Summary

When producing spoken discourse, *speakers* may use various prosodic means to convey the structure of the message they are uttering. Certain words or word groups will be marked as more important than others. The whole message will furthermore be ‘chunked’ into smaller pieces, such as paragraphs and sentences. *Listeners* of such spoken discourse have certain ideas about the structure of the text they are hearing. Certain words or word groups will be perceived as more important than others. They are also able to detect different types of boundaries, such as sentence boundaries and paragraph boundaries. The text uttered by the speaker, the *message* itself, also has a structure. This text is assumed to be more or less coherent, otherwise the listener will have trouble understanding it, and can be divided into paragraphs, sentences, clauses, phrases, etc. Apart from the acoustic means, the speaker also has a variety of linguistic means available to indicate the structure of the message. This structure of spoken or written texts in terms of important information and boundaries can be referred to as the *information structure* of a text.

In this thesis, we focus on two main research questions. First of all, from the speaker’s side, what acoustic means do speakers use to signal information structure in a spoken discourse? Secondly, from the listener’s side, how are the acoustic cues, as provided by the speaker, used by the listener to detect the structure of the spoken message? In order to answer these questions, spontaneously spoken discourses were collected from speakers of standard Dutch. The discourse text was obtained by making verbatim transcriptions of the discourses. This text was then analyzed for discourse structure using a purely text-based method called ‘Information Structure In Discourse’ (ISID), that we developed for this purpose. The material was then viewed from the three components of the communication chain: the message, the speaker, and the listener.

The method for discourse analysis, ISID, is described in chapter 2. On the basis of the different theories and approaches found in the literature, we developed a method suitable to analyze our spontaneous speech material. It was specifically designed to analyze spontaneous data. The results of a pilot experiment showed that indeed this method is useful to analyze the discourse structure of spontaneously spoken texts. Furthermore, it can be used to explain listeners’ judgements on perceived prominence.
The model has a ‘global’ level and a ‘local’ level of structure. The starting point of the analysis is a verbatim transcribed spoken text, without any punctuation and/or typographical lay-out. The first step is to identify the hierarchical level of global discourse structure (from the largest to the smaller functional unit): discourse → paragraphs → sentences → clauses → word groups. The discourse thus consists of paragraphs, which consist of sentences, which consist of clauses, which consist of word groups. The next step is to determine the local level of information structure. For each word group in the clause, the information status is labeled accordingly. The main category labels are: new, inferrable, evoked, discourse marker, modifier, and verb.

Chapter 3 gives an overview of the speakers and the material used in this thesis. The material is characterized in terms of various overall acoustic-prosodic measures, to see what means are used by the speakers to realize the essential information in spoken discourse.

Eight speakers of standard Dutch, four male and four female, were chosen. They were asked to read aloud a short story in Dutch at a normal speaking rate (‘Een triomf’, by Simon Carmiggelt, 1966; see appendix A and B). After a short break they were asked to retell the story they just read in their own words, with as many details as possible. The speech material was then described in temporal, pausal, and intonational features. The temporal aspects include clause, sentence, and paragraph durations, and speech tempo. Pausal features include type and duration of pauses, whereas the intonational aspects focus on global measures as minimum, maximum, and mean F0, range, peak heights, and high and low end tone.

A perception test was then carried out, in which listeners were asked to evaluate various prosodic aspects in the retold stories, using semantic scales. The aim was to see which features at the prosodic level listeners prefer when listening to a retold story in Dutch, and if ‘good’ and ‘bad’ speakers can be distinguished in this respect. The listeners were instructed to judge how well the speaker had performed the retelling task regarding the prosodic-acoustic realization. The results show that aspects related to the factor Intonational characteristics are most important, followed by the factors Dynamic features and Articulation quality. They furthermore show that there are clearly ‘good’ and ‘bad’ speakers when it comes to retelling a story. Still, all speakers were evaluated as sufficient. In general, the female speakers were preferred to the male speakers.

The next step was to see how the judgements from the perceptual evaluation test can be explained by looking at the acoustic aspects. The acoustic cue related to the judgements on Intonational characteristics appeared to be ‘mean F0’, in combination with ‘minimum F0’ and/or ‘maximum F0’. Listeners apparently based their judgements for the various scales on the ‘average F0’ of the speaker,
combined with either the ‘maximum F0’, the ‘minimum F0’, or both. For the Dynamic characteristics, all acoustic cues were relevant for the listener. A higher speaking rate, for example, caused speakers to be evaluated as more ‘self-confident’. Or, the shorter a speaker realized his/her pauses, the more he/she was judged to be ‘powerful’.

In conclusion, chapter 3 shows us that the female speakers were evaluated as ‘good’ altogether. However, low correlations with the ideal speaker for temporal and pausal features lower the overall judgement score. The male speakers vary much more. In general, male speakers were acceptable in retelling a story, without being very appreciated.

In chapter 4 we looked at how discourse boundaries (the global level in our ISID model) are realized acoustically by the speakers, and which of these acoustic cues are most important for the listener in the perception of discourse boundaries. A listening experiment was carried out to investigate where listeners perceive boundaries of different ‘depths’ in the discourse (non-final, sentence final, and paragraph final).

The data show that the relation between perceived and structural discourse boundaries is as follows: clause boundaries are, as was expected, associated with ‘non-final’ judgements. Sentence boundaries, however, are equally often associated with ‘non-final’ and with ‘sentence final’ judgements. Paragraph boundaries are mostly associated with ‘sentence final’ judgements, and not with the expected ‘paragraph final’ judgements. This suggests that in spoken discourse, the notion of ‘paragraph’ is not relevant for listeners. This seems to be a characteristic of written text that does not have perceptual relevance in spoken text, at least in the retold versions used in the present study.

The speakers realized discourse boundaries intonationally mainly with high boundary tones, also at places where low boundary tones were expected. Thus, in spontaneous discourse, speakers tend to mark continuity at boundaries rather than finality. Discourse boundaries are also marked by pausing. Heavier and more important boundaries are marked more often with a pause, pauses are also significantly longer in duration the stronger the boundary becomes. In this sense, pausing is a more reliable cue for boundary marking than the use of boundary tones. The combination of boundary tones and pausing, finally, shows that the heavier the boundary, the more often both cues are used by the speaker. This confirms the hypothesis that heavier boundaries are marked by more prosodic cues than shallower ones.

Perceived boundaries were classified as either ‘weak’, ‘strong’, or ‘extra strong’, depending on the number and type of judgement given by the listeners. Looking then at the acoustic characteristics of these perceived boundaries, we saw that boundaries perceived as strong were realized with low boundary tones, whereas boundaries perceived as being less strong were realized with high
boundary tones. However, pausing appeared to be more important for boundary perception than intonation. The stronger the perceived boundary, the more often it was realized with a pause, up to 100% for ‘extra strong’ boundaries. The combination of pausing and boundary tones, again, showed that the stronger a perceived boundary, the more often both prosodic cues were used by the speaker to realize this boundary.

Chapter 5 deals with the prosodic characteristics of information structure in discourse. The aim was to see how information status (the ‘local’ level in our ISID model) is prosodically realized by the speakers, using characteristics such as pitch accentuation and pausing. This acoustic realization of information structure was related to perceptual judgements by listeners, in this case perceived prominence.

The results for perceived prominence showed that there is a clear ordering in the relation between information status and perceived prominence, going from more to less prominent: new > inferrable > modifiers & verbs > discourse markers. The same ordering is found for the relation between information status and pitch accentuation. This same ordering is also visible if we look at the so-called prominent pitch accents, i.e. those words realized with a pitch accent and also perceived as being prominent. Pitch accent is the main cue for prominence perception.

However, pausing also appeared to have some relation with information status. Pauses realized within concepts occurred in concepts expressing new information in the discourse, i.e. new and inferrable information, and not in concepts expressing information that was already available. Pauses realized between concepts, i.e. on word group boundaries, did not have such a relation with information structure.

The ordering in the relation between information structure and pausing (either between or within concepts) is different from the one we saw before: discourse markers > new > inferrable > modifiers > verbs. The category discourse markers has moved to the first place in this ordering. This can be explained by the fact that discourse markers are located at the major turning points in the discourse. By pausing at these locations, the speaker indicates that he/she is planning the continuation of his/her story. To conclude, information structure in spoken discourse at the ‘local level’ is predominantly marked with intonational means, except for discourse markers, which are marked more with temporal means.

The conclusion of this thesis (Chapter 6) is that in spontaneous speech, discourse boundaries are realized with, in order of importance, silent pauses and high boundary tones. The stronger the boundary, the more often both cues are used by the speaker. Information status is realized by means of pitch
accentuation, except for the category 'discourse marker', which is realized by means of pausing. Listeners make use of prosodic cues to detect information structure in discourse: the main cue for boundary perception is pausing, for information status it is the presence of a pitch accent.

In this thesis we have focused on temporal and intonational features in relation to discourse structure, from the point of view of the speaker, the listener, and the text of the discourse itself. Other prosodic aspects, such as spectral features and segmental information are important as well, but have not been included here.

Apart from the prosodic means, speakers did, of course, also use linguistic means to signal the structure of their message, such as word order and grammatical structures. An interesting follow-up of the present research would be to combine linguistic and prosodic aspects of discourse structure, and to investigate the systematic properties (pragmatic, semantic, and syntactic) of acoustic chunks uttered by the speakers of a spontaneous discourse or dialogue. How important are these properties to the listener as indicators of efficient communication? The study of dialogues between two or more speakers, or human/machine dialogues, in relation to the linguistic and prosodic aspects of these dialogues would not only give more insight into the grammatical structure of spoken discourse and dialogue, but would furthermore be of importance to language and speech technological applications such as dialogue systems.
However, pausing also appeared to have some relation with information status. Pauses realized within concepts occurred in concepts expressing new information in the discourse, i.e., new and inferrable information, and not concepts expressing information that was already available. Pauses realized between concepts, i.e., on word group boundaries, did not have such a relation with information structure.

The ordering in the relation between information structure and pauses (either between or within concepts) is different from the one we saw earlier: discourse markers have moved to the first place in this ordering. This can be explained by the fact that discourse markers are located at the major move points in the discourse. By pausing at these locations, the speaker indicates what he/she is planning the continuation of a 'real' story. To conclude, information structure in spoken discourse is at the 'local level': it predominantly marked in intentional means, except for discourse markers, which are marked more intertemporal means.

The conclusion of this thesis (Chapter 6) is that in spontaneous spoken discourse boundaries are realized with, in order of importance, silent pauses and high boundary tones. The stronger the boundary, the more often both cues are used by the speaker. Information status is realized by means of plain