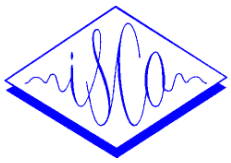

ESCA Workshop on Prosody
Lund, Sweden
September 27-29, 1993

On the Prosodic Prediction of Discourse Finality

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

ABSTRACT

It is investigated whether the approaching end of a spontaneously produced description is presignalled by prosodic means. Experiment I tries to determine to what extent listeners are able to estimate (on the basis of prosodic cues) how far a given utterance is situated from the end of a description. Experiment II is set up to systematically test - by means of manipulated synthetic prosody - the cue strength of speech melody and duration as predictors of discourse finality.

INTRODUCTION

One of the important functions of prosody is the demarcation of units of discourse. For instance, the end of discourse segments can be signalled by means of low boundary tones (e.g. Brown et al. 1980, Swerts et al. 1992). However, prosodic cues such as these tones are rather local in the sense of being positioned right before or at the actual boundary. It is a relevant question whether important breaks in the flow of information can also be presignalled, i.e. announced some time before they actually occur, so that listeners can anticipate them. Anecdotical evidence for the possibility of anticipation can e.g. be found in the small delay between the end of discourse segments and the onset of applause in political speeches.

Grosjean (1983) has already shown that subjects, basing themselves solely on prosodic cues, are surprisingly accurate at estimating the upcoming ending of a sentence. As Grosjean's study was limited to prosodic prediction at the sentence level in read-aloud speech, the present investigation is set up to test whether his findings can be generalized to (i) larger-scale discourse units (ii) in spontaneous speech. The analysis is centred on a particular type of descriptive language use, i.e. route descriptions.

EXPERIMENT I

In this experiment it is explored whether listeners can exploit prosodic cues of an utterance to estimate how far it is situated from the end of a (spontaneously produced) route description.

Speech materials, elicitation method

A Dutch speaker was asked to describe routes from given starting points to given end points on the basis of (schematic) city maps. Four of these descriptions were selected for further experimentation, the selection being based on the absence of clear non-prosodic (lexical) cues to discourse position. The seven last clauses of each description were isolated from their contexts, giving 28 utterances to be used in the listening experiment.

Perception test

Two lists of randomly ordered utterances were created, containing four repetitions of each utterance. Both lists were played to ten listeners each (chosen from students and staff of IPO) who were asked to estimate for each stimulus how many clauses (0 to 6), similar in length and nature to the one presented, followed in the original context. The interval between two utterances was 4 s, in which time period listeners had to respond. The listening test was preceded by the presentation of a short fragment of a route description (which was not used in the actual experiment) to get the subjects accustomed to the speaker's voice and speech register.

The listening results, averaged over the different types of clauses (-6 to -0) of the

four route descriptions, are given in figure 1, which is a three-dimensional plot of a confusion matrix.

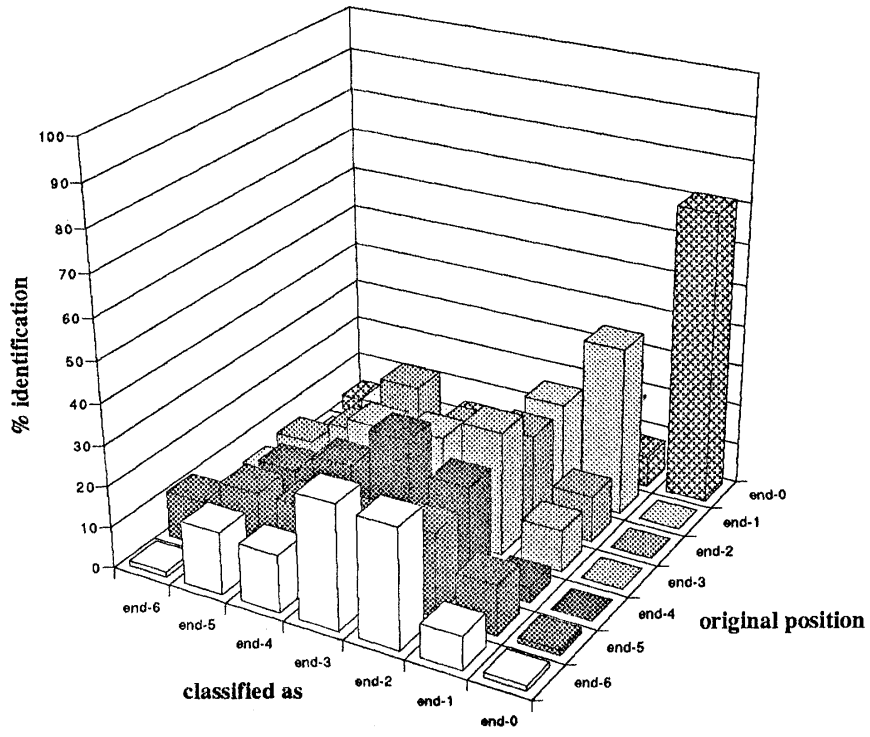

Figure 1. *3-D plot of perceptual results (averaged over the 4 route-descriptions)*

Figure 1 brings to light that in this speaker's route descriptions listeners only begin to hear that the end is approaching at one clause from the final one. As can be seen in the different bar chart diagrams, the answers to the clauses -6 to -2 are similarly distributed around the middle of the scale. The picture for clause -1 is different in that there is a clear preference of the listeners to label it - correctly - as the prefinal clause. Finally, the last clause of the total route description can most easily be classified correctly, which is reflected in the large amount of responses for the -0 class.

Acoustic measurements

Casual listening by two prosody researchers suggested that the most prominent cues were the pitch and duration of the last word of a clause. Therefore, fundamental frequency (Fo) was determined at the end of each clause and the relative duration of the last word of each clause was measured and compared to the average length of the same word read aloud four times in isolation by the same speaker. Average values were computed for each clause position over four descriptions (see figure 2).

It can be noticed that on the whole there is a close correspondence between the results of the perceptual test (see figure 1) and the acoustic variables measured. An ANOVA reveals that clauses -6 to -2 do not differ with respect to the two acoustic features (except for the difference in end frequency between -5 and -4), whereas the two other clauses, -1 and -0, differ from any of the others as to these two features, again with one exception: the difference in end frequency between -4 and -1 is not significant. From the prefinal clause onwards, a clear, though non-significant, tendency can be observed: end frequency gets lowered stepwise and the last word becomes increasingly shorter.

EXPERIMENT II

This experiment was set up to independently test the cue validity of both end frequency and relative duration of the last word in an utterance, keeping other (prosodic) variables

constant, in order to explore whether these parameters are sufficient to influence subjects' finality judgments.

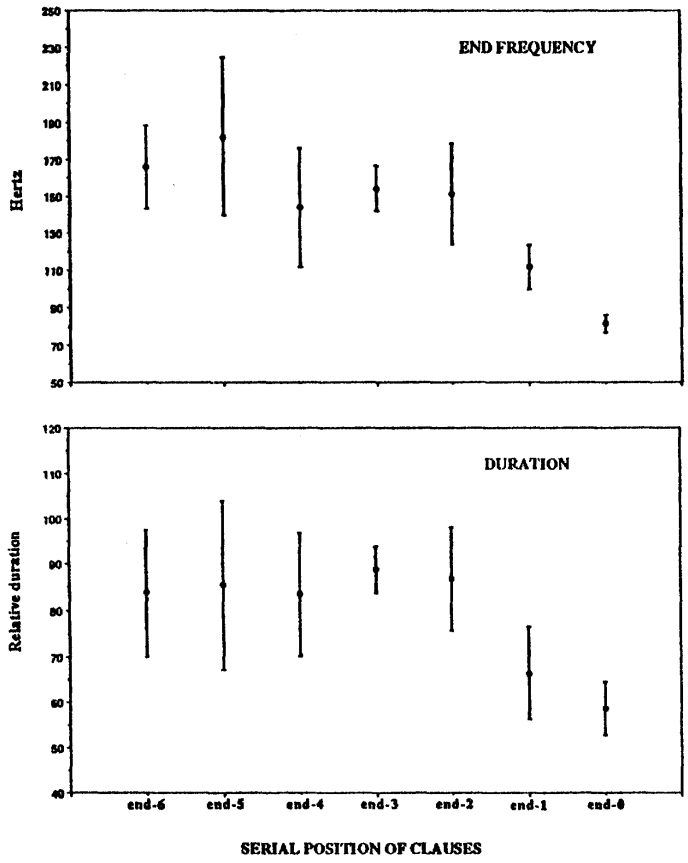

Figure 2. *Results of acoustic measurements (explanations in text)*

Speech materials, prosodic variables

The Dutch sentence "en dan gaan we rechtsaf" (and then we turn right), with accent on the syllable "rechts-", uttered by a male speaker at a normal speaking rate, was recorded with a 10 kHz sampling frequency at 12 bits. Melody and duration were varied by means of a wave-form manipulation technique (Charpentier and Moulines 1989), resulting in nine different prosodic patterns, as shown in figure 3.

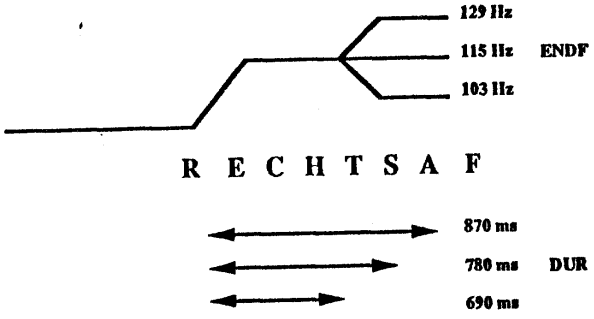

Figure 3. *Graphical representation of the stimuli used*

The different melodic contours that the utterance was provided with were identical from the beginning of the utterance up to the pitch accent in the syllable "rechts-". From that point, pitch could either rise on the syllable "-af" (and end in 129 Hz), fall (and end in 103 Hz) or remain at the higher declination line (and end in 115 Hz). The duration of the

word "rechtsaf" was either original (i.e. 870 ms), or reduced to 90% (780 ms) or 80% (690 ms), respectively.

Procedure

Two lists of randomized utterances were created, containing four repetitions of each utterance. Both lists were played to seven listeners each (all 14 subjects selected from students and staff of IPO). The interval between two successive utterances was 4 s, in which time period subjects had to respond. After having heard a particular sentence, listeners had to say how near to the end of a route description the sentence was uttered. They had to express their judgment on a ten-point scale with 10 meaning 'very close to the end' and 1 'far from the end'. The actual experiment was preceded by a random presentation of five of the utterances to give the listeners an impression of the stimuli.

The mean ratings, averaged over the responses of 14 listeners, are given in table 1, together with their standard deviations.

Table 1. *Results of perception test (explanations in text)*

Dur \ Endf	103 Hz	115 Hz	129 Hz
690 ms	7.32 (2.11)	6.67 (2.13)	5.39 (2.47)
780 ms	7.02 (2.49)	6.13 (2.18)	4.89 (2.51)
870 ms	6.30 (2.70)	5.48 (2.41)	4.32 (2.37)

Table 1 indicates that both prosodic variables had an effect on the listeners' perception of finality: a given utterance sounds more final the lower its end frequency (though all the boundary tones are not-low in the terminology of e.g. Brown et al. (1980)), and the faster the last word of the utterance is spoken.

DISCUSSION AND CONCLUSION

Summarizing the results of these experiments: speech melody, i.e. end frequency, and length, i.e. relative duration of the last word of an utterance, were shown to give information about the serial position of an utterance within a discourse unit. More specifically, in the monologues analyzed, intonational and durational properties distinguish between three classes of utterances: final, prefinal and non-final ones. The picture may be different for other speakers or for other speech genres, for instance dialogues. In the latter, prosodic predictors of finality may also play a role to indicate the ending of a turn: this phenomenon may then partly explain the often fluent transitions between speakers' turns, i.e. without much overlap nor delay (Levinson 1983).

ACKNOWLEDGEMENTS

The author is also affiliated with the University of Antwerp (UIA). The research was funded in part by the Belgian National Science Foundation (NFWO). René Collier, Ronald Geluykens, Angelien Sanderman, Margreet Sanders and Jacques Terken are thanked for stimulating discussions.

REFERENCES

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

F. Charpentier and E. Moulines (1989), "Pitch-synchronous waveform processing techniques for text-to-speech synthesis using diphones", *Proceedings EUROSPEECH '89, 2*, pp. 13-19.

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

S.C. Levinson (1983), *Pragmatics*, (CUP, Cambridge).

M. Swerts, R. Geluykens and J.M.B. Terken (1992), "Prosodic correlates of discourse units in spontaneous speech", *Proceedings ICSLP '92, Banff, Canada*, pp. 421-424.