General discussion

Abstract

In this final chapter the results from the various parts of this study are brought together and discussed. The main research questions in the present study were: Which prosodic cues do speakers use to signal the structure of a spoken discourse, and how do listeners use these cues to detect this structure? The results show that discourse boundaries are marked by the speakers predominantly with high boundary tones (also at locations where a low boundary tone was expected), and that heavier boundaries are marked with longer pauses. Listeners used acoustic pauses more than boundary tones as a cue for phrasing. Furthermore, there seems to be an ordering in the percentage ‘pitch-accented’ and ‘perceived as prominent’ relative to information structure: new > inferrable > modifier > discourse marker > verbs. Finally, the limitations of the present study are treated and suggestions for future research are provided.
6.1. Introduction

The present thesis dealt with two main research questions: first of all, what acoustic-prosodic means do speakers use to mark the information structure of a spontaneously spoken discourse, and secondly, which cues are most important for the listener to detect this structure. In other words, we want to know how *speakers* acoustically realize discourse structure, i.e. how do they indicate the important information of the discourse such as highly informative words and discourse boundaries. We also want to know how *listeners* use the acoustic cues provided by the speaker to determine the structure of the incoming discourse, i.e. on the basis of which acoustic cues do they decide which parts contain important information and where discourse boundaries are located?

We gave an overview of three different methods of textual discourse analysis. These methods served as a starting point for the development of our own framework for analyzing information structure in spoken discourse ('Information Structure In Discourse'), using a purely text-based method. The results of a pilot experiment showed that this framework is a very useful tool if one wants to obtain an independent analysis of global and local discourse structures, i.e. the location of different types of discourse boundaries and the information status of the various concepts that build up the discourse. This framework was therefore used in the remainder of this thesis.

The selection of speech material and the procedure of collecting and recording the material were presented next. Furthermore, an overall characterization of the speakers was given in terms of pausal and intonational features. Perceptual evaluations on the (prosodic) performance of the speakers in their production of the discourse completed this characterization. Speaker profiles, integrating the prosodic characterization in temporal and intonational terms, and the perceptual evaluation in terms of correlation with the ideal speaker, as well as the overall judgement, concluded this chapter.

Our next step was the investigation of prosodic characteristics of discourse boundaries. The location of boundaries in the text-based discourse structure was taken as a reference point, and the actual realization of boundaries by the speakers and the perception of boundaries by naive listeners was investigated, in terms of both pausal and intonational (boundary tones) aspects. Different strategies could be observed in the way speakers acoustically realize discourse boundaries, as well as in the use of acoustic cues by the listeners in perceiving discourse boundaries.

Finally, we presented a study on the focal structure in discourse, i.e. the relation between the text-based information structure (cf. the given/new distinction), on the one hand, and the realization (i.e. accentuation) and perception (i.e. prominence) of that structure, on the other hand. This relation was investigated by studying the location and realization of pauses and pitch
GENERAL DISCUSSION

accents relative to the information structure, the perceptual evaluation being obtained from prominence judgements by naive listeners.

This present chapter will bring together the results presented in the previous chapters. Section 6.2 presents a discussion of the main findings. In section 6.3 we formulate the answers to the research questions presented in chapter 1. Section 6.4 deals with the limitations of the present study and will present some suggestions for future research.

6.2. Discussion of main findings

Our study of the prosodic aspects of information structure in discourse was broken down into three parts: prosodic speaker characteristics, prosodic characteristics of boundaries, and prosodic characteristics of focal structure. Each of these parts will be discussed separately. Before that, however, we will go back to the model representing the various aspects in this study, reproduced below in Figure 6.1.

![Figure 6.1. Overview of the various aspects in this study and their relation.]

The central part in this figure is the discourse, i.e. the message between the speaker and the hearer. The speaker makes use of concepts to formulate his/her message. The structure of the message, in this case the spoken discourse, is realized by the use of both linguistic and acoustic means. The listener also makes use of these means to reconstruct the message of the speaker (the discourse), this time in terms of percepts.

The discourse can thus be seen from two sides: the structure as produced by the speaker and that perceived by the listener. On both sides, at least two levels can be distinguished: one to indicate the broad structure of the message in terms of where discourse units begin and end (‘discourse structure’ in the diagram) and
one to indicate the narrow structure in terms of important information ('information structure' in the diagram).

In this thesis, we focus on the path from the speaker via spoken discourse to the listener. The part from the writer through written discourse to the reader, as presented in chapter 2, lies outside the scope of the present study. The text-based discourse analysis is taken as reference point. The prosodic realization of the 'discourse structure' and 'information structure' by the speakers is investigated, as well as the perception of both types of structures by the listener.

In the following three subsections, the main results will be discussed. In section 6.2.1 the characterization of the speakers is presented. It shows what prosodic means were used by the speakers to realize spoken discourse, and how the use of these means was evaluated by the listeners. Section 6.2.2 discusses the results of the prosodic characteristics of discourse boundaries, both from the speakers' and the listeners’ point of view. The central issue here is which cues are used by the speaker to mark discourse boundaries in spoken discourse, and which cues are used by the listeners to detect these boundaries. Section 6.2.3, finally, discusses the findings on the prosodic characterization of focal structure. Here, information structure is investigated, again from the speakers’ and the listeners’ viewpoint. How do speakers prosodically mark information structure in spoken discourse, and how do listeners use the prosodic cues to detect important information?

6.2.1 Prosodic speaker characteristics

The perceptual evaluation of the ‘prosodic retelling performance’ of the eight speakers, as presented in chapter 3, showed that listeners have a clear idea about how the ideal speaker of a retold story should sound and that they for the most part on this. Furthermore, they also have specific ideas about the acoustic realization of retold stories. Some speakers are judged as ‘better’ retellers than others.

Listeners thus clearly make a distinction between ‘good’ and ‘bad’ speakers, which is related to the picture of the ideal speaker. The closer a real speaker is to this picture, the better he/she is evaluated. In this respect, aspects related to the factor Intonational characteristics are most important, followed by aspects related to the factors Dynamic features and Articulation. However, this ordering is not identical for all speakers: to characterize their speech, the Dynamic features are of more importance for some speakers, whereas for others the Intonational characteristics are more important.

The relation between these perceptual judgements and the acoustic realization by the speakers can be summarized as follows: The acoustic cue related to the Intonational characteristics factor is mean F0. Therefore, listeners based their judgements basically on the average F0 of the speaker, combined with the
maximum and/or the minimum F0. The combination of mean, maximum, and minimum F0 seemed to be important for the scales 'varied' and 'sonorous'. Mean and maximum F0 correlated positively with the scale 'melodious', whereas mean and minimum F0 appeared to be important for judgements on 'vivacious' and 'expressive'. Minimum and maximum F0 did not appear to be important for judgements on the scales 'active' and 'cheerful'.

As for the Dynamic features factor, all acoustic cues appeared to be relevant for the listener. Judgements for the scales 'brisk', 'self-confident', and 'quick' appeared to be related to speaking rate. Thus, higher judgements for 'self-confident' correlated with a higher speaking rate. Scales 'brisk' and 'powerful' were related to the duration of silent pauses, but in a negative way. The more a speaker was evaluated as 'powerful', the shorter he/she made his/her silent pauses. The data furthermore suggest that the more a speaker was evaluated as 'self-confident', the shorter he/she realized filled pauses. The scales 'brisk' and 'quick' also correlate highly with the ratio between speech and pause time.

The aim of the experiments described in chapter 3 was to find out which acoustic features speakers should have to be evaluated as a good speaker when retelling a story. In this respect we looked separately at male and female speakers. The reference point was the correlation with the ideal speaker, in terms of both the overall correlation and the correlation on the different scales and factors.

The female speakers were evaluated as 'good' collectively, but low correlations for temporal and pausal features lowered the overall score. The use of many filled pauses, rather than silent pauses, was not very appreciated, whereas high articulation rate and low pause ratio were.

Variation within the group of male speakers was much larger. A high number of filled pauses was reflected by a high negative correlation, and was in that respect literally far from ideal. Furthermore, a higher speaking rate and values close to the mean values of the male speakers appeared to have some positive effect on the evaluation. Listeners do not like characteristics indicating a narrow range (low sd of F0, low values for high end tones, and lower peak heights). In general, the male speakers' strategies were acceptable, without being appreciated very much.

6.2.2 Prosodic characteristics of discourse boundaries

In this part of the study, we wanted to investigate the production and perception of discourse boundaries in spontaneous speech. First of all, a listening experiment was run to see where listeners perceive boundaries in discourse. The results showed that variability in the judgements was high. Many of the judgements concerned cases where only a few of the 12 listeners agreed on the specific type of boundary (non-final, sentence final, or paragraph final). The
stronger the type of perceived boundary, the stronger the agreement, i.e. more
listeners agree on the type of boundary. Paragraph boundaries were not perceived
very often, and also not by all listeners. More boundaries were perceived than
was expected in the text-based discourse analysis, indicating that clause internal
boundaries were also perceived.

A comparison between the perceived type of boundary at locations where a
structural boundary (i.e. determined by the text-based discourse analysis) also
occurred, revealed that clause boundaries are, as expected, also associated with
non-final judgements. Sentence boundaries, however, are equally often
associated with non-final and with sentence final judgements. Paragraph
boundaries, on the other hand, show a tendency in most cases to be associated
with sentence finality.

The realization of boundary marking pitch movements in relation to the text-
based discourse structure showed that discourse boundaries are predominantly
realized with high boundary tones, even in places where a low boundary tone
was expected, such as at sentence and paragraph boundaries. When low
boundary tones are realized, they are associated with sentence and paragraph
boundaries. This indicates that in spontaneous discourse speakers tend to mark
boundaries as continuation (expressed by high boundary tones) rather than
finality (expressed by low boundary tones). In a monologue situation, as used in
our study, this continuation serves as a sign to the listener that the speaker has
not yet finished talking and thus makes sure that the speaker still has the
listener's attention. Half of all clause boundaries are realized with a boundary
marking pitch movement (51%), 91% of all sentence boundaries, and 80% of all
paragraph boundaries. These percentages vary for the individual speakers.
Discourse boundaries are also marked by means of pauses, in the majority of the
cases silent pauses. Forty percent of all clause boundaries, 66% of all sentence
boundaries, and 85% of all paragraph boundaries are marked with a pause.
Heavier and more important boundaries are thus marked with a pause more
often. Again, speaker differences were apparent. These data suggest that pausing
is a more reliable and important cue for boundary marking in spontaneous
speech than pitch movements. The combination of pauses and boundary marking
pitch movements was investigated next.

The heavier a boundary, the more often both pausing and pitch movements
are used. Clause boundaries are generally not marked by any prosodic cue, or
with only a pause or a boundary tone. Sentence boundaries are in most cases
marked with both a tone and a pause, or with only a tone. Paragraph boundaries,
finally, are always marked prosodically, in most cases with both a tone and a
pause, or with only a pause. These results confirm the hypothesis that heavier
boundaries are marked with more prosodic cues.

The final step in the investigation into discourse boundaries is relating the
judgements on perceived boundaries to specific prosodic cues as provided by the
speaker. Perceived boundaries were classified as weak, strong, or extra strong, dependent on the number and type of judgements given by the listeners. The results for boundary tones showed that high boundary tones are realized at boundaries perceived as weak, less at boundaries perceived as strong, and even less at boundaries perceived as extra strong. Low boundary tones are realized at boundaries perceived as strong and as extra strong. Thus, boundaries perceived as heavier are realized with low boundary tones, whereas boundaries perceived as shallower are realized with high boundary tones. However, relatively few perceived boundaries are realized with a boundary tone altogether, which indicates that boundary marking pitch movements are not the main cue for the perception of discourse boundaries.

Pausing appeared to be a more important cue for the listener. Boundaries perceived as weak were in 46% of the cases realized by the speaker with a pause. For boundaries perceived as strong, this is 89%, and for extra strong boundaries, it is even 100%. Again, we see a hierarchy in the weight of the boundary and the number of times it is marked with a pause. In total, 55% of all perceived boundaries is realized with a pause, which is a bit higher than the 47% for the boundary tones.

The use of both boundary tones and pausing in relation to perceived boundaries was looked at next. As could be expected, the heavier a boundary is perceived, the more often both prosodic cues were used by the speaker. Weak boundaries were in most cases not marked by any prosodic cue, whereas those perceived as strong were in the vast majority of the cases realized with both a pause and a boundary tone. Extra strong boundaries were always realized at least with a pause, and in the majority of the cases also with a boundary tone. Thus, in the perception of discourse boundaries in spontaneous speech, pausing appears to be the main cue for the listener, followed by boundary tones.

6.2.3 Prosodic characteristics of focal structure

In this part of our study on the prosodic aspects of information structure, the aim is to see how information status is marked prosodically by the speakers, using means such as pitch accentuation and pausing, and how this is related to perceived prominence by the listener.

The results from the experiment on perceived prominence show that there is an ordering in the relation between information structure and prominence. New information is always perceived as prominent. Inferrable information is perceptually prominent less often, but still in 96% of the cases. Half of all modifiers are perceived as prominent, and 28% of all discourse markers. Fifty-three percent of all verbs are perceptually prominent. Speakers may differ in this respect, but the overall ordering is more or less equal.
The same ordering is found for the relation between information structure and pitch accentuation. New information is accented more often than inferrable information, followed by verbs, modifiers, and, finally, discourse markers. Again, speaker differences were observed, but the ordering was roughly the same. Measurements of the actual realization of the pitch accents in fundamental frequency revealed that the type of information on which the pitch accent is realized has no effect. That is, pitch accents realized on new information are not realized differently from those realized on inferrable information or modifiers, for instance.

This ordering is also visible if one looks at the so-called ‘prominent pitch accents’, i.e. those words that are both accented and perceived as prominent. Overall, the vast majority of pitch accented words is also perceptually prominent. Of all new information 98% is prominent and pitch accented, for inferrable information this is 89%, it is 82% for both modifiers and verbs, and 67% for discourse markers.

The data presented so far clearly show that the various types of information structure were marked differently by the speakers, and were also picked up as such by the listeners. In this respect, the presence of a pitch accent appeared to be a major cue for the listener, as can be concluded from the high percentage ‘prominent pitch accent’.

Information structure could also be marked by pausing strategies. This was investigated by looking at clause-internal pauses (i.e. not at discourse boundaries) in relation to information structure. The majority of the pauses occurred between concepts (at word group boundaries) or after discourse markers; fewer pauses occurred within concepts.

The relation between the information status of the word group and the pause was investigated next. The occurrence of pauses between concepts did not appear to be dependent on the information status of the forthcoming concept. This may indicate that speakers just pause at these locations, for instance to give processing time to the listener, or as a ‘chunking device’ at word group level. Pauses within concepts, however, are associated with the information structure of the concept in which they are incorporated. Half of all pauses of this type occur in concepts that are new to the discourse (new and inferrable information). This also makes sense to the listener, providing an indication that he/she is listening to something important.

Compared to the total number of concepts in the discourse, we see that 20% of all new information is marked with a pause (either between or within), 18% of all inferrable information, 8% of all modifiers, 37% of all discourse markers, and only 6% of all verbs. Clearly, the ordering in this respect is different from what we have seen before. Here, discourse markers are marked with a pause, followed by new and inferrable information, and, finally, modifiers and verbs. We observed above that discourse markers are rarely accented or perceived as
prominent. The fact that discourse markers are realized so often (relative in the ordering) with a pause is not surprising, since these are the major turning points in the discourse, and may for that reason alone be marked with a pause. The other types of information structure are prosodically marked with intonational means, whereas the discourse markers are marked more often with temporal means.

6.3. Prosodic aspects of information structure in discourse

The main questions in the present research were formulated as follows:

- What acoustic-prosodic means are used by speakers to mark the structure of a spontaneously spoken discourse?
- Which acoustic-prosodic cues, as given by the speaker, do listeners use to detect the structure in spoken discourse?

The hypotheses concerning the acoustic realization and perceived structure of discourse were formulated in chapters 4 and 5. Table 6.1 gives a schematic overview, bringing the levels of discourse structure (phrasing) and information structure (prominence) together. In how far do our results meet these hypotheses? Table 6.2 presents an overview of the realization and perception of both discourse structure and information structure, according to the same setup as in Table 6.1.

On the level of discourse structure, the hypothesis that heavier discourse boundaries are realized with longer pauses is confirmed. An acoustic pause marking a sentence boundary is significantly longer than when realized at a clause boundary, and is significantly shorter than when realized at a paragraph boundary. As for the boundary tones, however, we predicted that clause boundaries would be marked with a high boundary tone, whereas boundaries at sentences and paragraphs would be marked with a low tone. This was confirmed for clause boundaries, but sentence and paragraph boundaries, contrary to our expectation, were realized with a high boundary tone as well in our corpus of spontaneous speech.
Table 6.1 Hypothesized realization of discourse structure and information structure by means of boundary tones, pitch accents, and pauses, and the hypothesized use of prosodic cues by the listeners in the perception of discourse boundaries and prominence.

<table>
<thead>
<tr>
<th>Structure</th>
<th>Realized</th>
<th>Perceived</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>DISCOURSE</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clause</td>
<td>Boundary tone</td>
<td>Pause</td>
</tr>
<tr>
<td></td>
<td>high</td>
<td>yes, shorter</td>
</tr>
<tr>
<td>Sentence</td>
<td>low</td>
<td>yes, longer</td>
</tr>
<tr>
<td>Paragraph</td>
<td>low</td>
<td>yes, still longer</td>
</tr>
<tr>
<td><strong>INFORMATION</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>New</td>
<td>Pitch accent</td>
<td>Pause</td>
</tr>
<tr>
<td></td>
<td>always</td>
<td>often</td>
</tr>
<tr>
<td>Inferrable</td>
<td>often</td>
<td>less often</td>
</tr>
<tr>
<td>Modifier</td>
<td>always</td>
<td>often</td>
</tr>
<tr>
<td>Disc. marker</td>
<td>always</td>
<td>always</td>
</tr>
<tr>
<td>Verb</td>
<td>never</td>
<td>never</td>
</tr>
</tbody>
</table>

Table 6.2 Actual realization of discourse structure and information structure, by means of boundary tones, pitch accents, and pauses, and of the prosodic cues used by the listeners in the perception of discourse boundaries and prominence.

<table>
<thead>
<tr>
<th>Structure</th>
<th>Realized</th>
<th>Perceived</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>DISCOURSE</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clause/Non-final</td>
<td>Boundary tone, pause, both, none</td>
<td>Boundary tone, pause, both, none</td>
</tr>
<tr>
<td>Sentence/Sent.final</td>
<td>high tone and (longer) pause</td>
<td>at least always pause, often tone</td>
</tr>
<tr>
<td>Paragraph/Par.final</td>
<td>high tone and (longest) pause</td>
<td></td>
</tr>
<tr>
<td><strong>INFORMATION</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>New</td>
<td>Pitch accent</td>
<td>Pause</td>
</tr>
<tr>
<td></td>
<td>always</td>
<td>few</td>
</tr>
<tr>
<td>Inferrable</td>
<td>very often</td>
<td>few</td>
</tr>
<tr>
<td>Modifier</td>
<td>often</td>
<td>very few</td>
</tr>
<tr>
<td>Discourse marker</td>
<td>few</td>
<td>often</td>
</tr>
<tr>
<td>Verb</td>
<td>often</td>
<td>very few</td>
</tr>
</tbody>
</table>

Our results are in accordance with what Blaauw (1995) has found for instruction monologues in Dutch. In her material, a little more than half of all paragraph-final phrases are marked with a low boundary tone, the rest being realized with a so-called non-low tone. However, Swerts & Geluykens (1994), using also instruction monologues, found that 81% of the paragraph final phrases were demarcated with a low boundary tone. Our results support the suggestion by Blaauw (1995) that “the connection between final paragraph position and low
boundary tunes in spontaneous speech is not as strong as has been suggested by Brown et al. (1980), Yule (1980) and Swerts & Geluykens (1994)” (p. 151). Furthermore, our results confirm findings by Blaauw (1995) that minor boundaries (i.e. our clause boundaries) are not prosodically marked in about half of the cases, whereas major boundaries (i.e. our sentence and paragraph boundaries) are realized in the vast majority of the cases in spontaneous speech. The actual realization of the prosodically marked boundaries found in our data differs from the results given by Blaauw. She finds that minor boundaries are predominantly marked with a pitch movement or a pitch movement and a pause, whereas our clause boundaries are mainly realized with a pitch movement or a pause, and seldom with both cues. Major boundaries are more comparable across both studies, i.e. marked with both a pause and a pitch movement.

The prosodic cues provided by the speaker are used by the listener to detect the structure of the spoken message. In this respect, the speakers use predominantly pausing rather than boundary tones. Since the speakers used high tones to mark discourse boundaries, our prediction that listeners make use of low tones cannot be confirmed. Speakers use high tones to mark discourse boundaries. A listener may still expect a low boundary tone to mark the end of a sentence or paragraph, but in the absence of such a cue in spontaneous speech, he/she relies on temporal cues instead.

The level of information structure shows that our hypotheses concerning new information are confirmed. This type of information is always marked with a pitch accent, and is also always perceived as being prominent. Inferrable information, however, is pitch accented and prominent more often than was expected, and behaves even similar to new information. This means that inferrable information should be treated as expressing additional information rather than the repetition of information (i.e. given) that is already available in the discourse, be it implicitly. The fact that inferrable information is recoverable from the discourse context or the knowledge of the world, does not mean that it is equal to given or evoked information, can thus be deaccented and is therefore not perceived as prominent. Both new and inferrable information are occasionally marked with a pause. Altogether, pauses are rarely used to mark information structure, but when this is the case, it concerns the most important information in the discourse (but see below for the discourse markers). Modifiers are marked with a pitch accent and are perceived as being prominent less often than was expected. This can be explained as follows: Modifiers occur mostly in combination with new and inferrable information, as adverbs of place or time. In this sense they are additional information, albeit newly introduced in the discourse, and were therefore expected to be pitch accented and prominent. But, since all new and almost all inferrable information is already pitch accented and prominent, these modifiers can very well do without pitch accent and
prominence. The speaker knows that the listener is paying attention to this type of information, and can therefore afford not to accent the modifiers. The category *discourse markers* is also much less pitch accented and prominent than was expected. However, this type of information is more often marked with a pause instead. The clear linguistic function of discourse markers may explain this. By pausing right after a discourse marker, the speaker signals his/her intention to continue the story, but he/she is apparently not quite sure about how to continue it. This signal is picked up by the listener, which explains the fact that discourse markers are still more often prominent than accented. *Verbs,* finally, were expected not to be pitch accented nor perceived as being prominent, following the ideas of Prince (1981) and also Chafe (1987) in which verbs form a so-called 'unitary concept' with the NP. Our data show that they quite often carry pitch accent, and are also often perceived as being prominent. They are in only a few cases accompanied by a pause. This shows that it is worthwhile to include verbs as a separate category in a model of information structure in spontaneous speech.

The results of the present study, in answer to our main research questions, can now be summarized as follow:

Speakers mark information structure in discourse in terms of phrasing or ‘chunking’ (discourse boundaries) and in terms of information status (accentuation). Structural discourse boundaries, i.e. those determined by textual information, are prosodically marked, but the specific means used to do this are dependent on the type of boundary: the heavier the boundary, the more a boundary is realized prosodically, and the more prosodic cues are used. These cues are, in order of importance, silent pauses and high boundary tones. Information expressing the lexical content of the discourse, such as new or inferrable information or modifiers, is marked predominantly by pitch accent. There is an ordering in the ‘accentability’ of information status in discourse: new > inferrable > verbs > modifiers > discourse markers. Information indicating the major transition points (discourse markers) in the discourse are mainly realized by means of a pause realized after this specific discourse marker. In terms of the scheme in Figure 6.1, these are the *acoustic means* used by the speaker to realize *information structure* and *discourse structure.* It was outside the scope of the present study to investigate also the speaker’s use of *linguistic means,* such as sentence form, sentence structure, word order, and/or the choice of specific lexical items, to structure discourse.

Listeners make use of the prosodic information provided by the speakers to detect the structure of spoken discourse. For the perception of discourse boundaries, pausing is more important than pitch movements, especially if it concerns heavier boundaries. Discourse boundaries are mainly perceived as being ‘non-final’. The perception of prominence (i.e. the important parts of the
discourse) is mainly triggered by pitch accents, and is also dependent on the type of information. The same ordering applies as for accentability: new > inferrable > verbs > modifiers > discourse markers. This is the panel on the right in the scheme in Figure 6.1, from the ‘spoken discourse’ to the LISTENER. The acoustic cues are provided by the speaker and used by the listener to decode the discourse structure and the information structure. Listeners may, of course, also have used linguistic cues. The investigation of these cues in relation to perceived discourse structure lies outside the scope of the present study.

Several strategies are possible in the use of prosodic means. As has become clear from our results on speaker characteristics, none of the speakers was really judged as ‘insufficient’ by the listeners in their prosodic realization of the retelling task, but the female speakers were preferred to the male speakers in general. However, the detailed analysis of the use of prosodic cues to mark both discourse boundaries and information status revealed no specific sex-related differences in strategy. The use of particular acoustic features, such as the frequent use of filled pauses, is related to the overall judgement and the similarity to the ‘ideal speaker’, but this is more an issue of individual speakers, irrespective of their sex.

6.4. Limitations and suggestions

The study of how prosody is used by speakers to structure their messages involves numerous aspects. In the present study, we have focused specifically on temporal and intonational aspects, but this does not mean that other aspects, such as spectral features and segmental information, are not important. Analyses of segment durations or of variations in intensity require a fully labeled database. We wanted to use a specific type of spontaneous speech material, one that comes close to an everyday communicative situation. The full labeling and segmentation of such a database would have been too time consuming, and was therefore not done. This implies that acoustic analyses on the segment level are not included in our study.

The relation between discourse structure and prosodic characteristics is very complex. The structure of the discourses was divided into a so-called discourse structure, being the division into smaller parts such as clauses, sentences, and paragraphs, and a so-called information structure, the classification of the various words and word groups according to their information status. One limitation of the present study lies in the fact that these two levels were studied and discussed separately. If both levels are integrated, information about the number of words expressing ‘new’ or ‘inferrable’ information per clause, sentence, or paragraph becomes available, as well as more detailed information about word order. As Terken & Hirschberg (1994) have suggested, information
about grammatical role, surface position, and form of referring expression should also be included in a discourse model. Labels to identify this type of information, thereby also integrating both levels of analysis, can, of course, easily be added to our model. Furthermore, a more elaborate analysis of hesitation and correction phenomena in relation to discourse structure would then become possible.

The database we recorded (see chapter 3) consisted of three spoken versions of the same discourse for each speaker. In the present research, we only used the so-called ‘retold’ version. The other two versions are the ‘original’ read version and the ‘re-read’ version. This ‘re-read’ version is a read aloud recording of the transcribed retold version, these two versions being in principle lexically identical. A natural follow-up would be a prosodic comparison of the ‘retold’ and ‘re-read’ versions. This would give insight into possible differences in the way speakers mark information structure in different speaking styles, and in how this is processed by the listener. A pilot experiment was set up in a first attempt to test this, but the results were not available in time to be included in this thesis.

In chapter 3 we tested how the speakers were evaluated in their prosodic performance in retelling a story. From a more psycholinguistic viewpoint, this could be completed by having listeners also evaluate the way the content of the original story is rendered. This would give valuable information about how discourse is processed in the mind of the speaker, and how the various types of information in the original discourse are transferred.

As has been pointed out several times, the use of linguistic means by the speaker to structure discourse was not investigated. It would be interesting to combine linguistic and prosodic aspects of discourse structure, to see what the systematic properties (pragmatic, semantic, and syntactic) are of the acoustic chunks produced by speakers in spontaneous discourse or dialogue. Which prosodic cues are used to mark these properties? And, how important are these means to the listener as indicators of efficient communication? Furthermore, what is the relation between the linguistic and the prosodic properties? Focusing on these aspects would not only give additional insight into the grammatical structure of spoken discourse and dialogue, but would also be relevant for language and speech technological applications such as dialogue systems.

Knowledge about the way human speakers structure their message can improve synthetic speech, especially where spontaneous output is required. Reading machines for the blind could benefit from good, naturally sounding synthetic speech output. Depending on the content and the type of material to be synthesized, a ‘spontaneous’ sounding voice can be more appropriate. Furthermore, more insight into the prosodic aspects of spontaneous speech may also help hearing-impaired people. Training programs could be developed to teach these people to produce their speech in a naturally sounding way, i.e. comparable to spontaneous speech as used by normally hearing people in everyday communicative situations.
The role of prosody is also of interest to the field of automatic speech recognition, for instance, in the disambiguation of different readings of a same sentence. The few prosody-driven recognizers that presently exist use automatic detection of pitch accents and phrase boundaries as indicators of discourse or dialogue structure. Knowledge about how humans build and perceive these structures, and use this structure to interpret the message, can help to improve the system’s performance.

Text-to-speech systems automatically place accent lending and boundary marking pitch movements in the spoken version of a text given to the system. Such systems require fairly structured texts as input, for instance news bulletins, weather reports or even books to be read aloud. A text in the form of a transcription of the retold stories as used in the present study, is not suitable as input for a text-to-speech system as described here. In the present research, the predicted location of pitch accents was based on the textual analysis. In a text-to-speech system, this is in principle also the case, but the knowledge is based on more strictly structured texts. It would therefore be interesting to compare the actual location of pitch accents, as determined by the intonation experts (chapters 4 and 5), with the location of accents determined by a text-to-speech system. We expect that where the two approaches differ, this will be at the level where spontaneous and read speech differ most, i.e. the use of modifiers, hesitations, corrections, false starts and the like.

In the present study, our speech material consisted of only monologues, i.e. one speaker without interference from other speakers. A natural follow-up would be to also look at dialogues between two or even more speakers, using, for instance, the Map Task (Anderson et al., 1991), or at human/machine dialogues. Preparations for such a follow up project are currently under way at our institute.
the same discourse for each speaker. In the present research, these differences were
identified both through analysis of the actual spoken language and through
observations of the way human speakers structure their messages. The
improvement of synthetic speech, especially with spontaneous speech, is
required. However, machines for the blind also benefit from good, naturally
sounding synthetic speech output. Depending on the context and the type of material to
be synthesized, a spontaneous-sounding voice can be more appropriate.
Furthermore, more insight into the phonic aspects of spontaneous speech may
even help hearing-impaired people. Training programs could be developed to
teach these people to produce their speech in a nearly-sounding way, i.e.
comparable to spontaneous speech as used by normally-hearing people, in
everyday communicative situations.