavior of the two main groups of speakers is strikingly different (1st vs 2nd rows). This is in itself a most interesting finding, as it indicates clearly that there are no clear-cut 'rules' for signalling discourse structure through accentuation. From the data of the 7 A-N speakers, we learn, however, that they can make subtle use of accent placement to provide an extra cue to the hearer as to discourse structure. Since this shift in prominence cannot really be due to interference from the given-new structure of these strings (as in each string both color and figure were different from, and unpredictable from, the preceding context), it must be concluded that this prosodic dimension is used as a device for bringing out topical structure: speakers use it, as it were, to highlight the extreme ends of a discourse unit.

Note also that neither topic structure (exp 1) nor turn structure (exp 2) in itself appears to be sufficient to cause a significant shift to A-accents (Table 3). When the two factors are combined, A-N speakers do show a striking increase in A-accents, which seems to imply that this is a matter of both information flow and interaction (the lower figures for experiment 4 are somewhat puzzling in this respect): the greater the complexity of the task involved,

the greater a need speakers appear to feel to exploit all prosodic variables to the full.

3. GLOBAL CUES TO FINALITY

A second aspect we wanted to address in this paper is the globality of finality-cues. The underlying question is whether finality of discourse units is signalled well before the actual end, so that hearers to some extent are enabled to predict when a speaker will round off a unit. In 3.1., we discuss a specific kind of intonation contour (a level tone), which appears to function as a non-local finality cue. In the subsequent sections, we present some results of acoustic measurements on data of experiment 3. In 3.2., the phenomenon of topline declination is treated. In 3.3., we embark on the gradual shift in prominence strength from the adjective to the noun over the course of a topical unit.

3.1. Analysis of non-final contours

Turning now to non-local signalling of discourse finality, there is one intonational cue which appears to be very prominent, and which we think may well be perceptually relevant. We have indicated in Table 1 that non-final figures are consistently marked with a rise-to-mid contour (RM). However, in experiments 3 through 5, each row of figures consisted of more than one series. In the final one of those strings, i.e. the string just before turn-taking occurs, internal figures tend to be marked differently, not with a low rise but with a kind of level tone. This pattern can be clearly distinguished from the RM: in the level tone, there is absence of outspoken pitch movement on the second accent, whereas in the 'real RMs' one observes a clear accent-lending fall or rise there. Note, though, that this level (L) tone has the same end-point as both RMs depicted in figure 2; in other words, we observe some prosodic similarity between all internal tones. Informal listening to these contours yield the strong impression that these contours pre-signal that the series is the last one in the turn, independentely from its final pitch contour. We have planned to investigate to what extent this factor is perceptually relevant.

3.2. Topline declination

Another prosodic dimension which we looked at, is the relative height of the Fo peaks in each string (topline declination). We limited the acoustic measurements to the speech materials from experiment 3. To calculate the values in table 4, we have selected the highest Fo on every A-N compound, irrespective of whether this occurred on the adjective or the noun, and compared this to all the other figures in the same string. Since strings consisted of two up to six figures (seven-figure strings were not used, as they are too rare), results were calculated for series of various lengths (see Table 4). The way this was done is the following: a mean peak height score in Hz for each speaker was calculated, and this was subtracted from actual peak heights in each position, to allow inter-speaker comparison. In Table 4, a positive figure thus indicates a peak height above mean peak height, a negative figure indicates peak height below mean peak height. Table 4 also distinguishes between turn-final (F) and non-turn-final (NF) strings, in order to assess the potential relevance of peak height variation in pre-signalling turn-finality.

position:	1st	2nd	3rd	4th	5th	6th
series of 2						
NF	+19.1	+5.9	—	—	—	—
F	+14.8	-37.4	—	—	—	—
series of 3						
NF	+16.4	+2.9	-1.5	—	—	—
F	+6.6	-9.1	-15.8	—	—	—
series of 4						
NF	+20.7	-0.5	-4.6	-2.3	—	—
F	+9.6	+2.5	-11.0	-23.2	—	—
series of 5						
NF	+19.5	-2.1	-5.4	-13.2	-5.6	—
F	+11.4	-15.7	-17.8	-6.7	-33.2	—
series of 6						
NF	+15.1	-7.5	-6.7	-11.6	-15.5	-6.1
F	+21.8	+11.8	+9.6	+9.2	-7.8	-23.8

Table 4: relationship between Fo peaks in final vs non-final series (see explanation in text)

Table 4 shows, first of all, that generally speaking there is indeed top-line declination present in each topical string; the first element in each string receives the highest Fo peak, and peak height then gradually declines. This appears to be independent of the actual length of the series. Moreover, the degree of declination seems to differ between turn-final and non-turn-final strings: whereas for turn-final strings there is indeed gradual declination up to and including the last item, in non-turn-final strings it is often the case that it is the before-last item which has the lowest peak; even in instances where this is not true in absolute terms, final peaks are still significantly higher than they are for turn-final strings. Note also that initial peak heights in non-final series are, generally speaking, higher than those in final series (apart from series of 6). Though results are not very conclusive, relative peak height of Fo peaks does appear to be important in two ways. Firstly,

peak height declination signals to some extent the topic structure of each turn, highest peaks occurring on the first item. Secondly, peak height at the end signals to some extent turn-finality (though peak height comparisons are by no means easy turn-finally, as we are dealing with a different intonation contour, viz. a FL). Once again, we are thus dealing with potential global prosodic cues for signalling discourse finality, as final series differ from non-final series with respect to some properties of the topline declination.

3.3. Relative differences in height of maxima in pitch accents

In our auditory analysis of accent positions (see 2.2.), it struck us that not all adjectival and nominal accents appeared equally strong. To give some acoustical support to our impression, we calculated, for each A-N compound, the difference in semitones in pitch height between the peak on the noun (if present) and the peak on the adjective (if present), assuming that this measure somewhat reflects the relative strength of the accents in each string. A high average peak on the adjective thus gives a negative value (N<A), a high peak on the noun a positive one (N>A). Measurements are only performed on the speech data of experiment 3 of 7 'A-N' speakers (see 2.2.). Results are represented in Table 5.

position:	1st	2nd	3rd	4th	5th	6th
series of 2						
F	-1.53	+0.07	—	—	—	—
NF	-3.07	+2.82	—	—	—	—
series of 3						
F	-1.35	-1.25	+1.04	—	—	—
NF	-1.92	-0.69	+4.00	—	—	—
series of 4						
F	-1.46	-0.04	-0.62	+1.34	—	—
NF	-1.31	+0.99	-0.76	+3.97	—	—
series of 5						
F	-2.35	-0.73	-0.52	-0.75	+0.95	—
NF	-2.42	-0.44	-0.36	+0.36	+2.44	—
series of 6						
F	-1.60	-1.57	+0.67	-1.57	+0.87	+1.70
NF	-1.17	-0.01	-0.13	-0.94	-0.01	+3.22

Table 5: relationship between A- and N-accent (in semitones) (7 speakers) (see text)

The results of this instrumental analysis are once again striking. Table 5 confirms the impression that, even in those cases where speakers, say, consistently place the accent on the adjective except for the last figure (resulting in e.g. an A-A-A-A-N series), it is the initial position which receives the strongest accent (highest Fo peak in relation to the noun), while subsequent accents appear to be less outspoken, and decrease gradually. In some cases, peaks on adjective and noun are about equally strong; although we forced our data into either an A- or an N-category for the auditory analysis, a subtler transcription method seems to be in order here. In other words, we observe a gradient shift in prominence strength from the accent to the noun over the course of a topic; we thus have found another global characteristic of a discourse unit. Moreover, there seems to be a difference between final and non-final series, in that the last NPs of final series have a less prominent noun-accent than the last NPs of non-final series (though, of course, still more prominent than the adjectival accent).

To supplement the data in Table 5, we have once again calculated the averages for all initial positions in a series, all final positions, and all intermediary positions, irrespective of the length of the series. This gives the situation presented in Table 6.

		initial position	mid-positions	final position
7 A-N speakers	F	-1.62	-0.55	+1.02
	NF	-1.98	-0.20	+3.29
3 N-N speakers	[F+NF]	+1.84	+1.65	+1.13

Table 6: relationship A- and N-accent in three major positions (in semitones)

Table 6 confirms, first of all, the findings of the auditory analysis for experiment 3 depicted in Table 3 above: the 7 A-N speakers show negative values in both initial and mid-position, while final position is highly positive, as a result of the much higher peak on the noun. Moreover, the average peak on the adjective is much higher in first position than it is in mid-position, confirming the finding that initial A-accents seem much more pronounced than intermediate A-accents. Those tendencies are more outspoken in NF- than in F-series, especially for the last positions in the rows. Table 6 thus brings out the results of Table 5 even more clearly. The 3 N-N speakers, which we have also included here for comparative reasons (we have added up F- and NF-series here), have positive values in all three locations, reflecting a higher average peak on the noun in all positions.

4. CONCLUSIONS

The preceding sections have provided both auditory and instrumental evidence for the claim that prosody is indeed used for structuring spoken discourse, both on the level of information flow (signalling topical units) and the level of interaction (signalling turn-taking). First of all, it has been shown that, from the speaker's point of view, prosodic

demarcation is largely dependent on the type of discourse setting: they clearly take into account whether his conversational partner is likely to interrupt him or not. Secondly, we found that both local (final intonation contours) and global (pitch range, accentuation) cues appear to be employed to structure spoken discourse. By combining information and interaction in a relatively simple interactive experimental setting, the pure contribution of prosody to the structuring of the discourse could be studied easily, although the precise impact of the different prosodic features requires further investigation.

This study can be extended in a number of ways. First of all, further perceptual research is needed to determine the relevance of these prosodic cues to the hearer. A kind of informal 'on-line' perceptual analysis was, of course, provided by the second participants in all five experimental settings, since, on top taking over turns, they also had to mark the discourse structure of the speech produced by their interlocutor. This gives us the chance to evaluate to what extent information flow was deduced successfully. Despite the difficulty of some of the tasks, success rate was quite high. Even for the most complex task in experiment 5, viz. deciding whether the turn-final series was 'complete' or not, subjects scored significantly above chance level (about 80 % correct). The perceptual efficiency of the turn-taking cues was also tested on-line, of course, by virtue of the fact that interlocutors had to react immediately by taking over the floor. Here, too, very few problems occurred. One could argue, however, that pause duration, or other factors, still functioned as important cues here apart from intonation, although the fluency as regards turn-taking is, on the whole, striking. Further perceptual experimentation is clearly needed here to determine the relative values of these different cues.

Secondly, it needs to be emphasized that both information flow and interaction have been kept relatively simple here; this research needs to be extended to more naturally occurring data. This will pose methodological problems, given the inherent contradiction between, on the one hand, collecting spontaneous data and, on the other hand, collecting data over which some variable control is possible. In another paper (Swerts & Geluykens 1992), we have shown how a compromise might be struck for a monologue setting. Similar methods wll have to be developed for the study of prosody in interactional settings.

(*) Both authors are also affiliated with the Belgian National Science Foundation (NFWO) and with the University of Antwerp (UFSIA and UIA, respectively).

REFERENCES

Brown, G., K. Currie & J. Kenworthy (1980). *Questions of intonation*. London: Croom Helm.

Chafe, W.L. (1987). Cognitive constraints on information flow. In: R.S. Tomlin, ed., *Coherence and grounding in discourse. Amsterdam: Benjamins, 21-55.*

Geluykens, R. (1988). Five types of clefting in English discourse. *Linguistics* 26: 823-841.

Geluykens, R. (1991). Topic management in conversational discourse: The collaborative dimension. *CLS* 27 (I).

Geluykens, R. (1992). *From discourse process to grammatical construction: On left-dislocation in English*. Amsterdam: Benjamins.

Johns-Lewis, C. (1986). *Intonation in discourse*. London: Croom Helm.

Levelt, W.J.M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.

Leroy, L. (1984). The psychological reality of fundamental frequency declination. *Antwerp Papers in Linguistics* 40, University of Antwerp, Belgium.

Levinson, S.C. (1983). *Pragmatics*. Cambridge [...]: C.U.P.

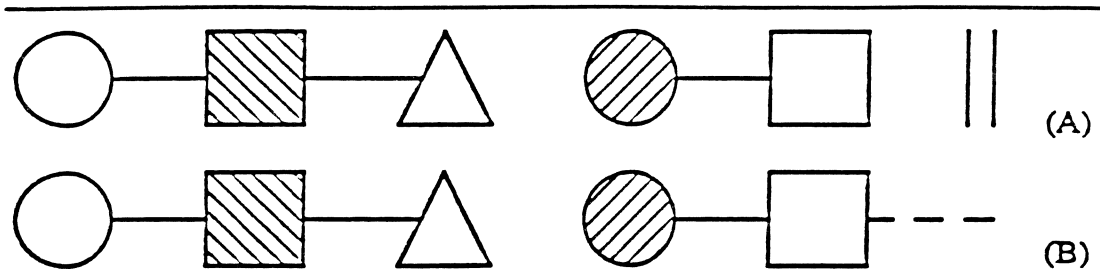
Sacks, H., E.A. Schegloff and G. Jefferson (1984). A simplest systematics for the organization of turn taking in conversation. *Language* 50: 696-735.

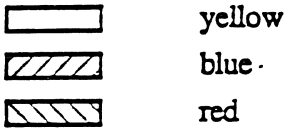
Swerts, M., D. Bouwhuis & R. Collier (in prep.). End(s) of intonation: A perceptual study of melodic cues to finality. Ms.

Swerts, M. & R. Collier (in press). On the controlled elicitation of spontaneous speech. To appear in *Speech Communication*.

Swerts, M. & R. Geluykens (1992). The prosodic structuring of information flow in spontaneous sppech. Paper presented at the Workshop on Prosody in Natural Speech Data, University of Pennsylvania, August 1992.

Swerts, M., R. Geluykens & J. Terken (in press). Prosodic correlates of discourse units in spoken monologue. To appear in *Proceedings of the 1992 ICSLP*, Banff, Canada.





1 2 3 I 4 5 II 6 7 8 9 10 ... (C)

1 2 3 I 4 5 > 6 7 8 9 10 ... (D)

Figure 1: example of production strings and perception tasks employed in exp 1-5

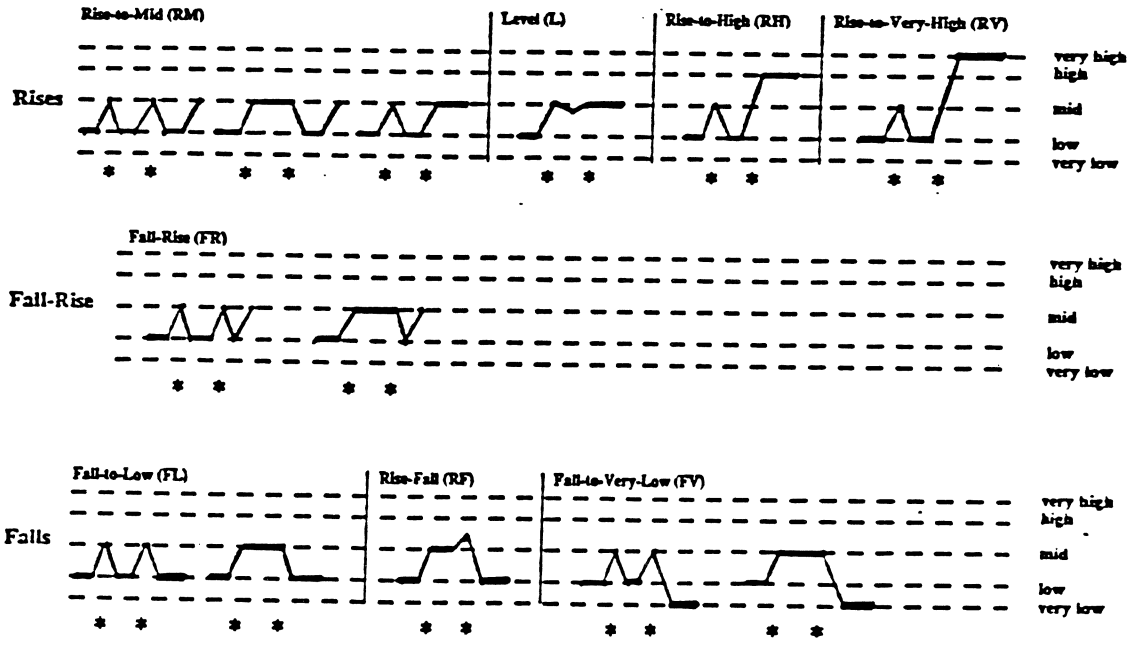


Figure 2: Detailed description of intonation contours. * represent the accents in the contours