Prosodic marking of semantic contrasts

Kaland, C.C.L.

Publication date:
2014

Document Version
Publisher's PDF, also known as Version of record

Prosodic marking of semantic contrasts

Do speakers adapt to addressees?

Published by
LOT
Trans 10
3512 JK Utrecht
The Netherlands
phone: +31 30 253 6111
e-mail: lot@uu.nl
http://www.lotschool.nl

Cover photo:
Taken by the author near Loch Etchachan, Cairngorms, Scotland (winter 2013).

ISBN: 978-94-6093-148-2
NUR 616

Prosodic marking of semantic contrasts

Do speakers adapt to addressees?

PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan Tilburg University,
op gezag van de rector magnificus,
prof. dr. Ph. Eijlander,
in het openbaar te verdedigen ten overstaan van een
door het college voor promoties aangewezen commissie
in de aula van de Universiteit
op woensdag 1 oktober 2014 om 10.15 uur

door

CONSTANTIJN CORNELIS LOURUS KALAND
geboren op 24 mei 1985 te Vlissingen

Promotores:	prof. dr. M.G.J. Swerts
	prof. dr. E.J. Krahmer
Promotiecommissie:	dr. Y. Chen
	dr. M.B. Goudbeek
	prof. dr. C.H.M. Gussenhoven
	prof. dr. B. Möbius
	prof. dr. S. Zerbian

voor mijn grootouders

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

1.1 Introduction

Speakers in a conversation generally adapt their utterances to their addressees. For example, when telling an anecdote about a friend, a speaker will probably take into account whether the addressee knows this friend or not. The speaker may choose to use a description such as “my friend” or “a friend” instead of a proper name (“John”) when telling the anecdote to an addressee who might not know John (Heller et al., 2012). Thus, speakers adapt their choice of words by making assumptions about the addressee’s knowledge, a process that has been called *audience design* (Clark & Murphy, 1982). In a somewhat similar vein, speakers may prefer to use words that have also been used by their addressee in previous speaking turns. For example, if the current addressee previously referred to a man of considerable height as “the tall guy”, the speaker is likely to use the same description to refer to this man and not a variant such as “the large guy” or a description including a different property, such as “the blond guy”. Adaptation by means of the choice of words would thus be evident from the fact that speakers *copy* crucial words from the ones used by their interlocutors.

Earlier work studying adaptation in language has often focused on the words and syntactic constructions that speakers use. However, less is known about the extent to which speakers adapt their prosody to their addressees as well. As we discuss in more detail below, some prior studies indicated that prosodic features may reveal audience design, and show that speakers sometimes copy prosodic features of their interlocutors. However, the process of prosodic adaptation is still far from completely understood, in part because previous studies on adaptation tended to ignore the various communicative functions of prosody. The current thesis is set up to provide more insight into the process of prosodic adaptation, focusing in particular on adaptive behavior regarding the use of pitch accents as markers of contrastive information.

Prosody refers to suprasegmental features of speech such as pitch, duration and intensity (Rietveld & Van Heuven, 2009). Various studies show that speakers adapt their prosody to their addressees. For example, speakers generally adapt their prosody when talking to infants or pets by using a higher pitch or a wider pitch range (Garnica, 1977; Fernald & Simon, 1984; Burnham et al., 2002) or when talking in a noisy environment by speaking with increased intensity (Lombard, 1911). Such cases can be seen as a form of audience design, given that speakers adapt to their addressees in order to successfully exchange information. It has also been shown that speakers in a dialogue adapt their speech to each other by taking over prosodic features (Pardo, 2006). Pardo (2006) showed that speakers in a dialogue start to sound more similar to each other in the course of an interaction because they tend to converge on a comparable pitch range, speech intensity and speech rate. Notice, though, that in these kinds of studies no attention is paid to the communicative functions of prosody; only global features of prosody over the course of an interaction are taken into account. As a result, it remains unclear how such global adaptations relate to specific communicative functions of prosodic features. We know that speakers may use prosody for a broad range of different functions, such as to mark whether an utterance is a declarative or a question (e.g., Haan, 2002), to regulate the turn-taking system (e.g., Caspers, 2003), to provide feedback (e.g., Granström et al., 2002), to express specific attitudes or emotions (e.g., Scherer, 1986) or to indicate whether an utterance should be interpreted in a literal or ironic sense (e.g., Bryant & Fox Tree, 2005).

It would seem logical to consider such functions of prosodic features as well when studying adaptive behavior. This can be illustrated with an example: Imagine a situation in which speaker A asks an addressee B a question (e.g., “should I turn left to get to the station?”), with the use of a high rise at the end of the utterance. If B then provides an answer to that question (e.g., “yes, turn left!”), B is very likely to provide the utterance with a declarative intonation pattern, including the use of a final fall, especially if B is certain about the correctness of that response. In such a communicative context, it would be odd if speaker B would adapt to speaker A by copying A’s high final rise, as that might create the incorrect suggestion that B is insecure about the response. By contrast, one could also think of a context in which two speakers indeed copy each other’s intonation patterns for communicative functional reasons, for example when two speakers compose a list of items to take with them on a trip. In such a situation A could refer to one item producing a final rise (e.g., “a toothbrush”) and B is likely to do the same when referring to another item (e.g., “a comb”), especially when there are more items on the list that have to be referred to. In other words, copying prosodic features of an addressee could be infelicitous in certain communicative contexts because it would interfere with specific functions of those features, whereas in other communicative contexts copying prosodic features would be perfectly acceptable. The question that arises is to what extent the realization of a speaker’s prosody is governed by its communicative function and to what extent by its similarity to the prosody of the interlocutor. This is a central question addressed in the current thesis. While the example above dealt with the way speakers use boundary tones, the studies on prosodic adaptation presented in this thesis zoom in on another prosodic function, namely the use of pitch accents as markers of important information (i.e., Pierrehumbert & Hirschberg, 1990). In particular, in a series of experiments we study to what extent the speaker’s assumptions about the addressee influence the prosodic marking of information structure in general and contrastive intonation in particular, and to what extent the speaker’s use of pitch accents can be explained on the basis of copying behavior.

The remainder of this introduction first presents a short, general overview of speaker adaptation and prosody, followed by a discussion of previous work on contrastive intonation. Then, the research questions addressed in the current thesis are discussed, together with a brief preview of the studies presented in the remainder of this thesis.

1.2 Prosody and adaptation

So far, most studies on prosodic adaptation have mainly focused on global characteristics of prosody, for instance showing that interlocutors start sounding more similar as the conversation progresses (e.g., Pardo, 2006). This form of adaptation has been linked to behavioral mimicry (Bargh & Chartrand, 1999; Chartrand & Bargh, 1999), and has been referred to as a form of *adaptation* (Gregory & Hoyt, 1982), *alignment* (Pickering & Garrod, 2004) or *convergence* (Pardo, 2006). In particular, speakers have been shown to gradually converge on a similar pitch range (Gregory, 1986; Couper-Kuhlen, 1996). That is, the fundamental frequency (F0) of the speech of different interlocutors tends to become similar within certain frequency ranges once they engage in a dialogue. Furthermore, when one speaker talks with a soft voice, the interlocutor is likely to adjust the intensity of his or her speech to a lower level as well (Natale, 1975). Similarly, it has been shown that interlocutors adjust their speech rate to each other when engaging in a dialogue (Giles et al., 1991; Szczepek-Reed, 2010). In all of these studies prosodic adaptation was observed, but typically this was reflective of general aspects of the interaction and did not relate to communicative functions of prosody. Features such as global pitch range, intensity and speech rate merely relate to the form of an utterance (i.e., they are paralinguistic) and have been argued to be of social rather than linguistic relevance (Shepard et al., 2001; Pardo, 2006; Ladd, 2008).

In contrast, it is well known that speakers may use prosody for communicative (and hence linguistic rather than paralinguistic) purposes as well. An example of linguistic prosody (in Germanic languages) is the prosodic marking of information structure, in that pitch accents are used to mark which words are important, because they express new or contrastive information in a certain utterance (Pierrehumbert & Hirschberg, 1990). Pitch accented words are generally produced with a higher F0, longer duration and more intensity and are perceived as more prominent compared to words that are not accented (Ladd, 2008). Consider the English utterances in (1a) to (1c) as if they were produced in a sequence. In (1a) the word "car" is likely to be produced with a pitch accent, because the default location for pitch accents in English is the rightmost word in a phrase (Ladd, 2008). In (1b), however, "yellow" is likely to be accented and not "car", since the speaker may want to indicate that the car's color is different compared to the car described in the previous utterance. Note that in (1c) the pitch accent would again be on the noun "van" since in the context of (1b) this word refers to contrastive information.

(1) (a) Today I saw a blue car.
(b) Yesterday I saw a yellow car.
(c) Last week I saw a yellow van.

While prosodic adaptation is argued to be a universal phenomenon in human communication (Pickering & Garrod, 2004; Pardo, 2006), the linguistic use of prosody is known to differ across languages. For example, in its context the pattern in (1b) is a common intonation pattern for speakers of Germanic languages such as English, Dutch or German. However, not all languages use pitch accents to indicate semantic contrasts, such as the one between “yellow” and “blue” in the example above. Romance languages such as Spanish, Italian or Romanian, for example, have more fixed intonation patterns and may use word order to focus on contrastive information. For instance, in the Italian versions of the examples in (1), given in (2) in their respective order, speakers would produce a pitch accent on the rightmost word in the phrase in all cases, irrespective of whether that word refers to contrastive information or not.

(2) (a) Oggi ho visto una macchina blu.
(b) Ieri ho visto una macchina gialla.
(c) La settimana scorsa ho visto un forgune giallo.

We can distinguish these two groups of languages according to their ‘plasticity’ (Vallduví, 1991); so-called plastic languages (such as Germanic languages) allow a speaker to shift the default intonation pattern (as we have seen above) to indicate the information status of words. Non-plastic languages (such as Romance languages) do not allow shifting of pitch accents within noun phrases (NPs) as a function of discourse context. It is known that a non-plastic language such as Italian may express contrastive information in prosody, but it does so by varying the type of pitch accent and not by varying the position of the pitch accent (Bocci & Avesani, 2006). In this thesis we investigate to what extent language differences in linguistic prosody cause differences in the possibilities for prosodic adaptation. Therefore, we make a distinction between typical prosody which is in accordance with the rules of prosodic marking of information status in the language of interest and atypical prosody, which is not in accordance with those rules. As adaptation cannot be observed on the basis of typical prosody alone, speakers in our experiment interacted with a confederate who uses atypical prosody. We measured adaptation in terms of the extent to which participants deviated from typical prosody (in Dutch or Italian) in situations where they interacted with a confederate who used atypical prosody.

Furthermore, speakers may differ in their use of linguistic prosody depending on their mental development. In particular, Peppé et al. (2003) have shown that intonation patterns marking semantic contrasts are more likely to be produced erroneously by speakers with an autism spectrum disorder, such as high functioning autism (HFA). It has been argued that speakers with HFA may experience difficulties when making assumptions about how important or contrastive certain information is for their addressee, which (according to some researchers) is caused by an impaired theory of mind (Baron-Cohen, 1995; Shriberg et al., 2001). However, it remains to be investigated to what extent HFA speakers take into account the (different) perspective of their addressee when producing contrastive intonation. This is done in the current thesis by comparing adaptation of contrastive intonation of typically-developing speakers with HFA speakers.

1.3 Contrastive intonation

The use of pitch accents in general and contrastive intonation specifically has been topic of discussion in many studies investigating different languages. However, little or no research has dealt with the question to what extent speakers adapt their use of pitch accents to their conversation partner (either as a form of audience design, or because a speaker is copying prosodic features of a partner). Instead, most studies on pitch accents have focused on phonetic and phonological features of such accents, and on the question to what extent the position of such accents can be predicted on the basis of information structural factors. In this section, we give a brief summary of the major insights that can be drawn from these studies.

In a phonological theory about the intonation of English, the claim is that specific pitch accent types signal a specific information status of words in a discourse, in that pitch accents that signal new information are different from pitch accents that signal contrastive information (Pierrehumbert & Hirschberg, 1990).

Reconsider the sequence of example sentences in (1a) and (1b). The word “car” is new information in (1a) and given information in (1b). A generalizing claim for Germanic languages such as (American) English or Dutch is that new information is accented and given information deaccented (but see Terken & Hirschberg, 1994). However, there has been much discussion about the extent to which different categories of information status are expressed in prosodically different ways, as suggested by Pierrehumbert and Hirschberg (1990).

To illustrate one of the issues in the discussion on the prosody-information status interface in contrastive contexts, consider the pitch accented word “van” in (1c). That pitch accent could either signal a contrast with (1b) when both were uttered in sequence, or, when there is no such contrast (i.e., without any preceding discourse context), signal a default statement of new information, like the accent on “car” in (1a). According to Ladd (2008), the question is whether the ‘contrastive’ reading and the ‘new’ reading correspond to different underlying phonological categories.

Ladd (2008) distinguishes between three views on the problem of contrastive intonation, all using a specific terminology (in italics). These views have dealt with the extent to which the occurrence of pitch accents in contrastive contexts is driven by speakers’ intentions and/or by phonological rules. Following the ‘normal stress’ view (e.g., Newman, 1946) phonological rules determine which words are acoustically most prominent, without assigning any intrinsic meaning to this prominence. The prosodic marking of semantic contrasts (*contrastive stress*) is not seen as linguistic under this view, as it is unpredictable from the sentence structure. Contrastive stress is therefore different from normal stress, as normal stress is assigned by predictable phonological rules. According to the ‘highlighting view’ (e.g., Bolinger, 1972) every word in an utterance, whether referring to new or contrastive information, can be produced with a pitch accent to signal its information status (i.e., *contrastive focus*). Crucially, this depends entirely on what the speaker has to say and not on any linguistic rules. The highlighting view therefore does not see an essential difference between the prosodic marking of contrastive and new information. Ladd (2008) sees the Focus-to-Accent (FTA) approach (e.g., Gussenhoven, 1983) as a third and bridging view between the 'normal stress' and the 'highlighting' approach. The FTA view distinguishes between the unpredictable notion of focus, which is determined by the speaker, and the predictable notion of accent, which follows from phonological rules. Crucially, focus may concern individual words as well as larger parts of a phrase. Following the FTA view, the pitch accent on "van" in (1c) signals *narrow focus* when it marks the contrast with (1b) and broad focus when it is used as a non-contrastive statement. In short, the three views briefly discussed illustrate the importance of distinguishing phonological rules from speaker intentions when it comes to the prosodic marking of contrastive information.

To what extent contrastive information can actually be seen as a different category of information status when compared to new information has been investigated both from phonetic and phonological as well as syntactic and semantic perspectives. Some have argued that contrastive and new information indeed differ, in that they are expressed by different types of pitch accents (Pierrehumbert & Hirschberg, 1990; Selkirk, 2008). The pitch accent that marks new information in American English has been described as H* (Pierrehumbert & Hirschberg, 1990), a high pitch that is aligned with the lexically stressed syllable of the accented word. The contrastive pitch accent of American English has been described as L+H* (Pierrehumbert & Hirschberg, 1990), a contour in which the contrastive word is marked with a high pitch following immediately from a lower pitch on the preceding part of the phrase. It has also been mentioned that the contrastive pitch accent is directly followed by a drop in pitch (Couper-Kuhlen, 1984; Chafe, 1974). Other work has claimed a less clear distinction between the new and contrastive category of information status, both in terms of prosody (Watson et al., 2008) as well as underlying in semantic and syntactic terms (Schwarzschild, 1999; Féry & Samek-Lodovici, 2006). Much discussion concerns the question to what extent the prosodic marking of contrastive information in Germanic languages like English is essentially different from the prosodic marking of new information, as suggested by the different transcriptions in Pierrehumbert and Hirschberg (1990). Krahmer and Swerts (2001) investigated this question using Dutch NPs consisting of one adjective and one noun (such as *blauwe vierkant*, “blue square”). Their study showed the importance of taking into account both production and perception data in the study of intonation. In an acoustic analysis of the produced NPs they found that the shape of the pitch accent does not differ between new or contrastive information. However, Krahmer and Swerts (2001) found in a perception experiment that pitch accents signaling contrastive information were perceived as more prominent than pitch accents signaling new information. This difference was only measureable when participants listened to the entire NP, not when they listened to the new or contrastive word in isolation.

1.4 Adaptation

Most of the models of contrastive intonation discussed so far argue that the contrastive intonation patterns relate to speaker intentions. In particular, as we have seen above, it is the speaker who decides to focus on specific words to indicate a semantic contrast with previously mentioned information. This raises the question whether this is because the contrast is important for the speaker or (also) for the addressee. Thus, to better understand what drives contrastive intonation one should disentangle speaker- and addressee-related factors in experimental approaches. In many previous studies, it was assumed that what is contrastive (or given or new, for that matter) is identical for the speaker and for the addressee, and in many situations this is plausible. However, the perspectives of speaker and addressee need not always be identical, for instance because they may have slightly different perspectives on a shared visual scene (in such a way that some objects are only visible for the speaker but not for the addressee) or because they have experienced the preceding discourse in a slightly different manner (because the speaker heard some utterances that were not audible for the addressee). In such contexts, the perspective of the speaker and addressee about information status starts to diverge, and the question is how this influences prosodic marking of information status. Do speakers only take their own perspective into account, or do they (also) consider the perspective of the addressee? What if speakers are inherently less capable of taking the addressee’s perspective (as has been argued to be the case for HFA speakers)? In this thesis we choose an interactional approach in which we focus on the assumptions of the speaker about the addressee's knowledge state. In particular, we approach this issue by means of experimental situations where the information status of particular words is different for the speaker and for the addressee.

Previous psycholinguistic work on adaptation in language has been concerned with the question to what extent the speakers' production of utterances is based on their own or their addressees' knowledge state. This work has led to different views on the production of language in interaction. Some researchers have claimed that utterances are initially produced egocentrically, i.e., without taking the addressee into account (Brown & Dell, 1987; Dell & Brown, 1991; Horton & Keysar, 1996; Keysar, Barr, & Horton, 1998; Pickering & Garrod, 2004; Barr & Keysar, 2007), while other studies have argued that speakers produce utterances explicitly for their addressees (Clark & Murphy, 1982; Clark & Wilkes-Gibbs, 1986; Clark & Brennan, 1991; Nadig & Sedivy, 2002). More recent work suggests that utterances are the result of both speaker- and addressee-related features (Brennan & Hanna, 2009; Galati & Brennan, 2010). Many of the studies just mentioned have looked at lexical evidence, investigating the words speakers utter in a communication task.

For example, Clark and Wilkes-Gibbs (1986) studied a communication 'game' in which one player (the matcher) was instructed to put tangram figures in the same order as those of the other player (the director). It was the task of the director to describe the figures as quickly and accurately as possible in such a way that the matcher could successfully rearrange the figures. Both director and matcher could talk as much as needed. Crucially, the task consisted of six different trials so that all the figures were described repeatedly. Clark and Wilkes-Gibbs (1986), among other things, counted the number of words the director used per trial to describe the figures. They found that with every repetition fewer words were used. This finding was explained by a collaborative model, which assumes that speakers design their utterances for their addressees: in the first trial a description could have been "all right, the next one looks like a person who's ice skating, except they're sticking two arms out in front", whereas in the sixth and last trial the same figure could be described as "the ice skater" (Clark & Wilkes-Gibbs, 1986, p.12).

However, work by Horton and Keysar (1996) suggests that speakers are sometimes more egocentric when producing language. They report an experiment in which speaker and addressee each saw one half of a computer screen. On the computer screen a ball was moving from the speaker-side towards the addressee-side. The task of the speaker was to describe the moving ball. Crucially, another non-moving smaller or bigger ball was visible on the speakers' part of the screen and not on the addressees'. It was counted how often speakers referred to the moving ball using an adjective ("small ball" or "big ball"). Each time a speaker did this (thereby 'leaking' information about the size of the ball not visible for the addressee), it was taken as evidence that the speaker did not take the perspective of the addressee into account, since for the addressee "the ball" would have been a sufficient description in all the cases. Horton and Keysar (1996) found that when speakers were put under time pressure they produced more adjectives than under normal time conditions. They argued that under time pressure the speakers were less able to monitor the addressee's perspective and thus were more egocentric in their speech production (Horton & Keysar, 1996).

So far, adaptation has been studied mainly by looking at the verbal utterances produced by speakers (e.g., Horton & Keysar, 1996; Clark & Wilkes-Gibbs, 1986, among many others). In this thesis we argue that a better understanding of adaptation in language production is obtained by investigating non-verbal aspects of language as well, such as the prosodic marking of semantic contrasts, which relies at least partly on the communicative intentions of the speaker.

1.5 Current thesis

This thesis focuses on contrastive intonation in interactions between speaker and addressee and aims to answer two research questions. The first question is to what extent speakers take into account the knowledge state of their addressee when producing a phrase with contrastive intonation. The second question is to what extent the prosodic marking of semantic contrasts is affected by interactional processes, such as whether two interlocutors take over each other's prosody. Contrastive intonation patterns are elicited by manipulating the order of stimuli presented to participants in experiments, which enables us to control the information status of words (i.e., contrastive or given). Contrastive intonation patterns thus act as an ideal basis to study intonation in interaction. In the current thesis, all studies make use of communication tasks with speaker-addressee interactions to elicit contrastive intonation patterns. Furthermore, all experiments elicit NPs consisting of an adjective and a noun, with one of the two referring to contrastive information. The pitch accent that speakers of Germanic languages such as Dutch use to indicate this contrast is therefore on the default (noun) or non-default (adjective) location. It is crucial to distinguish pitch accents on default locations from pitch accents on non-default locations (cf. (1b) and (1c)), as these can correspond to different linguistic functions, plausibly with differences in linguistic prosody as well. And importantly, in the present methodology all analyses concern production and perception measures of pitch (F0) and prominence respectively.

The current thesis presents four self-contained studies, three of which were published in scientific journals and one which is currently under review. Due to the self-contained nature of each of these studies, a small amount of redundancy in the respective introductions was unavoidable. In addition, due to the requirements of the different journals, there may be minor differences in the way results are presented and analysed. The studies approach the general research question of this thesis in two ways. The first two studies investigate to what extent speakers mark semantic contrasts prosodically based on their own knowledge state or based on that of their addressee. Both investigations make use of experimental settings in which not all information is shared between speaker and addressee, similar to Horton and Keysar (1996) and Galati and Brennan (2010). Crucially, these studies investigate how the speakers' prosody is affected by their knowledge of which information is available for their addressee. The aim of the first two studies is to shed light on whether contrastive intonation is primarily speaker- or addressee-driven in speakers with different mental developments, comparing HFA speakers with typically-developing ones (Chapter 2) and to what extent speakers' contrastive information may reveal ('leak') information about objects only the speaker can see (Chapter 3) respectively. The last two studies investigate to what extent speakers adapt to addressees in their use of contrastive intonation, and how this is related to adaptation for prosodic aspects that do not have a linguistic function. This comparison is crucial for an investigation of adaptation, as both prosody with a linguistic function and prosody without a linguistic function manifest themselves in the same acoustic features, such as pitch, duration or loudness (Ladd, 2008; Pardo, 2006). Specifically, the last two studies question whether contrastive intonation is a cue for speaker adaptation (Chapter 4) and to what extent speakers of different languages differ in prosodic adaptation (Chapter 5).

1.6 The studies

This section discusses the respective research questions (RQ) for each study presented in the following chapters.

1.6.1 RQ1: To what extent is contrastive intonation speaker- or addressee-driven? (Chapter 2)

The first study investigated to what extent contrastive intonation is speaker- or addressee-driven. To this end a typically-developing group of speakers and a group of speakers having high functioning autism (HFA) were compared (native speakers of Dutch in both groups). The HFA group was chosen because HFA speakers have been argued to have more difficulties accounting for another's mental state (Baron-Cohen, 1995), which is plausibly reflected in their use of prosody (Peppé et al., 2003). Both groups took part in a communication task where sequences of figures had to be described to two different addressees. The sequence of figures was manipulated in such a way that in the critical situation one figure was described for one addressee and the following, contrastive figure was described to another addressee, which made sure that the relevant property was contrastive for the speaker but not for the addressee. Results show that both speaker groups used a less clearly marked contrastive intonation for sequences uttered to different addressees compared to sequences of contrastive figures uttered to the same addressee. This result was taken as evidence that contrastive intonation is both speaker-driven (because prosodic marking of contrasts occurred whenever there was such a contrast for the speaker) as well as addressee-driven (because prosodic marking of contrasts differed as a function of same or different addressee). An additional acoustic analysis showed that typically-developing speakers and speakers with autism do differ on features that relate to the form of their prosody (global features such as pitch range) rather than on features that relate to the function of prosody (contrastive intonation).

1.6.2 RQ2: To what extent do speakers leak contrastive information to their addressees by their use of prosody? (Chapter 3)

The second study investigated the prosody of native Dutch speakers who referred to information that was visually available only for themselves, not for the addressee. In particular, in this experiment speakers saw four figures of which one was occluded for the addressee. The occluded figure (e.g., a small triangle) crucially contrasted with a unique target figure that was visible for both the speaker and the addressee (e.g., a big triangle). In this case, a description such as "the triangle" would have been sufficient from the perspective of the addressee; when speakers would produce a description such as "the big triangle" they arguably leak information about the size of the occluded figure. It has been shown in earlier research that speakers used more adjectives referring to such contrasts when explicitly instructed *not* to leak information about the occluded figure (Wardlow Lane et al., 2006). These results suggest that speakers do not take into account the perspective of the addressee when explicitly instructed not to leak certain information. This paradoxical result has been explained in terms of 'ironic processes', known from psychological experiments where participants, when asked not to think about a white bear, could not avoid doing so (Wegner et al., 1984). The results of this study extended these findings; we showed that not only the speakers' reported thoughts (Wegner et al., 1984) and the choice of words (Wardlow Lane et al., 2006), but also the speakers' prosody is affected by ironic processes. That is, speakers mark contrastive information that they should ignore even more strongly than contrasts that they need not to ignore.

1.6.3 RQ3: To what extent is contrastive intonation a cue for speaker adaptation? (Chapter 4)

The third study is an exploration of the extent to which contrastive intonation is a perceptual cue for speaker adaptation in Dutch. So far, much work considered speaker adaptation in prosody as the copying of certain global prosodic features (i.e., Pardo, 2006), while neglecting the role of linguistically functional prosody. This study reports a perception experiment in which listeners are presented with manipulated dialogue segments. The dialogue segments consisted of interactions between two speakers who referred to contrastive information. In one half of the stimuli the speaker pairs used contrastive intonation coherently, with words referring to new or contrastive information being accented and words referring to given information being deaccented (speaker A: "blue BALL" followed by speaker B: "RED ball"). In the other half of the stimuli the speakers copied each other's intonation pattern in contrastive contexts, even if that would have meant a conflict with the linguistic function of pitch accents (speaker A: "blue BALL" followed by speaker B: "red BALL"). This study investigated whether listeners perceived speakers as better adaptors to their interlocutors when they used contrastive intonation coherently (i.e., linguistically functional) or when they copied the form of the intonation pattern of their interlocutor (i.e., not linguistically functional). Results showed that speakers are perceived as better adaptors when they used contrastive intonation coherently compared to when they merely copied the prosodic intonation pattern of their interlocutor.

1.6.4 RQ4: To what extent do speakers of Dutch and Italian adapt to atypical prosody in contrastive contexts? (Chapter 5)

The fourth and last study investigated to what extent speakers adapt their prosody in contrastive contexts across languages. To this end speakers of Dutch and Italian, a plastic and a non-plastic language (Vallduví, 1991), were compared on the extent to which they adapted their prosody to interlocutors who use prosody in an atypical way. Therefore, one half of the speakers in the experiment interacted with a confederate who used prosody which was in accordance with the rules of prosodic marking of information status in the language of interest (typical). The other half of the speakers interacted with a confederate who used prosody which was not in accordance with those rules (atypical). In this study we questioned to what extent the plasticity difference was reflected in the extent to which speakers of Dutch or Italian adapted to atypical prosody in contrastive contexts. Adaptation was measured as the difference between the prosody produced by the speakers in the typical and atypical condition. We predicted that Dutch speakers, because of their 'plastic' prosody were more likely to adapt than speakers of Italian who exhibit a 'non-plastic' prosody. Results showed that speakers of Dutch indeed adapted more than speakers of Italian. This study shows that the possibilities for prosodic adaptation can thus be related to characteristics of the language one speaks.

CHAPTER 2

ACCOUNTING FOR THE LISTENER: COMPARING THE PRODUCTION OF CONTRASTIVE INTONATION IN TYPICALLY-DEVELOPING SPEAKERS AND SPEAKERS WITH AUTISM

ABSTRACT

The present research investigates what drives the prosodic marking of contrastive information. For example, a typically-developing speaker of a Germanic language like Dutch generally refers to a pink car as a "PINK car" (accented words in capitals) when a previously mentioned car was red. The main question addressed in this chapter is whether contrastive intonation is produced with respect to the speaker's or (also) the listener's perspective on the preceding discourse. Furthermore, this research investigates the production of contrastive intonation by typically-developing speakers and speakers with autism. The latter group is investigated because people with autism are argued to have difficulties accounting for another person's mental state and exhibit difficulties in the production and perception of accentuation and pitch range. To this end, utterances with contrastive intonation are elicited from both groups and analyzed in terms of function and form of prosody using production and perception measures. Contrary to expectations, typically-developing speakers and speakers with autism produce functionally similar contrastive intonation, as both groups account for both their own and their listener's perspective. However, typically-developing speakers use a larger pitch range and are perceived as speaking more dynamically than speakers with autism, suggesting differences in their use of prosodic form.

This chapter is based on:

Kaland, C. C. L., Swerts, M. G. J., & Krahmer, E. J. (2013b). Accounting for the listener: Comparing the production of contrastive intonation in typically-developing speakers and speakers with autism. *Journal of the Acoustical Society of America*, *134*(3), 2182-2196.

Kaland, C. C. L., Krahmer, E. J., & Swerts, M. G. J. (2012b). Contrastive intonation in autism: The effect of speaker- and listener-perspective. *Proceedings 13^th^ Annual Conference of the International Speech Communication Association (Interspeech)*, Portland (OR), USA, 1047-1050.

2.1 Introduction

Imagine a situation in which John and Mary are having a conversation about cars they saw lately. John says that he saw a red Ferrari last month. Peter, who is also in the room, cannot hear the conversation because he is listening to music and wears headphones. Actually, John knows that Peter is more interested in cars than Mary. Before John can go on telling Mary which other Ferrari he saw, Peter turns off his music and puts away his headphones. At that moment John is about to address Peter to say: "... and this week I saw a pink Ferrari". Given this situation, an interesting mismatch occurs between Peter and John's perspective on the information John is conveying. For John the phrase "the pink Ferrari" contrasts with the preceding phrase, as there is a semantic opposition in terms of the cars' color. For Peter, however, the mentioning of the pink Ferrari represents entirely new information, because he did not hear the preceding phrase referring to the differently colored car.

In such a setting, how would John utter this sentence? The two Ferraris can be distinguished on the basis of just their color. Therefore, the information status of *pink* can be called contrastive with respect to the previously mentioned alternative color. Current models of intonation would therefore predict that the speaker prosodically marks the contrastive information by means of increased prominence (e.g., by producing a pitch accent in Germanic languages like German; Pechmann, 1984a, 1984b). By doing so the speaker signals that the given information *Ferrari* is still the topic of discourse, but that the color is different. In turn, the pitch accent draws the listener's attention specifically towards the contrastive information. In the scenario above, would John produce, "And this week I saw a pink Ferrari" with a pitch accent on the adjective *pink*? Following his own perspective it makes sense to prosodically mark the contrast, as John himself knows about the red Ferrari. However, Peter does not know which, if any, other Ferraris have been mentioned in the preceding discourse that he did not witness. Therefore, from Peter's perspective it makes no sense for *pink* to be prosodically marked.

The aim of the present research is to shed light on what drives the prosodic marking of contrastive information. We investigate to what extent speakers take into account their own and/or their listener's perspective when producing contrastive intonation. For this reason we analyze the production of contrastive intonation when speaker and listener have different perspectives on the preceding discourse. The following section provides a review of previous work on the effects of speaker and listener perspective on language production. Thereafter, a specific section discusses aspects of prosody and contrastive intonation. The final section of the introduction argues why it is revealing to conduct this kind of research both with typically-developing speakers and speakers with autism, given that the latter have been argued to have difficulties with perspective taking and with producing appropriate intonation.

2.1.1 Background

A central question in current research on language production focuses on the extent to which speakers take into account the perspective of their listener. Some studies claimed that speakers do not always incorporate information about the listener (Horton & Keysar, 1996; Bard et al., 2000; Pickering & Garrod, 2004). Other studies argued that speakers do take the listener's perspective into account at the phonetic level and higher (Clark & Murphy, 1982) as soon as that information is available (Brennan & Hanna, 2009; Galati & Brennan, 2010). When perspectives of discourse partners differ, speakers may have particular difficulty accounting for what their listeners know. Studies looking into this often investigated the production of referring expressions using experimental tasks in which speaker and listener have different perspectives on the objects to be described. For instance, in these tasks speakers needed to refer to objects that are only visible to them and not to their listener (among others Horton & Keysar, 1996; Keysar et al., 1998). These referring expressions were taken as evidence that speakers fail to account for the differing perspective of their listeners. However, previous work in this area, especially studies that use a paradigm in which utterances were collected from speakers and listeners with different perspectives, tends to focus mainly on lexical aspects of language production. In these studies, the analysis of speakers' behavior was limited to the question whether a certain referential attribute was uttered or not. Fewer studies investigated how speakers refer to information *prosodically* in these situations. The underlying question in this line of research is whether speakers prosodically reduce repeated information for themselves or whether they do it for their listeners. So far, studies have mainly investigated the intelligibility and duration of information that is repeated by the speaker (Bard et al., 2000; Galati & Brennan, 2010). The crucial manipulation in these studies was whether the repeated information is uttered to the same or a different listener as compared to the initial mention. Studies compared the amount of reduction measured in the repeated mention to the same listener with the amount of reduction in the repeated mention to a different listener. This paradigm has led to different views on the role of perspective-taking in the production of prosody. Some work found that speakers reduced repeated information even if the listener did not hear the previous mention (Bard et al., 2000; Bard & Aylett, 2000; but see Gregory et al., 2001). Models of speech production following from this evidence claimed that the incorporation of listener information in speech production is a cognitively costly process (Bard et al., 2000; Bard & Aylett, 2000). Other work found that speakers indeed reduced repeated information to a different listener, albeit to a lesser extent compared to information uttered to the same listener (Galati & Brennan, 2010). The one-bit model proposed by Galati & Brennan (2010) claims that it is computationally easy for the speaker to incorporate listener information, which happens as soon as that information is available. That is, for speakers it may require just one bit of information: the listener heard certain information or not. This model is in line with several studies of English showing that speakers' prosodic cues to syntactic disambiguation are used when necessary and that these cues are helpful for listeners (Schafer et al., 2000; Snedeker & Trueswell, 2003; Kraljic & Brennan, 2005).

The present research aims to contribute to this debate by investigating the use of intonation. Like prosodic reduction of repeated information, speakers may use intonation to signal the importance of discourse information. That is, intonation can make important words acoustically prominent, as in the case of contrastive intonation in the Ferrari example above. So far, no research within the perspective-taking debate has investigated the use of intonation. It remains to be investigated whether speakers highlight important words because they are important (i.e., contrastive) for themselves or because they are important for their listeners, or whether both speaker- and listener-oriented factors have to be taken into account. Contrastive intonation is especially useful to study perspective-taking, as this pattern is argued to be relevant both from a listener and speaker perspective.

2.1.1.1 Semantics-intonation interface of contrastive intonation

Speakers typically use contrastive intonation to indicate a semantic contrast. Rooth (1992) defines a semantic contrast as the presupposition of a set of alternatives to the contrastively focused word. Following the example given above, producing a pitch accent on *pink* in "...and this week I saw a pink Ferrari" presupposes the existence of a set of one or more differently colored Ferraris. In the example given above this set is given in the discourse context (and consists of a red Ferrari). There is considerable debate in the literature about the interface between the semantics of a contrast and how it is expressed in intonation. These studies have shown that this interface differs substantially between Germanic and Romance languages, like Dutch and Italian (Swerts et al., 2002) as well as Dutch and Romanian (Swerts, 2007). Some authors have argued that contrastive information in English is semantically and intonationally different from new information (Selkirk, 2008; Pierrehumbert & Hirschberg, 1990) whereas others favored a less clear distinction between these categories of information status in English (Schwarzschild, 1999; Féry & Samek-Lodovici, 2006; Watson et al., 2008). Specifically, it has been discussed whether pitch accents indicating new information are phonetically distinct from pitch accents indicating contrastive information (see Krahmer & Swerts, 2001 for a discussion). While this issue is beyond the focus of the present study, it is undisputedly the case that words referring to contrastive information are prosodically the most prominent within the domain of their scope in languages like English and Dutch (Rooth, 1996; Calhoun, 2009). Krahmer and Swerts (2001) showed that the perceived prominence of a contrastively focused word in a Dutch noun phrase is the result of both accentuation of the contrastive information (i.e., *pink*) and deaccentuation of the given information (i.e., *Ferrari*).

2.1.1.2 Listener-driven contrastive intonation

Studies showed that speakers who use a contrastive intonation pattern help their listeners to interpret the discourse structure. According to Levelt (1989) listeners may use initial mentions (*red Ferrari* in the example above) as a 'gestalt'. In particular, subsequent utterances can be interpreted with respect to already mentioned referents and their properties. In the case of contrastive intonation, the property that changed with respect to the gestalt (*pink* in the example above) is marked prosodically. Following Levelt (1989), speakers using contrastive intonation indicate to listeners that they can hold on to the gestalt they had in mind instead of creating a new one. In Levelt's (1989) interpretation, listeners can use the gestalt strategy most effectively when the noun is mentioned. Evidence from object naming tasks indeed support the gestalt view. For example, Pechmann (1984a, 1984b) found that a noun is almost always included when referring to objects in German, although speakers could say "... and this week I saw a pink one". Psycholinguistic research points out that the contrastive intonation indeed helps listeners and allows them to use a gestalt strategy. Weber et al. (2006) used eye-tracking to investigate German listeners' eye-gaze at contrastive and non-contrastive referents. For example, with respect to *purple scissors* a contrastive referring expression may be *red scissors* and a non-contrastive referring expression may be *red vase*. The intonation of the references was manipulated such that they either occurred with a contrastive or neutral intonation. Results of Weber et al. (2006) showed that a German contrastive intonation pattern results in more looks to contrastive than to non-contrastive referents. In two reaction-time experiments, Braun and Tagliapietra (2010) investigated to what extent contrastive intonation in Dutch facilitates the retrieval of contextual alternatives. It has been argued that when alternatives are not mentioned explicitly in a previous utterance, listeners may accommodate for this by presupposing that information (Lewis, 1979). The mechanism of accommodation holds for contrastive intonation in that listeners recognize contextual alternatives more rapidly than generic (non-contrastive) alternatives (Braun & Tagliapietra, 2010). Furthermore, English listeners remember words with contrastive intonation better than words with neutral intonation (Fraundorf et al., 2010). This effect holds up to one day after the pattern was heard and concerns the memorization of both the contrastive and the alternative information.

2.1.1.3 Speaker-driven contrastive intonation

The aforementioned studies may suggest that contrastive intonation is listener-driven in that speakers use this pattern to facilitate their listeners' perception of contrastive information. Chafe (1976) indicated that contrastive intonation in English can also be produced with regard to the speakers' perspective only, without taking the knowledge of the listener into account. In Chafe's (1976) example, Sherlock Holmes thinks for a long time about possible perpetrators of a crime and then suddenly says: "The BUTLER did it!" (with a pitch accent on *butler*). At the moment Holmes utters his thoughts the listener may not be aware with which alternative perpetrators *butler* might contrast. Crucially, the contextual alternatives for the butler are not explicitly mentioned and are therefore not explicitly shared with the listener. Chafe (1976) called this 'quasi-given' information in that givenness of the alternative information only holds from the speakers' perspective. Such a speaker-driven contrastive intonation does not necessarily harm the process of communication. That is, the accommodation mechanism of listeners plausibly allows speakers to produce a contrastive intonation pattern when the alternative information remains unmentioned. Furthermore, there is evidence that speakers use prosody to disambiguate information structure, even if disambiguation is not needed from the perspective of the listener (Schafer et al., 2000). This evidence favors a speaker-driven account of prosody. There is no experimental evidence so far that contrastive intonation is speaker-driven. As such, this is one of the issues we address in the current study.

2.1.1.4 Theory of Mind and contrastive intonation in autism

The ability to recognize and account for another person's perspective has often been analyzed in terms of Theory of Mind (ToM) models. This ability is claimed to be impaired in people with an autism spectrum disorder (Baron-Cohen, 1995, 2001). A functioning ToM is crucial for communication in general and for pragmatic ways of language use in particular. Pragmatic use of language, including contrastive intonation, depends highly on the intentions of the speaker. These are per definition unpredictable for the listener and require the speaker to account for that. To produce contrastive intonation the speaker has to rely on the discourse context at hand, which is crucially determined by what was said previously by both interlocutors. When unable to account for the other person's perspective, it may be difficult to use contrastive intonation in a way that is understandable for the listener. Research indeed showed that this pattern is particularly problematic in autism (see McCann & Peppé, 2003, for an overview). Studies showed that English speakers with autism place accents on more than one syllable (Baltaxe, 1984) or on inappropriate words (McCaleb & Prizant, 1985; Fine et al., 1991; Shriberg et al., 2001). An interesting finding is reported by Peppé et al. (2007), who investigated both the perception and production of contrastive intonation in English. They showed that children with autism have difficulties interpreting contrastive information when the adjective is accented. Production data showed that those children often accentuate the adjective when accentuation is not necessary. This finding is in line with Baltaxe and Guthrie (1987) who found a general tendency in English speaking children with autism to emphasize words in the primary sentence position. Difficulties in the production of contrastive intonation have been explained by impaired perspective-taking in autism by Shriberg et al. (2001). They argued that a speaker has to keep track of what is new for the listener in order to appropriately use contrastive intonation. As this is difficult for speakers with autism, their contrastive intonation is deviant (Shriberg et al., 2001).

Only a few studies have directly investigated the extent to which disordered prosody relates to perspective-taking difficulties in autism. Participants in these studies had to recognize mental states or emotions on the basis of vocal cues. Although English listeners with autism performed worse than typically-developing speakers on these tasks, the difference was not always found to be significant (cf. Rutherford et al., 2002 and Chevallier et al., 2011). To our knowledge there is no research that used the production of contrastive intonation to assess the extent to which speakers with autism take the perspective of their listeners into account.

Presumably, it is more difficult for speakers with autism than for typically-developing speakers to adapt their intonation when the perspective of their listener is different from their own. Therefore, the question remains as to how speakers with autism produce contrastive intonation when their listeners did not hear the previous utterance containing alternative information. Answering this question would shed more light on the relation between intonation and perspective-taking abilities in autism. This issue will be addressed in the current study.

2.1.1.5 Disordered prosody in autism: function versus form

As suggested in the literature on autism discussed above, it is plausible to relate the problematic use of contrastive intonation to difficulties in perspective taking. However, it is also possible that contrastive intonation problems in autism stem from general deficits in prosody. It has been noted for instance, that speakers with autism sound different from typically-developing speakers. Acoustic impressions vary to a large extent. Speakers with autism have been described as sounding 'monotonous' (Von Benda, 1983) and 'singsong' (Baltaxe & Simmons, 1985). This heterogeneity can be ascribed to the diversity of impairments in autism (Diehl et al., 2009) or to varying language abilities among autistic participants (DePape et al., 2012). Studies that investigated prosodic form in autism in more acoustic detail mostly focused on pitch range. A common finding is that speakers with autism use a larger pitch range than typically-developing speakers (for an overview see Nadig & Shaw, 2011). A larger pitch range means that speakers produce speech with more tonal variability. We cannot rule out the possibility that this variability influences the way speakers with autism produce pitch accents when using contrastive intonation. To investigate this possibility, the present study distinguishes prosodic function from prosodic form. That is, contrastive intonation will be considered as a functional property of prosody because of its communicative function and close relation to the semantics of an utterance. Pitch range will be considered as a feature that primarily relates to the form of an utterance rather than to its function or semantics. In the current study, we primarily analyze contrastive intonation as a functional property of prosody that speakers may use to account for their listeners. In addition, an analysis of pitch range is carried out to shed more light on the alleged prosodic deficits in autism and the extent to which they are related to perspective taking.

2.1.2 Research goals

Following the discussion of the literature, the research presented here has two major aims. First, we investigate the extent to which contrastive intonation is speaker- or listener-driven. Although the literature provides evidence for both explanations, there has been a lack of experimental approaches to this issue. Second, we investigate the production of contrastive intonation by both typically-developing speakers and speakers with autism. The latter group has been argued to have difficulties in accounting for the perspective of another person and has been shown to exhibit atypical prosody. However, the relationship between perspective-taking difficulties and the problematic use of contrastive intonation in autism has not been investigated directly. To explore this relationship, we distinguish between functional and formal aspects of prosody. In line with our aims, the present study investigates the extent to which typically-developing speakers and speakers with autism account for their listeners when producing contrastive intonation. To this end, we conduct a production experiment with speakers of Dutch who produce noun phrases (NPs) referring to information that is contrastive with respect to a previously uttered alternative NP. As Dutch is a Germanic language, prosody is used to indicate semantic contrasts like in English and German. Speakers in the production experiment either utter the alternative NP and contrastive NP to the same listener or to different listeners. This manipulation allows a comparison of utterances when the speaker and the listener have the same or different perspectives on the information. In the Ferrari example above, this would be a comparison between John's utterance to Mary, who heard the previous contrasting utterance and John's utterance to Peter, who did not hear the previous contrasting utterance. The speakers' utterances produced when their perspective differs from their listeners' are crucial for an investigation of perspective-taking abilities. Presumably, speakers with autism account for their listeners to a lesser extent than typically-developing speakers and, moreover, they may produce contrastive intonation irrespective of whether they utter the contrastive NP to the same listener or to a different one. Furthermore, speakers with autism are expected to make more accent placement errors than typically-developing speakers (McCann & Peppé, 2003). For example, speakers with autism may accent a word that needs to be deaccented or vice versa. Elicited NPs are analyzed in terms of production measures of F0 and pitch range and in terms of perception measures of prominence and speech dynamicity.

The remainder of this chapter is structured as follows: Section 2.2 describes a production experiment with typically-developing speakers and speakers with autism; Section 2.3 describes the results of perception experiments using elicited NPs of typically-developing speakers and speakers with autism; this chapter ends with a discussion of the results and concluding remarks.

Table 2.1: *Schematic overview of example stimuli for each level of the variables listener and focus. For illustrational purposes the alternative NP here is always uttered to listener A. In the actual experiment listeners are balanced over conditions.*

Listener	Focus	Alternative NP	Contrastive NP
Same	Adjective	“blue triangle” to A	“red triangle” to A
	Noun	“blue triangle” to A	“blue drop” to A
Different	Adjective	“blue triangle” to A	“red triangle” to B
	Noun	“blue triangle” to A	“blue drop” to B

2.2 Production

2.2.1 Method

To elicit references to contrastive information, participants acted as speakers in a referential communication task (performed in Dutch). In this task they instructed two different listeners to put figures that were printed on paper cards on a piece of paper. The order of instructions was manipulated so that successive instructions referred to figures that could be distinguished by just their color or just their shape (test stimuli) or by both their color and their shape (fillers). A test stimulus concerned the NP in the latter of two successive instructions, as the present study investigates contrastive intonation with respect to the previous utterance (Table 2.1). Two successive instructions, part of larger sequence of utterances, were either uttered to the same listener or to a different listener. The setup ensured that contrastive intonation patterns produced for different listeners only made sense from the speaker’s perspective. In particular, speakers were told that when addressing one listener, the other listener heard music via a headphone so that the instruction could not be heard. This ensured that speakers believed that they did not share the previous utterances with different listeners. In reality, listeners were confederates and heard all instructions (see section 2.2.1.3). Because contrastive information in the test stimuli concerned either the color or shape of the target figure, the focused word was either the adjective or the noun.

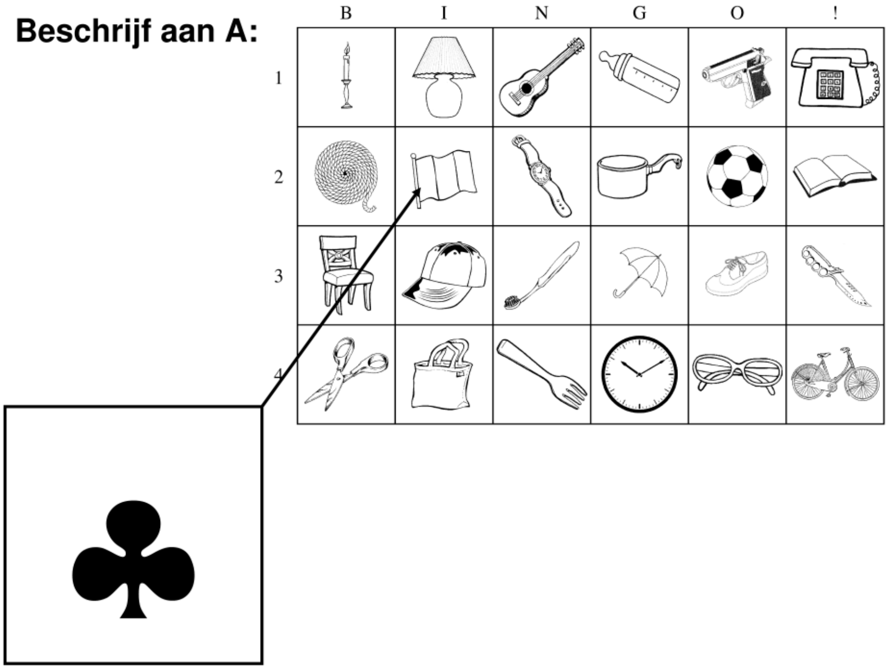

Figure 2.1: *Example of the speaker's screen, showing in Dutch "Beschrijf aan A" (describe to A), the target figure (bottom left) and A's bingo card. A typical instruction would be: "put the red clover on the flag".*

2.2.1.1 Design and materials

The communication task was played as a bingo game with the speaker as the game leader and listeners as players. Each listener had a different bingo card displaying 24 common objects (e.g., fruit, tools, means of transport, see Figure 2.1) and small cards, each the size of a grid-square, displaying coloured shapes: a drop, a clover, a canoe or a triangle (in Dutch *druppel*, *klaver*, *kano* and *driehoek*, respectively) colored red, yellow, green or blue (in Dutch *rood*, *geel*, *groen* and *blauw* respectively). Bingo cards were 6 x 4 grids with rows numbered from 1 to 4 and columns marked by each character of the word "bingo!" (Figure 2.1). Note that the Dutch words referring to the color or shape of the figures all had two syllables with lexical stress on the first syllable. The speaker instructed the listeners to put a colored shape on top of one of the objects on the bingo card, e.g., "leg de blauwe driehoek op de banaan" (*put the blue triangle on the banana*). The phrase involving the object was included to prevent the use of boundary tones on the noun referring to the shape. Six game rounds were played, which began with the speaker's announcement of what the goal of that round was. This could be, for example, to have each cell of row 2 on the bingo card covered with a figure, for example. The listener who first achieved the right pattern would shout "bingo!", upon which that listener received a point and the round ended. The speaker switched 20 times between listeners at random places in the game. The speaker kept the score. The first instruction of each new round was a filler to account for speakers' pitch reset upon switching discourse contexts (Brown et al., 1980). The stimulus order occurred in two randomizations; each of which was presented to 10 participants. Speakers uttered 48 instructions in total (equally spread over listeners, crossed for the factors listener and focus) of which 24 were fillers and not taken into account for analysis.

2.2.1.2 Participants

20 typically-developing participants (TYP) acted as speaker in the production experiment (17 women, 3 men, M_{age} = 21.8 years, age range: 18-29 years). They were all native Dutch speakers and students of Tilburg University who participated for course credit. None of them were diagnosed with autism at the moment the experiment took place.

Additionally, 20 participants with an autism spectrum disorder acted as speaker in the production experiment (6 women, 14 men, M_{age} = 28.9 years, age range: 18-51 years). They were all native speakers of Dutch with high functioning autism (HFA), diagnosed between November 2005 and October 2011. All participants met the requirements for an autism spectrum disorder as described in DSM-IV (American Psychiatric Association, 2000). They were either diagnosed by a psychiatrist or by a psychologist as having Asperger Syndrome (1 woman, 6 men) or Pervasive Developmental Disorder - Not Otherwise Specified (PDD-NOS; 5 women, 8 men)[1]. The population of participants with autism did not allow for a match on age or education level with the typically-developing participants. 11 participants (3 Asperger, 8 PDD-NOS) had one or more comorbid disorders of which Attention Deficit Hyperactivity Disorder (ADHD), anxiety disorders and depression were most frequent. They were given a small present for their effort.

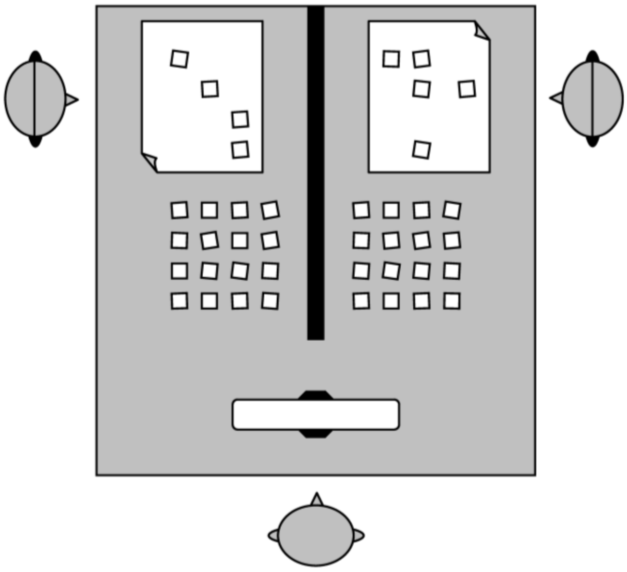

Figure 2.2: *Birdseye view of the experimental setup showing the speaker facing the screen (bottom) and the listeners, at opposite sides of a partition, facing their bingo cards and figures (top).*

2.2.1.3 Procedure

The speaker was seated at one end of a table, facing the listeners, who were seated at the other end and could not see each other (Figure 2.2). Before the game began speakers received instructions and played a training round. Listeners wore open-ear headphones to enhance the speaker's illusion that the listener who was not addressed heard music. The choice for open-ear headphones made sure that the listener who was addressed could indeed hear the speaker. During debriefing speakers were asked whether they believed that listeners heard music when they were not addressed and not the instruction to the other listener (all responded affirmatively). Thereafter, the actual setup of the experiment was explained.

Speakers (not listeners) saw a screen displaying the target figure and the bingo card of the listener to be addressed (Figure 2.1). The screen's lay-out indicated when speakers had to switch between listeners. In particular, for listener A the target figure was displayed on the screen's left side and for listener B the target figure was displayed on the right side. In accordance, speakers had to look past the left side of the screen when addressing listener A and past the right side of the screen when addressing listener B (Figure 2.2). Additionally, speakers were told that the software responsible for the instruction slides on the screen also switched music between listeners. Speakers' speech was recorded digitally and saved as wave-files.

2.2.1.4 Prosodic analysis

NPs referring to target figures in the test stimuli (N_{TYP} = 480, N_{HFA} = 480) were extracted from the wave-file recordings using Praat (Boersma & Weenink, 2011). They were analyzed in terms of pitch (F0). Pitch was taken as a correlate of the produced prominence (Ladd, 2008), which in this study is seen as a functional correlate of prosody. It has to be noted that prominence also manifests itself in other acoustic features, such as duration and intensity (i.e., Kochanski et al., 2005). It is beyond the scope of the current study to consider all possible correlates of prominence. We therefore take just one acoustic measure of production, pitch, to be verified with perceptual ratings of prominence (section 2.3).

For pitch measures, F0 maxima in Hertz on the stressed syllable of the adjective and the noun were measured in Praat (Boersma & Weenink, 2011). Some speakers ended the NP with a high boundary tone on the last syllable of the noun referring to the shape of the target figure (N_{TYP} = 101, N_{HFA} = 45)[2]. However, that syllable was never the stressed one (see section 2.2.1.1).

As shown by Krahmer and Swerts (2001), the contrastively focused word in Dutch obtains prominence *both* by its accentuation *and* by the deaccentuation of the unfocused word. To account for this a difference score was computed. That is, the F0 maximum of the unfocused word was subtracted from the F0 maximum of the focused word. In this way, positive difference scores indicated that the focused word had a higher pitch than the unfocused word and negative scores indicated that the unfocused word had a higher pitch than the focused word.

NPs were further analyzed for pitch range, taken as a correlate of prosodic form. We calculated pitch range as the mean of the standard deviations of F0 movement taken from both the adjective and the noun using Praat (Boersma & Weenink, 2011). This method closely resembles the one used by Nilsonne et al. (1988) which was explicitly designed for clinical acoustic measures and was used previously to measure the pitch range of HFA speakers (Diehl et al., 2009). The standard deviation used in those studies was calculated from mean F0 measurements for every 250 milliseconds in larger stretches of speech. As the current study focuses on NPs which are per definition short, we obtained a standard deviation directly from Praat (Boersma & Weenink, 2011). This standard deviation is based on mean F0 measurements every 10 milliseconds and is therefore arguably more suitable for an analysis of NPs. To abstract over speakers' gender differences standard deviations were measured in ERB (Glasberg & Moore, 1990), which uses a logarithmic scale for higher frequencies and better represents human pitch perception in speech than, for example, a non-logarithmic Hertz scale.

2.2.1.5 Statistical analysis

Repeated Measures Analyses of Variance (RM-ANOVAs) were performed on F0 difference scores and pitch range values as dependent variables with listener (2 levels: same, different) and focus (2 levels: adjective, noun) as within-subject factors and with development (2 levels: typical, HFA) as between subject factor.

Furthermore, RM-ANOVAs were performed for each development group separately. That is, on the data of the typically-developing speakers RM-ANOVAs were performed with the F0 difference scores and pitch range values as dependent variables with listener (2 levels: same, different) and focus (2 levels: adjective, noun) as within-subject factors. On the data of the speakers with autism RM-ANOVAs were performed with F0 difference scores and pitch range values as dependent variables with listener (2 levels: same, different) and focus (2 levels:

adjective, noun) as within-subject factors and diagnosis (2 levels: Asperger, PDD-NOS) as a between subject factor.

To explore individual differences between speakers, univariate ANOVAs were performed for each development group separately with F0 difference scores and pitch range values as dependent variables, listener (2 levels: same, different) and focus (2 levels: adjective, noun) as independent variables and speaker (20 levels) as a random independent variable. Because main effects of listener and focus are given by RM-ANOVA, only (interaction) effects involving the variable speaker are reported.

2.2.2 Results

Results are discussed following the order of statistical tests given in Table 2.2.

2.2.2.1 Both development groups taken together (TYP and HFA)

Analysis of both development groups together on F0 difference scores showed no main effects of listener or focus. There was a significant interaction between those factors in that addressing the same listener resulted in larger F0 differences for a focused adjective and addressing a different listener resulted in slightly larger F0 differences for a focused noun. Development also showed an effect, in that typically-developing speakers produced significantly larger F0 differences (M = 33.80) than HFA speakers (M = 8.49). For pitch range no (interaction) effects were found for listener and focus. Furthermore, typically-developing speakers produced NPs with a significantly larger pitch range (M = .47) than HFA speakers (M = .36).

Table 2.2: *Results of all ANOVAs carried out with F0 difference scores and pitch range values as dependent variables for both development groups taken together (TYP & HFA) and separately (TYP, HFA). Interaction effects not listed here were not significant on either of the dependent variables.*

	F0 difference	Pitch range
TYP & HFA		
Listener	n.s.	n.s.
Focus	n.s.	n.s.
Listener*Focus	$F(1,38) = 5.29$, $p < .05$, $\eta_p^2 = .12$	n.s.
Development	$F(1,38) = 10.01$, $p < .01$, $\eta_p^2 = .21$	$F(1,38) = 4.76$, $p < .05$, $\eta_p^2 = .11$
TYP		
Listener	n.s.	n.s.
Focus	n.s.	n.s.
Listener*Focus	$F(1,19) = 7.21$, $p < .05$, $\eta_p^2 = .28$	n.s.
Speaker	n.s.	$F(1,19) = 6.74$, $p < .001$, $\eta_p^2 = .87$
Speaker*Focus	$F(1,19) = 13.41$, $p < .001$, $\eta_p^2 = .93$	$F(1,19) = 3.01$, $p < .05$, $\eta_p^2 = .75$
HFA		
Listener	n.s.	n.s.
Focus	$F(1,19) = 23.32$, $p < .001$, $\eta_p^2 = .55$	n.s.
Listener*Focus	$F(1,19) = 9.65$, $p < .01$, $\eta_p^2 = .34$	n.s.
Diagnosis	$F(1,18) = 3.16$, $p = .092$, $\eta_p^2 = .15$	$F(1,18) = 4.07$, $p = .059$, $\eta_p^2 = .18$
Speaker	n.s.	$F(1,19) = 21.16$, $p < .001$, $\eta_p^2 = .97$
Speaker*Listener	n.s.	$F(1,19) = 2.68$, $p < .05$, $\eta_p^2 = .73$
Speaker*Focus	$F(1,19) = 3.88$, $p < .01$, $\eta_p^2 = .80$	n.s.

Table 2.3: *Mean F0 maximum (Hz), mean pitch range (ERB) and standard deviations as a function of development, listener and focus. Individual scores for adjective and noun as well as difference scores are given for F0 maxima. Pitch range values cover entire NPs.*

Development	Listener	Focus	F0 maximum *M* (*SD*)			Pitch range *M* (*SD*)
			Adjective	Noun	Difference	NP
Typical	Same	Adjective	272.16 (*60.78*)	232.30 (*47.45*)	39.86 (*50.62*)	.48 (*.17*)
		Noun	248.58 (*59.69*)	252.70 (*52.71*)	4.12 (*59.81*)	.46 (*.16*)
	Different	Adjective	276.36 (*61.48*)	241.75 (*49.76*)	34.61 (*45.40*)	.48 (*.16*)
		Noun	249.47 (*53.75*)	258.40 (*55.88*)	8.93 (*52.71*)	.45 (*.15*)
HFA	Same	Adjective	208.84 (*55.83*)	193.77 (*59.74*)	15.07 (*32.70*)	.36 (*.17*)
		Noun	196.96 (*53.61*)	200.20 (*52.29*)	3.24 (*27.22*)	.33 (*.17*)
	Different	Adjective	214.10 (*58.73*)	194.62 (*56.67*)	19.48 (*32.52*)	.36 (*.18*)
		Noun	203.18 (*55.57*)	199.36 (*54.19*)	-3.83 (*26.51*)	.37 (*.19*)

2.2.2.2 Typically-developing speakers (TYP)

Zooming in on the typically-developing speakers, no main effect of listener or focus on F0 difference scores were found. However, there was a significant interaction between the two factors in that addressing the same listener resulted in larger difference scores for a focused adjective, whereas addressing a different listener resulted in larger difference scores for a focused noun. The univariate ANOVA showed no effect of the random variable speaker on the F0 difference scores. However, a significant interaction effect of the variables focus and speaker was found, indicating that speakers differed individually in the way they marked focus by means of F0 differences. The pitch range measures revealed no main effects of listener or focus nor any interaction effects. The factor speaker as well as its interaction with focus revealed significant effects for pitch range, which suggests that there are individual differences in speakers' pitch ranges and that these differences depend on whether the adjective or the noun is in focus.

2.2.2.3 Speakers with autism (HFA)

Analysis of the speakers with autism on F0 difference scores showed no effect of listener. Results showed an effect of focus in that focused adjectives resulted in larger F0 differences (M = 17.27) than focused nouns (M = -.29). Furthermore, the factors listener and focus interacted such that difference scores for focused adjectives became larger when a different listener was addressed than when the same listener was addressed. Difference scores for focused nouns were small when the same listener was addressed and negative when a different listener was addressed, revealing that speakers produced focused nouns with a lower maximum F0 than unfocused adjectives when they addressed a different listener. Inspection of the F0 scores of adjective and noun separately (Table 2.3) revealed that speakers produced focused nouns with a lower maximum F0 than unfocused adjectives when they addressed a different listener. The factor diagnosis showed a trend towards significance in that speakers with Asperger syndrome (M = 13.94) produced larger F0 differences than speakers with PDD-NOS (M = 5.56). The factor speaker showed no main effect on the F0 differences, although its interaction with focus was significant, indicating that speakers differed individually in the way they marked focus by means of F0 differences. Pitch range was not significantly affected by listener or focus nor their interaction. The factor diagnosis revealed that speakers with Asperger produced larger pitch ranges (M = .46) than speakers with PDD-NOS (M = .31), which showed a trend. The factor speaker had a significant effect on pitch range and showed an interaction effect with listener, which suggests that there are individual differences in pitch ranges between speakers with autism and that these differences depend on whether the same or a different listener was addressed.

2.2.3 Discussion

The production measures of F0 showed general differences between development groups in that typically-developing speakers produced larger differences between focused and unfocused words. This may suggest that accentuation was more clearly realized by typically-developing speakers than speakers with autism. Such an outcome is compatible with the finding that typically-developing speakers used a larger pitch range than speakers with autism.

F0 differences were not affected by whether speakers addressed the same or a different listener, which seems to indicate that speakers from both development groups used a similar intonation irrespective of the listener's perspective. In particular, they did not adapt F0 differences or pitch range when addressing a different listener. However, for both development groups interactions were found for listener and focus, indicating that addressing a different listener affected the F0 differences depending on which word in the NP indicated the contrast. We return to this issue in section 2.3.3.1.

Pitch range was not affected by which listener speakers addressed, nor by whether the adjective or the noun was in focus. However, in both development groups individual differences in the use of pitch range were found. These differences can be explained when we take into account that pitch range is a feature that closely relates to the individual characteristics of a speaker. That is, factors like speaking style or mood affect pitch range (i.e., Scherer, 1986). One is therefore likely to find individual differences in pitch range. Also, the individual differences found for participants relate to the way in which focus is marked. This indicates that participants varied significantly in the strength with which they produced contrastive intonation.

So far, the results provided a first impression of the production of contrastive intonation by means of F0 differences and of a general measure of prosodic form: pitch range. Most importantly, no differences between typically-developing speakers and speakers with autism were found in the way they realize their prosody when addressing a different listener. However, these results need perceptual verification to provide a better insight into the communicational relevance of the produced prosody. This verification is provided in section 2.3, which also gives a more elaborate discussion of all the prosodic analyses.

2.3 Perception

2.3.1 Method

The following sections report on two perception experiments that elicit judgments of prominence and speech dynamicity by naive listeners. Prominence ratings are taken as a perceptual verification of the F0 measures reported in section 2.2, as pitch is argued to be a main correlate of prominence (Ladd, 2008). Judgments of speech dynamicity are taken as perceptual evaluation of pitch range. Pitch range closely correlates with how monotonous or dynamic a speaker sounds. When using a small pitch range speakers presumably sound more monotonous than when using a large pitch range. With respect to the distinction between prosodic function and prosodic form, we see prominence as a functional correlate and speech dynamicity as a formal correlate. Prominence in the perception experiment is judged on the basis of form. Nevertheless, we assume that its fluctuations are the result of the prosodic marking of information structure, which is a functional use of prosody (Krahmer & Swerts, 2001; Ladd, 2008).

2.3.1.1 Prominence

NPs collected in the production experiments (N_{TYP} = 480, N_{HFA} = 480) were presented in a web-based task (WWStim; Veenker, 2003) to three intonation experts. They rated the strength of the accent on a three point scale (0 = no accent, 1 = weak accent, 2 = strong accent). Adjectives were rated in the first part of the task, nouns were rated in the second part. Experts heard the entire NP in each part. The presentation order of NPs was randomized so that experts were blind for condition. To abstract over the experts' ratings, the prominence scores per word were added up so that they ranged from 0 to 6 (0 when all experts rated the accent as absent, 6 when all experts rated the accent as strong). The experts' ratings were consistent as indicated by Pearson's correlation coefficients: [$r_{TYP}(478)$ range = .62 - .72, p < .001] and [$r_{HFA}(478)$ range = .59 - .69, p < .001].

A difference score was computed in the same way as was done for the F0 measures. That is, the prominence score of the unfocused word was subtracted from the prominence score of the focused word. Again, we investigated possible effects of a high boundary tone on the prominence difference scores using a subset of the collected NPs (N_{TYP} = 101, N_{HFA} = 45)[3].

2.3.1.2 Speech dynamicity

A subset of the NPs collected in the production experiments was taken (N = 160) such that the speech of each typically-developing speaker and each speaker with autism in each condition was represented once (i.e., 40 (speaker) x 2 (listener: same/different) x 2 (focus: adjective/noun)). NPs were presented to listeners in an online judgment task. The participant's task was to judge how monotonous or dynamic the speech sounded. Participants were asked to score this on a five point scale; ranging from 'monotonous' (1) to 'dynamic' (5). Participants did the perception experiment in a sound proof booth and wore headphones so that they could not hear any surrounding noise. Before the start of the actual experiment participants had the opportunity to adjust the audio volume, received instructions and had to judge an example stimulus. Stimuli were presented on html-pages designed using WWStim (Veenker, 2003). Stimuli were presented in a random order which was different for each participant. During the experiment each stimulus could be played as often as required. The judgment could be altered before proceeding to the next stimulus. However, participants could no longer alter judgments once they had moved on to the next stimulus. The task lasted about 25 minutes and the results were collected on a web server.

30 different participants completed the judgment task (22 women, 8 men, M_{age} = 22.6 years, age range: 18-60 years). They were all native Dutch speakers and students of Tilburg University who had no hearing problems and who participated for course credit. None of them had participated in the production experiments.

2.3.1.3 Statistics

RM-ANOVAs were performed on prominence difference scores values as dependent variable with listener (2 levels: same, different) and focus (2 levels: adjective, noun) as within-subject factors and with development (2 levels: typical, HFA) as between subject factor.

Concerning the speech dynamicity scores RM-ANOVAs were performed with development (2 levels: typical, HFA), listener (2 levels: same, different) and focus (2 levels: adjective, noun) as within-subject factors.

Furthermore, RM-ANOVAs were performed for each development group separately. On the data of the typically-developing speakers RM-ANOVAs were performed with the prominence difference scores and speech dynamicity scores as dependent variables with listener (2 levels: same, different) and focus (2 levels: adjective, noun) as within-subject factors. On the data of the speakers with autism RM-ANOVAs were performed with prominence difference scores as dependent variables with listener (2 levels: same, different) and focus (2 levels: adjective, noun) as within-subject factors and diagnosis (2 levels: Asperger, PDD-NOS) as between subject factor. Concerning speech dynamicity, the data of the speakers with autism was analyzed using an RM-ANOVA with diagnosis (2 levels: Asperger, PDD-NOS), listener (2 levels: same, different) and focus (2 levels: adjective, noun) as within-subject factors.

To explore individual differences between speakers univariate ANOVAs were performed for each development group separately with prominence difference scores and speech dynamicity values as dependent variables, listener (2 levels: same, different) and focus (2 levels: adjective, noun) as independent variables and speaker (20 levels) as random independent variable. Because main effects of listener and focus are given by RM-ANOVA, only (interaction) effects involving the variable speaker are reported.

In addition, Pearson correlation coefficients were calculated between the F0 difference and prominence difference scores as well as between the pitch range and speech dynamicity scores.

2.3.2 Results

Results are discussed following the order of statistical tests given in Table 2.5.

2.3.2.1 Both development groups taken together (TYP and HFA)

Data for both development groups taken together (Table 2.4) showed that addressing the same listener (M = 2.52) resulted in significantly larger prominence difference scores than addressing a different listener (M = 1.64). Concerning focus, speakers produced larger differences when the adjective was focused (M = 3.31) than when the noun was focused (M = .84). The between-subject factor of development showed a trend in that typically-developing speakers produced larger (M = 2.42) differences between focused and unfocused words than HFA speakers (M = 1.74). For speech dynamicity, no (interaction) effects of listener and focus were found. Concerning development, listeners perceived the speech of typically-developing speakers as more dynamic (M = 3.22) than the speech of speakers with autism (M = 3.14).

The correlation measures showed that prominence ratings and F0 maxima were correlated positively for both typically-developing speakers ($r_{adjective}(478) = .25$, $p < .01$ and $r_{noun}(478) = .10$, $p < .05$) and speakers with autism ($r_{adjective}(478) = .18$, $p < .01$ and $r_{noun}(478) = .12$, $p < .01$). Furthermore, a correlation was found between pitch range and speech dynamicity for both typically-developing speakers ($r(78) = .33$, $p < .01$) and speakers with autism ($r(78) = .60$, $p < .001$).

Table 2.4: *Mean prominence score, mean speech dynamicity and standard deviations as a function of development, listener and focus. Individual scores for adjective and noun as well as difference scores are given for prominence scores. Speech dynamicity values cover entire NPs.*

Development	Listener	Focus	Prominence score *M* (*SD*)			Speech dynamicity *M* (*SD*)
			Adjective	Noun	Difference	NP
Typical	Same	Adjective	5.21 (*1.00*)	1.32 (*1.44*)	3.89 (*2.23*)	3.28 (*.31*)
		Noun	2.42 (*1.96*)	4.30 (*1.83*)	1.88 (*3.63*)	3.17 (*.32*)
	Different	Adjective	4.79 (*1.43*)	1.67 (*1.63*)	3.13 (*2.84*)	3.18 (*.35*)
		Noun	2.94 (*1.84*)	3.71 (*1.92*)	0.76 (*3.57*)	3.25 (*.33*)
HFA	Same	Adjective	5.03 (*1.42*)	1.76 (*1.40*)	3.27 (*2.61*)	3.08 (*.35*)
		Noun	2.96 (*2.06*)	3.98 (*1.78*)	1.03 (*3.62*)	3.14 (*.30*)
	Different	Adjective	4.86 (*1.48*)	1.90 (*1.49*)	2.96 (*2.67*)	3.18 (*.36*)
		Noun	3.66 (*1.98*)	3.36 (*1.90*)	-0.30 (*3.55*)	3.16 (*.32*)

Table 2.5: *Results of all ANOVAs carried out with prominence difference scores and speech dynamicity values as dependent variables for both development groups taken together (TYP & HFA) and separately (TYP, HFA). Interaction effects not listed here were not significant on either of the dependent variables.*

	Prominence difference	Speech dynamicity
TYP & HFA		
Listener	$F(1,38) = 25.75$, $p < .001$, $\eta_p^2 = .40$	n.s.
Focus	$F(1,38) = 26.52$, $p < .001$, $\eta_p^2 = .41$	n.s.
Listener*Focus	n.s.	n.s.
Development	$F(1,38) = 2.96$, $p = .093$, $\eta_p^2 = .07$	$F(1,29) = 7.34$, $p < .05$, $\eta_p^2 = .20$
TYP		
Listener	$F(1,19) = 16.48$, $p < .001$, $\eta_p^2 = .46$	n.s.
Focus	$F(1,19) = 11.81$, $p < .01$, $\eta_p^2 = .38$	n.s.
Listener*Focus	n.s.	$F(1,29) = 8.72$, $p < .01$, $\eta_p^2 = .23$
Speaker	n.s.	$F(1,19) = 15.35$, $p < .05$, $\eta_p^2 = .99$
Speaker*Focus	$F(1,19) = 4.41$, $p < .01$, $\eta_p^2 = .82$	n.s.
HFA		
Listener	$F(1,19) = 10.05$, $p < .01$, $\eta_p^2 = .35$	$F(1,29) = 4.07$, $p = .053$, $\eta_p^2 = .12$
Focus	$F(1,19) = 14.72$, $p < .01$, $\eta_p^2 = .44$	n.s.
Listener*Focus	$F(1,19) = 5.11$, $p < .05$, $\eta_p^2 = .21$	n.s.
Diagnosis	n.s.	$F(1,29) = 5.36$, $p < .05$, $\eta_p^2 = .16$
Speaker	n.s.	$F(1,19) = 24.70$, $p < .01$, $\eta_p^2 = .99$
Speaker*Focus	$F(1,19) = 10.15$, $p < .001$, $\eta_p^2 = .91$	n.s.

2.3.2.2 Typically-developing speakers (TYP)

For the prominence difference scores, no negative means were found (Table 2.4), revealing that overall the focused word was perceived as more prominent than the unfocused word. For the factor listener, prominence difference scores were larger when the same listener was addressed (M = 2.89) than when a different listener was addressed (M = 1.95). Furthermore, the difference between the focused word and the unfocused word was larger when the focused word was the adjective (M = 3.51) than when the focused word was the noun (M = 1.33). How the prominence difference scores relate to the prominence scores of the adjective and noun individually becomes clear from inspection of the data in Table 2.4. These reveal that the focused word was less prominent and the unfocused word was more prominent when the listener was different than when the listener was the same, as shown by the two main effects. Prominence difference scores showed no effects of the random variable speaker. However, a significant interaction effect of the variables focus and speaker was found. With respect to speech dynamicity, no effects of listener or focus were found, although their interaction was significant in that addressing the same listener results in larger dynamicity scores for focused adjectives whereas addressing a different listener results in larger dynamicity scores for a focused nouns. The factor speaker was found significant for the speech dynamicity scores, indicating that there were significant differences among speakers concerning how dynamic their speech was perceived.

2.3.2.3 Speakers with autism (HFA)

Prominence difference scores in one condition showed a negative mean, which indicated that the focused word was not always perceived as the most prominent word (Table 2.4). In particular, this was the case when speakers addressed a different listener and the noun was focused. In this particular situation intonation experts perceived the adjective as more prominent than the noun. Regarding the factor listener, prominence difference scores were significantly larger when speakers addressed the same listener (M = 2.15) than when they addressed a different listener (M = 1.33). For focus, prominence difference scores were significantly larger when the adjective was focused (*M* = 3.11) than when the noun was focused (*M* = .36), see Table 2.4. Furthermore, the factors listener and focus interacted in that for focused adjectives difference scores were higher than for focused nouns and that this difference was even larger when the speaker was addressing a different listener. The prominence scores of the adjective and noun individually revealed that, generally, addressing a different listener resulted in less prominent focused words and more prominent unfocused words. Participants with Asperger showed smaller prominence difference scores (*M* = 1.61) than those with PDD-NOS (*M* = 1.80). The factor diagnosis was, however, not significant for the prominence difference scores. Results of the univariate ANOVA showed no main effects of the random variable speaker on the prominence difference scores. Interaction effects of the variables focus and speaker were found to be significant. For the speech dynamicity scores, a trend for the factor listener was found in that addressing the same listener (*M* = 3.11) results in lower scores than addressing a different listener (*M* = 3.20). No (interaction) effects for focus were found. Listeners perceived speakers with Asperger (*M* = 3.20) as significantly more dynamic than speakers with PDD-NOS (*M* = 3.10). With respect to individual differences, the factor speaker showed a main effect on the speech dynamicity scores, indicating that there were significant differences among speakers concerning how dynamic their speech was perceived.

2.3.3 Discussion

2.3.3.1 Prominence

The perception measures of prominence showed that contrastive intonation is perceived differently when the speaker had previously mentioned information that had not been shared with the listener, both for typically-developing speakers and for speakers with autism. That is, smaller prominence differences between focused and unfocused words were found when speakers address different listeners compared to when speakers utter both the alternative and the contrastive NP to the same listener. This indicates that all speakers, to some extent, accounted for the knowledge of their listener by producing contrastive intonation less clearly when it had no function for the listener. However, speakers did not fully abandon contrastive intonation when the listener did not hear the alternative NP. In particular, as we have seen, there still was a reduced form of contrastive intonation which can only be explained from the speaker's perspective. From the speaker's perspective there was always a contrast, even when talking to a different listener. This fact was reflected in the attenuated contrastive intonation pattern that speakers produce in such a situation. Here, attenuation of contrastive intonation is defined by both the decreased prominence of focused words and the increased prominence of unfocused words. For the nouns, this could mean that speakers produced a default intonation pattern with a standard accent on the noun. However, it does not explain the results of the adjectives, which were still higher in F0 and perceived as more prominent than the noun when a different listener is addressed. Therefore, it is more likely that speakers produced a reduced but functional contrastive intonation pattern when addressing a different listener.

Concerning the effect of focus, the present results parallel with previous findings in the literature. In particular, this holds for the unfocused words in the current experimental setup. Result of the unfocused words, which represented given information for the speaker, are compatible with Galati and Brennan (2010) and Gregory et al. (2001). Like the present results, those studies found that speakers reduce given information to a larger extent when addressing the same listener than when addressing a different listener. Furthermore, the effect of focus indicates that speakers produced smaller prominence differences for a contrastive intonation pattern that had the noun in focus than for one that had the adjective in focus. This is in line with Krahmer and Swerts (2001) who found that accents in a non-default position (adjective) were perceived as more prominent than accents in a default position (noun). The prominence scores in the current study were obtained via a rating task and depend on human judgments only. Nevertheless, prominence judgments were acoustically sound, as shown by correlating F0 measures[4].

It is interesting that speakers with autism, like the typically-developing speakers, produced some form of focus marking when addressing a different listener. We interpret this as a speaker-related result, because accounting for the informational needs of the listener would have resulted in not marking a contrast prosodically as the listener did not hear the previous utterance. The fact that HFA speakers did mark a contrast to some extent suggests that they took into account their own perspective. Concerning listener-factors, results showed that HFA speakers produce contrastive intonation more clearly when addressing the same listener. Such a result is not compatible with an impaired ToM that predicts HFA speakers to produce contrastive intonation patterns on the basis of their own perspective. In this experiment that would have been to produce contrastive intonation irrespective of whether the previous utterance was shared with the same or a different listener. It has to be noted that participants in our study were high functioning and their results may therefore not easily generalize to all speakers with autism. Furthermore, an unexpected result is that the unfocused adjective was perceived as more prominent when the noun is focused. This is the case for HFA speakers, however only when they addressed a different listener, and seems to indicate that focus was marked incorrectly. Focus was marked correctly when they addressed the same listener. If HFA speakers would have made accent placement errors in general, these errors should also have appeared when addressing the same listener. Results showed a more prominent adjective when the noun was focused, but only when HFA speakers addressed a different listener. We therefore see this result not as an indication of accent placement errors, but rather as a side-effect of the variable *listener*. Also, the correlation between F0 and prominence showed that the adjective is more emphasized than the noun in the case where a speaker addressed a different listener and where the noun was focused.

The investigations of individual differences among participants also showed similarities between the typically-developing speakers and the speakers with autism. That is, in both groups participants generally behaved the same with respect to produced F0 and perceived prominence. Interaction effects were also found in both groups, indicating that some participants showed larger prominence or F0 differences to express a contrastive intonation than others. It has to be noted that these effects were found for both typically-developing speakers and speakers with autism. In particular, both development groups showed that the way focus was marked varied among speakers. As there was no interaction with the factor listener, individual differences between participants are not to be related to perspective taking. This holds for both the typically-developing speakers and for speakers with autism, indicating that these groups are similar with respect to individual variation in the production of contrastive intonation.

To conclude, both typically-developing speakers and speakers with autism produced contrastive intonation less clearly when they addressed a different listener. Thus, both groups showed evidence for taking into account both speaker- and listener-perspectives. Furthermore, speakers with autism produced smaller prominence differences between accented and deaccented words compared to typically-developing speakers, although this effect was only marginally significant. The prominence results provide evidence only for subtle differences between the two speaker groups. That is, speakers with autism had a tendency to produce the adjective more prominently when addressing a different listener, even when the noun was focused. Typically-developing speakers did not show this tendency.

2.3.3.2 Speech dynamicity

Results indicated that speech dynamicity was generally not influenced by whether the speaker addressed the same or a different listener or by whether the adjective or the noun was in focus. Two exceptions to this generalization need to be reported. First, typically-developing speakers sounded more dynamic when addressing the same listener and focusing on the adjective compared to when addressing a different listener and focusing on the noun. Possibly, this interaction effect relates to the prominence difference scores, which were generally smaller when addressing a different listener. An interaction effect between listener and focus in the same direction as the speech dynamicity scores was also found for typically-developing speakers' F0 differences. However, it remains an open question as to why this interaction effect was not found for speech dynamicity scores of speakers with autism. Second, the perceived speech dynamicity of speakers with autism was lower when addressing the same listener compared to when addressing a different listener. Initially, this result seems counter to what was observed for the F0 and prominence difference scores, which were both smaller when addressing a different listener. One would expect to perceive speech as more dynamic when differences between accented and deaccented words are larger. However, an explanation of the effect of listener on speech dynamicity of speakers with autism may be found in the fact that they produced the adjective more prominently, even when the noun was in focus. This exceptional way of prominence marking happened only when addressing a different listener and resulted in more NPs in which the adjective was more prominent than the noun. As this pattern is counter to the default intonation pattern on Dutch NPs, it may have given listeners a more dynamic impression of the speech.

Concerning the difference between development groups, typically-developing speakers used a larger pitch range and were perceived as more dynamic than speakers with autism. These findings appear to be related as shown by correlation measures. The present findings offer no evidence for the earlier claims that speakers with autism use a larger pitch range than typically-developing speakers, though they are reminiscent of studies that report that speakers with autism speak more monotonous than typically-developing speakers (Von Benda, 1983). The absence of a large pitch range effect in the data of the speakers with autism cannot be explained by the size of our sample (cf. Bonneh et al., 2011) nor by the way in which we measured pitch range (cf. Diehl et al., 2009). That is, as in our study, Bonneh et al. (2011) used relatively small speech samples and Diehl et al. (2009) used a similar way of measuring pitch range. Those similarities can therefore not explain why our study finds contradicting results on pitch range. However, there is additional evidence that the results in this study are not simply an artifact of the methodology used. In particular, differences in speech dynamicity showed an effect of diagnosis for speakers with autism in that speakers with Asperger were perceived as more dynamic than speakers with PDD-NOS, which is compatible with the findings on F0 differences and pitch range.

To conclude, the present results showed that, besides individual differences in both development groups, typically-developing speakers and speakers with autism used prosodic structures which were formally distinct, in terms of variation in pitch range and speech dynamicity.

2.4 General discussion and conclusions

The production experiment showed that typically-developing speakers produced contrastive intonation less clearly when their listener did not hear the previously mentioned utterance. We take this result as evidence that contrastive intonation is both speaker- and listener-driven. We found similar results for speakers with autism, contrary to our expectations. In a prosodic analysis of pitch range and speech dynamicity we found differences between typically-developing speakers and speakers with autism at the level of prosodic form, as in Rutherford et al. (2002) and Chevallier et al. (2011). In sum, we showed that contrastive intonation was produced by taking into account the perspective of the listener, both by typically-developing speakers and speakers with autism. This conclusion is in line with theories about the production of prosody that incorporate the listener's perspective (Clark & Murphy, 1982; Galati & Brennan, 2010). The outcomes also refine previous work on contrastive intonation (Chafe, 1976; Pechmann, 1984a, 1984b; Braun & Tagliapietra, 2010). That is, speakers produced this pattern by blending both their own and their listener's perspective. In the current study, the contrastive intonation pattern that speakers produced to a different listener was not identical to the pattern they produced to the same listener. When addressing a different listener, speakers used an attenuated contrastive intonation pattern. This means that speakers can mark a semantic contrast in their prosody in an attenuated way. This could be explained if we assume that speakers indeed took into account both their own and their listener's need in this situation. In particular, there was no need to mark a contrast prosodically for their listener. Thus, the attenuation of contrastive intonation stops at a point that the intonation pattern still satisfies the need for contrast marking for the speaker. To conclude, cognitive mechanisms behind the production of contrastive intonation keep track of the informational needs of both interlocutors in a dialogue.

It has to be noted that the current study explicitly distinguishes prosodic function from prosodic form. This study has shown that this distinction is crucial when it comes to an investigation of prosody in autism. Taking into account the perspective of the listener in this study concerns accounting for the informational needs of the listener. This is most probably signaled by the functional use of prosody, which closely relates to the semantics of an utterance. We did not find evidence that speakers with autism behave differently from typically-developing speakers concerning perspective taking in functional prosody. This is not in line with Baron-Cohen (1995) and Shriberg et al. (2001). If the ToM of speakers with autism were impaired and there were resulting prosodic difficulties, this study could have plausibly found that speakers with autism behaved similarly when addressing the same or a different listener. However, in the current study we were not able to find evidence that an impaired ToM was related to the functional use of prosody in autism. Again, a potential factor explaining the contradictory findings is the participant population in this study. That is, we only tested speakers with high functioning autism, who may have been more similar to typically-developing speakers compared to speakers with autism in general. With respect to the form of prosody, this study found that typically-developing speakers were different from speakers with autism. Speakers with autism were perceived as more monotonous than typically-developing speakers, as reflected in the pitch range these groups produced. This result is in line with acoustic impressions by Von Benda (1983). It remains an open question why other work on pitch range in autism found effects in the opposite direction (i.e., Nadig & Shaw, 2011).

Our study thus shows the importance of distinguishing prosodic form from prosodic function. Prosodic differences between typically-developing speakers and speakers with autism found in this study relate to form rather than function. However, it is plausible that the two levels interact, as they are both reflected by F0 measurements. In our results, contrastive intonation patterns showed a statistical trend in that they exhibited smaller prominence differences for speakers with autism than for typically-developing speakers. The analysis of speech dynamicity and pitch range explains this difference. In particular, speakers with autism used a smaller pitch range and sounded less dynamic than typically-developing speakers. This corresponds with smaller prominence differences between focused and unfocused words. Thus, the statistical trend at the level of prosodic function is plausibly a reflection of significant differences at the level of prosodic form.

Why would speakers with autism in the current study be able to account for their listener when producing contrastive intonation? An important explanation may be found in the distinction between function and form in prosody. The few studies that directly investigated the relation between perspective taking and prosody in autism focused on the recognition of emotions or mental states (Rutherford et al., 2002; Chevallier et al., 2011). To recognize a happy speaker, for example, it makes no sense to attend to the functional use of prosody, such as the pitch accents that mark important information. One rather has to attend the formal aspects of prosody, such as how dynamic a speaker sounds. If formal aspects are indeed the core prosodic deficit in autism, it would explain why Rutherford et al. (2002) and Chevallier et al. (2011) found that their participants with autism had difficulties recognizing emotions and mental states. Such an explanation is compatible with the current differences between typically-developing speakers and speakers with autism on pitch range and speech dynamicity. Furthermore, the fact that speakers with autism in the present study were found to account for their listeners may be explained by the methodological approach taken in this study. That is, it was relatively easy for participants in both production experiments to account for the listener. Accounting for the listener in this study was defined as knowing whether that listener heard certain information or not. As Galati and Brennan (2010) argued, such a mechanism can easily be accounted for with a one-bit model. Therefore, it is plausible that speakers with autism have no difficulties in accounting for the listener in this setting. Thus, the one-bit model (Galati & Brennan, 2010) could be seen as the most minimal form of ToM. A more challenging setting would have been when speakers did not know whether the listener they addressed knows certain information or not. Although it is difficult to find a suitable experimental setup for this manipulation, it would have been a more demanding test for perspective taking abilities of the participant. It could be expected that typical speakers are more likely to choose an intonation pattern that by default takes the listener into account to avoid miscommunication (i.e., the listener did not hear anything) compared to speakers with autism. A follow-up study could use such a paradigm.

In the current study we did not find evidence for general accent placement errors in speakers with autism. We did find that the adjective was erroneously more prominent than the noun when speakers with autism addressed a different listener. The fact that speakers with autism did not show this tendency when addressing the same listener suggests that an explanation could be found in impaired perspective-taking, rather than accentuation errors in general. If that explanation holds, there is reason to relate the functional use of prosody to impaired perspective-taking. This relation remains to be investigated. Most importantly, it is unclear why the adjective showed erroneous prominence patterns and not the noun, both in Peppé et al. (2007) and in the current study. With respect to this issue three factors should be taken into account. First, the adjective is the first word in an NP in most Germanic languages, but not (necessarily) in Romance languages. Therefore, the present data for Dutch do not allow to conclude that the first word in an NP is problematic nor that the word class of adjectives is problematic in autism. Second, the asymmetry between adjective and noun in the current study was to some extent related to the boundary tones produced on the noun. The results have shown that nouns with a high boundary were perceived as more prominent. Although boundary tones were not of primary interest in this study, it is important to take into account effects of other aspects of prosody when considering the current results. Third, the default position of an accent in an NP in Germanic languages is the noun. It could be that accents in a non-default position (adjective) are problematic for people with autism. Other languages use different intonational grammars and may therefore mark semantic contrasts in a different way in their intonation. A follow-up study could test languages from different families to investigate whether speakers with autism show accentuation errors and whether these errors consistently relate to phrase position, word class, intonational grammar or a combination of those factors. Most crucially, future research should take into account to what extent difficulties in the functional use of prosody are the default in autism or are a result of difficulties at the level of prosodic form. Results of the current study suggest an explanation on the basis of the latter.

Endnotes Chapter 2

[1] In the current study DSM-IV (American Psychiatric Association, 2000) was followed and distinguishes between Asperger and PDD-NOS. After this study had been carried out, DSM-V (American Psychiatric Association, 2013) appeared in which this distinction is no longer maintained.

[2] To investigate the effect of the high boundary tone a multivariate analysis of variance with F0 maxima of the noun as dependent variables and listener (2 levels: same, different), focus (adjective, noun) and boundary tone (2 levels: high, low) as independent variables was carried out for the typically-developing speakers and speakers with autism separately. This analysis showed no effects of listener. The variable focus showed an effect only for the typically-developing speakers in that F0 maxima were higher when the noun was in focus (M_{TYP} = 256.82) compared to when the adjective was in focus (M_{TYP} = 236.88): [$F(1,472) = 11.45$, $p < .01$, $\eta_p^2 = .02$]. Boundary tone had no effect for the typically-developing speakers, but did have an effect for the speakers with autism in that F0 maxima were higher when the noun ended with a high boundary tone (M_{HFA} = 224.23) compared to when the noun ended with a low boundary tone (M_{HFA} = 195.15): [$F(1,472) = 8.05$, $p < .01$, $\eta_p^2 = .02$]. No interaction effects were found. These results indicated that the presence of a boundary tone on the noun did affect the maximum pitch of the noun in speakers with autism and not in typically-developing speakers.

[3] To investigate the effect of the high boundary tone a multivariate analysis of variance with prominence scores of the noun as dependent variables and listener (2 levels: same, different), focus (adjective, noun) and boundary tone (2 levels: high, low) as independent variables was carried out for the typically-developing speakers and speakers with autism separately. This analysis showed no effects of listener. The variable focus showed an effect for the typically-developing speakers scores [$F(1,472) = 150.42$, $p < .001$, $\eta_p^2 = .24$] and for the speakers with autism [$F(1,472) = 40.83$, $p < .001$, $\eta_p^2 = .08$] in that prominence scores were higher when the noun was in focus (M_{TYP} = 4.25, M_{HFA} = 4.31) compared to when the adjective was in focus (M_{TYP} = 1.93, M_{HFA} = 2.45). Boundary tone had an effect for the typically-developing speakers [$F(1,472) = 38.06$, $p < .001$, $\eta_p^2 = .08$] and for the speakers with autism [$F(1,472) = 30.00$, $p < .001$, $\eta_p^2 = .06$] in that those were higher when the noun ended with a high boundary tone (M_{TYP} = 3.67, M_{HFA} = 4.81) compared to when the noun ended with a low

boundary tone (M_{TYP} = 2.51, M_{HFA} = 2.59). No interaction effects were found. These results indicated that the presence of a boundary tone on the noun did affect the perceived prominence of the noun in speakers from both development groups.

[4] It has to be noted that despite their correlation, F0 measures differ from the prominence scores in several ways. First, F0 measures do not show main effects of listener or focus, whereas prominence scores do. Furthermore, F0 differences are not always smaller when speakers address a different listener. This can be seen in the case where the noun is focused (Table 2.3, Table 2.4). The differences between produced F0 and perceived prominence may be explained by the fact that F0 values are more variable than prominence scores (see standard deviations in Table 2.3, Table 2.4), due to individual differences in the voices of participants. Second, an F0 difference score is susceptible to side-effects in pitch like the declination effect (Breckenridge, 1977). That is, F0 measured at the beginning of an utterance is generally higher than the F0 at the end of an utterance. Listeners account for this declination when perceiving prominence by taking into account the prosodic context of an F0 movement (Gussenhoven et al., 1997). This explains why large differences in F0 do not always result in large differences in perceived prominence. Therefore, we may see a correlation between the perceived prominence and F0 at the individual word level, but not see similar main effects at the noun phrase level. Third, F0 measurements show significant differences between typically-developing speakers and speakers with autism in that the latter produce smaller F0 differences than the former. This is in line with the trend observed for the prominence scores and compatible with the difference in pitch range between the development groups. It is plausible that the smaller pitch range used by speakers with autism results in smaller prominence and F0 differences when compared to typically-developing speakers.

CHAPTER 3

WHITE BEAR EFFECTS IN LANGUAGE PRODUCTION: EVIDENCE FROM THE PROSODIC REALISATION OF ADJECTIVES

ABSTRACT

A central problem in recent research on speech production concerns the question to what extent speakers adapt their linguistic expressions to the needs of their addressees. It is claimed that speakers sometimes leak information about objects that are only visible for them and not for their listeners. Previous research only takes the occurrence of adjectives as evidence for the leakage of privileged information. The present study hypothesizes that leaked information is also encoded in the prosody of those adjectives. A production experiment elicited adjectives that leak information and adjectives that do not leak information. An acoustic analysis and prominence rating task showed that adjectives that leak information were uttered with a higher pitch and perceived as more prominent compared to adjectives that do not leak information. Furthermore, a guessing task suggested that the adjectives' prosody relates to how listeners infer possible privileged information.

This chapter is based on:

Kaland, C. C. L., Krahmer, E. J., & Swerts, M. G. J. (to appear). White bear effects in language production*:* evidence from the prosodic realisation of adjectives. *Language and Speech, 58*(1).

Kaland, C. C. L., Krahmer, E. J., & Swerts, M. G. J. (2011). Salient in the mind, salient in prosody. *Proceedings 33rd Annual Conference of the Cognitive Science Society* Boston (MA), USA, 261-266.

3.1 Introduction

Consider a situation in which a girl and boy are looking out of a window. They both see a street with cars passing by. The girl describes what is happening. She could say: “I like that big car!” In this example, the girl may use the adjective *big* either in a restrictive way to distinguish the car under description from another smaller car or in a non-restrictive way to express just the size of the car (Haeseryn et al., 1997). For the boy who sees the same street and cars *big* may help to identify the car that the girl describes, no matter whether the adjective was used in a restrictive or non-restrictive way. Now, consider a situation in which the girl and the boy do not have the same view, as often occurs in ordinary language use. That is, the girl looks from a specific angle at the street and she sees both a big and small car. However, from where the boy is standing, only the big car is visible. When the girl now says: “I like that big car!”, *big* provides information that the boy does not necessarily need to identify the car, since he sees no smaller car. In principle, the boy may therefore interpret the adjective in a non-restrictive sense, since he only has a partial view on the complete visual scene that the girl is looking at. Interestingly, however, it is claimed that adjectives in these situations can leak information about the speakers’ perspective, that is, they can provide cues to the listener about information that is only accessible to the speaker (Wardlow Lane et al., 2006). In the example above, *big* would then be argued to reveal to the boy that the girl sees another smaller car.

The present research sheds more light on this claim. So far, research provides evidence for information leaking on the basis of the lexical use of adjectives. The basic question in these studies is whether a speaker utters an adjective or not. Whenever a speaker utters an adjective (in a restrictive way) this is taken as evidence for the leaking of information. The aim of the present study is to investigate how these adjectives are realised prosodically. We hypothesize that the prosody of adjectives that leak information signals their information status. Therefore, we compare adjectives taken from a setting where speaker and addressee have the same perspective (i.e., adjectives that do not leak information) with adjectives taken from a setting where speaker and addressee have different perspectives (i.e., adjectives that potentially leak information). Furthermore, we conduct a perception test in order to investigate whether listeners can hear that an adjective provides leaked information.

These questions are in line with a recent line of research on speech production concerned with the question to what extent speakers adapt their linguistic expressions to the needs of their addressees. In the literature, one can distinguish different viewpoints on this issue, with some work claiming that speakers' utterances are initially designed egocentrically, that is, without taking the addressee into account (Brown & Dell, 1987; Dell & Brown, 1991; Horton & Keysar, 1996; Keysar, Barr, & Horton, 1998; Pickering & Garrod, 2004; Barr & Keysar, 2007), others maintaining that speakers explicitly design utterances for specific audiences (Clark & Murphy, 1982; Clark & Wilkes-Gibbs, 1986; Clark & Brennan, 1991; Nadig & Sedivy, 2002) and still others arguing that utterances are the result of both speaker- and addressee-related features (Brennan & Hanna, 2009; Galati & Brennan, 2010). When referring to objects in the world, it has repeatedly been shown that speakers may give more information than is necessary from the perspective of an addressee (Engelhardt, Bailey, & Ferreira, 2006). This violates a strict interpretation of the Gricean maxim of quantity (Grice, 1975; but see Davies & Katsos, 2013, for discussion). Along these lines, work by Pechmann (1984a, 1984b) shows that speakers may overspecify their references by mentioning properties of objects that are redundant from the perspective of the addressee. A theory of incremental language production potentially explains why speakers overspecify. This theory claims that a speaker can start articulating without being aware of the whole visual context that is under discussion (Fry, 1969; Levelt, 1989; Pechmann, 1989). That is, speakers often do not have a full visual representation of all the objects and their features before they start describing one of them. Therefore, speakers may not know which features are relevant for their addressee when they formulate their description.

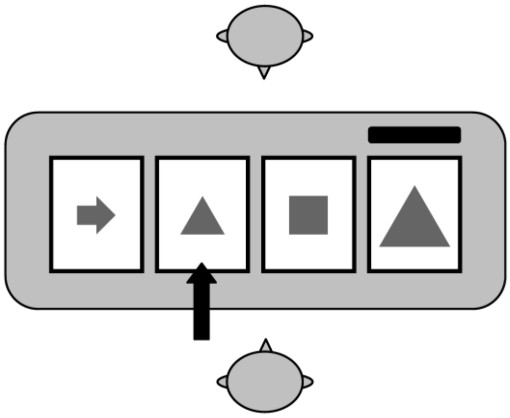

Figure 3.1: *Experimental setup following Wardlow Lane et al. (2006) with the speaker (bottom) and addressee (top) sitting at opposite sides of a table. The arrow indicates the target object, the bar indicates the occluder (present in the privileged and conceal setting, not in the shared setting).*

The extent to which referring expressions actually show whether speakers take into account the perspective of their addressee is often investigated using specific communication tasks. In these communication tasks speakers were instructed to refer to figures of geometrical objects. Some figures were shared by all interlocutors (common ground; Stalnaker, 1978; Clark, 1996), while other figures were only available for the speaker (privileged ground). The basic question addressed in these tasks was whether speakers in their referring expressions take into account what their addressees know, whether they only rely on information in common ground, or also use information from privileged ground (Horton & Keysar, 1996; Keysar et al., 1998; Nadig & Sedivy, 2002; Keysar, Lin, & Barr, 2003; Wardlow Lane, Groisman, & Ferreira, 2006). A specific series of studies investigated explicit descriptions of figures in common ground that were claimed to implicitly leak information about figures in privileged ground (Wardlow Lane et al., 2006; Wardlow Lane & Ferreira, 2008; and recently Wardlow Lane & Liersch, 2012). Along the lines of the ‘car’ example sketched above, these studies showed that to describe a mutually visible figure, speakers may produce “the small triangle” even when a big triangle was occluded for the addressee (Figure 3.1). In this situation *small* provided the addressee with more information than needed to identify the figure (e.g., "the triangle" would have been sufficient). Interestingly, when speakers were explicitly instructed not to give information about the occluded figure, the target figure was even more often described with an adjective that referred to the contrast between the target and the occluded figure.

Wardlow Lane et al. (2006) explained this finding according to the theory of ironic processes (Wegner, Schneider, Carter, & White, 1987; Wegner, 1994). In a series of experiments Wegner et al. (1987) asked participants to report their conscious thoughts during two periods of five minutes. In one condition participants were asked not to think about a white bear in the first five minutes, whereas in the other condition participants were told that they were allowed to think about a white bear. The amount of times participants reported to have thought about a white bear in the second period of five minutes was measured. It was found that participants thought significantly more about white bears when they first had to suppress the thought compared to when they were allowed to have that thought (Wegner et al., 1987). The theory of ironic process accounts for these findings and distinguishes two cognitive processes: an operator process which is responsible for running actions and a monitor process which is in constant search for failures of the first. Instructions of the type "do not..." cause unsuccessful scenarios to be more salient for the speaker. As a consequence, the monitor process is triggered to check for this specific failure, causing interference with the operator process. As a result, people exhibit exactly the behaviour they are instructed to avoid. In this way, when told not to think about a white bear, a mental representation of such a bear automatically comes up (original example from Dostoyevsky and/or Tolstoy, see Wegner & Schneider, 2003). Along these lines, Wardlow Lane et al. (2006) show that ironic processes affect the language production process up to the grammatical encoding stage in Levelt's model (1989, 1999). As a result of the boosted salience of a contrast relation between figures in common ground and privileged ground, adjectives referring to this contrast are more likely to be included in the grammatical structure. Moreover, as a consequence of ironic processes the speaker has more difficulties formulating expressions that provide just enough information from the perspective of the addressee.

Studies investigating privileged information commonly focus on lexical production data only. More specifically, previous work is usually limited to counting the relative frequency with which speakers incorporate an adjective in their references (Horton & Keysar, 1996; Wardlow Lane et al., 2006). A strict definition of information leaking suggests that listeners hearing an utterance with an adjective that leaks information are able to deduce that leaked information. Thus, by listening to utterances that leak information listeners should be able to restrict the set of possible objects in privileged ground. To continue the ‘car’ example sketched above, the main question is whether “the big car” uttered in the situation where the interlocutors have different views on the street indeed reveals to the boy that the girl sees another smaller car. There is at least one complicating factor. In particular, adjectives used in situations where they potentially leak information have an equivocal information status. On the one hand they provide explicit information about the object under description and refer to properties that belong to objects in common ground. This would be non-restrictive use of the adjective as it does not provide information about objects in privileged ground. On the other hand adjectives that leak information may provide implicit information about other objects visible for the speaker, which would be restrictive use of the adjective. That is, by uttering *big* a speaker could implicitly acknowledge the existence of a smaller object. Hence, the leaking of information depends entirely on a contrast between two objects. This contrast is crucially only available from the speaker’s perspective and not from the listener’s perspective. Thus, just by the verbal interpretation of the adjective the listener cannot tell whether the adjective is an implicit reference to privileged information. Therefore, the question remains how listeners interpret this equivocal status of adjectives in circumstances where those adjectives presumably leak information.

The present study argues that non-verbal aspects of those adjectives, in particular prosody, may reveal to listeners that a speaker leaks information. At the sentence level, for example, it is known that prosodic cues distinguish restrictive from non-restrictive relative clauses (Kaland & Van Heuven, 2010). At the level of NPs, adjectives may provide contrastive information, which is often prosodically marked by means of a pitch accent in Germanic languages like Dutch, German and English (Pechmann, 1984a, 1984b; Pierrehumbert & Hirschberg, 1990; Krahmer & Swerts, 2001; Calhoun, 2009). Pitch accented words in these languages are acoustically realised with an increased pitch and longer duration and perceived as more prominent compared to unaccented words (e.g., Gussenhoven et al., 1997; Streefkerk, 2002; Ladd, 2008). Thus, when speakers indeed prosodically mark adjectives that leak information, this could be a cue for listeners to interpret those adjectives as referring implicitly to privileged information. For example, when hearing "the big triangle" with *big* being acoustically prominent the addressee may know that the speaker sees another smaller triangle as well.

No previous study investigated how adjectives that leak information are realised prosodically and whether listeners are able to use prosodic cues to infer leaked information. An investigation of these issues could shed more light on the production of speech prosody and to which extent speakers use it to account for their listener's perspective. This is taken as the aim of the current study. To this end we investigate the speaker's production of adjectives by means of both production and perception measures. We hypothesize that ironic processes (Wegner et al., 1987) influence the articulatory stage of language production in Levelt's model (Levelt, 1989 p. 9; 1999 p. 87). It is known that when a contrast is less salient for speakers, such as when two figures in common ground are minimally distinguishable and when speakers do not receive additional "do not…" instructions, speakers generally do not mark that contrast prosodically (Pechmann, 1984a, 1984b, 1989). However, we hypothesize that after a "do not…" instruction the contrast between a figure in common ground and a figure in privileged ground may be extra salient for a speaker. This may cause the speaker to mark that contrast prosodically. In languages like Dutch this would be realised by making the word that refers to the contrastive information more prominent (Krahmer & Swerts, 2001). Therefore, speakers may produce adjectives that leak information with different prosodic marking compared to non-leaking adjectives, and this difference might be perceivable and interpreted by listeners. In order to investigate this, we collect data from speakers that produced adjectives that potentially leak and adjectives that do not leak information. Adjectives that leak privileged information are elicited in a communicative setting similar to Wardlow Lane et al. (2006) in that there is an occluded figure that speakers have to ignore. In another setting, not used in Wardlow Lane et al. (2006), all figures are visible for both interlocutors such that uttered adjectives never provide leaked information. Produced adjectives are prosodically analysed by acoustic measures of pitch (F0) and by perceptual measures of prominence. In a subsequent perception experiment listeners are asked to guess the privileged information on the basis of the produced adjectives.

3.2 Production data

3.2.1 Participants

A total of 42 different participants acted as speaker (31 women, 11 men, M_{age} = 21.3 years, age range: 18-29 years). Among them, 29 participants acted also as addressee, of which 17 acted as speaker first and 12 acted as addressee first. All participants were native speakers of Dutch and students at Tilburg University who took part as a course requirement.

3.2.2 Design

Utterances were collected using a variant of the paradigm of Wardlow Lane et al. (2006) where speakers were asked to describe figures for their addressees (Figure 3.1). Wardlow Lane et al. (2006) elicited adjectives in two communicative settings (*baseline* and *conceal*). In those settings, there was always an occluded figure. In the baseline setting participants received no additional instructions related to the occluded figure. In the conceal setting participants were explicitly instructed not to give information about the occluded figure. Both Wardlow Lane et al.'s (2006) settings were replicated in the current study (the baseline setting, which we call the *privileged setting*, and the *conceal setting*). As a control condition, the current setup added a third communicative setting, which we call the *shared setting*, in which all four figures were mutually visible for the participants. In this way, it was made sure that adjectives elicited in this setting do not leak privileged information. The participants' task was to describe a mutually visible figure in such a way that the addressee could indicate the intended one out of a set of four figures (for details on the procedure of this task see Wardlow Lane et al., 2006).

3.2.3 Materials and procedure

In total, the experiment consisted of 48 stimuli (16 per communicative setting). The setup ensured that adjectives refer to a contrast between the mutually visible target figure and another figure in the set. This other figure was either mutually visible (*shared*) or visible for the speaker only (*privileged*, *conceal*). The contrast concerned either the size or the color of the figures. That is, in one half of the stimuli the target figure was smaller/bigger than the other figure, whereas in the other half of the stimuli the target figure was colored differently from the other figure. The communicative settings were presented in three different orders following a reduced latin-square design. Participants were randomly assigned to one of these orders in such a way that an equal number of participants per order was maintained. The visual displays of the target figures were identical for each communicative settting. Each stimulus was prevented from contrasting minimally with the previous stimulus, which could have evoked prosodic marking (Pechmann, 1984a, 1984b). Therefore, two successive targets were never similarly shaped. In this way the adjective never provided contrastive information with respect to the previously described target figure. Participants were instructed about the course of the experiment before the start of the experiment. No instructions were given about the phrasal construction that participants could use to describe the target figure. They played a training round before the actual stimuli were presented. One trial consisted of four figures, which were placed on a table in front of the speaker. Two figures of each trial exhibited a minimal contrast in terms of size or color. A Marantz M-600 solid state recorder recorded the speaker's utterances digitally as a wave-file.

3.2.4 Statistical analysis

Elicited noun phrases (NPs) contained either zero, one or two adjectives. We counted only the adjectives that matched the contrast (N = 1198). Thus, color adjectives uttered when the contrast concerned size and size adjectives uttered when the contrast concerned color were not taken into account (N_{shared} = 190, $N_{\text{privileged}}$ = 363, N_{conceal} = 265). This method was used because speakers were free with respect to the verbal construction of their utterances, which resulted in different numbers of adjectives produced in the communicative settings. In addition, this method excludes adjectives that do not leak any potential information about another object. The relatively high proportion of overspecification compared to Wardlow Lane et al. (2006) may be explained by the existence of a contrast between figures in each trial. This could have stimulated the use of adjectives in each condition. Adjective use was calculated as a proportion. That is, for each case where an adjective was uttered a score of 1 was counted, whereas for each case where no adjective was uttered a score of 0 was counted. Proportions reported here are averages of these values. Repeated measures Analyses of Variance (ANOVAs) were performed on logit transformed proportion measures of adjective use as dependent variable and with communicative setting (3 levels: shared, privileged, conceal) as within subject factor and with order of communicative setting (3 levels) as between subjects factor. Pairwise comparisons (Bonferroni corrected) were performed between each level of communicative setting. We report an analysis by subjects (F1), by items (F2) and by both subjects and items simultaneously (minF'), see Clark (1973). For the sake of readability, means are reported for untransformed proportions of adjective use.

3.2.5 Results

Results show that communicative setting has a main effect on the frequency with which speakers use adjectives: $F_1(2, 78) = 16.05$, $p < .001$, $\eta_p^2 = .29$, $F_2(2, 30) = 14.89$, $p < .001$, $\eta_p^2 = .50$, $minF'(2, 84) = 7.72$, $p < .001$. Speakers most often use adjectives in the shared setting (M = .72) followed by the conceal setting (M = .61) and the privileged setting (M = .46), see Table 3.1. Bonferroni corrected pairwise comparisons show that the privileged setting differs significantly from both the shared and the conceal setting (F_1: $p_{shared\text{-}privileged} < .001$, $p_{privileged\text{-}conceal} < .05$, F_2: $p_{shared\text{-}privileged} < .01$, $p_{privileged\text{-}conceal} < .001$), while the shared and conceal setting differ only marginally: F_1: $p = .06$, F_2: $p = .05$. The difference between the privileged and conceal setting confirms results by Wardlow Lane et al. (2006). The effect of order was found significant only in the F2 analysis: $F_1(2, 39) = 1.02$, *n.s.*, $F_2(2, 30) = 15.76$, $p < .001$, $\eta_p^2 = .51$, $minF'(2, 44) = .96$, *n.s.* The F2 analysis further shows an interaction effect between order and setting: $F_2(4, 60) = 5.58$, $p < .01$, $\eta_p^2 = .27$. Pairwise comparisons (Bonferroni corrected) in the F2 analysis show that participants used significantly more adjectives when the order of communicative settings was conceal-shared-priviliged ($M = .66$) compared to when the order was privileged-conceal-shared ($M = .51$). This may be explained when considering that adjective use was influenced by the previous communicative setting. In particular, when in the previous communicative setting participants were stimulated to use adjectives (e.g., shared, conceal), this could have stimulated the use of adjectives in following settings even when there was no need to use adjectives in that setting (i.e., privileged). This explanation is supported by the interaction effect between order and setting, which shows that the fewest adjectives were used in the privileged setting compared to both the shared and the conceal settings and that this effect was even stronger when the privileged setting was the first in the order of settings.

Table 3.1: *Mean adjective frequency (untransformed) and standard deviation as a function of communicative setting.*

Setting	*M*	*SD*
Shared	.72	.22
Privileged	.46	.42
Conceal	.61	.39

3.2.6 Conclusions

The production data show that results of Wardlow Lane et al. (2006) are successfully replicated in that speakers use significantly more adjectives to refer to privileged information when instructed not to do so. In general, our study shows a higher overall rate of adjectives (*M* = .59) than the study of Wardlow Lane et al. (2006) (*M* = .1). This difference could be the result of an additional color contrast in our study, whereas in Wardlow Lane et al. (2006) only size contrasts are tested. As shown by Koolen et al. (2013), more variation in terms of object features leads to more overspecification, and our results are consistent with this. Adjectives collected in the shared and the conceal setting are analysed further in the next sections in order to compare adjectives that leak information (conceal) with adjectives that do not leak information (shared).

3.3 Acoustic analysis

The acoustic analysis investigates whether speakers utter adjectives that leak information prosodically in a more prominent fashion than adjectives that do not leak information. It is known that Dutch NPs with a neutral intonation pattern tend to exhibit a pitch accent on the noun whereas NPs with contrastively focused adjectives usually exhibit a pitch accent on the adjective and a deaccented noun (Krahmer & Swerts, 2001). Pitch accented words are acoustically prominent, as can be measured by their F0 (e.g., Streefkerk, 2002). These relations are explored in the current acoustic analysis.

3.3.1 Method

A subset of adjectives collected in the shared and conceal setting and produced by 33 different speakers were acoustically analysed (N_{shared} = 141, N_{conceal} = 80). Adjectives in this subset met a number of criteria that made them suitable for analysis of pitch (F0). That is, analysis was done on NPs that consisted of one adjective and one noun, so that the NPs were comparable in word length. Adjectives included were always bisyllabic, that is *kleine* (small), *grote* (big), *groene* (green) and *grijze* (gray) while NPs with monosyllabic adjectival counterparts such as *klein* (small), *groot* (big), *groen* (green) or *grijs* (grey) were excluded from analysis. Nouns were either monosyllabic (*ster*, star) or bisyllabic (*driehoek,* triangle). Furthermore, utterances including fillers such as “uhhmm”, hesitations, atypical word order or a wrong target word were excluded as well (N_{shared} = 18, $N_{conceal}$ = 8). In addition, utterances in which vowel reduction does not allow for F0-measurement were excluded from pitch analysis (*N* = 4). Selected utterances were manually segmented into adjectives and nouns by auditory inspection and spectral analysis in Praat (Boersma & Weenink, 2011). Spectral analysis also included checking for pitch-tracking errors, which were not found. A script extracted the maximum F0 (Hz) from each word (adjective and noun) on the basis of the autocorrelation method implemented in Praat (Boersma & Weenink, 2011). To abstract over gender differences and to better represent perceived prominence by the human ear, Hertz values were converted into ERB values using the formula by Glasberg and Moore (1990) where *f* is the value in Hertz: $21.4*\log_{10}(0.00437*f+1)$. The pitch value (ERB) of the noun was subtracted from that of the adjective resulting in a relative measure which accounts for the fact that accents are perceived relative to each other (Gussenhoven, Repp, Rietveld, Rump, & Terken, 1997).

3.3.2 Statistical analysis

Univariate ANOVAs were performed with relative pitch measures of adjectives as dependent variables and with communicative setting (2 levels: shared, conceal) as within subjects factor.

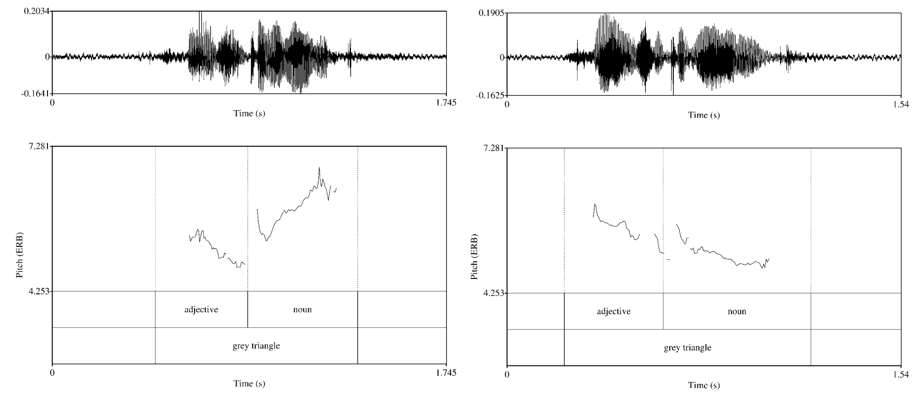

Figure 3.2: *Waveforms and F0 contours of the NP grijze driehoek (grey triangle) taken from the shared (left) and the conceal (right) setting.*

3.3.3 Results

Results show that adjectives in the conceal setting are uttered with a higher relative pitch (M = .37) than in the shared setting (M = -.11): $F_1(1, 50) = 12.22$, $p < .01$, $\eta_p^2 = .20$, $F_2(1, 10) = 11.02$, $p < .01$, $\eta_p^2 = .52$, $minF'(1, 31) = 5.79$, $p < .05$, see Table 3.2 and Figure 3.2. Figure 3.2 shows the waveforms and pitch contours of two utterances taken from the shared and the conceal setting respectively.

Table 3.2: *Mean relative pitch (ERB) and standard deviation as a function of communicative setting.*

Setting	*M*	*SD*
Shared	-.11	.07
Conceal	.37	.09

3.3.4 Conclusions

Speakers produced adjectives in the conceal setting with a higher relative pitch. It remains unclear whether the higher relative pitch of adjectives in the conceal condition is strong enough to exceed a perceptual threshold. This question is addressed in a prominence rating task.

3.4 Prominence judgements

When acoustic differences, as found in relative pitch, affect the prominence of adjectives, it can be expected that addressees perceive adjectives taken from the conceal setting as more prominent than adjectives taken from the shared setting. To investigate the perceived prominence of adjectives a rating task is carried out.

3.4.1 Method

The same NPs used for acoustic analysis (N = 225) were presented to listeners in a prominence rating task. To avoid misunderstanding, the instructions included the more common term *emphasis* (Dutch: *nadruk*) instead of *prominence* (Dutch: *prominentie*). Participants were asked to rate the emphasis on either the adjective or the noun on a seven point scale. In the first part of the task participants rated the emphasis on all the adjectives, whereas in the second part of the task they rated the emphasis of all the nouns. In both parts the entire NP was presented. The task was web-based and designed using WWStim (Veenker, 2003). Participants signed up and completed the experiment online. Before the start of the actual experiment participants could adjust their audio volume, receive instructions and had to complete an example stimulus. Stimuli were presented on html-pages on which the NP could be played as many times as needed using a button. Participants could not alter previous ratings when they had proceeded to the next stimulus. NPs were presented in a random order which was different for each participant. The entire task lasts about 40 minutes. Results were collected on a web server.

3.4.2 Participants

A total of 13 participants (10 men, 3 women, M_{age} = 29.8 years, age range: 24-44) completed the task. All of them were native speakers of Dutch without hearing problems who participated voluntarily. None of them participated in other experiments reported in this chapter.

3.4.3 Statistical analysis

Prominence scores were again computed as a relative measure, for which the prominence value of the noun was subtracted from the prominence value of the adjective. This measure accounts for the fact that the perception of prominence is dependent on surrounding material in a phrase (Gussenhoven et al., 1997). Furthermore, such a measure abstracts over possible individual differences in the use of the rating scale (e.g., tendencies to use only one end of the scale). Repeated measures ANOVAs were performed with relative prominence measures of adjectives as dependent variable and with communicative setting (2 levels: shared, conceal) as within-subject factor. Furthermore, Pearson's correlation coefficient was calculated between pitch measures and prominence ratings.

Table 3.3: *Mean relative prominence for the adjective and standard deviation as a function of communicative setting.*

Setting	*M*	*SD*
Shared	.03	.14
Conceal	.86	.15

3.4.4 Results

Results indicate that adjectives in the conceal setting were perceived as relatively more prominent (M = .86) than adjectives in the shared setting (M = .03): $F_1(1, 12) = 18.28$, $p < .01$, $\eta_p^2 = .60$, $F_2(1, 5) = 38.92$, $p < .01$, $\eta_p^2 = .89$, $minF'(1, 17) = 12.44$, $p < .01$, see Table 3.3. Pearson's correlation coefficient was computed to investigate whether the relative prominence ratings match the relative pitch measures. Results indicate that they indeed correlate moderately for pitch; $r = .58$, $N = 221$, $p < .001$. The fact that this correlation is not stronger may be explained by the fact that prominence depends on other correlates, such as duration and loudness, as well (Kochanski et al., 2005).

3.4.5 Conclusions

So far, results have shown that speakers in the conceal setting produced adjectives with a higher relative pitch which correlates with a higher relative prominence perceived by listeners. This suggests that speakers encode an implicit reference to privileged information prosodically. That is, adjectives that leak information are produced more prominently than adjectives that do not leak information. This result indicates that ironic process (Wegner et al., 1987) indeed affect the articulatory stage of language production. That is, the salient contrast between the figure in common ground and the figure in privileged ground in the conceal condition is prosodically expressed by speakers. It remains to be seen, however, whether listeners perceive this prominence as a reference to privileged information. To this end, a forced choice guessing task was carried out.

Table 3.4: *Number of stimuli per NP and communicative setting as presented in the guessing task (total N = 136).*

NP	NP (translated)	Shared	Conceal
grijze driehoek	grey triangle	9	9
groene ruit	green square	8	5
groene ster	green star	5	5
grote driehoek	big triangle	8	7
grote kruis	big cross	2	3
kleine cirkel	small circle	5	4
kleine ruit	small square	6	7
kleine ster	small star	7	7
paarse kruis	purple cross	4	8
paarse pijl	purple arrow	4	3
rode kruis	red cross	5	5
zwarte ruit	black square	5	5
TOTAL		68	68

3.5 Guessing task

In order to investigate whether listeners perceive NPs from the conceal setting as an implicit reference to privileged information, a guessing task was carried out. In this experiment participants had to guess the privileged information on the basis of a target description taken from the shared or the conceal setting. When the prosody of the adjectives uttered in the conceal setting indeed is a cue that signals leaked information, we expect listeners to be better able to guess the privileged information when listening to adjectives uttered in the conceal setting than when listening to adjectives uttered in the shared setting.

3.5.1 Method

136 NPs were taken from the production data, of which 60 NPs were also used in the acoustic analysis and 76 NPs were not used in the acoustic analysis. This was done to create a balanced design in which the NPs were equally divided over communicative settings. All NPs consisted of one adjective and one noun and were free from production errors like the ones described in section 3.1.1. In total 12 different NPs occurred several times in each condition (Table 3.4). These 12 items matched with the 16 target figures from the production task, because 4 target figures in the production task elicited the same NP twice (*kleine ruit, kleine ster, grote driehoek* and *grijze driehoek*). These target figures were considered as different items because their color or size differed. That is, figures of a small red star and a small blue star may both have elicited the NP *small star* when the contrast concerned size (section 3.2.3), while they were essentially different target figures. Participants were instructed to listen to the NPs and saw a set of four figures of which one was occluded (presented with a question mark, see Figure 3.3). Participants had to guess the occluded figure by making a choice between two possible figures on the basis of the NP they heard. Either the occluded figure differed in size or color from the target (contrastive adjective) or the occluded figure differed in shape from the target (contrastive noun). For example, when hearing “the small circle” participants guessed whether the occluded figure was the “the big circle” or “the small star”. In fact, the ‘contrastive adjective’ type of answer always corresponded with the figure that was occluded in the production task.

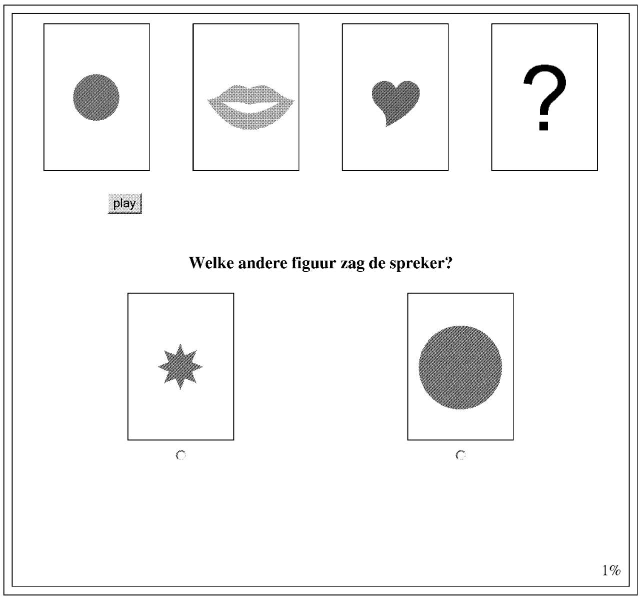

Figure 3.3: *Screenshot showing an example stimulus and the Dutch question "Which other figure did the speaker see?", which is presented to addressees in the guessing task.*

3.5.2 Procedure

Participants did the task online. Each stimulus was presented on an html-page designed using WWStim (Veenker, 2003) on which the set of figures was shown. Participants had to click a button underneath the target figure to play its description (Figure 3.3). Two figures presented as choice options were shown in random order at the bottom of the html-page. Stimuli were presented in a random order different for each participant to prevent any order effects. Participants signed up and completed the experiment online. Before the start of the actual experiment participants could adjust their audio volume, received instructions and had to complete an example stimulus. During the experiment each description could be played as often as needed. The choice could be altered before proceeding to the next stimulus. Participants could not alter previous choices when they proceeded to the next stimulus. The task lasted about 15 minutes. Results were collected on a web server.

3.5.3 Participants

A total of 72 participants (11 men, 61 women, M_{age} = 22.03 years, age range: 18-43) completed the task. All of them were native speaking Dutch students of Tilburg University without hearing problems who participated for a course credit.

Table 3.5: *Mean proportions (untransformed) of 'contrastive adjective' choices and standard deviation as a function of communicative setting.*

Setting	*M*	*SD*
Shared	.66	.02
Conceal	.69	.02

3.5.4 Statistical analysis

A proportion of participants' choices for the 'contrastive adjective' answer was calculated. That is, for each 'contrastive adjective' answer, a score of 1 was assigned, whereas for each 'contrastive noun' answer, 0 was assigned. Proportions reported here are averages of these values. All participants rated all stimuli. Repeated measures ANOVAs were performed with logit transformed proportion measures of adjective choices as dependent variable and with communicative setting (2 levels: shared, conceal) as within-subject factor. Means are reported for untransformed proportions measures of adjective choices. Furthermore, a one sample t-test was carried out to test whether listeners made their choice above chance level and a correlation measure was taken to investigate the relation between relative pitch, relative prominence and the proportion of choices in the guessing task. The correlation measure was carried out on the 60 NPs that matched the ones in the acoustic analysis.

3.5.5 Results

Results suggest that listeners slightly more often chose the option that the adjective leaked contrastive information in the conceal condition than in the shared condition, even though this effect is only significant in the analysis by subject and shows a trend towards significance in the analysis by item: $F_1(1, 71) = 5.86$, $p < .05$, $\eta_p^2 = .08$, $F_2(1, 15) = 3.86$, $p = .07$, $\eta_p^2 = .21$, $minF'(1, 38) = 2.33$, $p = .14$, see Table 3.5. Possibly, the F2 analysis (and consequently the minF') failed to reach significance because of lack of power, as there were far less item classes than subjects[1]. The one sample t-test shows that overall addressees chose above chance level for the 'contrastive adjective' answer ($M = .67$): $t(71) = 7.1$, $p < .001$. This indicates that participants have an overall tendency to interpret the adjective as contrastive. Pearson correlation measures indicate that proportions of choice for the 'contrastive adjective' answer correlated with relative prominence: $r = .71$, $N = 60$, $p < .001$, but not with relative pitch.

3.5.6 Conclusions

The effects found in the guessing task were subtle. This is reflected in the statistical analyses, which shows that the tests only reached significance in the analysis by subject and not in the analysis by item nor in the minF' analysis. Furthermore, it is interesting to state that the numeric differences are in the expected direction, and also that the proportion of times that people guessed that the NP contrasted with the information in the adjective correlated significantly with earlier prominence scores. However, we have to note that overall, participants have a tendency to interpret the adjective as contrastive. This possibly is the result of the fact that no NPs in which the noun exhibited some prosodic marking of a contrast were presented to participants. In addition, these findings suggest that upon hearing an adjective, addressees are likely to assume that it was uttered to refer to a contrast. This is compatible with Weber, Braun, and Crocker (2006) who find the same bias in addressees' contrastive interpretation of adjectives. Furthermore, the choices participants made in the guessing task correlated with relative prominence, suggesting that an acoustic cue was responsible for a contrastive interpretation of the adjective. The choices did not correlate with measures of pitch. This may be explained by the fact that pitch was a measure of production and a rough approximation of prominence (Kochanski et al., 2005), while both the judged prominence and the proportion of choices in the guessing task were perceptual measures. However, we cannot conclude that there is no relation at all between the choices in the guessing task and the pitch of the speakers in the production task, because results of the prominence rating task show a correlation between prominence and relative measures of pitch. The current results may indicate that the choices in the guessing task were at least indirectly affected by the produced pitch. Furthermore, the bias to interpret the adjective contrastively may have masked possible effects of communicative setting, as participants had a tendency to choose the 'contrastive adjective' type of answer regardless of communicative setting. This may explain why the effect of communicative setting was subtle in the current analysis.

3.6 Discussion and conclusions

The findings presented in this study extend our knowledge about how speakers produce leaked information in their references. We have shown that adjectives that leak information about the speaker's perspective are realised prosodically differently from adjectives that do not leak information. In particular, adjectives that leak information have a higher relative pitch and are perceived as more prominent than adjectives that do not leak information.

The current study adds a new insight to the theory of ironic processes (Wegner, Schneider, Carter, & White, 1987; Wegner, 1994). We show that due to ironic processes privileged information is reflected both in the production and the perception of speech. The results of our production task confirm Wardlow Lane et al.'s finding that these processes affect the grammatical encoding stage of language production (Levelt, 1989; 1999), and the results of our acoustic analysis indicate that ironic processes also affect the articulatory encoding stage. Furthermore, we show that speech perception is affected, in particular the way listeners perceive prominence patterns of NPs that leak information. This indicates that the perceptual stage of audition is also affected by ironic processes. This conclusion adds a new perspective to the literature on privileged information (Horton & Keysar, 1996; Nadig & Sedivy, 2002; Wardlow Lane et al., 2006). In particular, we show that it is worth investigating how privileged information is realised prosodically, rather than limit the analysis to whether that information is realised lexically. We find that lexically identical NPs are prosodically different according to whether they refer to privileged information or not and that this difference is possibly the result of ironic processes.

Results in the current study are compatible with an incremental theory of speech production (Fry, 1969; Levelt, 1989; Pechmann, 1989). This theory possibly explains why speakers overspecify and when contrastive intonation is used. The basic assumption of incremental speech production is that speakers may start their utterance without being aware of the full visual context they describe. Therefore, contrasts between objects within one visual context, as in the shared setting in the current study's paradigm, are usually not prosodically marked (Pechmann, 1989). Only when speakers have a full mental representation of the discourse context under description will they mark contrastive information prosodically. In the current study ironic processes possibly cause speakers to pay attention to the contrast between the target figure and the occluded figure. It is therefore plausible that speakers have a representation of this contrast before they start talking. Along these lines, speakers are likely to prosodically mark that contrast by means of an increased pitch. Such an explanation could be supported by additional evidence from eye-tracking. That is, speakers presumably look more often at the occluded figure when instructed not to do so. Importantly, the acoustic evidence for prosodic marking is found in small differences in pitch, which suggests that the prosodic encoding of privileged information is a subtle process. This is confirmed by results of the guessing task. Differences in addressees' ability to guess the privileged information reached significance. The guessing results positively correlate with the perceived prominence, indicating that prosody may have influenced participants' guesses. It has to be noted that this subtle effect most plausibly, but not necessarily, originates from pitch differences found in this study. The pitch effects correlate with the perception of prominence, which in turn correlates with the guesses for the privileged figure. It has to be remarked that other prosodic cues, such as voice quality of vowels, duration and intensity which were not taken into account in the current analysis, may also have played a role in the observed differences between the communicative settings.

To conclude, as shown in Wegner et al. (1987) and Wegner (1994), when people are instructed not to think of a white bear, they cannot avoid doing so. The same effect was found in speakers' use of adjectives in Wardlow Lane et al.'s (2006) study when instructed to ignore privileged information. The present study extends these findings by showing that ironic processes affect how speakers encode privileged information prosodically and how addressees perceive that information.

Endnotes Chapter 3

[1] Note that items in the current design are nested under the factor communicative setting, for which it has been argued that F1 analysis would be sufficient (Raaijmakers et al., 1999).

CHAPTER 4

ON HOW ACCENT DISTRIBUTION CAN SIGNAL SPEAKER ADAPTATION

ABSTRACT

Some dialogues are perceived as running more smoothly than others. To some extent that impression could be related to how well speakers adapt their prosody to each other. Adaptation in prosody can be signaled by the use of pitch accents that indicate how utterances are structurally related to those of the interlocutor (prosodic function) or by copying the interlocutor's prosodic features (prosodic form). The same acoustic features, such as pitch, are involved in both ways of adaptation. Furthermore, function and form may require a different prosody for successful adaptation in certain discourse contexts. In this study we investigate to what extent interlocutors are perceived as good adapters, depending on whether the prosody of both speakers is functionally coherent or similar in form. This is done in two perception tests using prosodically manipulated dialogues. Results show that coherent functional prosody can be a cue for speaker adaptation and that this cue is more powerful than a similar prosodic form.

This chapter is based on:

Kaland, C. C. L., Krahmer, E. J., & Swerts, M. G. J. (2013a). On How Accent Distribution Can Signal Speaker Adaptation. *Phonetica*, *69*, 216-230.

Kaland, C. C. L., Krahmer, E. J., & Swerts, M. G. J. (2012a). Accentuation as a cue for speaker adaptation. *Proceedings Speech Prosody 2012*, Shanghai, China.

4.1 Introduction

Speakers in a dialogue adapt to each other. In particular, speaker adaptation is often analyzed regarding the extent to which interlocutors copy certain verbal or non-verbal features of their conversation partners. This kind of adaptation behavior has been modeled in different, but related accounts, most notably in terms of alignment (Pickering & Garrod, 2004), convergence (Pardo, 2006) or mimicry (Bargh & Chartrand, 1999). As we discuss in this introductory section, there is a large body of work to support the claim that people indeed adapt to each other by copying each other's conversational characteristics, which also holds - more specifically - for the way speakers copy formal aspects of each other's prosody. In addition, while adaptation has mostly been studied in terms of such copying behavior, we will argue that interlocutors may also adapt to each other in a different way, namely in signaling that their utterances are meaningfully related to the utterances of their conversation partners. As we will see, the latter type of adaptation can be reflected in the functional use of prosody, in particular in how speakers distribute pitch accents in their utterances. Interestingly, the formal and functional forms of prosodic adaptation can be in conflict with each other, which begs the question whether one is more important than the other. To answer this general question, we will use a perceptual approach to investigate whether different kinds of prosodic adaptation have a different impact on the way listeners perceive the smoothness of an interaction between two interlocutors.

Adaptation may appear from the fact that the form of speakers' utterances is determined by that of the utterances they received from their conversation partners. This happens at different linguistic levels and is referred to as alignment (Pickering & Garrod, 2004; Garrod & Pickering, 2004). For example, at the syntactic level it has been shown that speakers repeat prepositions from their conversation partner when answering a question (Levelt & Kelter, 1982). In their experiment shopkeepers were asked the question "What time do you close?" or "At what time do you close?". They frequently responded "Five o'clock" to the former and "At five o'clock" to the latter question. Furthermore, Brennan and Clark (1996) showed that conversation partners in a referential task often formed an agreement on how to refer to an object, i.e., *conceptual pact*, and ended up using the same lexical description when repeatedly referring to a concept. For example, speakers may agree to refer to a tangram figure that vaguely resembled a skater by “the ice skater”. These studies reveal that by taking over lexical features, speakers show that they share information with their interlocutor (common ground; Clark, 1996). It has also been shown that people in interaction are likely to take over non-verbal aspects. For example, in a conversation speakers may copy each other’s body postures (Bargh & Chartrand, 1999), such as having the arms crossed. The process in which people in interaction exhibit postural similarity is called *mimicry*. According to some models, mimicry happens automatically and is of great importance in that people ground themselves in their social environment by taking over the interactional behavior that surrounds them (Bargh & Chartrand, 1999). In this way mimicry supports the social bonding process between people. The mechanism behind alignment in dialogue is assumed to require a tight link between perception and behavior (Garrod & Pickering, 2004; Dijksterhuis & Bargh, 2001). In other words, speakers’ behavior is a product of the behavior they perceive. The studies just discussed show that adaptation by means of copying certain behavioral features usually concerns the formal characteristics of that behavior. That is, taking over linguistic features from an interlocutor rather affects *how* (form) speakers say what they mean rather than *what* (function) they say. For example, “Five o’clock” and “At five o’clock” are semantically equivalent in the contexts sketched above. However, at the level of prosody the distinction between form and function may be more vague as they both manifest themselves in the same acoustic features, such as pitch, duration and intensity. The aim of the present study is to explore to what extent speaker adaptation at the level of prosody is related to function, form or a combination of both.

There is evidence that speakers copy the prosodic form of each others’ utterances. For example, dialogue interlocutors are likely to converge on a global pitch range (Gregory, 1986; Couper-Kuhlen, 1996). That is, speakers’ F0 tends to become similar within certain frequency bands once they engage in a dialogue. Also speech intensity (Natale, 1975) and speech rate (Giles et al., 1991; Szczepek Reed, 2010) are likely to converge to similar levels once speakers interact in a dialogue.

These processes are attested for both young and old age groups (Nilsenová et al., 2009; Coulston et al., 2002). It is debatable whether prosodic adaptation processes like the ones just mentioned are fully automatic. As shown by Pardo (2006), the extent to which interlocutors sound similar on repeated words is also largely determined by the gender and the task specific role of the speaker. That is, in a map-task male speakers converged more than female speakers and instruction givers converged more than instruction receivers (Pardo, 2006). In addition, it has been shown that speakers are likely to take over prosodic features that are somehow related to the type of dialogue act in a conversation. For example, Okada et al. (2012) investigated dialogues in which speakers described maps that either matched or mismatched the map of their interlocutor in a map task. In the former type of dialogues interlocutors exhibited agreement, whereas in the latter type of dialogue interlocutors exhibited disagreement. Okada et al. (2012) found that the pitch intervals (rounded off to the nearest semitone) of successive utterances had the same key in agreement-dialogues and a different key in disagreement-dialogues. Furthermore, Nilsenová et al. (2009) tested whether speakers are more likely to take over prosody when it was related to the semantics (i.e., function) of the utterance compared to when it was not. Participants in their study interacted with a manipulated computer voice. The manipulated prosodic features were boundary tones, which mark phrase beginnings and endings, and pitch range, which has no direct linguistic function. Results showed that speakers prefer to take over boundary tones over pitch range (Nilsenová et al., 2009).

Speakers may also use prosodic features like pitch, duration and intensity in a functional way to adapt to their interlocutor. Most clearly, they can use pitch accents to highlight particular words that are important to the discourse (Pierrehumbert & Hirschberg, 1990). In most Germanic languages, pitch-accented words are phonetically realized by a certain pitch contour, generally accompanied by an increase in duration and intensity, which lends prominence to the accented word (Pierrehumbert & Hirschberg, 1990). In Dutch, as well as many other languages, speakers use such accents to make new or contrastive information more prominent, whereas given information is typically deaccented. In interaction the appropriate use of accentuation has to do with the fact that such accents signal coherence between successive utterances by different speakers. That is, by accenting new information and by deaccenting given information speakers show that they account for the information they share with their partners.

At first sight it appears to be the case that speakers often copy the accentuation pattern of their interlocutor when interacting in dialogue. This becomes especially clear when considering contrastive intonation. For example, a first speaker may ask "Would you like to have a green or a red apple?" with a pitch accent on *green* and on *red* to signal the distinguishing feature (color) between the proposed apples. A second speaker may respond "I would like to have a red apple" with a pitch accent on *red*. In the latter utterance the second speaker copies the contrastive intonation to acknowledge that he or she is aware of the color contrast as well. In that respect, this example is analogous to the one discussed above where speakers tend to use a preposition in a response when that preposition was also present in the syntactic structure of the conversation partner's question. However, there are other dialogue contexts in which it is not appropriate to copy the interlocutor's accentuation pattern. For example, speakers may start a conversation with their partner by introducing the phrase: "I have a green apple". When this utterance represents entirely new information, speakers will typically produce that sentence with a main accent on the word *apple*. Suppose that their addressee continues the interaction by stating: "And I have a red apple". A felicitous use of accents would lead to a realization in which only the word *red* is accented in the final noun phrase (henceforth: NP) of that sentence, to signal the difference in color between the piece of fruit in that sentence and the one mentioned in the preceding utterance. Thus, in this type of context it may be appropriate for speakers to use an accentuation pattern that has a different form compared to the one of the other speaker, as a single accent on the noun *apple* in the second sentence would signal that the second speaker does not fully take into account what has been said in the preceding sentence. Pechmann (1984a, 1984b, 1989) showed that speakers indeed systematically use a contrastive intonation pattern to describe an object that is minimally distinguishable from another object.

With regard to perception, it has been shown that listeners update the information status of both members of a contrast set (i.e., in the example *red* and *apple* as contrastive and given information respectively) when hearing a contrastive intonation (Braun & Tagliapietra, 2010). Thus, like copying certain prosodic features, accentuation is a cue for the adaptation process of speakers, as it signals coherence between utterances of both interlocutors. In general, there is much work in which it was shown that accent distributions have an impact on the way speech is perceived. For example, the quality of the spoken output of a speech synthesis system is related to the extent to which the accent distribution in the generated utterances is appropriate (Quené & Kager, 1989). Similarly, Nooteboom and Kruyt (1987) found that sequences of utterances are perceived as more natural if the distributions of the accents in these consecutive utterances correctly match given/new distinctions. Terken and Nooteboom (1987) showed that inappropriately accented information confuses listeners, resulting in longer comprehension times. Finally, Swerts and Marsi (2012) found that the perceived difference between "good" and "bad" newsreaders is related to how these speakers distributed accents in their spoken texts. However, such studies on the perception of accent distributions have mostly been based on analyses of single speaker productions (monologues), so that it remains to be seen how accent distributions across turns from different interlocutors are perceived. More specifically, the current study focuses on how pitch accents that cue differences in information status in sequences of turns from different speakers may be indicative of how well these speakers adapt to each other.

In sum, when it comes to prosody, speakers have different ways to adapt to each other. On the one hand, speakers may achieve similarity at the level of prosodic form by copying each other's acoustic characteristics. On the other hand, speakers may achieve coherence at the level of prosodic function by means of pitch accents to link their utterances semantically to what has been uttered by their conversation partner. Although form and function may apply to different levels of prosody they are both signaled by the same acoustic features. Research has shown that speakers start to sound similar to each other in the course of their interaction, but it is not clear - as we argued above - whether this easily generalizes to accent distributions.

Speakers might not automatically copy each other's accent patterns because that could interfere with informational constraints, like the existence of contrasts across speaking turns. Much previous research investigated speaker adaptation processes from a production side (i.e., Natale, 1975; Couper-Kuhlen, 1996; Pardo, 2006; Nilsenová et al., 2009). In the current chapter, we aim to investigate how listeners perceive dialogue fragments that contain different accent distributions across speaking turns. Our perceptual approach is in line with that of earlier studies, even when these were not concerned with prosody. For example, it has been shown that the level of adaptation between conversation partners can have an effect on the subjective impression of that interaction (e.g., Chartrand & Bargh, 1999). As speaker adaptation is commonly defined with respect to copying certain formal linguistic features (Pickering & Garrod, 2004), we first explore whether functional prosody by means of accentuation is a perceptual cue for speaker adaptation. This is tested in experiment I. Thereafter, in experiment II we test the role of formal prosody in addition to functional prosody as perceptual cues for speaker adaptation. We hypothesize that functional prosody, due to its systematic use (Pechmann, 1984a, 1984b; 1989) and its close relation to semantics (Nilsenová et al., 2009) is a cue for speaker adaptation and that it is stronger compared to similarity in prosodic form. Both experiment I and II are perception tasks eliciting listeners' judgments about the adaptation process between speakers who use contrastive intonation in prosodically manipulated dialogues. Listeners hear the dialogues passively, as overhearers (Schober & Clark, 1989).

4.2 Experiment I: Prosodic function

Experiment I elicits judgments about the speaker adaptation process by presenting listeners prosodically manipulated dialogues. Dialogues consist of utterances that are elicited in a data collection task. Elicited utterances have a natural contrastive intonation. Prosodic manipulation is obtained by the order in which utterances are put together in a dialogue.

4.2.1 Data collection

We used a specific semi-spontaneous procedure to obtain naturally-sounding utterance stimuli for our perception test. That is, to elicit utterances with a contrastive intonation, a cooperative object naming task was carried out. The participants were instructed to construct a specific shape by using geometrical figures in different forms (triangle, square, parallelogram) and colors (red, blue, yellow). The order in which the geometrical figures should be put together was manipulated in such a way that two successive figures were minimally distinguishable. For example, a blue square may be followed by a blue triangle (shape contrast) or a yellow square (color contrast). In this way shape contrasts elicited accented nouns and color contrasts elicited accented adjectives. Twelve participants carried out the object naming task together with an experimenter (6 males, 6 females, M_{age} = 27.8 years, age range: 20-33). All participants were native speakers of Dutch. The experimenter and the participant alternately put a geometrical figure in place. Participants were instructed to describe their act using the same matrix sentence for each description: *Ik leg de* [*blauwe driehoek*] *hier* (I put the [blue triangle] here). In this way the grammatical position of the NP referring to the geometrical figure was kept constant. Furthermore, each utterance ended with the word *hier* (here) so that boundary tones did not occur within the target NP. A computer screen showed how the target shape consisted of the geometrical figures. In addition, geometrical figures were numbered according to the order in which they had to be constructed (Figure 4.1). In total 40 shapes were constructed such that each description of a geometrical figure occurred repeatedly in each accent condition. All utterances were recorded as wave-files using a Marantz PMD-600 solid state recorder.

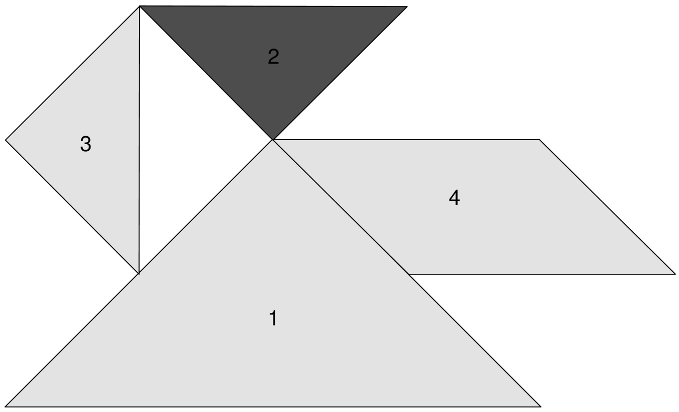

Figure 4.1: *Example shape consisting of geometrical figures.*

Table 4.1: *Mean pitch maxima in ERB and standard deviations as a function of accent and word type (N = 288).*

Word type	Accented	Deaccented
Adjective	5.21 (*.19*)	4.40 (*.27*)
Noun	5.32 (*.36*)	4.66 (*.18*)

4.2.2 Acoustic analysis

A total of 288 out of 480 NPs were selected from the twelve speakers that were recorded during the data collection. To meet the requirements of the perception experiment six speaker pairs consisting of one male and one female speaker were formed (see section 4.2.3). NPs produced by the experimenter were omitted. On the basis of auditory judgments NPs with the clearest accent on either the adjective or the noun were selected. A software program (Audacity Team, 2006) was used to extract the NPs from the wave-file recordings. As a sanity check, we investigated the effect of accent on the NPs by carrying out an acoustic analysis. That is, for each NP the pitch maximum of the adjective and the noun was measured automatically with a script using Praat (Boersma & Weenink, 2011). Pitch was measured in ERB (Glasberg & Moore, 1990) for two reasons. First, ERB closely resembles the perception of prominence due to accentuation. Second, its logarithmic scale abstracts largely over gender differences. A repeated measurements analysis of variance was carried out with maximum pitch as dependent variable and with accent (2 levels: accented, deaccented) and word type (2 levels: adjective, noun) as independent variables.

Speakers produce both accented adjectives and accented nouns with a higher pitch than their deaccented counterparts: $F(1,5) = 65.55$, $p < .001$, $\eta_p^2 = .93$ (Table 4.1). Although nouns exhibit slightly higher pitch maxima than adjectives, there is no significant effect of word type nor an interaction effect of accent and word type. Furthermore, it is known that in Dutch the perception of contrastive intonation depends on both the prominence of the adjective and the noun (Krahmer & Swerts, 2001). Therefore, a pitch maxima difference measure is computed by subtracting the maximum pitch value of the deaccented word from the maximum pitch value of the accented word. The difference measure covers the values of both adjectives and nouns. Thereafter, values are computed per speaker pair to investigate whether they differ individually. For example, the value of pair AB is the average pitch maxima difference of speaker A and of speaker B (Table 4.2). In this way we obtain one value per speaker pair that resembles the clarity of their contrastive intonation. Results indicate that speakers from different pairs differ in the degree to which they use pitch to mark the difference between accented and deaccented words. Table 4.2 shows that pitch maxima differences are the smallest for pair EF and the largest for pair GH. A repeated measures analysis of variance (RM-ANOVA) with the difference measure as dependent variable and with speaker pair as independent variable shows an effect of speaker pair: $F(5,55) = 3.56$, $p < .007$, $\eta_p^2 = .25$. Pairwise comparisons (after Bonferroni correction) show significant differences between the pairs EF-GH [*MD* (*SE*) = .62 (*.08*), 95% CI = (.32, .91), $p < .001$] and GH-IJ [*MD* (*SE*) = .54 (*.08*), 95% CI = (.25, .84), $p < .001$].

Table 4.2: *Mean pitch maxima differences and standard deviations per speaker and per speaker pair.*

Speaker	*M* (*SD*)	Speaker pair	*M* (*SD*)
A	.83 (*1.01*)	AB	.76 (*.49*)
B	.69 (*.28*)		
C	.63 (*.33*)	CD	.80 (*.38*)
D	.97 (*.59*)		
E	.47 (*.18*)	EF	.45 (*.26*)
F	.45 (*.42*)		
G	.85 (*.67*)	GH	1.07 (*.25*)
H	1.29 (*.48*)		
I	.28 (*.42*)	IJ	.53 (*.22*)
J	.78 (*.26*)		
K	1.44 (*1.40*)	KL	.85 (*.76*)
L	.25 (*.35*)		

4.2.3 Stimuli

Short dialogues were constructed consisting of four NPs (Table 4.3). NPs were taken alternately from one male and one female speaker (two NPs each) to balance out possible effects of participants preferences for male or female voices. Background noise was added to the dialogue to mask edges at turn shifts which favors the perception of the dialogue as a whole. Dialogues were presented as fragments cut from the middle of an entire dialogue. Therefore, fragments of NPs (approximately one syllable) were added to dialogue edges of which the volume faded in or faded out respectively. NPs were ordered in such a way that accents were either used coherently or incoherently throughout the dialogue. For functionally coherent dialogues, the contrastive word in the NP was accented and the given word in the NP was deaccented. Contrastiveness and givenness were defined with respect to the previous utterance (Table 4.3). That is, in *blue triangle* after *red triangle* the adjective *blue* would be contrastive and the noun *triangle* would be given. For functionally incoherent dialogues the contrastive word in the NP was deaccented and the given word was accented. The accentuation pattern was balanced over the dialogues to compensate for habituation effects. That is, in one half of the dialogues the first two NPs of a dialogue had an accented adjective whereas the last two NPs had an accented noun (AANN) and vice versa in the second half of the dialogues (NNAA). The duration of one dialogue was approximately 22 seconds. In total 72 dialogues were constructed using 6 speaker pairs. Per speaker pair 6 coherent and 6 incoherent dialogues were constructed.

Table 4.3: *Examples of functionally coherent (+) and incoherent (-) dialogues per accentuation pattern (accented words in capitals).*

Function	Pattern	NP1	NP2	NP3	NP4
+	AANN	RED triangle	BLUE triangle	yellow SQUARE	yellow TRIANGLE
	NNAA	yellow SQUARE	yellow TRIANGLE	RED triangle	YELLOW triangle
-	AANN	BLUE square	YELLOW triangle	red TRIANGLE	yellow TRIANGLE
	NNAA	red TRIANGLE	blue TRIANGLE	YELLOW square	YELLOW triangle

4.2.4 Procedure

Coherent dialogues were coupled with incoherent dialogues in fixed pairs (6 in total). A web-based perception task presented all dialogue pairs per speaker pair (36 pairs in total). Each pair of dialogues originated from the same speaker pair. In order to reduce effects of other sound sources participants were instructed to do the experiment in a quiet room or to wear headphones. The entire experiment lasted about 20 minutes. Dialogue pairs occurred in a different random order for each participant. Each dialogue pair was presented on an html-page designed using WWStim (Veenker, 2003). Each pair of dialogues was represented by two buttons which participants could click to hear them. Participants were told that speakers in the dialogue constructed a tangram figure together while describing the geometrical figures they used. Instructions to participants noted that some speaker pairs collaborated better than others. The participants' task was to choose the dialogue in which speakers account for each other's utterances the best. For each stimulus participants were shown the question *In welke dialoog houden sprekers het beste rekening met elkaar?* (*In which dialogue do speakers account for each other the best?*). They were instructed to pay close attention to intonation. Participants were allowed to listen to the dialogues as often as they liked. Once they had made their choice this was registered digitally and could not be altered. The data was analyzed in terms of proportions where 1 corresponds with a choice for a functionally coherent dialogue and 0 corresponds with a choice for a functionally incoherent dialogue.

4.2.5 Participants

The perception experiment was done by 20 naïve participants (9 males, 11 females, M_{age} = 27.5 years, age range: 20-35). None of the participants of experiment I took part in the data collection task. All participants were native speakers of Dutch without hearing problems and had no phonetic or linguistic training.

Table 4.4: *Means and standard deviations of proportions of choices for functionally coherent dialogues and chi-square tests per speaker pair.*

Speaker pair	*M* (*SD*)	$\chi^2(df = 1, N = 120)$
AB	.76 (*.22*)	32.03 ($p < .001$)
CD	.72 (*.28*)	22.53 ($p < .001$)
EF	.63 (*.26*)	7.5 ($p < .01$)
GH	.82 (*.16*)	48.13 ($p < .001$)
IJ	.78 (*.22*)	36.30 ($p < .001$)
KL	.63 (*.29*)	7.5 ($p < .01$)

4.2.6 Statistical analysis

To investigate participants' preference for coherent or incoherent dialogues a chi-square test is performed on the proportion of choices (overall and per speaker pair). Furthermore, we test whether participants' choices depend on the speaker pair they heard. Therefore, an RM-ANOVA and pairwise comparisons with mean proportion of choices as dependent variable and speaker pair (6 levels: pairs AB to KL) as independent variable is carried out.

4.2.7 Results

Participants have a preference for indicating the functionally coherent dialogue as the dialogue in which speakers account for each other's utterances the best. That is, the coherent dialogues are chosen in 71.94% of the cases. This rate is statistically above the chance level of 50%: $\chi^2(1, N = 720) = 138.69$, $p < .001$, for all speakers pairs (Table 4.4). Results further indicate that the factor speaker pair has an effect on participants choices: $F(5,95) = 4.78$, $p < .01$, $\eta_p^2 = .20$ (Table 4.4). Pairwise comparisons (after Bonferroni correction) show that the pairs EF-GH and GH-KL differ significantly: *MD* (*SE*) = .19 (*.05*), 95% CI = (.02, .37), $p < .05$. Other pairs do not differ significantly. This shows that to some extent participants' choices depend on which speaker pairs they had heard.

4.2.8 Conclusions

Experiment I shows that listeners can use the functional coherence of contrastive intonation patterns at a dialogue level to judge the interlocutors' adaptation process. That is, our judges appear to assess dialogue partners in a conversation as being more cooperative if the accent distributions match the given-new distinctions across speaking turns. Thus, functional prosody is not only a cue within discourse to signal its structure. It also exhibits information about the discourse process itself as a collaborative interaction between speakers.

Interestingly, there is an effect of speaker pair in that listeners judge some pairs as adapting better than other pairs. Inspection of both Table 4.2 and Table 4.4 suggests that this side-effect is not simply a listeners' preference for certain speakers. Rather, speaker pairs that show large acoustic differences between accented and deaccented words seem to be perceived as better adapters compared to pairs showing small acoustic differences (cf. pair EF and GH). Thus, the effect of speaker pair seems to be grounded in the acoustic characteristics of accentuation. Such a conclusion favors the view that functional prosody plays an active role in the perception of speaker adaptation. The question remains, however, how this role relates to prosodic form, as both function and form are expressed in the same acoustic features. In particular, it remains to be investigated to what extent speaker adaptation is signaled by functional or formal prosody. This was tested in experiment II.

4.3 Experiment II: Prosodic function versus prosodic form

4.3.1 Data collection

NPs used in experiment II were collected in the same data collection procedure as described for experiment I (section 4.2). Selected NPs for experiment II partly overlap those of experiment I.

Table 4.5: *Mean pitch maxima (ERB) and standard deviations as a function of accent and word type (N = 96).*

Word type	Accented	Deaccented
Adjective	5.27 (*.38*)	4.09 (*.78*)
Noun	5.01 (*1.00*)	4.70 (*.40*)

4.3.2 Acoustic analysis

A total of 96 out of 480 NPs were selected from the twelve speakers (six pairs) that were recorded during the data collection (see section 4.2) to meet the requirements of the perception experiment (see section 4.3.3). Acoustic analysis was carried out as described for experiment I (section 4.2.2).

Speakers produce both accented adjectives and accented nouns with a higher pitch than their deaccented counterparts: $F(1,5) = 61.03$, $p < .001$, $\eta_p^2 = .92$ (Table 4.5). Although nouns exhibit slightly higher pitch maxima than adjectives, there is no significant effect of word type nor an interaction effect of accent and word type.

Table 4.6: *Dialogues used in experiment II as a function of prosodic function and prosodic form (accented words in capitals, + = coherent/similar, - = incoherent/dissimilar).*

Function	Form	NP1	NP2	NP3	NP4
+	+	BLUE triangle	YELLOW triangle	RED triangle	BLUE triangle
+	-	BLUE square	blue TRIANGLE	YELLOW triangle	yellow SQUARE
-	+	YELLOW square	YELLOW triangle	BLUE square	BLUE triangle
-	-	blue TRIANGLE	YELLOW square	red TRIANGLE	BLUE square

4.3.3 Design

We constructed short dialogues that consist of four NPs collected during the object naming task. The construction of dialogues is similar with respect to balancing male and female speakers, the addition of background noise and volume fading at dialogue edges, as described in section 4.2.3. However the constructed dialogues were not identical to the ones of experiment I as in those dialogues only one variable (functional coherence) was manipulated. In the current design two variables were manipulated: *functional coherence* (2 levels: coherent, incoherent) and *formal similarity* (2 levels: similar, dissimilar), see Table 4.6. For functionally coherent dialogues the accentuation pattern is coherent with respect to the content of what the previous speaker said. Functionally incoherent dialogues follow the opposite pattern. This manipulation of functional coherence is identical to experiment I. For formally similar dialogues speakers are perceived as copying the accentuation pattern of the previous speaker. That is, when the previous speaker accentuates the adjective and deaccentuates the noun, the next speaker does so too, and similarly for cases where only the noun is accented. In formally dissimilar dialogues speakers accentuate the word that the previous speaker deaccentuated and deaccentuate the word that the previous speaker accentuated, see Table 4.6. The four types of dialogues were constructed for each speaker pair, resulting in a total of 24 dialogues.

4.3.4 Materials

The perception task was again web-based and designed using WWStim (Veenker, 2003). Stimuli were presented online as html-pages. One stimulus was considered as an html-page displaying two buttons with which two different dialogues of the same speaker pair could be played (Figure 4.2). Participants made a forced choice between the two dialogues. Dialogue pairs were constructed in such a way that the four dialogue types were uniquely combined (6 per speaker pair). This setup had a total of 36 stimuli with all individual dialogues having an equal chance to be chosen.

4.3.5 Procedure

The experimental procedure is identical to the one described for experiment I (section 4.2.4).

4.3.6 Participants

A total of 20 naive participants completed the experiment (5 men, 15 women, M_{age} = 26.6 years, age range: 21-39). They were all native speakers of Dutch without hearing problems. None of the participants acted in the data collection task nor in experiment I.

4.3.7 Statistical analysis

For each dialogue a score was computed as a proportion for which 1 corresponds with a chosen dialogue and 0 with an unchosen dialogue. Chi-square tests are performed on the proportion data for the four dialogue types, both overall and per speaker pair. Furthermore, chi-square tests are performed on the proportion data of the variables *functional coherence* and *formal similarity* individually to explore participants' preferences in more detail.

4.3.8 Results

A chi-square on the proportion data of all four dialogue types shows a significant effect [$\chi^2(3, N = 720) = 34.68$, $p < .001$] indicating that participants' preferences clearly differ from a random distribution (25% preference for all dialogue types). Table 4.7 shows that there are differences per speaker pair in that for some speaker pairs listener's judgments are randomly distributed (AB, IJ, KL) whereas for other speaker pairs that is not the case (CD, EF and GH). The data in Table 4.7 further suggest that listeners are especially sensitive to functional coherence. Indeed, if we do focused analyses, it turns out that participants choose the functionally coherent dialogues more often than the functionally incoherent dialogues: $\chi^2(1, N = 720) = 32.09$, $p < .001$. To compare experiment I and experiment II, Figure 4.2 shows the proportions of choices for functionally coherent dialogues. Note that values of experiment I are identical to Table 4.4. Values of experiment II are taken from Table 4.7 by taking the sum of values of choices for functionally coherent dialogues which abstracts over formal similarity. Therefore, the computation for speaker pair AB is .24+.31 = .55. The results of experiment I and II are comparable, with similar patterns across speaker pairs (Figure 4.2). As for formal similarity, participants choose the dialogues with formally similar prosody equally often as dialogues with formally dissimilar prosody: $\chi^2(1, N = 720) = .14$, $p = .71$.

Table 4.7: *Proportion of chosen dialogues and standard deviations crossed for prosodic function and prosodic form and chi-square tests overall and per speaker pair (+ = coherent/similar, - = incoherent/dissimilar).*

Speaker pair	Function +		Function -		$\chi^2(df = 3)$
	Form +	Form -	Form +	Form -	
AB (N = 120)	.24 (*.43*)	.31 (*.46*)	.27 (*.44*)	.18 (*.39*)	3.93 (n.s.)
CD (N = 120)	.29 (*.46*)	.36 (*.48*)	.20 (*.40*)	.15 (*.36*)	12.47 ($p < .01$)
EF (N = 120)	.32 (*.47*)	.28 (*.45*)	.24 (*.43*)	.16 (*.37*)	6.73 ($p = .08$)
GH (N = 120)	.35 (*.48*)	.33 (*.47*)	.17 (*.37*)	.15 (*.36*)	16.27 ($p < .01$)
IJ (N = 120)	.24 (*.43*)	.34 (*.48*)	.19 (*.40*)	.23 (*.42*)	6.00 (n.s.)
KL (N = 120)	.31 (*.46*)	.26 (*.44*)	.23 (*.42*)	.21 (*.41*)	2.80 (n.s.)
Total (N = 720)	.29 (*.46*)	.31 (*.46*)	.22 (*.41*)	.18 (*.38*)	34.68 ($p < .001$)

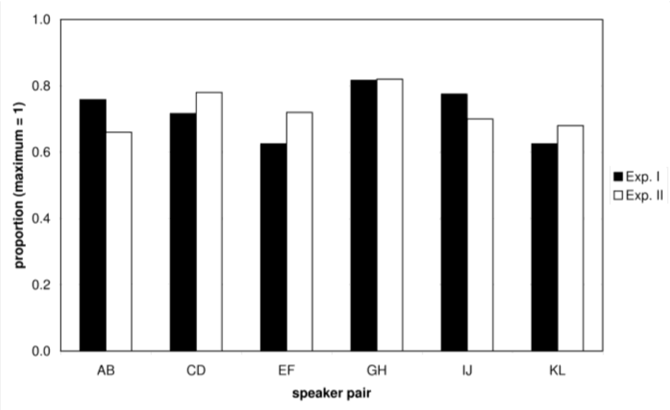

Figure 4.2: *Proportion of choices for functionally coherent dialogues in experiment I and II. Note that values of experiment I are identical to Table 4.4. For experiment II values in Table 4.7 are added up (see section 4.3.8).*

4.3.9 Conclusions

Experiment II shows that speakers are perceived as better adapters when they use prosody functionally coherent. Prosodic form, i.e., whether or not speakers literally copy the accent distribution of their partners', does not have an effect on the subjective impression of the interaction. This result is in line with experiment I, in which it is shown that accentuation is a cue for the adaptation process between interlocutors. Experiment II extends this finding by showing that functional prosody between speakers' successive utterances is a stronger cue for the adaptation process than mere prosodic form.

4.4 Discussion

The present study shows that accentuation patterns that are functionally coherent with respect to the previous speaker's utterance can act as a cue for the adaptation process between interlocutors. That is, when speakers use their accents to mark contrasts with information of their partner's preceding turn, then this leads to an overall impression of the dialogue being coherent, when compared to cases in which the accent distributions are inconsistent with such contrasts. Moreover, the current results indicate that functional prosody, as far as accent distributions are concerned, is a stronger cue for speaker adaptation than formal prosody. Such an outcome is in accordance with studies indicating that speaker adaptation in prosody is more likely when the prosody does relate to the semantics of an utterance (Nilsenová et al., 2009; Okada et al., 2012) compared to when it does not. That is, when compared to formal prosody, functional prosody is a stronger indication that dialogue partners are well adapted to each other.

Both experiment I and II indicate that listeners' adaptation judgments differ per speaker pair. Although listeners were instructed to judge adaptation on the basis of intonation, we cannot rule out the possibility that certain preferences for speaker pairs did play a role in the judgments. Those preferences may relate to the sound of a certain voice or a combination of voices in particular pairs. For example, in the current approach we did not control for pitch range differences between speakers, a feature that often converges to similar levels when speakers interact (Couper-Kuhlen, 1996). Therefore, it may have been the case that certain speaker pairs matched acoustically better than others, which could have caused differences in how participants judged the adaptation process. However, such a conclusion cannot be drawn from the current results. Our findings do suggest that speaker pair differences found in experiment I are grounded in the acoustic realization of pitch accents. This favours the view that the prosodic form did play a minor role as a cue for speaker adaptation. A closer investigation of individual differences is left for future research.

The current findings confirm claims in previous work on the use of accents in discourse. That is, functionally coherent accentuation provides cues for how successive utterances are related (Pierrehumbert & Hirschberg, 1990; Pechmann, 1984a, 1984b, 1989). The structure that can be derived from accents thus is a useful tool for both speakers and listeners to build new information on top of what has been said before. The current study seems to suggest that the clearer the accents are used to signal discourse structure the better the interlocutors are judged to account for each other. Thus, pitch accents signal both structure and speakers' interactional behavior in discourse.

Before concluding, three points about the current methodology should be noted explicitly. First, listeners in experiment II did not show a preference for formally similar dialogues compared to the formally dissimilar dialogues. This does not allow us to conclude that prosodic form is not a cue for speaker adaptation at all. Rather, the outcome may be related to how the variables function and form were manipulated. That is, for both variables the location of the pitch accent was the cue on which listeners should have based their judgments. Typically, the location of pitch accents signals the information status of words rather than whether two speakers sound similar. In other words, adaptation at the level of prosodic form may be less bound to a specific location in an utterance (i.e., Natale, 1975; Gregory, 1986; Giles et al., 1991; Couper-Kuhlen, 1996; Szczepek Reed, 2010) than adaptation at the level of prosodic function. The choice for a manipulation of the pitch accent's location in the current approach may have restrained an effect of prosodic form.

Second, listeners in the current perception experiments are not addressees. In the perception experiment listeners hear the dialogue passively. That is, they are overhearers (Schober & Clark, 1989) and do not actively build common ground with an interlocutor. With respect to this distinction one could think of a variant of the perception experiment in which participants are both interlocutor and judge. For example, they act in a dialogue with a partner that either responds with a coherent or incoherent intonation pattern. Participants may then be expected to judge the adaptation process in a more radical way due to their active role in dialogue. This paradigm is carried out in a follow-up study presented in Chapter 5 (Kaland et al., submitted).

Third, it has to be noted that in the current setup dialogues are reduced to combinations of short utterances (NPs). In spontaneous speech there may be more complex discourse structures and therefore other prosodic cues to partner adaptation. One such cue could be the use of boundary tones. It is known that those tones mark whether the speaker has more to say, asks for response or has finished the utterance. Nilsenová et al. (2009) showed that speakers are likely to adapt to their interlocutor by copying their boundary tone in the course of the interaction between them. With respect to utterance finality, however, copying the interlocutor's boundary tone may not be appropriate in the middle of a dialogue. Future work could investigate, along the lines of experiment II, to what extent speaker adaptation in boundary tones is signaled by its functional or formal use. We conjecture that, as shown in the current study for accentuation, the function of prosody is a stronger cue for the interlocutor's adaptation than the form of prosody.

To conclude, this study has shown that adaptation is signaled merely by speakers' coherence in prosodic function and not by speakers' similarity in prosodic form. This nuances a common definition of prosodic adaptation that relies on the speakers' tendency to copy their interlocutors' prosodic form. A stronger form of adaptation may be found in the way speakers use prosody to signal coherence between their own and their interlocutor's utterances in a functional way.

CHAPTER 5

ADAPTING TO ATYPICAL PROSODY: CONTRASTIVE NOUN PHRASES IN DUTCH AND ITALIAN

ABSTRACT

This chapter investigates to what extent the prosody produced by speakers in a conversation is dependent on the prosody of their interlocutor and to what extent on constraints imposed by the prosodic rules of the speakers' native language. We know from earlier work that speakers may adapt to the pitch level (F0) of their interlocutors. In addition, we know that the speaker's native language requires a certain prosody that may be language-specific, for instance regarding the distribution of pitch accents. In particular, languages differ in the extent to which intonation patterns can be shifted, with so-called plastic languages such as Dutch being more flexible than non-plastic languages such as Italian. In this study we investigate how these differences relate to the extent to which speakers adapt their prosody to their interlocutor. Therefore, a production experiment elicited contrastive noun phrases from Dutch and Italian speakers, interacting with an interlocutor who produced prosodic structures that were either typical or atypical in the language. Analysis of the produced pitch and perceived prominence of the NPs indicated that speakers of Dutch adapt their accent structure to that of an interlocutor, while speakers of Italian do not adapt in this way.

This chapter is based on:

Kaland, C. C. L., Avesani, C., Krahmer, E. J., Swerts, M. G. J., & Zappoli, A. (submitted). Adapting to atypical prosody: Contrastive noun phrases in Dutch and Italian.

5.1 Introduction

When two people interact in a conversation they are likely to adapt to each other. It has been shown that speakers may take over various forms of verbal and non-verbal behavior of their interlocutor (Pickering & Garrod, 2004; Chartrand & Bargh, 1999). For example, when one speaker has the arms crossed, the other is more likely to do so too, which may be seen as a non-verbal way of adaptation. Chartrand and Bargh (1999) refer to this type of copying behavior as the Chameleon-effect and argue that it has a social function. By acting like others people ground themselves in their social environment, a process that is argued to happen unconsciously and automatically. The mechanism behind the Chameleon-effect is assumed to rely on a tight link between perception and behavior: people do what they observe others doing (Bargh & Chartrand, 1999; Chartrand & Bargh, 1999). In the case of the crossed arms, the way people cross their arms does not appear to be determined by specific cultural rules, and does not have an obvious, conventionalized pragmatic function. In that sense, this kind of nonverbal copying behavior is different from other kinds where two or more people copy each other's gestures in a more ritualized manner, such as in the case of greeting, where a person may stretch his or her arm to shake hands with another person who simultaneously produces the same action. Here, someone who appears to copy the behavior of another person is mainly doing this because that person is aware of the culture-specific conventions that determine how one is supposed to greet another person. The fact that one has to know the culture-specific rules of such greeting patterns becomes immediately clear when someone is exposed to a new culture, where one may feel awkward because of a lack of knowledge on how to hug or bow to greet another person.

The current study looks at the extent to which speakers copy each other's prosodic behavior, in particular comparing cases where an interlocutor does or does not adhere to the specific prosodic rules of the language they are speaking. As we will argue in more detail below, the main motivation for conducting this study is that results on the amount of prosodic adaptation can shed light on typological differences in prosodic structure between languages, in particular on the extent to which languages can flexibly adopt these structures.

In general, there is evidence that speakers indeed adapt to each other at the level of prosody, and that they take over their interlocutor's prosody in several ways. For example, when one speaker talks with a soft voice, the interlocutor is likely to adjust the intensity of his or her speech to a lower level (Natale, 1975). Furthermore, in the course of a conversation speakers' pitch ranges tend to converge to similar levels (Gregory, 1986; Couper-Kuhlen, 1996). This means that the vibrations of speakers' vocal folds tend to become similar within certain frequency bands once they engage in a dialogue. In a similar way, interlocutors adjust their speech rate to each other (Giles et al., 1991; Szczepek Reed, 2010). There has been some debate on whether taking over prosodic characteristics of the interlocutor is an unconscious and automatic process. Some researchers have argued that it is indeed largely unconscious and automatic, and have claimed that priming is the main automatic mechanism responsible for this behavior (e.g., Pickering & Garrod, 2004), whereas others argued that the social setting and gender of the interlocutors mediate the degree to which speakers copy each other's prosody (e.g., Pardo, 2006). It has also been argued that copying the interlocutor's prosody has a social function (Shepard et al., 2001; Pardo, 2006), like the behavioral mimicry described by Bargh and Chartrand (1999). Most of these previous studies, however, have focused on features that serve paralinguistic functions of prosody (Ladd, 2008); the variations in pitch range, speech rate and intensity (e.g., Couper-Kuhlen, 1996; Pardo, 2006; but see Ladd, 1993) discussed above generally do not signal the linguistic structure of an utterance, but rather are used to signal information which is more attitudinal or emotional in nature. In other words, copying such kinds of prosodic features in general can be done without having to take into account the semantic or syntactic structure of an utterance. It remains to be investigated to what extent speakers adapt to the prosody of their conversation partner, when dealing with more linguistic uses of prosody, which as we will argue below, are probably more dependent on language-specific constraints on the interface between prosodic structure and other forms of linguistic structure. The next section discusses the distinction between linguistic and paralinguistic prosody in more detail.

5.1.1 Linguistic and paralinguistic prosody

Copying linguistic behavior is argued to play a central role in the acquisition of both first and second languages (e.g., Allott, 2003). Children acquire their mother language in part by imitating the sounds produced by their parents, which is important to learn the rules that determine the prosody of their language. Even though it has been suggested that the basics of the prosodic patterns of a specific language are already acquired while a baby is still in the mother's womb (Partanen et al., 2013), these prosodic structures become more language-specific in the course of child's development, as the result of intensive interaction with other speakers of the same language. In other words, copying behavior is one of the mechanisms behind both the interlocutors' convergence towards a similar prosody (during a conversation) and the acquisition of a native prosody (during childhood). In the present study we argue that the extent to which speakers can prosodically adapt to each other depends on characteristics of the prosodic structure of their native language. As we will argue below, languages can be quite distinct in how strongly they dictate what prosodic structures are allowed in specific contexts. This is especially true for the use of pitch accents, which make specific information acoustically more prominent due to an increase in pitch, duration and intensity (Fry, 1995; De Jong, 1995; Sluijter & Van Heuven, 1996; Turk & Sawusch, 1996; Rietveld & Van Heuven, 2009). For languages like Dutch, pitch accents are typically used to mark which words in an utterance refer to new or contrastive information (i.e., information status), whereas given information is usually deaccented. Thus, the prosodic marking of information status is governed by rules that dictate which words can and which words cannot be realized with a pitch accent in a particular discourse context.

In previous research, we have shown that these rules in Dutch are relevant for questions about prosodic adaptation. In a series of perception experiments in Chapter 4 (Kaland et al., 2013a), the participants' task was to judge whether two speakers producing an utterance pair such as (1) were better adapters than two speakers producing an utterance pair such as (2), where both are English translations of Dutch examples (accented words in capitals).

(1) A: blue TRIANGLE
B: YELLOW triangle

(2) A: blue TRIANGLE
B: yellow TRIANGLE

Note that in (1) the accentuation pattern of B is ‘functionally coherent’, in the sense that the given word (triangle) is not accented, while the contrastive word (yellow) is. In contrast, in example (2) speaker B merely copies the intonation pattern of A’s utterance, irrespective of information structure. It was found that coherent accentuation as in (1) was a stronger cue to speaker adaptation than merely copying the intonation pattern as in (2). These results are in line with Nilsenová et al. (2009) who found that speakers of Dutch are more likely to take over prosodic features that have a linguistic function, such as boundary tones (one of the functions of which is to mark the utterance finality), compared to prosodic features that were not used with a linguistic function (i.e., paralinguistic), such as global pitch range. This preference may be explained on the basis of a cognitive semantic boost, which listeners receive from linguistic prosody and not from paralinguistic prosody (Pickering & Garrod, 2004). That is, for linguistic prosody the semantic level in speech perception gets activated, due to its close relation with the meaning of an utterance. For paralinguistic prosody no such activation occurs, as this type of prosody does not have a linguistic meaning. It was also found that adaptation to boundary tones occurred immediately after the last utterance of the interlocutor (Nilsenová et al., 2009), unlike adaptation to paralinguistic prosody, which usually is more apparent from observing a longer stretch of conversation (Pardo, 2006). Therefore, in the current study, we also take into account whether the degree of adaptation changes during the experiment. That begs the question to what extent prosodic adaptation occurs immediately (e.g., in the initial turns of a spoken interaction between two people), or whether it develops as a conversation progresses. In sum, studies suggest that taking over the linguistic rules behind the interlocutor’s prosody is a stronger form of adaptation than taking over paralinguistic prosodic features.

5.1.2 Prosody across languages

It is important to stress that both the experiments in Chapter 4 (Kaland et al., 2013a) and the ones of Nilsenová et al. (2009) have studied Dutch, which shares the characteristic with other Germanic languages that pitch accents serve to mark important information, but this use of prosody does not necessarily generalize to other languages. Indeed, the literature has shown that there are substantial differences in this respect between so-called plastic and non-plastic languages (Vallduví, 1991). For example, consider the NP *yellow car* in two different languages like Dutch (*gele auto*) and Italian (*macchina gialla*); note that the word order in Dutch and Italian is reversed in these NPs. The default way to indicate which words in an NP refer to new information depends on the phonological possibilities of the language. In Dutch and Italian the rightmost word in the phrase is considered as the phonological head and is therefore typically pitch accented in neutral contexts (Nespor, 1993; Nespor & Vogel, 2008; Truckenbrodt, 2007). In these contexts, both languages also allow a pitch accent on the preceding word in the NP (Swerts et al., 2002; Rasier & Hiligsmann, 2007). The plasticity difference between Germanic languages like Dutch (plastic) and Romance languages like Italian (non-plastic) lies in the ability to shift accents to a non-default position in a phrase, for example to mark a semantic contrast. To illustrate the plasticity difference, consider a situation in which two speakers are talking about cars. The first speaker refers to a black car after which the second speaker refers to a yellow car. In a plastic language like Dutch, a speaker generally utters the latter referring expression with a single pitch accent on the word that indicates the semantic contrast with the previous utterance (i.e., the adjective *gele*), see also example (1) in Chapter 4 (Kaland et al., 2013a) mentioned above and Krahmer and Swerts (2001). However, in a non-plastic language like Italian both the noun *macchina* and the adjective *gialla* generally receive a pitch accent. This pattern is used irrespective of the (contrastive) information status of the words in the NP (for Italian see Swerts et al.

2002; for a similar phenomenon in Romanian see Swerts, 2007). It has to be noted that a pattern in which the adjective is accented and not the noun is allowed by the intonational grammar of Italian. However, behavioral data shows that this pattern does not occur in contrastive contexts in experimentally elicited utterances in paradigms like the one used in Swerts et al. (2002). In general, it is not the case that Italian cannot express semantic contrasts, but it achieves this through different strategies. Within an NP such as *macchina gialla*, Italian speakers could, but not necessarily have to, use a specific type of pitch accent on the word that expresses contrastive information (Bocci & Avesani, 2006). At a sentence level, the same goal is achieved by variation in word order as well as by shifting pitch accents. Plastic languages, on the contrary, generally use a more fixed word order (Vallduví, 1991). Thus, while intonation patterns of NPs in Dutch can be shifted to express the discourse status of words, this would be highly marked in Italian.

The prosodic differences between languages become especially clear when considering speakers learning a second language. When native speakers of Italian learn Dutch, one of the aspects they need to acquire is prosody. It has been shown that a speaker's native prosody influences the prosody produced in the second language (for Dutch and French, see Rasier et al., 2010; for Dutch and Spanish, see Van Maastricht, Krahmer & Swerts, submitted; for Italian and German see Avesani et al., in press). This process is called prosodic transfer, in that the speaker's native prosody is transferred to the second language. Prosodic transfer may result in the production of atypical intonation patterns, which may be less or not at all acceptable for native listeners (Mennen, 2007; Swerts & Zerbian, 2010). It may be that an Italian native speaker who utters a Dutch NP, produces a pitch accent on the noun in a situation where the adjective expresses contrastive information or vice versa. Or it may be that a Dutch native speaker who utters an Italian NP produces one pitch accent instead of two pitch accents. Such non-native intonation patterns can have negative consequences for speech perception, given that listeners have been shown to be sensitive to pitch accents in atypical positions in Germanic languages. In particular, Dutch listeners have been shown to process words faster when pitch accents match the given-new distinction of information status compared to when pitch accents do not match this distinction (Terken & Nooteboom, 1987; Birch & Clifton, 1995). Speech perception even seems to be sensitive to different types of atypicality in the placement of pitch accents, as shown by Hruska et al. (2001). In their ERP-study, native speakers of German exhibited larger mismatch negativities for accents that were missing than for accents that were superfluous at a certain sentence location.

The crosslinguistic differences just discussed may have interesting consequences when it comes to prosodic adaptation. Imagine a speaker of Dutch learning Italian (or vice versa) who uses atypical prosody because this speaker does not yet master the second language. As we will argue below, this situation challenges the adaptive behavior of speakers, with possible differences in the extent to which speakers of Dutch and Italian adapt to the atypical prosody of such a speaker. Given that speakers have the tendency to take over their interlocutors' form of prosody (Pardo, 2006), the question arises whether speakers would still copy prosodic characteristics of an interlocutor who uses atypical prosody, which is especially interesting from a crosslinguistic perspective. Roughly, there are two options for speakers whose interlocutors use atypical prosody. On the one hand, speakers could indeed adapt their prosody by taking over the atypical pattern. This possibility is supported by studies indicating that speaker adaptation in prosody is an automatic process (Pickering & Garrod, 2004; Pardo, 2006). Similarly, there is evidence at the lexical level that speakers tend to take over less idiomatic terms (i.e., *tires* instead of *wheels*) if these were introduced by a non-native interlocutor (Bortfeld & Brennan, 1997). On the other hand, speakers may detect that the interlocutor uses contextually atypical pitch accents (Hruska et al., 2001) and therefore hold on to the demands of their native intonational grammar. As these demands differ across languages, one is likely to find differences in the extent to which speakers of different languages adapt to atypical prosody. One could argue that, because a plastic language like Dutch is able to shift pitch accents to non-default positions, speakers of Dutch are more likely to adapt to atypical prosody compared to speakers of Italian. In a similar vein, it could be that Italian speakers are less sensitive to deviations from default accent patterns. As we know from segmental phonology, speakers may loose their ability to hear certain phonological contrasts if that contrast does not bear any functional load in their native language, and consequently become unable to also produce that contrast. A classical example is the contract between /l/ and /r/ in English, which is hard to perceive *and* produce for speakers of Japanese (e.g., Goto, 1971). Likewise, it is known that speakers of Italian are insensitive to eyebrow movements that mark prominence within NPs, while such eyebrow movements do matter for Dutch speakers (Krahmer & Swerts, 2004), suggesting that Italian speakers (unlike Dutch speakers) do not interpret eyebrow movements as markers of contrastive information. Along the same lines, if pitch accent distributions in an Italian NP are not exploited for marking information status, it could be that speakers of that language find it harder to perceive such pitch accent distributions, and cannot easily 'copy' them into their own speech productions. Theoretically, one could also argue for the opposite, in that speakers of Italian are more likely to adapt to a speaker who uses atypical prosody compared to speakers of Dutch. In Italian pitch accents at non-default locations are less likely to harm the semantic content of an utterance compared to Dutch. Therefore speakers of Italian could take over atypical prosody without communicative consequences, whereas speakers of Dutch cannot do so. Finally, when plasticity and prosodic adaptation would be unrelated phenomena, it could be expected that both speakers of a plastic language and speakers of a non-plastic language are similar in the extent to which they adapt to the atypical prosody of their interlocutor. These research questions will be investigated in the current study.

5.1.3 Research goals

To sum up the issues just discussed, the general question addressed in the current study is whether speakers of Dutch and speakers of Italian differ in the extent to which they adapt to an interlocutor who uses atypical prosody. This question is investigated by means of two production experiments carried out with speakers of Dutch and speakers of Italian respectively, who were interacting with an interlocutor and describing figures in contrastive discourse contexts. In one condition the prosody of the confederate was typical according to the native intonational grammar.

On the basis of typical prosody only, we would be unable to tell to what extent languages differ in prosodic adaptation, since it is hard to find out whether speakers produce prosodic patterns because they adapt to their partners, or because they simply follow the rules of their language. A second condition in which the interlocutor uses prosody which is atypical in both Dutch and Italian allows for such a crosslinguistic comparison. Therefore such a condition is added to both the Dutch and Italian experiments. The NPs collected in each experiment were analyzed in terms of production and perception measures of pitch and prominence respectively. In this way, the current study sheds more light on the way prosody is produced, both in interaction and crosslinguistically, and how different levels of prosody relate to each other.

5.2 Method

5.2.1 Participants

A total of 20 native speakers of Dutch carried out the experiment (all women, M_{age} = 21.9, age range = 18-24). For Italian, 20 native speakers carried out the experiment (6 men, 14 women, M_{age} = 21.7, age range = 19-27). None of the participants had hearing problems or was color blind.

5.2.2 Design

To elicit utterances a referential communication task was carried out in the form of a bingo game. In this game a confederate and a participant speaker instructed each other in turn (as players) to put figures on a piece of paper (the bingo card, see Figure 5.1). The confederate and participant had different bingo cards displaying commonly used objects (e.g., ball, bicycle, umbrella). Their task was to instruct each other to put a figure on their bingo card until the figures cover a certain pattern (e.g., a complete row) on the bingo card (example: “put the green triangle on the ball”). Figures occurred in different shapes and colors such that for Dutch the shapes *driehoek*, *druppel*, *kano* and *klaver* (triangle, drop, canoe and clover) in the colors *blauw*, *geel*, *groen* and *rood* (blue, yellow, green and red) were used and for Italian the shapes *croce*, *luna*, *mela*, and *stella* (cross, moon, apple and star) in the colors *giallo*, *nero*, *rosso* and *verde* (yellow, black, red and green) were used. In order to obtain a homogeneous collection of NPs suitable for acoustic analysis all the words referring to the shapes and colors were bisyllabic with lexical stress on the first syllable in both languages. Instructions to place a figure on the bingo card were given by the confederate and the participant in turn. Both the participant and the confederate executed each instruction by placing the relevant object on the relevant figure. Each participant played a bingo game that consisted of four different rounds. Per round the confederate and the participants each produced four utterances. Whenever one of the players had a bingo the round ended. Players played for a different pattern each round.

The order of instructions which the confederate and the participant gave each other was manipulated in such a way that two successive utterances referred to figures that could be distinguished by only one property. This property was either the color or the shape of the figure and was defined with respect to the previous utterance. For example, when the confederate was the instructing player an utterance could be "put the green triangle on the ball". Thereafter, the participant was the instructing player and could utter "put the blue triangle on the bicycle" (i.e., a color contrast). In one half of the stimuli the contrastive property was the color and in the other half of the stimuli the contrastive property was the shape. Instructions consisted of NPs with one adjective and one noun referring to the color or shape of the target figure respectively. The task was carried out with the confederate who produced typical prosody for one half of the participants and atypical prosody for the other half of the participants. For the typical prosody, the Dutch confederate produced pitch accents on words referring to contrastive information and deaccented words referring to given information. The Italian confederate produced the typical prosody with a pitch accent on both words in the NP, a pattern produced in all discourse contexts in Swerts et al. (2002). As for the atypical prosody, the confederates of both languages produced pitch accents on words referring to given information and deaccented words referring to contrastive information (see Table 5.1).

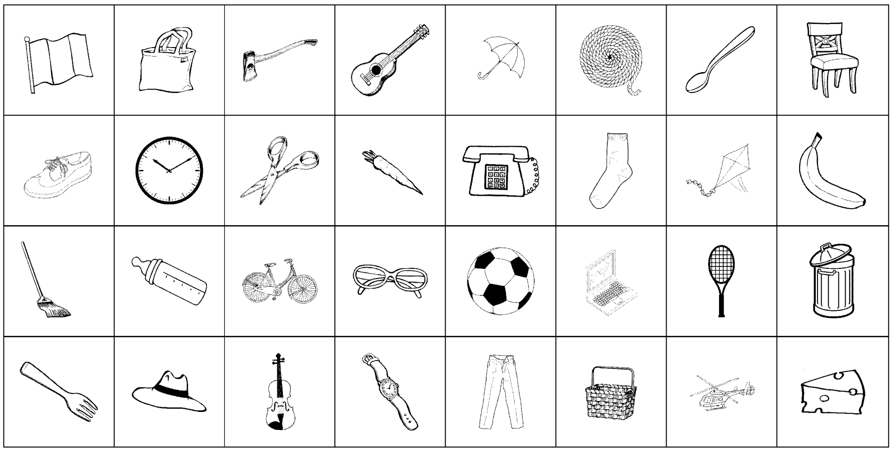

Figure 5.1: *Example of a bingo card displaying commonly used objects.*

Table 5.1: *Stylized pitch contours on NPs (Dutch: blue canoe and Italian: black apple) produced by the Dutch and Italian confederates. Note that the word order in Dutch NPs is adjective-noun, whereas in Italian NPs this is noun-adjective.*

Confederate prosody	Contrastive word	Dutch		Italian	
		BLAUWE	KANO	MELA	NERA
Typical	Adjective	----/^\-----------		----/^\-----/^\---	
	Noun	-------------/^\---		----/^\-----/^\---	
Atypical	Adjective	-------------/^\---		---/^\----------	
	Noun	----/^\-----------		----------/^\--	

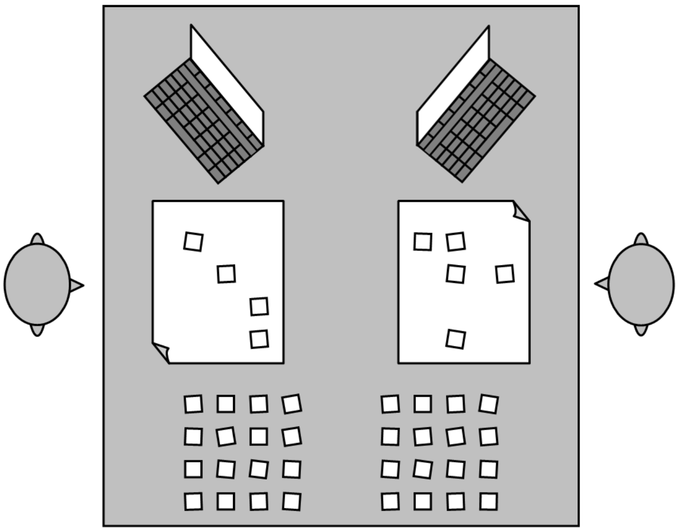

Figure 5.2: *Birdseye view of the experimental setup with the confederate and participant at opposite sides of a table facing each other.*

5.2.3 Procedure

The confederates were instructed about the goal and course of the experiment. There was one Dutch and one Italian confederate. As a training, the actual experiment was simulated with the confederates, separately for each language. This was done to make the confederate familiar with the experimental setup. In the actual experiment, confederate and participants were seated at opposite sides of a table facing each other (Figure 5.2). Both had access to a laptop computer that displayed the turn in the game (Figure 5.3) for which they were the instructing player. The participant's laptop displayed a visual representation of the target figures. The confederate's laptop showed a written description of the target figure. For the condition with atypical prosody, the description showed the word to be accented as underlined and in upper case and showed the word to be deaccented not underlined and in lower case. This was done to facilitate the confederate's production of atypical prosody. Before the start of the experiment the confederate and the participant were both instructed about the course of the game. Note that the confederate acted as if she heard the instructions for the first time. This was done to support the participants' idea that they were playing against another naïve participant. The first round of the game was a training round to make the participants familiar with the course of the experiment. Thereafter four rounds were played. When the game was finished, participants were debriefed. None of the participants reported any suspicions about the experimental setup and all believed they were interacting with another participant. Speech of both players was recorded digitally and saved as a wave-file.

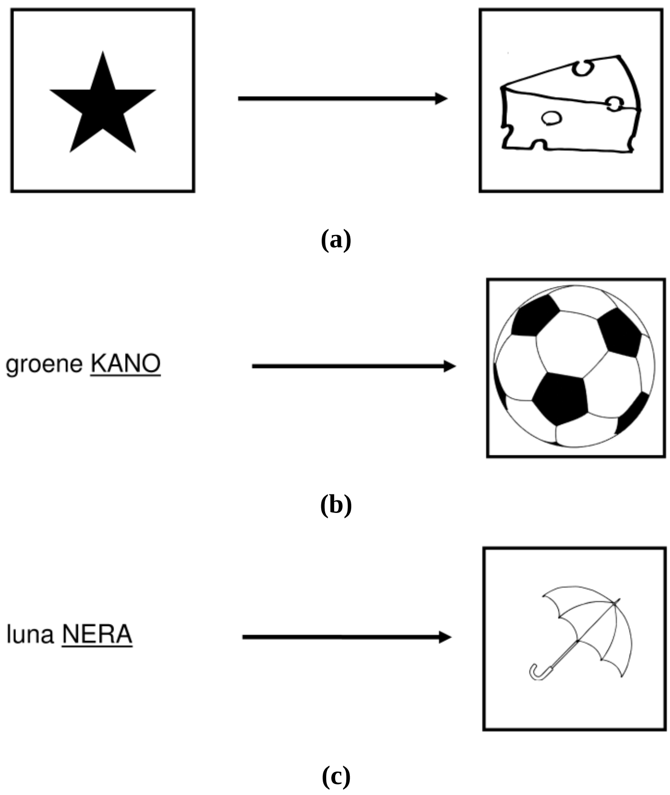

Figure 5.3: *Examples of turns displayed by the laptop computer to the instructing player. Panel (a) displays a visual representation as seen by the participants. Panels (b) and (c) display a written description in Dutch and Italian respectively indicating to the confederate which word in the condition with atypical prosody should have been accented.*

5.2.4 Data processing: confederate prosody

In this section an acoustic analysis of the confederates' prosody is described in order to investigate whether the confederates produced the intended prosody. The confederate speaker of Dutch was a 20 year old female. The confederate speaker of Italian was a 26 year old female. The confederate speakers were selected on the basis of whether they exhibited a clear intonation and articulation.

The total amount of NPs produced by both confederates in the production experiments of Dutch and Italian taken together was 640. Half of the confederates' NPs (160 in each language) were not taken into account as they did not refer to figures that contrasted with previously mentioned figures. These NPs were fillers referring to entirely new colors or shapes and occurred either at the beginning or in the middle of a round. Thus, for each confederate 160 NPs were analysed.

The acoustic analysis concerns a production and a perception measure of prominence. As for production, we measured the maximum pitch (F0) of the first syllable of the adjective and noun in the NPs produced in the experiment using Praat (Boersma & Weenink, 2011). Pitch is chosen as a correlate of prominence (Ladd, 2008) and is perceptually most closely approximated by ERB measurements (Glasberg & Moore, 1990).

In order to verify the ERB measures perceptually a prominence rating task was also carried out for each language. In this task NPs were presented in a web-based task using WWStim (Veenker, 2003) to three intonation experts of Dutch and three intonation experts of Italian who were asked to listen to NPs that were produced in their native language. The intonation experts all had an understanding of prominence and pitch accents at an academic level. They rated the strength of the accent on a three point scale (0 = no accent, 1 = weak accent, 2 = strong accent). Adjectives were rated in the first part of the task, nouns were rated in the second part. Experts heard the entire NP in both parts. The presentation order of NPs was randomized so that experts were blind for condition. To abstract over the experts' ratings, the prominence scores per word were added up so that they ranged from 0 to 6 (0 when all experts rated the accent as absent, 6 when all experts rated the accent as strong).

It is known that prominence is perceived with respect to surrounding material in a phrase (Gussenhoven et al., 1997). Therefore, a difference score was computed for both the pitch measures and the prominence ratings by subtracting the score of the semantically non-contrastive word from the score of the semantically contrastive word, a method previously used in Chapter 4 (Kaland et al., 2013a) and Chapter 2 (Kaland et al., 2013b) for Dutch. This method obtains positive scores when contrastive words have a higher pitch and are perceived as more prominent than non-contrastive words. Such an outcome could be expected for Dutch, whereas for Italian adjective and noun may have similar pitch and prominence levels when the NP is double accented, similar to Swerts et al. (2002), probably resulting in difference scores close to zero. Pearson's correlation coefficient showed that pitch measures and prominence ratings correlated for Dutch: $r(160) = .75$, $p < .001$, and for Italian: $r(160) = .60$, $p < .001$.

A repeated measures analysis of variance (RM-ANOVA) was carried out on the pitch and prominence difference scores with confederate prosody (2 levels: typical, atypical) as between-subjects factor and with contrastive word (2 levels: adjective, noun) as within-subjects factor separately for Dutch and Italian. Confederate prosody had a significant effect on the pitch differences scores for Dutch: $M_{\text{typ}} = .67$, $M_{\text{atyp}} = -.43$, $F(1,18) = 260.89$, $p < .001$, $\eta_p^2 = .94$ and for Italian: $M_{\text{typ}} = -.04$, $M_{\text{atyp}} = -1.11$, $F(1,18) = 342.79$, $p < .001$, $\eta_p^2 = .95$ as well as on the prominence difference scores for Dutch: $M_{\text{typ}} = 4.97$, $M_{\text{atyp}} = -4.90$, $F(1,18) = 3590.82$, $p < .001$, $\eta_p^2 = .99$ and for Italian: $M_{\text{typ}} = -.06$, $M_{\text{atyp}} = -5.26$, $F(1,18) = 1262.59$, $p < .001$, $\eta_p^2 = .99$. This indicates that the confederates produced the typical and atypical prosody as intended (see also Table 5.2). Contrastive word had a significant effect on the pitch differences scores in Dutch: $M_{\text{adj}} = .23$, $M_{\text{noun}} = .01$, $F(1,18) = 7.14$, $p < .05$, $\eta_p^2 = .28$ and for Italian on both the differences scores of pitch: $M_{\text{adj}} = -.77$, $M_{\text{noun}} = -.37$, $F(1,18) = 66.36$, $p < .001$, $\eta_p^2 = .79$ and prominence: $M_{\text{adj}} = -1.28$, $M_{\text{noun}} = -4.04$, $F(1,18) = 195.88$, $p < .001$, $\eta_p^2 = .92$. The effects of contrastive word confirm the differences between Dutch and Italian in the condition with typical prosody. For Dutch, shifting a pitch accent to a non-default position has a larger impact on its acoustic realization, resulting in larger pitch and prominence

Table 5.2: *Means and standard deviations for pitch maxima per word, pitch differences in ERB, prominence per word and mean prominence differences of the NPs produced by the confederate.*

Language	Confederate prosody	Contrastive word	Pitch maximum		Pitch difference	Prominence		Prominence difference
			Adjective	Noun		Adjective	Noun	
Dutch	Typical	Adjective	6.10 (*0.30*)	5.26 (*0.85*)	0.84 (*0.85*)	5.03 (*0.89*)	0.20 (*0.48*)	4.83 (*1.23*)
		Noun	5.45 (*0.22*)	5.96 (*0.17*)	0.51 (*0.23*)	0.08 (*0.27*)	5.18 (*0.56*)	5.10 (*0.71*)
	Atypical	Adjective	5.46 (*0.23*)	5.84 (*0.15*)	-0.38 (*0.29*)	0.17 (*0.46*)	5.07 (*0.52*)	-4.90 (*0.66*)
		Noun	6.13 (*0.26*)	5.65 (*0.30*)	-0.48 (*0.27*)	5.02 (*0.74*)	0.12 (*0.39*)	-4.90 (*0.95*)
Italian	Typical	Adjective	5.86 (*0.22*)	5.80 (*0.20*)	0.05 (*0.23*)	3.37 (*0.96*)	0.83 (*0.91*)	2.53 (*1.31*)
		Noun	5.89 (*0.26*)	5.77 (*0.26*)	-0.12 (*0.23*)	3.50 (*0.89*)	0.84 (*0.84*)	-2.66 (*1.21*)
	Atypical	Adjective	5.19 (*0.19*)	6.79 (*0.22*)	-1.59 (*0.27*)	0.77 (*0.77*)	5.87 (*0.51*)	-5.10 (*0.88*)
		Noun	6.43 (*0.24*)	5.81 (*0.27*)	-0.62 (*0.29*)	5.78 (*0.55*)	0.36 (*0.56*)	-5.42 (*0.76*)

differences for focused adjectives than for focused nouns. For Italian, the pitch and prominence difference between adjective and noun in absolute terms is similar regardless of contrastive word. As for the typical prosody, positive difference scores are found for focused adjectives and negative difference scores are found for focused nouns, indicating that the most emphasis was produced and perceived on the adjective in all conditions. In sum, the confederates' prosody was produced as intended both in Dutch and Italian with the expected differences between the languages and between the levels of confederate prosody (typical and atypical).

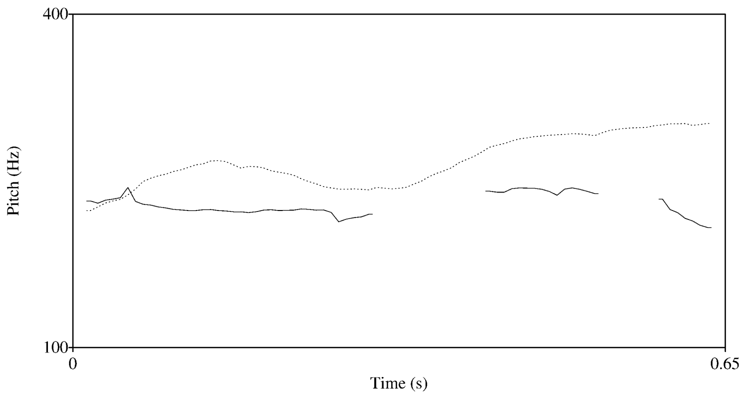

Figure 5.4: *Pitch contours of default intonation patterns produced on "blauwe drupel" (blue drop) by a female Dutch speaker (solid line) and on "mela nera" (black apple) by a female Italian speaker (dotted line).*

5.2.5 Data processing: participants' prosody

The total amount of NPs produced by participants of Dutch and Italian taken together was 640. As some participants by mistake used words not referring to the target figure's color and/or shape, these NPs were not taken into account for analysis (this concerns 11 NPs produced by 8 Dutch participants and 4 NPs produced by 4 Italian participants). Thus, in total the acoustic analysis concerned 309 NPs from Dutch participants and 316 NPs from Italian participants.

The acoustic analysis of pitch and prominence for the participants' data was identical to the analysis described for the confederates' data. It appeared from the acoustic analysis that Dutch and Italian speakers differed in the use of high boundary tones on the last syllable of the target NP (Figure 5.4). For Dutch, a high boundary tone was produced on 66/309 NPs (participants) and 54/320 NPs (confederate), whereas for Italian a high boundary tone was produced on 266/316 NPs (participants) and 167/320 NPs (confederate). Similar observations were done in other studies comparing Dutch with a Romance language like Romanian (Swerts, 2007) or Spanish (Van Maastricht, Krahmer & Swerts, 2013). The differences could be related to varying ways of prosodic phrasing (Ladd, 2008) as some languages favor smaller phrases (e.g., Egyptian Arabic: Hellmuth, 2007) than others (e.g., Lisbon Portuguese: Elordieta et al., 2003). This particular difference in the use of boundary tones seems worth studying in future work. In the current acoustic analysis the use of high boundary tones was abstracted over by taking pitch measurements only from the accented (non-final) syllable of the last word in the target NPs.

Pearson's correlation coefficient showed that pitch measures and prominence ratings correlated for Dutch: $r(309) = .57$, $p < .001$, and for Italian: $r(316) = .53$, $p < .001$. Note that these correlations are lower than those reported for the confederates, which presumably is due to the respective homogeneity of the datasets. In particular, the confederates' data consisted of speech from one speaker whereas the participants' data consisted of speech from 20 speakers.

Furthermore, the collected NPs of the participants were labeled according to the moment they were uttered in the experiment. This means that the NPs were divided over two time blocks corresponding with the first two rounds and the last two rounds of the bingo game respectively. In this way the statistical analysis allows us to investigate whether possible effects of prosodic adaptation were immediate or occurred in the course of the interaction.

5.2.6 Statistical analysis

Repeated measures analyses of variance (RM-ANOVAs) were performed on the participants' pitch and prominence difference scores separately for Dutch and Italian, with confederate prosody (2 levels: typical, atypical) as between-subjects factor and with time block (2 levels: first, second) and contrastive word (2 levels: adjective, noun) as within-subjects factors.

For a comparative analysis between the Dutch and the Italian results, separate RM-ANOVAs were carried out for the typical and atypical prosody. This was done to distinguish between results that were expected to differ between the two languages (typical prosody) and results that were not necessarily expected to differ (atypical prosody). The RM-ANOVAs were carried out on the participants' pitch and prominence difference scores, with language (2 levels: Dutch, Italian) as between-subjects factor and time block (2 levels: first, second) and contrastive word (2 levels: adjective, noun) as within-subjects factors.

5.3 Results

This section discusses the results on Dutch and Italian in their respecetive order.

5.3.1 Dutch

The results of the Dutch participants (Table 5.3, Table 5.4) showed an effect of confederate prosody for pitch difference scores (M_{pi}) and a trend for the prominence difference scores (M_{pr}) in that difference scores were smaller when participants interacted with a confederate who used atypical prosody (M_{pi} = .04, M_{pr} = 1.11) compared to when participants interacted with a confederate who used typical prosody (M_{pi} = .16, M_{pr} = 1.88). Time block showed no effects. Furthermore, contrastive word had an effect on both pitch and prominence in that difference scores were higher when the adjective was the contrastive word (M_{pi} = .33, M_{pr} = 2.93) compared to when the noun was the contrastive word (M_{pi} = -.12, M_{pr} = .06). The interaction between time block and contrastive word was found significant for the pitch difference scores in that they were negative and decreasing when the noun was the contrastive word (M_{block1} = -.08, M_{block2} = -.16) and positive and increasing when the adjective was the contrastive word (M_{block1} = .28, M_{block2} = .37). This indicates that in the second time block pitch differences were larger in absolute terms and that overall the adjective was produced with a higher pitch when compared to the noun, irrespective of which word was contrastive.

Table 5.3: *Means and standard deviations for pitch maxima per word, pitch differences in ERB, prominence per word and mean prominence differences of the NPs produced by participants.*

Language	Confederate prosody	Contrastive word	Pitch maximum		Pitch difference	Prominence		Prominence difference
			Adjective	Noun		Adjective	Noun	
Dutch	Typical	Adjective	6.61 (*0.63*)	6.23 (*0.62*)	0.38 (*0.43*)	4.19 (*1.15*)	1.01 (*1.07*)	3.17 (*1.86*)
		Noun	6.41 (*0.50*)	6.37 (*0.50*)	-0.04 (*0.41*)	2.24 (*1.65*)	3.01 (*1.87*)	0.78 (*3.35*)
	Atypical	Adjective	6.57 (*0.59*)	6.28 (*0.58*)	0.29 (*0.42*)	4.01 (*1.31*)	1.06 (*1.07*)	2.95 (*1.94*)
		Noun	6.49 (*0.56*)	6.30 (*0.60*)	-0.20 (*0.42*)	2.75 (*1.86*)	2.12 (*1.49*)	-0.63 (*3.07*)
Italian	Typical	Adjective	5.88 (*0.98*)	5.70 (*1.02*)	0.18 (*0.48*)	4.10 (*1.10*)	1.16 (*0.98*)	2.94 (*1.44*)
		Noun	5.93 (*1.05*)	5.68 (*1.04*)	-0.24 (*0.42*)	3.99 (*0.94*)	1.24 (*1.15*)	-2.75 (*1.51*)
	Atypical	Adjective	5.88 (*1.07*)	5.66 (*1.00*)	0.22 (*0.37*)	3.91 (*1.07*)	1.28 (*1.20*)	2.63 (*1.42*)
		Noun	5.90 (*1.02*)	5.73 (*1.00*)	-0.17 (*0.33*)	3.83 (*1.08*)	1.41 (*1.09*)	-2.42 (*1.48*)

Table 5.4: *Results of RM-ANOVAs per independent variable on the participants' pitch and prominence difference scores split for language. Interactions that are not reported in the table are not significant.*

Language	Factor	Pitch	Prominence
Dutch	Confederate prosody	$F(1,18) = 5.65$, $p < .05$, $\eta_p^2 = .24$	$F(1,18) = 3.56$, $p = .08$, $\eta_p^2 = .17$
	Time block	n.s.	n.s.
	Contrastive word	$F(1,18) = 86.41$, $p < .001$, $\eta_p^2 = .83$	$F(1,18) = 40.58$, $p < .001$, $\eta_p^2 = .69$
	Time block * Contrastive word	$F(1,18) = 6.10$, $p < .05$, $\eta_p^2 = .25$	n.s.
Italian	Confederate prosody	n.s.	n.s.
	Time block	$F(1,18) = 9.37$, $p < .01$, $\eta_p^2 = .34$	$F(1,18) = 4.07$, $p = .06$, $\eta_p^2 = .18$
	Contrastive word	$F(1,18) = 14.65$, $p < .01$, $\eta_p^2 = .45$	$F(1,18) = 295.60$, $p < .001$, $\eta_p^2 = .94$
	Conf. prosody * Time block	$F(1,18) = 3.49$, $p = .08$, $\eta_p^2 = .16$	n.s.

5.3.2 Italian

The Italian participants showed no effect of confederate prosody on both the pitch and the prominence difference scores. In Table 5.3 and Table 5.4 it can be observed that the difference scores were not significantly affected by whether the confederate used typical or atypical prosody. Time block had a significant effect on the pitch difference scores and showed a trend for the prominence difference scores. This result indicates that differences scores are negative in the first time block ($M_{pi} = -.06$, $M_{pr} = -.07$) and positive in the second time block ($M_{pi} = .04$, $M_{pr} = .23$). Furthermore, contrastive word had a significant effect on both the pitch and prominence difference scores in that they were positive when the adjective was the contrastive word (M_{pi} = .18, M_{pr} = 2.71) and negative when the noun was the contrastive word (M_{pi} = -.19, M_{pr} = -2.62). This indicates that in Italian the adjective was produced with a higher pitch and perceived as more prominent compared to the noun, irrespective of which word was contrastive. In absolute terms, the difference scores were similar, which indicates that both the adjective and the noun were accented. The interaction between confederate prosody and time block revealed a trend for the pitch difference scores, in that these varied to a larger extent across time block when the confederate used typical prosody (M_{block1} = -.11, M_{block2} = .05) compared to when the confederate used atypical prosody (M_{block1} = .00, M_{block2} = .04).

5.3.3 Comparative analysis

In this section the results on typical prosody for both Dutch and Italian are discussed first, after which the results on atypical prosody of both languages are discussed.

5.3.3.1 Typical prosody

The data of both Dutch and Italian participants who interacted with a confederate who used typical prosody showed an effect of the factor language on both the pitch and prominence difference scores (Table 5.3, Table 5.5). That is, difference scores were larger in Dutch (M_{pi} = .16, M_{pr} = 1.88) compared to Italian (M_{pi} = -.03, M_{pr} = .02). Time block showed no significant effects. The effect of contrastive word was found significant, in that both pitch and prominence difference scores were positive for focused adjectives (M_{pi} = .26, M_{pr} = 2.92) and negative for focused nouns (M_{pi} = -.13, M_{pr} = -1.03) in both languages. Furthermore, the interaction of language and time block was significant in that the pitch difference scores were larger and positive for Dutch (M_{block1} = .19 and M_{block2} = .14) and smaller and negative for Italian (M_{block1} = -.11 and M_{block2} = .05). The interaction between language and contrastive word was found significant for the prominence difference scores, in that for Dutch these were larger for focused adjectives (*M* = 2.98) than for focused nouns (*M* = .78) whereas for Italian focused adjectives (*M* = 2.92) and focused nouns (*M* = -2.77) resulted in difference scores of similar size in absolute terms. In addition, the interaction between time block and contrastive word showed a trend for both the pitch and prominence difference scores. In particular, when the adjective was focused difference scores were more variable over time blocks (M_{pi_block1} = .20, M_{pi_block2} = .33, M_{pr_block1} = 2.74, M_{pr_block2} = 3.16) compared to when the noun was focused (M_{pi_block1} = -.12, M_{pi_block2} = -.15, M_{pr_block1} = -.93 and M_{pr_block2} = -1.07).

5.3.3.2 Atypical prosody

When participants interacted with a confederate who used atypical prosody (Table 5.3, Table 5.5) a significant effect of the factor language is found for the prominence difference scores. This effect showed larger difference scores for Dutch participants (*M* = 1.11) than for Italian participants (*M* = .07). No significant effects of time block were found. Furthermore, the effect of contrastive word was found significant for both the pitch and prominence in that the difference scores were larger (and positive) when the adjective was the contrastive word (M_{pi} = .25, M_{pr} = 2.72) compared to the negative values when the noun was the focused word (M_{pi} = -.19, M_{pr} = -1.54). The interaction language and time block revealed a trend for the prominence difference scores in that for Dutch these showed a decrease across time blocks (M_{block1} = 1.33, M_{block2} = .90) and for Italian these showed an increase across time blocks (M_{block1} = -.06, M_{block2} = .22). Additionally, the interaction language and contrastive word revealed a trend for the prominence difference scores in that for Dutch these were larger for focused adjectives (*M* = 2.88) than for focused nouns (*M* = -.65) whereas for Italian focused adjectives (*M* = 2.57) and focused nouns (*M* = -2.40) resulted in difference scores of similar size in absolute terms. The interaction time block and contrastive word was found significant for the pitch difference scores, in that the adjective was produced with a higher pitch than the noun and that this difference was smaller in the first time block ($M_{adjective}$ = .19, M_{noun} = -.18) compared to second time block ($M_{adjective}$ = .31, M_{noun} = -.19).

Table 5.5: *Results of RM-ANOVAs per independent variable on the participants' pitch and prominence difference scores of both Dutch and Italian NPs taken together and split for confederate prosody. Interactions that are not reported in the table are not significant.*

Confederate prosody	Factor	Pitch	Prominence
Typical	Language	$F(1,18) = 15.53$, $p < .01$, $\eta_p^2 = .46$	$F(1,18) = 26.86$, $p < .001$, $\eta_p^2 = .60$
	Time block	n.s.	n.s.
	Contrastive word	$F(1,18) = 22.91$, $p < .001$, $\eta_p^2 = .56$	$F(1,18) = 106.70$, $p < .001$, $\eta_p^2 = .86$
	Language * Time block	$F(1,18) = 7.30$, $p < .05$, $\eta_p^2 = .29$	n.s.
	Language * Contrastive word	n.s.	$F(1,18) = 20.89$, $p < .001$, $\eta_p^2 = .54$
	Time block * Contrastive word	$F(1,18) = 2.29$, $p = .07$, $\eta_p^2 = .17$	$F(1,18) = 2.95$, $p = .10$, $\eta_p^2 = .14$
Atypical	Language	n.s.	$F(1,18) = 12.31$, $p < .01$, $\eta_p^2 = .41$
	Time block	n.s.	n.s.
	Contrastive word	$F(1,18) = 33.65$, $p < .001$, $\eta_p^2 = .65$	$F(1,18) = 118.59$, $p < .001$, $\eta_p^2 = .87$
	Language * Time block	n.s.	$F(1,18) = 3.23$, $p = .09$, $\eta_p^2 = .15$
	Language * Contrastive word	n.s.	$F(1,18) = 3.43$, $p = .08$, $\eta_p^2 = .16$
	Time block * Contrastive word	$F(1,18) = 4.77$, $p < .05$, $\eta_p^2 = .21$	n.s.

5.4 Conclusions

5.4.1 Dutch

The results indicate that Dutch participants' prosody, as measured by pitch and prominence difference scores, was affected by the confederates' prosody. The speakers' prosody changed as a function of whether the confederate used typical or atypical prosody. Participants produced utterances with smaller pitch differences and with smaller perceived prominence differences when they interacted with a confederate who used atypical prosody. In particular, the difference scores for the adjectives became smaller for the setting with atypical prosody. The difference scores of the nouns showed less variation in size when compared to the difference scores of the adjective. Nouns showed negative difference scores in the atypical condition. This indicates that when the noun was the contrastive word participants produced adjectives with a higher pitch and these adjectives were perceived as more prominent compared to the noun. Furthermore, time block showed no effects which indicated that the extent to which participants adapted their prosody did not differ between the two blocks.

5.4.2 Italian

The prosody of Italian participants was not significantly affected by the prosody of the confederate. In other words, the Italian speakers produced essentially the same prosodic patterns, irrespective of whether the confederate produced speech with typical or atypical prosody. Their difference scores were affected by time block, however only significantly for the pitch difference scores. It has to be noted that the differences in absolute terms were similar across time blocks in Italian. Furthermore, the differences scores of around zero were plausibly the result of the double accentuation in Italian, which was used in all discourse contexts. Therefore, a small deviation from this pattern, either resulting in positive or negative difference scores, could have been enough for a significant effect of time block in Italian. In Dutch however, there is *a priori* more variation due to the shift of single pitch accents. Therefore, the effect of time block in Italian may be the result of the way this language uses linguistic prosody rather than the result of adaptation processes, especially when compared with the results of Dutch. Pitch and prominence differences had about the same size for focused adjectives and focused nouns in absolute terms. However, when looking at the direction of the difference (i.e., positive or negative values) the results show negative values for focused nouns, not for focused adjectives. This indicates that the adjective was the word with the highest pitch and the most prominence irrespective of which word was contrastive. In sum, these results confirm that Italian uses double accentuation in NPs with the rightmost pitch accent being the most prominent. This pattern does not differ as a function of semantic contrasts nor as a result of adaptation to the interlocutor's prosody.

5.4.3 Comparative

The participants' data confirm the differences between Dutch and Italian in typical prosody. That is, the larger pitch and prominence differences found for Dutch are the result of single accentuation on either the adjective or noun, whereas in Italian smaller difference scores reveal that both adjective and noun were accented. Furthermore, the (interaction) effects of contrastive word indicates that in both Dutch and Italian a focused adjective results in larger difference scores compared to a focused noun. For Dutch such an outcome confirms that shifting accents to a non-default position (the adjective) leads to the perception of more prominence compared to accents in a default position (Krahmer & Swerts, 2001). For Italian, the results confirm that the adjective generally is the most prominent of the two accents in an NP. The interaction between language and time block also confirms that in Dutch the focused word has a higher pitch and is perceived as being more prominent than the unfocused word as shown by the positive difference scores across all time blocks. Italian, however, has smaller pitch and prominence differences confirming that both words in the NP received an accent during the experiment. Time block did not show any consistent (interaction) effects across languages.

The effect of the factor language on the participants' prominence difference scores confirms that Dutch and Italian participants reacted differently to the atypical prosody of the confederate. The difference between focused adjectives and focused nouns, as indicated by the significant effects of contrastive word, is again confirmed in the participants' data in the atypical prosody.

5.5 Discussion

This study has shown that speakers of Dutch and Italian differ in adaptation to atypical prosody. Dutch speakers adapt their contrastive intonation to an interlocutor who uses atypical prosody by producing utterances with smaller differences in produced pitch and perceived prominence between accented and unaccented words, when compared to the typical prosody. Italian speakers use an accent on both the adjective and the noun irrespective of whether the interlocutor uses typical or atypical prosody. These results confirm the hypotheses that speakers can adapt their prosody to an interlocutor who uses atypical prosody and that the extent to which they do so depends on the plasticity of the language.

It has to be noted that the atypical prosody produced by the confederates was different for Dutch and Italian (Table 5.2). The main difference was in the size of the difference scores for pitch and prominence, which were smaller for Dutch than for Italian. This was confirmed by informal reports of the confederates. In particular, the Italian confederate reported to have produced a corrective-like pitch accent on either the noun or the adjective in the condition with atypical prosody. This type of accent may have been more exaggerated than the accents in the atypical prosody of the Dutch confederate, for whom it was more natural to produce single accents on non-default positions. However, it is important to note that if this difference would have biased the current results, Italian participants should have shown *larger* difference scores in their prosody when interacting with an interlocutor who used atypical prosody compared to when interacting with an interlocutor who used typical prosody. The fact that the difference scores for Italian participants were similar in both conditions indicates that the difference in atypical prosody did not have this biasing effect.

The plasticity of a language may explain the current results. It is not the case that Italian speakers were more likely to take over the atypical prosody because of the small communicative consequences it would have had in their language; i.e., the alternative hypothesis given in the introduction. More plausibly, Italian speakers were insensitive for the (atypical) prosodic marking of semantic contrasts by means of pitch accent distribution, like they were insensitive for eyebrow movements in these contexts (Krahmer & Swerts, 2004). This explanation can be seen as a direct result of the characteristics of the Italian prosodic system with regard to NPs, which does not allow to shift pitch accents to non-default locations. More concretely, the current results can be explained by the fact that the atypical prosody had a different perceptual effect in Dutch and Italian. In particular, in Dutch the atypical prosody was inappropriate in the context at hand, but could have been appropriate in a different discourse context. Thus, in the condition with atypical prosody the Dutch confederate consequently marked semantic contrasts that did not exist in the experiment, which could have made participants reconsider previous semantic contrasts and be more hesitant when producing contrastive intonation themselves. Italian participants, however, may not have been sensitive for the atypical prosody of the confederate and therefore had no reason to reconsider previous semantic contrasts. It has to be noted that in the current analysis only the pitch accent *distribution* was analysed. This method has been proven successful in earlier work (Krahmer & Swerts, 2001; Swerts et al., 2002). However, since Italian allows for a different *type* of pitch accent in contrastive contexts, future work could investigate whether Italian speakers select a different type of pitch accent in settings where they have to adapt their prosody to their interlocutors.

In general, time block did not show consistent effects on the pitch and prominence differences. We did not find any evidence that the Dutch speakers adapted to a more or lesser extent depending on the amount of time that elapsed in the interaction with their interlocutor. Therefore, it seems that the effects of adaptation were immediate. Such a conclusion is in line with work showing that linguistic prosody has an immediate effect on perception (i.e., Steinhauer et al., 1999). This is also in line with plasticity differences between the tested languages which are per definition related to the linguistically relevant prosody (Vallduví, 1991). The type of adaptation found in this study is therefore different from adaptation of paralinguistic prosody, such as taking over the pitch range of the interlocutor, which typically arises in the course of a dialogue (Gregory & Hoyt, 1982).

The adaptation difference between Dutch and Italian also shows that the linguistic and the paralinguistic levels of prosody interact. Specifically, flexibility in terms of shifting linguistically relevant pitch accents turns out to affect speakers' adaptation to the surface form of their interlocutors' prosody. This suggests that it is indeed useful to distinguish these two levels of prosody. However, a strict separation of the two levels does not hold, as there is interaction between them. A model of speech production that distinguishes phonological encoding from mere articulation (Levelt, 1989; 1999) should allow for spreading of features like 'prosodic flexibility' from the former to the latter. This nuances the interactive alignment account of Pickering and Garrod (2004; 2013), where alignment on one language level leads to alignment at other language levels. As this study has shown, the interaction between two levels of prosody within a speaker, and in particular the possibilities of a speaker's native prosody, also determine to what extent that speaker adapts to the interlocutors' speech.

As mentioned earlier, behavioral copying mechanisms may underlie learning a native (linguistic) prosody as well as taking over phonetic features (paralinguistic) of the interlocutor. It is therefore crucial to note that the current adaptation differences between Dutch and Italian do not generalize to language learning in general. Thus, we cannot conclude on the basis of the current results that Dutch speakers learn languages or prosody more easily, because they adapt to atypical prosody more than Italian speakers. To acquire a foreign language a speaker needs to learn other linguistic features such as a lexicon, morphology and syntactic constructions. The current results only hold for the level of prosody and show that the degree of prosodic adaptation depends on the speaker's native intonation patterns.

CHAPTER 6

GENERAL CONCLUSION AND DISCUSSION

6.1 Conclusions

This thesis investigated to what extent speakers adapt to their interlocutor in the use of a specific type of linguistic prosody: contrastive intonation. On the one hand we investigated to what extent speakers account for the knowledge state of their addressee when using contrastive intonation. This was investigated in Chapter 2 and Chapter 3 which both reported experiments with speakers describing figures to their addressees. Both experiments were one-directional in the sense that the speakers were the only persons describing while the addressees only had to perform some action with the figure described by the speaker, but did not have to produce any descriptions themselves. In both studies we used settings in which a speaker had access to more or different kinds of information than the addressee, either because the addressee could not see parts of the scene that a speaker was describing, or because the addressee had not witnessed the preceding discourse that the speaker had produced.

On the other hand we investigated to what extent prosodic adaptation between interlocutors is a result of them trying to make the produced accent patterns consistent with the functional use of those accents, or a result of interlocutors merely copying each other's prosodic patterns without taking into account the functions of these patterns. This was investigated in Chapter 4 and Chapter 5, which, unlike the two preceding studies, consisted of dialogues that were more interactive in nature with interlocutors describing figures in turns. The study in Chapter 4 consisted of a perception experiment in which participants were asked to rate various recorded and manipulated dialogue fragments between two speakers. Chapter 5 presented a cross-linguistic study that compared adaptive processes in Dutch and Italian speakers.

Since this thesis consists of a collection of individual essays, the findings of each individual study were also discussed in the preceding chapters. Here we briefly discuss the research questions presented above and their theoretical implications. Thereafter, a general discussion is given as well as suggestions for future research.

6.2 Prosodic marking of semantic contrasts: speaker- versus addressee-related factors

6.2.1 RQ1: To what extent is contrastive intonation speaker- or addressee-driven? (Chapter 2)

The first study addressed the question to what extent contrastive intonation is speaker- or addressee-driven. For this purpose, a typically-developing group of speakers and a group of speakers diagnosed with high functioning autism (HFA) were compared. Both groups participated in a communicative task in which sequences of figures had to be described to different addressees. The order in which the addressees were addressed by the speaker was manipulated in such a way that in the critical trials one figure was described to one addressee and the subsequent one (which contained a semantic contrast) to the other. In this way, the critical property (e.g., the color or the shape of a specific figure) was contrastive for the speaker but not the addressee. The results of this study indicated that speakers produce contrastive intonation less clearly when the semantic contrast is not shared with their addressee, compared to when the semantic contrast is shared with their addressee. Since speakers still produced some form of prosodic marking (i.e., attenuated contrastive intonation) when the contrast was not shared with the addressee, we concluded that speakers account for both their own and their addressee's knowledge state. We assumed that HFA speakers would have had more difficulties to account for the knowledge state of their addressee because of their impaired Theory of Mind (Baron-Cohen, 1995). We found no evidence to support this hypothesis, however, because our HFA speakers were similar to typically-developing speakers in their use of contrastive intonation. However, the speech of HFA participants was found to differ from that of typically-developing speakers, in particular where the pitch range and perceived speech dynamicity was concerned. This study showed that it is essential to distinguish aspects of prosody that have a linguistic function such as the prosodic marking of contrastive information, from aspects of prosody that do not have a linguistic function, such as with the way pitch range and speech dynamicity varied in this study. Note that this is not say that pitch range can never be used for linguistic purposes, since it has for instance been shown that variations in pitch range can be used to demarcate syntactic units (Ladd, 1988).

6.2.2 RQ2: To what extent do speakers leak contrastive information to their addressees by their use of prosody? (Chapter 3)

In this study, speakers saw four figures of which one was occluded for the addressee. The occluded figure contrasted with a unique target figure that was visible for both the speaker and the addressee. In earlier work, Wardlow Lane et al. (2006) showed that speakers used more adjectives referring to such contrasts when explicitly instructed *not* to leak information about the occluded figure. This type of instructions has been shown to make certain information cognitively extra salient (Wegner et al., 1987). Our study showed that speakers prosodically mark contrasts in a visual scene that were cognitively extra salient for them and which were unavailable for their addressee, when compared to contrasts in a visual scene that were not extra salient and shared with the addressee. This study showed that speakers produce some form of prosodic marking even when not all information is shared with the addressee. This prosodic marking is most plausibly the result of speaker-internal ironic processes (e.g., Wegner et al., 1987). Presumably, the instructions made semantic contrasts extra salient for speakers in such a way that speakers were more likely to refer to these contrasts prosodically when compared to a condition in which ironic processes were not evoked.

The first two studies (Chapter 2 and Chapter 3) both show that the form of the contrastive intonation pattern is at least determined by the assumptions of the speaker about the addressee’s knowledge state. When the speaker knows that certain information is not available for the addressee contrastive intonation is produced in an attenuated way. In this sense, both studies have shown that the production of pitch accents to mark semantic contrasts depends on both speaker- and addressee related factors. Such a conclusion is in line with previous research on speaker adaptation that showed that speakers account for the perspectives of both themselves and their addressees when producing utterances (Brennan & Hanna, 2009; Galati & Brennan, 2010). As most clearly shown in Chapter 2, the prosodic marking of semantic contrasts appears to be a composite result of speakers' perspective on the discourse, and their assumptions about what the addressees' perspective is on this discourse. Therefore, even in cases where this does not seem warranted from the point of view of the addressee, speakers produce intonation patterns that still mark the contrastive information, albeit in a less clear way when compared to situations in which there is a match between these two perspectives. In situations where contrasts in the speaker's private knowledge state are made extra salient (Chapter 3), speakers have a harder time finding this balance between the different perspectives on contrastive information, and even tend to exaggerate the prosodic marking for cases in which the addressee has no access to the semantic contrast.

Models on speaker adaptation processes (e.g., Pickering & Garrod, 2004) generally only consider global features of prosody, without a linguistic function, such as how speakers converge on pitch range. The current thesis explicitly focused on the linguistic use of prosody (contrastive intonation) and showed that speakers also adapt this type of prosody to account for the knowledge state of the listener. The next section discusses the second research question of this thesis on how adaptation by means of linguistically functional prosody relates to adaptation by means of copying prosodic features.

6.3 Prosodic marking of semantic contrasts and prosodic adaptation processes

6.3.1 RQ3: To what extent is contrastive intonation a cue for speaker adaptation? (Chapter 4)

Chapter 4 showed that contrastive intonation in Dutch is a cue for speaker adaptation. We carried out a perception experiment in which listeners were presented with manipulated dialogue segments. These segments consisted of interactions between two speakers who referred to contrastive information. In one half of the stimuli the speaker pairs used contrastive intonation in a coherent way for Dutch, by accentuating new or contrastive information and deaccentuating given information. In the other half of the stimuli the speakers copied each other's intonation pattern in contrastive contexts and could therefore violate the coherent use of contrastive intonation. Results showed that speaker pairs that matched in their linguistically coherent use of contrastive intonation were perceived as better adapters than interlocutors that copied each other's intonation pattern. Therefore, we concluded that the functional use of prosody plays a more important role in speaker adaptation than accent patterns that were copied from the interlocutor and did not mark contrasts across speaker turns.

6.3.2 RQ4: To what extent do speakers of Dutch and Italian adapt to atypical prosody in contrastive contexts? (Chapter 5)

In Chapter 5 we compared speakers of Dutch and Italian, because these languages are claimed to differ in the way they use pitch accents to mark information status (Vallduví, 1991). Dutch, a so-called plastic language, allows speakers to shift pitch accents to non-default locations to mark semantic contrasts, whereas Italian, a so-called non-plastic language, does not allow speakers to shift pitch accents in such a way, since in Italian pitch accents occur on the same phrasal locations irrespective of whether there is contrastive information or not. Speakers of Dutch and Italian were compared regarding the extent to which they adapted their prosody to interlocutors who use prosody in an atypical way. One half of the speakers interacted with a (Dutch or Italian) confederate who used prosody in accordance with the rules of prosodic marking of information status in the language of interest (typical prosody). For the other half of the speakers the confederate used prosody which was not in accordance with those rules (atypical prosody). Results showed that speakers of Dutch, which allows shifting pitch accents in a flexible way, adapted to atypical prosody to a larger extent than speakers of Italian, who use pitch accents in a more fixed way.

The outcomes of the studies in Chapter 4 and Chapter 5 highlight the relevance of the distinction between aspects of prosody that have a linguistic and aspects of prosody that do not have a linguistic function. In particular, the linguistic functionality of prosody was shown to be a crucial factor for the degree of adaptation between speakers. This appeared from the study in Chapter 4 which investigated the perception of prosody and from the study in Chapter 5 that investigated the production of prosody. Prosody had a linguistic function in both studies in that it signaled coherence between utterances in half of the interactions presented to listeners in Chapter 4 and in half of the interactions of Dutch speakers in Chapter 5. In the other conditions of these studies, in which the accent distributions did not serve this linguistic function of marking semantic contrasts, as in the incoherent interactions in Chapter 4 and the interactions of Italian speakers in Chapter 5, the degree of (judged) adaptation was smaller or adaptation did not occur at all.

It has to be stressed that the outcomes of the studies in Chapter 4 and Chapter 5 supplement earlier claims on speaker adaptation in prosody, and we may even need to nuance (implicit) claims made in the literature on prosodic adaptation. Traditionally, adaptation in prosody has been studied as a process whereby global prosodic features (such as pitch range or loudness) are copied, without taking into account whether such features served a specific linguistic function (Gregory & Hoyt, 1982; Pickering & Garrod, 2004; Pardo, 2006). In this thesis we have shown that, in order to understand prosodic adaptation between conversation partners, we need to take into account the linguistic use of prosody. It is known that speakers automatically converge to a similar surface form of prosody, most likely for social reasons (Pardo, 2006). However, from the current results it appears that prosodic convergence is just one way of adaptation, and as far as pitch accents are concerned, it appears to be less important than adaptation by means of the functional use of prosody. Therefore, future investigations of adaptation in prosody should take into account the linguistic functionality of prosody.

Taking all studies in this thesis together, we can conclude that the way speakers mark information structure prosodically by means of contrastive intonation, is highly determined by the interaction with the interlocutor. That is, speakers adapt their prosody depending on their assumptions about the knowledge state of the addressee as well as their own (Chapter 2 and Chapter 3). Furthermore, we find a relation between prosodic adaptation and the extent to which prosody is used in a functional way. In particular, the degree of perceived (Chapter 4) or produced

(Chapter 5) adaptation depends on the extent to which prosody contributes to the linguistic meaning of utterances.

As discussed in the introduction, speaker's intentions play an important role in the production of contrastive intonation (Ladd, 2008). In this thesis we have shown that speakers' intentions are best understood when speaker- and addressee-perspective are torn apart, as done most clearly in the studies presented in Chapter 2 and Chapter 3. Plausibly, the speakers estimate the discourse coherence required by both their own and their addressees' knowledge state. When a semantic contrast is unavailable for the addressee (Chapter 2) there is less need for the speaker to mark the contrast prosodically. A similar result was found in the study presented in Chapter 5, where speakers of Dutch interacting with a confederate who used atypical prosody, also produced contrastive intonation less clearly. The atypical prosody of the confederate crucially did not mark the semantic contrasts as required in Dutch and therefore did not signal the coherence between two successive utterances. Either consciously or unconsciously, speakers in the study presented in Chapter 5 may have interpreted the atypical prosody of the confederate as a decreased need for coherence. In turn, speakers themselves produced an attenuated form of contrastive intonation.

In sum, the prosodic marking of information status can be seen as a way in which speakers adapt to their addressees. We have shown that this way of adaptation is language dependent and stronger than copying prosodic features of the interlocutor. We have argued that this is the case because the prosodic marking of information status closely relates to the semantics of utterances, which are at the core of mutual understanding between speaker and addressee in a conversation.

6.4 Perspectives for future research

6.4.1 Limitations

The studies conducted for this thesis have certain limitations, which we briefly want to highlight here, since they could be addressed in follow-up research. First, in the current thesis we only investigated pitch, because this is usually considered a major acoustic correlate of prosodic prominence (e.g., Ladd, 2008). However, it would be interesting to also consider adaptation for other acoustic correlates such as intensity or duration. One reason why it would be interesting to conduct such additional analyses would be that there is some debate as to whether or not these variables are also good indicators of prosodic prominence (Kochanski et al., 2005). If they are as important as pitch, one would expect similar patterns in adaptive behaviour as reported in the thesis, and variation in loudness and durational patterns to be determined by functional considerations. If these variables are less clear indicators of prominence, then speakers may be more flexible in how they copy such patterns, as those would then not interfere with the linguistic functions of prominence.

Furthermore, in the current study the main focus was on Dutch, and it would be interesting to investigate other languages as well. The study in Chapter 5 compared Dutch with Italian as these two languages are different in the prosodic marking of contrastive information, in particular analyzing variation in pitch. However, to show that these findings are not tied to this particular language pair, it would be interesting to repeat this study with a different language pair. It would seem useful to broaden the scope of these analyses to include other languages. For instance, it is known that speakers of Bulgarian, a Slavic language, mark semantic contrasts in prosody and generally use intensity to a larger extent in the production of prominence when compared to speakers of other languages (Andreeva et al., 2013). Therefore, future work in speaker adaptation in prosody could take into account Bulgarian to investigate how speakers of this language compare to speakers of Dutch. For example, it can be expected that in Bulgarian prosodic adaptation by means of contrastive intonation manifests itself more in intensity cues compared to Dutch.

Additionally, in the study presented in Chapter 5 we carried out an experiment using confederates. Confederates are frequently used in experimental studies, but recently there has been some discussion about possible, undesired side-effects that confederates might have (see, e.g., Kuhlen & Brennan, 2013). For example, participants may become suspicious about the confederate’s behavior during an experiment (Kuhlen & Brennan, 2013); in theory, these could have led to some experimental artefacts, and speakers might have shown forms of prosodic adaptation that are not representative of interactions with genuine participants. Although in the study in Chapter 5 we asked participants whether they had any suspicions (none of the participants reported having them), it would be useful to replicate this study using more spontaneous dialogue without confederates in order to check to what extent results will differ when compared to a study using confederates. Next to these general suggestions for follow-up research, we highlight two issues in particular for further study in the next sections.

6.4.2 Default versus non-default locations

Chapter 2, Chapter 4 and Chapter 5 have also shown that in Dutch, pitch accents on default locations differ from pitch accents on non-default locations. In particular, these studies showed that differences between the accented word and the deaccented word were larger (i.e., higher difference scores) when the adjective referred to contrastive information than when the noun referred to contrastive information. This appeared from the difference score of both the produced pitch (F0) and the perceived prominence. It is known that the lineair location of pitch peaks in an utterance, as well as information structural expectations on the location of pitch accents, influence the perceived prominence (Gussenhoven et al., 1997; Cole et al., 2010; Bishop, 2012). Therefore, the acoustic realization of pitch accents, as measured in pitch height, may not always correspond with the perceived prominence. In our studies, however, we found that the height of the pitch peaks on the words in the NP correlates with the perceived prominence. Note, however, that the correlation was not strong, suggesting that other factors such as intensity or duration might also play a role. The correlation suggests that pitch accents on non-default locations are different from pitch accents on default locations and, crucially, both in production (pitch) and perception (prominence). Specifically, pitch accents on non-default locations (the adjective in Dutch) are more prominent than their default-located counterparts, because the pitch accent's production is different in terms of pitch height.

It was beyond the scope of this thesis to investigate the exact relation between categories of information status and to what extent they are marked with different pitch accents. At the discourse level, two pitch accents are claimed to be different whenever *both* their phonetic realization *and* their meaning differs (e.g., Pierrehumbert & Hirschberg, 1990). That two pitch accents are sometimes hard to distinguish is clear from the discussion on the alleged existence of contrastive accents that have a unique melodic shape, when compared to pitch accents as markers of new information (e.g., Krahmer & Swerts, 2001; Watson et al., 2008). While the current thesis focused on pitch accents in contrastive contexts only, our results suggest interesting directions for future research on pitch accents. In particular, the pitch accent difference related to the location in the NP found in Chapter 2, Chapter 4 and Chapter 5 hints at the possibility that Dutch speakers select a different pitch accent on adjectives than on nouns within NPs that refer to contrastive information. Our results show that the acoustics of contrastive adjectives and contrastive nouns differ in terms of production and perception. In addition, it is known that contrastive adjectives and contrastive nouns have different functions. Generally, nouns refer to a gestalt (Levelt, 1989) and adjectives refer to certain properties of that gestalt. Thus, both the acoustics of the pitch accents and the meaning of adjectives and nouns in contrastive NPs differ.

More research is needed to investigate whether speakers of Dutch indeed use different pitch accents *in order to* convey the different meanings of adjectives and nouns in contrastive contexts. This issue needs more investigation because pitch accents are traditionally claimed to signal information status at the discourse level and not at the word level (Pierrehumbert & Hirschberg, 1990). Future work should investigate whether the pitch accent difference between adjectives and nouns is limited to contrastive NPs or occurs in NPs that have a different information status as well. In addition, future research should also take into account other phonetic details of the pitch accent, such as intensity, duration, the specific shape of the F0 contour and the timing of the F0 maximum.

6.4.3 Pitch accent variability

Furthermore, the current thesis supports the idea that pitch accent distribution is variable, in the sense that pitch accents are sometimes less clearly realized (see also Swerts & Marsi, 2012; Swerts & Zerbian, 2010). Traditional analyses of intonation investigated how the placement of pitch accents follows from the semantics, syntax or pragmatics of utterances (i.e., Gussenhoven; 1983; Selkirk, 1984; Pierehumbert & Hirschberg, 1990; Terken & Hirschberg, 1994; Calhoun, 2009). The aim of these analyses often was to find the (linguistic) rules that govern the way pitch accents are distributed, in terms of their presence or absence at certain sentence locations. The assumptions of such approaches tended to be quite deterministic in nature, in the sense that they clearly prescribe what accent patterns are needed for specific utterances. In contrast, in the current thesis, the less clearly realized pitch accents occurred especially in situations where there was a mismatch between the information available to the speaker and the information available to the addressee (Chapter 2 and Chapter 5). The attenuated accentuation pattern produced by the speaker seems therefore the outcome of both speaker- and addressee-related discourse factors, rather than a result of purely semantic, syntactic or pragmatic factors. In the current thesis we have shown that in these mismatching situations the realization of pitch accents can be variable. Future accounts that model pitch accents should therefore allow for this variability. One of the requirements to the experimental paradigm in future studies is therefore to tear apart speaker- and addressee-perspectives and investigate both the production and the perception of pitch accents, as done in the studies presented in this thesis.

REFERENCES

Allott, R. (2003). Imitation in language and speech: Roles and functional base. *Proceedings AISB Convention* University of Aberystwyth, UK, 105-112.

Andreeva, B., Wolska, M., Koreman, J., & Barry, W. (2013). Production of prosodic prominence across languages. *Proceedings pS-prominenceS: Prominences in Linguistics*, Viterbo, Italy.

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington: APA.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Washington: APA.

Audacity Team. (2006). *Audacity* (version 1.2.6). Retrieved from http://audacity.sourceforge.net.

Avesani, C., Bocci, G., Vayra, M., & Zappoli, A. (2013). All'interfaccia tra prosodia e struttura informativa. La realizzazione prosodica dell'informazione data in tedesco e in italiano L2. In V. Galatà (Ed.), *Multimodalità e Multilingualità la Sfida più Avanzata della Comunicazione Orale*, 53-71. Rome: Bulzoni Editore.

Baltaxe, C. (1984). Use of contrastive stress in normal, aphasic, and autistic children. *Journal of Speech and Hearing Research*, *27*, 97-105.

Baltaxe, C. A. M., & Guthrie, D. (1987). The use of primary sentence stress by normal, aphasic and autistic children. *Journal of Autism and Developmental Disorders*, *17*(2), 255-271.

Baltaxe, C. A. M., & Simmons, J. Q. (1985). Prosodic development in normal and autistic children. In E. Schopler & G. B. Mesibov (Eds.), *Communication Problems in Autism*, 95-125. New York: Plenum.

Bard, E. G., Anderson, A. H., Sotillo, C. F., Aylett, M., Doherty-Sneddon, G., & Newlands, A. (2000). Controlling the intelligibility of referring expressions in dialogue. *Journal of Memory and Language*, *42*(1), 1-22.

Bard, E. G., & Aylett, M. P. (2000). Accessibility, duration, and modelling the listener in spoken dialogue. *Proceedings Götelog 2000, Fourth workshop on the semantics and pragmatics of dialogue*, Göteborg, Sweden.

Bargh, J. A., & Chartrand, T. L. (1999). The unbearable automaticity of being. *American Psychologist*, *54*(7), 462-479.

Baron-Cohen, S. (1995). *Mindblindness: An essay on autism and theory of mind*. Cambridge, MA.: MIT Press.

Baron-Cohen, S. (2001). Theory of mind and autism: A review. *International Review of Mental Retardation*, *23*, 169-184.

Barr, D. J., & Keysar, B. (2007). Perspective taking and the coordination of meaning in language use. In M. J. Traxler & M. A. Gernsbacher (Eds.), *Handbook of psycholinguistics* (2 ed.), 901-938. New York: Academic Press.

Benda, U. von. (1983). Zur auditiven Beurteilung der Intonation autistischer Kinder: Ein Hörexperiment. In D.-W. Allhoff (Ed.), *Mündliche Kommunikation, Störungen und Therapie* (Vol. 10). Frankfurt am Main: Scriptor.

Birch, S., & Clifton Jr., C. (1995). Focus, accent, and argument structure: Effects on language comprehension. *Language and Speech*, *38*(4), 365-391.

Bishop, J. (2012). Information structural expectations in the perception of prosodic prominence. In G. Elordieta & P. Prieto (Eds.), *Prosody and Meaning (Interface Explorations)*. Berlin: Mouton de Gruyter.

Bocci, G., & Avesani, C. (2006). Focus contrastivo nella periferia sinistra della frase: Un solo accento ma non solo un accento. In R. Savy & C. Crocco (Eds.), *Analisi prosodica. Teorie , modelli e sistemi di annotazione. Atti del II Convegno Nazionale AISV 2005* (Vol. 2): Associazione Italiana di Scienze della Voce.

Boersma, P., & Weenink, D. (2011). *Praat: Doing phonetics by computer* (version 5.2.19). retrieved from http://praat.org.

Bolinger, D. (1972). Accent is predictable (if you're a mind-reader). *Language*, *48*(3), 633-644.

Bonneh, Y. S., Levanon, Y., Dean-Pardo, O., Lossos, L., & Adini, Y. (2011). Abnormal speech spectrum and increased pitch variability in young autistic children. *Frontiers in Human Neuroscience, 4*, 237.

Bortfeld, H., & Brennan, S. E. (1997). Use and acquisition of idiomatic expressions in referring by native and non-native speakers. *Discourse Processes*, *23*(2), 119-147.

Braun, B., & Tagliapietra, L. (2010). The role of contrastive intonation contours in the retrieval of contextual alternatives. *Language and Cognitive Processes, 25*(7-9), 1024-1043.

Breckenridge, J. (1977). The declination effect. *Journal of the Acoustical Society of America*, *61*(S1), S90.

Brennan, S. E., & Clark, H. H. (1996). Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory and Cognition, 22*(6), 1482-1493.

Brennan, S. E., & Hanna, J. E. (2009). Partner-specific adaptation in dialog. *Topics in Cognitive Science, 1*(2), 274-291.

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

Brown, P. M., & Dell, G. S. (1987). Adapting production to comprehension: The explicit mention of instruments. *Cognitive Psychology*, *19*(4), 441-472.

Bryant, G. A., & Fox Tree, J. E. (2005). Is there an ironic tone of voice? *Language and Speech*, *48*(3), 257-277.

Burnham, D., Kitamura, C., & Vollmer-Conna, U. (2002). What's new, pussycat? On talking to babies and animals. *Science, 296*, 1435.

Caspers, J. (2003). Local speech melody as a limiting factor in the turn-taking system in Dutch. *Journal of Phonetics, 31*(2), 251-276.

Calhoun, S. (2009). What makes a word contrastive? Prosodic, semantic and pragmatic perspectives. In D. Barth-Weingarten, N. Dehé & A. Wichmann (Eds.), *Where prosody meets pragmatics: Research at the interface* (Studies in Pragmatics 8), 53-78. Bingley: Emerald.

Chafe, W. (1976). Givenness, contrastiveness, definiteness, subjects, topics, and point of view. In C. N. Li (Ed.), *Subject and Topic*, 25-55. New York: Academic Press.

Chafe, W. (1984). Language and consciousness. *Language, 50*(1), 111-133.

Chartrand, T. L., & Bargh, J. A. (1999). The chameleon effect: The perception-behavior link and social interaction. *Journal of Personality and Social Psychology, 76*(6), 893-910.

Chevallier, C., Noveck, I., Happé, F., & Wilson, D. (2011). What's in a voice? Prosody as a test case for the Theory of Mind account of autism. *Neuropsychologia, 49*(3), 507-517.

Clark, H. H. (1973). The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior*, *12*(4), 335-359.

Clark, H. H. (1996). *Using language*. Cambridge; New York: Cambridge University Press.

Clark, H. H., & Brennan, S. A. (1991). Grounding in communication. In L. B. Resnick, J. M. Levine & S. D. Teasley (Eds.), *Perspectives on socially shared cognition*, 127-149. Washington: APA Books.

Clark, H. H., & Murphy, G. L. (1982). Audience design in meaning and reference. In J.-F. L. Ny & W. Kintsch (Eds.), *Language and comprehension*. New York: North-Holland.

Clark, H. H., & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition*, *22*(1), 1-39.

Cole, J., Mo, Y., & Hasegawa-Johnson, M. (2010). Signal-based and expectation-based factors in the perception of prosodic prominence. *Laboratory Phonology*, *1*(2), 425-452.

Coulston, R., Oviatt, S., & Darves, C. (2002). Amplitude convergence in children's conversational speech with animated personas. *Proceedings International Conference of Spoken Language Processing*, Denver (CO), USA, 2689-2692.

Couper-Kuhlen, E. (1984). A new look at contrastive intonation. In R. Watts & U. Weidman (Eds.), *Modes of Interpretation: Essays presented to Ernst Leisi*, 137-158. Tübingen: Gunter Narr Verlag.

Couper-Kuhlen, E. (1996). The prosody of repetition: On quoting and mimicry. In E. Couper-Kuhlen & M. Selting (Eds.), *Prosody in Conversation: Interactional Studies*, 366-405. Cambridge: Cambridge University Press.

Davies, C., & Katsos, N. (2013). Are speakers and listeners 'only moderately Gricean'? An empirical response to Engelhardt et al. (2006). *Journal of Pragmatics*, *49*(1), 78-106.

Dell, G. S., & Brown, P. M. (1991). Mechanisms for listener-adaptations in language production: Limiting the role of the 'model of the listener'. *Bridges between psychology and linguistics: A Swarthmore Festschrift for Lila Gleitman*, 105-129. Hillsdale, NJ: Lawrence Erlbaum.

DePape, A.-M. R., Chen, A., Hall, G. B. C., & Trainor, L. J. (2012). Use of prosody and information structure in high functioning adults with autism in relation to language ability. *Frontiers in Psychology, 3*, 72.

Diehl, J., Watson, D., Bennetto, L., McDonough, J., & Gunlogson, C. (2009). An acoustic analysis of prosody in high-functioning autism. *Applied Psycholinguistics, 30*(3), 385-404.

Dijksterhuis, A., & Bargh, J. A. (2001). The perception-behavior expressway: Automatic effects of social perception on social behavior. *Advances in Experimental Social Psychology, 33,* 1-40.

Elordieta, G., Frota, S., Prieto, P., & Vigário, M. (2003). Effects of constituent weight and syntactic branching on intonational phrasing in Ibero-Romance. *Proceedings 15th International Congress of Phonetic Sciences*, Barcelona, Spain, 487-490.

Engelhardt, P. E., Bailey, K. G. D., & Ferreira, F. (2006). Do speakers and listeners observe the Gricean Maxim of Quantity? *Journal of Memory and Language, 54*(4), 554-573.

Fernald, A., & Simon, T. (1984). Expanded intonation contours in mothers' speech to newborns. *Developmental PsychologyDevelopmental Psychology, 20*(1), 104-113.

Féry, C., & Samek-Lodovici, V. (2006). Focus projection and prosodic prominence in nested foci. *Language, 82*(1), 131-150.

Fine, J., Bartolucci, G., Ginsberg, G., & Szatmari, P. (1991). The use of intonation to communicate in pervasive developmental disorders. *Journal of Child Psychology and Psychiatry, 32*(5), 771-782.

Fraundorf, S. H., Watson, D. G., & Benjamin, A. S. (2010). Recognition memory reveals just how CONTRASTIVE contrastive accenting really is. *Journal of Memory and Language, 63*(3), 367-386.

Fry, D. B. (1955). Duration and intensity as physical correlates of linguistic stress. *Journal of the Acoustical Society of America, 27*(4), 765-768.

Fry, D. B. (1969). The linguistic evidence of speech errors. *BRNO Studies of English, 8*, 69-74.

Galati, A., & Brennan, S. E. (2010). Attenuating information in spoken communication: For the speaker, or for the addressee? *Journal of Memory and Language, 62*(1), 35-51.

Garnica, O. (1977). Some prosodic and paralinguistic features of speech to young children. In C. E. Snow & C. A. Ferguson (Eds.), *Talking to children. Language input and acquisition*. Cambridge, MA: Cambridge University Press.

Garrod, S., & Pickering, M. J. (2004). Why is conversation so easy? *Trends in Cognitive Sciences, 8*(1), 8-11.

Giles, H., Coupland, N., & Coupland, J. (1991). Accommodation theory: Communication, context and consequences. In H. Giles, N. Coupland & J. Coupland (Eds.), *Contexts of accommodation: Developments in applied linguistics*, 1-68. Cambridge: Cambridge University Press.

Glasberg, B. R., & Moore, B. C. J. (1990). Derivation of auditory filter shapes from notched-noise data. *Hearing Research, 47*(1-2), 103-138.

Goto, H. (1971). Auditory perception by normal Japanese adults of the sounds "L" and "R". *Neuropsychologia, 9*(3), 317-323.

Granström, B., House, D., & Swerts, M. (2002). Multimodal feedback cues in human-machine interactions. *Proceedings Speech Prosody*, Aix-en-Provence, France, 347-350.

Gregory, M. L., Jurafsky, D., & Healy, A. F. (2001). The role of the hearer in durational shortening. *Proceedings AMLaP Conference*, Saarbrücken, Germany.

Gregory, S. W. (1986). Social psychological implications of voice frequency correlations: Analyzing conversation partner adaptation by computer. *Social Psychology Quarterly, 49*(3), 237-246.

Gregory, S. W., & Hoyt, B. R. (1982). Conversation partner mutual adaptation as demonstrated by Fourier series analysis. *Journal of Psycholinguistic Research, 11*(1), 35-46.

Grice, H. P. (1975). Logic and conversation *Syntax and semantics* (Vol. 3). New York: Academic Press.

Gussenhoven, C. (1983). Focus, mode and nucleus. *Journal of Linguistics, 19*(2), 377-417.

Gussenhoven, C., Repp, B. H., Rietveld, A., Rump, H. H., & Terken, J. (1997). The perceptual prominence of fundamental frequency peaks. Journal of the Acoustical Society of America, *102*(5), 3009-3022.

Haan, J. (2001). *Speaking of questions. An exploration of Dutch question intonation*. Dissertation, Katholieke Universiteit Nijmegen.

Haeseryn, W., Romijn, K., Geerts, G., De Rooij, J., & Van den Toorn, M. C. (1997). Algemene Nederlandse Spraakkunst (2 ed.). Groningen/Deurne: Martinus Nijhoff/Wolters Plantyn.

Heller, D., Skovbroten, K., & Tanenhaus, M. K. (2009). To name or to describe: Shared knowledge affects referential form. *Topics in Cognitive Science, 4*(2), 290-305.

Hellmuth, S. (2007). The relationship between prosodic structure and pitch accent distribution: Evidence from Egyptian Arabic. *The Linguistic Review, 24*(2-3), 291-316.

Horton, W., & Keysar, B. (1996). When do speakers take into account common ground? *Cognition, 59*(1), 91-117.

Hruska, C., Alter, K., Steinhauer, K., & Steube, A. (2001). Misleading dialogues: Human's brain reaction to prosodic information. In C. Cavé, I. Guaïtella & S. Santi (Eds.), *Orality and Gestures. Interactions et comportements multimodaux dans la communication*, 425-430. Paris: l'Harmattan.

Jong, K. de (1995). The supraglottal articulation of prominence in English: Linguistic stress as localized hyperarticulation. *Journal of the Acoustical Society of America, 97*(1), 491-504.

Kaland, C. C. L., Avesani, C., Krahmer, E. J., Swerts, M. G. J., & Zappoli, A. (submitted). Adapting to atypical prosody: Contrastive noun phrases in Dutch and Italian. Submitted for journal publication.

Kaland, C. C. L., Krahmer, E. J., & Swerts, M. G. J. (to appear). White bear effects in language production: Evidence from the prosodic realisation of adjectives. *Language and Speech, 58*(1).

Kaland, C. C. L., Krahmer, E. J., & Swerts, M. G. J. (2011). Salient in the mind, salient in prosody. *Proceedings 33rd Annual Conference of the Cognitive Science Society* Boston (MA), USA, 261-266.

Kaland, C. C. L., Krahmer, E. J., & Swerts, M. G. J. (2012a). Accentuation as a cue for speaker adaptation. *Proceedings Speech Prosody 2012*, Shanghai, China.

Kaland, C. C. L., Krahmer, E. J., & Swerts, M. G. J. (2012b). Contrastive intonation in autism: The effect of speaker- and listener-perspective. *Proceedings 13th Annual Conference of the International Speech Communication Association (Interspeech)*, Portland (OR), USA, 1047-1050.

Kaland, C. C. L., Krahmer, E. J., & Swerts, M. G. J. (2013a). On how accent distribution can signal speaker adaptation. *Phonetica, 69*(4), 216-230.

Kaland, C. C. L., Swerts, M. G. J., & Krahmer, E. J. (2013b). Accounting for the listener: Comparing the production of contrastive intonation in typically-developing speakers and speakers with autism. *Journal of the Acoustical Society of America, 134*(3), 2182-2196.

Kaland, C. C. L., & Van Heuven, V. J. (2010). The structure-prosody interface of restrictive and appositive relative clauses in Dutch and German. *Proceedings Speech Prosody*, Chicago (IL), USA, 100061-100064.

Keysar, B., Barr, D. J., & Horton, W. S. (1998). The egocentric basis of language use: Insights from a processing approach. *Current Directions in Psychological Sciences, 7*(2), 46-50.

Keysar, B., Lin, S., & Barr, D. J. (2003). Limits on theory of mind use in adults. *Cognition, 89*(1), 25-41.

Kochanski, G., Grabe, E., Coleman, J., & Rosner, B. (2005). Loudness predicts prominence: Fundamental frequency lends little. *Journal of the Acoustical Society of America, 118*(2), 1038-1054.

Koolen, R., Goudbeek, M., & Krahmer, E. (2013). The effect of scene variation on the redundant use of color in definite reference. *Cognitive Science, 37*(2), 395-411.

Krahmer, E., & Swerts, M. (2001). On the alleged existence of contrastive accents. *Speech Communication, 34*(4), 391-405.

Krahmer, E., & Swerts, M. (2004). More about brows: A cross-linguistic study via analysis-by-synthesis. In Z. Ruttkay & C. Pelachaud (Eds.), *From brows to trust: Evaluating embodied conversational agents*, 191-216. Dordrecht: Kluwer.

Kraljic, T., & Brennan, S. E. (2005). Prosodic disambiguation of syntactic structure: For the speaker or for the addressee? *Cognitive Psychology, 50*(2), 194-231.

Kuhlen, A. K., & Brennan, S. E. (2013). Language in dialogue: When confederates might be hazardous to your data. *Psychonomic Bulletin and Review, 20*(1), 54-72.

Ladd, D. R. (1988). Declination reset and the hierarchical organization of utterance. *Journal of the Acoustical Society of America, 84*(2), 530-544.

Ladd, D. R. (1993). Constraints on the gradient variability of pitch range (or) pitch level 4 Lives! In P. Keating (Ed.), *Papers in Laboratory Phonology III* (Vol. 43-63). Cambridge: Cambridge University Press.

Ladd, D. R. (2008). *Intonational phonology* (2nd ed.). Cambridge; New York: Cambridge University Press.

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

Levelt, W. J. M. (1999). Producing spoken language: A blueprint of the speaker. In P. Hagoort & C. M. Brown (Eds.), *The neurocognition of language*, 94-122. Oxford: Oxford University Press.

Levelt, W. J. M., & Kelter, S. (1982). Surface form and memory in question answering. *Cognitive Psychology*, *14*(1), 78-106.

Lewis, D. (1979). Scorekeeping in a language game. *Journal of Philosophical Logic, 8*(1), 339-359.

Lombard, É. (1911). Le signe de l'élévation de la voix. *Annales des Maladies de L'Oreille et du Larynx, 37*(2), 101-109.

Maastricht, L. J. van, Krahmer, E. J., & Swerts, M. G. J. (2013). Acquiring native-like intonation in Dutch and Spanish: Comparing the L1 and L2 of native speakers and second language learners. *Proceedings Phonetics, phonology and languages in contact: Contact varieties, multilingualism, and second language learning*, Paris, France.

McCaleb, P., & Prizant, B. M. (1985). Encoding of new versus old information by autistic children. *Journal of Speech and Hearing Disorders, 50*, 230-240.

McCann, J., & Peppé, S. (2003). Prosody in autism spectrum disorders: A critical review. *International Journal of Language and Communication Disorders, 38*(4), 325-350.

Mennen, I. (2007). Phonological and phonetic influences in non-native intonation. In J. Trouvain & U. Gut (Eds.), *Non-native Prosody: Phonetic Descriptions and Teaching Practice (Nicht-muttersprachliche Prosodie: Phonetische Beschreibungen und didaktische Praxis)*, 53 - 76. Berlin: Mouton De Gruyter.

Nadig, A., & Sedivy, J. (2002). Evidence of perspective-taking constraints in children's on-line reference resolution. *Psychological Science, 13*(4), 329-336.

Nadig, A., & Shaw, H. (2011). Acoustic and perceptual measurement of expressive prosody in high-functioning autism: Increased pitch range and what it means to listeners. *Journal of Autism and Developmental Disorders, 42*(4), 499-511.

Natale, M. (1975). Converge of mean vocal intensity in dyadic communication as a function of social desirability. *Journal of Personality and Social Psychology, 32*(5), 790-804.

Nespor, M. (1993). *Fonologia*. Bologna: Il Mulino.

Nespor, M., & Vogel, I. (2007). *Prosodic phonology* (2nd ed.). Berlin & New York: De Gruyter.

Newman, S. S. (1946). On the stress system of English. *Word, 2*, 171-187.

Nilsenová, M., Swerts, M. G. J., Houtepen, V., & Dittrich, H. (2009). Pitch adaptation in different age groups: Boundary tones versus global pitch. *Proceedings Interspeech*, Brighton, UK, 1015-1018.

Nilsonne, Å., Sundberg, J., Ternström, S., & Askenfelt, A. (1988). Measuring the rate of change of voice fundamental frequency in fluent speech during mental depression. *Journal of the Acoustical Society of America, 83*(2), 716-728.

Nooteboom, S. G., & Kruyt, J. G. (1987). Accents, focus distribution, and the perceived distribution of given and new information. *Journal of the Acoustical Society of America, 82*(5), 1512-1524.

Okada, B. M., Lachs, L., & Boone, B. (2012). Interpreting tone of voice: Musical pitch relationships convey agreement in dyadic conversation. *Journal of the Acoustical Society of America*, *132*(3), 208-214.

Pardo, J. S. (2006). On phonetic convergence during conversational interaction. *Journal of the Acoustical Society of America*, *119*(4), 2382-2393.

Partanen, E., Kujala, T., Näätänen, R., Liitola, A., Sambeth, A., & Huotilainen, M. (2013). Learning-induced neural plasticity of speech processing before birth. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(37), 15145-15150.

Pechmann, T. (1984a). Accentuation and redundancy in children's and adults' referential communication. In H. Bouma & D. G. Bouwhuis (Eds.), *Attention and Performance 10*. Hillsdale: Lawrence Erlbaum.

Pechmann, T. (1984b). *Überspezifizierung und Betonung in referentieller Kommunikation.* Dissertation, Universität Mannheim.

Pechmann, T. (1989). Incremental speech production and referential overspecification. *Linguistics*, *27*(1), 89-110.

Peppé, S., McCann, J., Gibbon, F., O'Hare, A., & Rutherford, M. (2007). Receptive and expressive prosodic ability in children with high-functioning autism. *Journal of Speech, Language, and Hearing Research*, *50*, 1015-1028.

Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences, 27*(2), 169-226.

Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral and Brain Sciences*, *36*(4), 329-392.

Pierrehumbert, J., & Hirschberg, J. (1990). The meaning of intonational contours in the interpretation of discourse. In P. Cohen, J. Morgan & M. Pollack (Eds.), *Intentions in communication*, 271-312. Cambridge, MA: MIT Press.

Quené, H., & Kager, R. (1989). Automatic accentuation and prosodic phrasing for Dutch text-to-speech conversion. *Proceedings European Conference on Speech Communication and Technology (Eurospeech)*, Edinburgh, UK, 214-217.

Raaijmakers, J. G. W., Schrijnemakers, J. M. C., & Gremmen, F. (1999). How to deal with "the language-as-fixed-effect fallacy": Common misconceptions and alternative solutions. *Journal of Memory and Language, 41*(4), 416-426.

Rasier, L., Caspers, J., & Van Heuven, V. J. (2010). Accentual marking of information status in Dutch and French as foreign languages. Production and perception. In M. Wrembel, M. Kul & K. Dziubalska-Kolaczyk (Eds.), *Polish Studies in English Language and Litterature 32*, 379-385.

Rasier, L., & Hiligsmann, P. (2007). Prosodic transfer from L1 to L2. Theoretical and methodological issues. *Cahiers de Linguistique Francaise, 28*, 41-66.

Rietveld, A. C. M., & Van Heuven, V. J. (2009). *Algemene fonetiek* (3 ed.). Bussum: Coutinho.

Rooth, M. (1992). A theory of focus interpretation. *Natural Language Semantics, 1*(1), 75-116.

Rooth, M. (1996). On the interface principles for intonational focus. *Proceedings SALT VI*, Ithaca (NY), USA, 202-226.

Rutherford, M. D., Baron-Cohen, S., & Wheelwright, S. (2002). Reading the mind in the voice: A study with normal adults and adults with Asperger syndrome and high functioning autism. *Journal of Autism and Developmental Disorders, 32*(3), 189-194.

Schafer, A. J., Speer, S. R., Warren, P., & White, S. D. (2000). Intonational disambiguation in sentence production and comprehension. *Journal of Psycholinguistic Research, 29*(2) 169-182.

Scherer, K. R. (1986). Vocal affect expression: A review and a model for future research. *Psychological Bulletin, 99*(2), 143-165.

Schober, M. F., & Herbert , H. C. (1989). Understanding by addressees and overhearers. *Cognitive Psychology, 21*(2), 211-232.

Schwarzschild, R. (1999). GIVENness, Avoid F and other constraints on the placement of focus. *Natural Language Semantics, 7*(2), 141-177.

Selkirk, E.O. (2008). Contrastive focus, givenness and the unmarked status of “discourse-new”. *Acta Linguistica Hungarica, 55*(3-4), 331-346.

Selkirk, E. O. (1984). *Phonology and syntax: The relation between sound and structure*. Cambridge, Mass.: MIT Press.

Shepard, C. A., Giles, H., & Le Poire, B. A. (2001). Communication accommodation theory. In W. P. Robinson & H. Giles (Eds.), *The New Handbook of Language and Social Psychology*, 33-56. New York: Wiley.

Shriberg, L. D., Paul, R., McSweeny, J. L., Klin, A., Cohen, D. J., & Volkmar, F. R. (2001). Speech and prosody characteristics of adolescents and adults with high-functioning autism and Asperger’s syndrome. *Journal of Speech, Language, and Hearing Research, 44*, 1097-1115.

Sluijter, A. M. C., & Van Heuven, V. J. (1996). Spectral tilt as an acoustic correlate of linguistic stress. *Journal of the Acoustical Society of America*, *100*(4), 2471-2485.

Snedeker, J., & Trueswell, J. (2003). Using prosody to avoid ambiguity: Effects of speaker awareness and referential context. *Journal of Memory and Language, 48*(1), 103-130.

Stalnaker, R. (1978). Assertion. *Syntax and Semantics, 9*, 315-332.

Steinhauer, K., Alter, K., & Friederici, A. D. (1999). Brain potentials indicate immediate use of prosodic cues in natural speech processing. *Nature Neuroscience*, *2*, 191-196.

Streefkerk, B. M. (2002). *Prominence. Acoustic and lexical/syntactic correlates.* Dissertation, Universiteit van Amsterdam.

Swerts, M. (2007). Contrast and accent in Dutch and Romanian. *Journal of Phonetics, 35*(3), 380-397.

Swerts, M., Krahmer, E. J., & Avesani, C. (2002). Prosodic marking of information status in Dutch and Italian: A comparative analysis. *Journal of Phonetics, 30*(4), 629-654.

Swerts, M., & Marsi, E. (2012). Prosodic evaluation of accent distributions in spoken news bulletins of Flemish newsreaders. *Journal of the Acoustical Society of America*, *132*(4), 2616-2624.

Swerts, M., & Zerbian, S. (2010). Intonational differences between L1 and L2 English in South Africa. *Phonetica, 67*(3), 127-146.

Szczepek-Reed, B. (2010). Speech rhythm across turn transitions in cross-cultural talk-in-interaction. *Journal of Pragmatics*, *42*(4), 1037-1059.

Terken, J., & Hirschberg, J. (1994). Deaccentuation of words representing 'Given' information: Effects of persistence of grammatical role and surface position. *Language and Speech, 37*(2), 125-145.

Terken, J., & Nooteboom, S. G. (1987). Opposite Effects of accentuation and deaccentuation on verification latencies for given and new information. *Language and Cognitive Processes, 2*(3-4), 145-163.

Truckenbrodt, H. (2007). The syntax-phonology interface. In P. d. Lacy (Ed.), *The Cambridge handbook of phonology*, 435-456. Cambridge: Cambridge University Press.

Turk, A., & Sawusch, J. (1996). The processing of duration and intensity cues to prominence. *Journal of the Acoustical Society of America, 99*(6), 3782-3790.

Vallduví, E. (1991). The role of plasticity in the association of focus and prominence. *Proceedings Eastern States Conference in Linguistics 7*, 295-306.

Veenker, T. J. G. (2003). WWStim: A CGI script for presenting web-based questionnaires and experiments (version 1.4.4): Utrecht University.

Wardlow Lane, L., & Ferreira, V. S. (2008). Speaker-external versus speaker-internal forces on utterance form: Do cognitive demands override threats to referential success? *Journal of Experimental Psychology: Learning, Memory and Cognition, 34*(6), 1466-1481.

Wardlow Lane, L., Groisman, M., & Ferreira, V. S. (2006). Don’t talk about pink elephants! Speakers’ control over leaking private information during language production. *Psychological Science, 17*(4), 273-277.

Wardlow Lane, L., & Liersch, M. J. (2011). Can you keep a secret? Increasing speakers' motivation to keep information confidential yields poorer outcomes. *Language and Cognitive Processes, 27*(3), 462-473.

Watson, D. G., Arnold, J. E., & Tanenhaus, M. K. (2008). Tic tac TOE: Effects of predictability and importance on acoustic prominence in language production. *Cognition, 106*(3), 156-157.

Weber, A., Braun, B., & Crocker, M. W. (2006). Finding referents in time: Eye-tracking evidence for the role of contrastive accents. *Language and Speech, 49*(3), 367-392.

Wegner, D. M. (1994). Ironic processes of mental control. *Psychological Review, 101*(1), 34-52.

Wegner, D. M., & Schneider, D. J. (2003). The white bear story. *Psychological Inquiry, 14*(3-4), 326-329.

Wegner, D. M., Schneider, D. J., Carter, S., & White, T. (1987). Paradoxical effects of thought suppression. *Journal of Personality and Social Psychology, 53*(1), 5-13.

AUTHOR INDEX

A

B

C

ACKNOWLEDGEMENTS

Laat ik beginnen Marc Swerts en Emiel Krahmer te bedanken. Jullie hebben me op uiteenlopende manieren geholpen: van het vele commentaar op mijn schrijven tot jullie persoonlijke interesse. Het meest dankbaar ben ik voor jullie vertrouwen in mij.

Unassailably, this thesis would not have its current form without the valuable comments of the committee members. I am thankful to Yiya Chen, Martijn Goudbeek, Carlos Gussenhoven, Bernd Möbius and Sabine Zerbian.

Colleagues, all of you, thank you! It is a pleasure working with you, especially because not all conversations are academic. Or better, the academic conversations are sometimes hard to find, especially during lunch breaks. Thank you Adriana, Alain, Alex, Anja, Anne, Annemarie, Anne-Marie, Carel, Emmelyn, Eric, Fons, Harold, Jacintha, Jacqueline, Jan, Janneke, Joost, Jorrig (Mr. Birds), Juliette, Karin, Kiek, Lauraine, Lennard, Leonoor, Lieke, Lisanne, Lisette, Maaike, Mandy, Maria, Mariana (buongiorno), Mariek, Marieke, Marije, Marjolijn, Martijn G (hai hai), Monique, Noami, Per, Phoebe, Rein, Rick, Ruud K (jow!), Ruud M, Suleman, Sylvia, Véronique and Yan.

Tegen sommige collega’s zeg ik graag een extra word van dank voor de fijne samenwerking. Martijn B, het was erg leuk experimenten te doen met “smarties”! Hans, bedankt voor de oneindige stroom interessante onderwerpen en voor het relativeren van het universitaire leven (als dit niet bij samenwerking hoort, dan maar bij vriendschap ;). Marie, hartelijk dank dat je mijn interesse in de psychoakoestiek een por hebt gegeven. Vincent van Heuven ben ik dankbaar voor zijn (fonetische) adviezen. Verder bedank ik Hanneke, Kitty, Madelene, Tom en Margot voor hulp bij diverse experimenten.

One of the things I enjoyed a lot was collaborating with researchers from abroad. It was nice and fruitful to have worked with Verònica, Joan and Pilar; thanks to all of you! I am thankful to Cinzia Avesani for welcoming me in her labs. I owe many thanks to Alessandra for her support and for relating phonology to bumper cars.

Realisme gebiedt me te zeggen dat verhuizen van Leiden naar Tilburg ondoorgrondelijk is. Ik ben daarom erg blij met de vrienden die bleven. Jeroen, bedankt voor je levendige karakter en goede gezelschap, op elke hoogte. Niall en Ilona, bedankt voor jullie onvoorwaardelijke vriendschap en gastvrijheid, om nog maar te zwijgen (en dat is maar beter ook) over de “Bube-avonden” met Robbert en Charlie. Marieke, een taalkundige interesse hebben is één ding, maar hoe combineer je dat met bergsport? Bedankt voor die onvergetelijke top-avonturen!

Es war sehr schön, Joona und Gianluca, euch nach unserer Zeit in Freiburg noch mehrmals zu treffen. Kiitos und grazie! Es gibt Freundschaften, die sind weniger an Zeit oder Distanz gebunden.

Thank you Jette and Smithy for the good times in Tilburg (and beyond) and for your open-heartedness. I miss it.

Een nieuwe plek brengt nieuwe vriendschappen met zich mee. Rawie, dankjewel dat je me introduceerde bij de TilSAC en voor de toffe tijd in de touwen. Ik ben blij deel uit te maken van klimgezelschap “De Tekenbeet”. Bedankt Geert, Joery, Maarten, Marc, Martin, Rob en Twan dat we onze passie konden delen. Maar het meest dankbaar ben ik voor het feit dat jullie me met twee voeten op de grond houden.

Mijn ouders, Ger en Will, ben ik dankbaar voor liefde, zorg en vertrouwen die ik nodig had om dit proefschrift te maken.

Extra dankbaar ben ik Anna en Jorn, omdat ze zo’n leuke zus en zwager zijn!

Reuzeblij ben ik met de familie, vrienden en (oud-)huisgenoten die me op welke manier dan ook bijstaan: (speciale tante) Leuni, Daniël, Wout, Jan en Erna, Erika en Martijn, Manouk en Debby.

Graag wil ik ook mijn grootouders bedanken, aan wie ik dit proefschrift opdraag; opa Lou en oma Pie, opa Kees en oma Nee. Jullie blijven me inspireren.

One very nice girl remains to be thanked for being so supportive and for the wonderful times I spent with her. Thank you for that, Sarah. And of course, for that we are on the same page.

SUMMARY

Prosodic marking of semantic contrasts: do speakers adapt to addressees? describes a series of psycholinguistic experiments on the extent to which speakers adapt to addressees by using contrastive intonation. Four studies have been carried out to investigate prosodic adaptation processes of different kinds. Chapter 2 and Chapter 3 report studies on the extent to which speakers, in the way they prosodically mark semantic contrasts, account for the knowledge state of their addressees. Chapter 4 and Chapter 5 report studies on the extent to which paralinguistic prosodic adaptation processes between interlocutors affect the prosodic marking of semantic contrasts. The experimental approach throughout the thesis consists of eliciting speech recordings which are analyzed by production and perception measures.

The first study (Chapter 2) addressed the question to what extent contrastive intonation is speaker- or addressee-driven by comparing typically-developing speakers and speakers with autism. The results of this study indicated that both speaker groups produce contrastive intonation more clearly when the semantic contrast is shared with their addressee, compared to when the semantic contrast is not shared with their addressee. Prosodic differences between typically-developing speakers and speakers with autism were found only for produced pitch range and perceived speech dynamicity.

The second study (Chapter 3) investigated to what extent speakers, in their prosody, account for the knowledge state of their addressees when prompted to leak contrastive information. To this end, a specific experimental paradigm elicited speakers' utterances that leaked contrastive information from a visual scene. Crucially, some information in the scene was only visible for the speaker, but not for the addressee. Results indicated that speakers who leaked certain privileged contrastive information used a specific prosody (i.e., pitch). It was shown that speakers produce some form of prosodic marking even when not all information is shared with the addressee.

The third study (Chapter 4) explored to what extent contrastive intonation is a cue for speaker adaptation and how the functional use of this intonation pattern

relates to cues to speaker adaptation which do not have a linguistic function, such as the copying of prosodic features. To this end, perception experiments were carried out, in which manipulated dialogues of speaker pairs were presented to participants. In one half of the stimuli the speaker pairs used contrastive intonation in a coherent way. In the other half of the stimuli the speakers copied each other's intonation pattern in contrastive contexts. Results showed that speaker pairs that used contrastive intonation coherently were perceived as better adapters than interlocutors that copied each other's intonation pattern.

The fourth study (Chapter 5) investigated the extent to which speakers of Dutch and Italian adapt to atypical prosody in contrastive contexts. Experiments with speakers of either Dutch or Italian were carried out. One half of the speakers interacted with a (Dutch or Italian) confederate who used prosody in accordance with the rules of prosodic marking of information status in the language of interest (typical prosody). For the other half of the speakers the confederate used prosody which was not in accordance with those rules (atypical prosody). Results showed that speakers of Dutch, a language which allows shifting pitch accents in a flexible way, adapted to atypical prosody to a larger extent than speakers of Italian, a language which uses pitch accents in a more fixed way.

Taking all studies in this thesis together, we can conclude that the way speakers mark information structure prosodically, by means of contrastive intonation, is highly determined by the interaction with the interlocutor. That is, speakers adapt their prosody depending on their assumptions about the knowledge state of the addressee as well as their own (Chapter 2 and Chapter 3). Furthermore, we find a relation between prosodic adaptation and the extent to which prosody is used in a functional way. In particular, the degree of perceived (Chapter 4) or produced (Chapter 5) adaptation depends on the extent to which prosody contributes to the linguistic meaning of utterances.

SAMENVATTING

Prosodische markering van semantische contrasten: passen sprekers zich aan luisteraars aan? beschrijft een reeks experimenten die onderzoeken in welke mate sprekers, in hun gebruik van contrastieve intonatie, zich aanpassen aan hun luisteraars. Vier studies zijn uitgevoerd om verschillende vormen van prosodische aanpassing te onderzoeken. Hoofdstuk 2 en 3 rapporteren studies naar de mate waarin sprekers, in hun gebruik van contrastieve intonatie, rekening houden met de kennis van hun luisteraars. Hoofdstuk 4 en 5 rapporteren studies naar de mate waarin paralinguïstische prosodische aanpassingsprocessen tussen sprekers het gebruik van contrastieve intonatie beïnvloeden. De experimentele benadering die is gehanteerd door het gehele proefschrift bestaat uit het uitvragen van spraak die vervolgens is geanalyseerd door middel van productie- en perceptie-metingen.

De eerste studie (Hoofdstuk 2) stelt de vraag in welke mate contrastieve intonatie spreker- of luisteraar-gedreven is door typische sprekers te vergelijken met sprekers met autisme. De resultaten van deze studie lieten zien dat beide sprekergroepen contrastieve intonatie minder duidelijk produceerden wanneer een semantisch contrast niet was gedeeld met hun luisteraar ten opzichte van wanneer een semantisch contrast wel was gedeeld met hun luisteraar. Prosodische verschillen tussen typische sprekers en sprekers met autisme zijn alleen gevonden in het geproduceerde toonhoogtebereik en de waargenomen spraakdynamiek.

De tweede studie (Hoofdstuk 3) onderzocht in welke mate sprekers in hun prosodie, wanneer zij werden uitgelokt contrastieve informatie te lekken, rekening houden met de kennis van hun luisteraars. Daartoe zijn door middel van een specifiek experimenteel paradigma uitingen van sprekers uitgevraagd die contrastieve informatie uit een visuele scène lekten. Belangrijk hierbij was dat sommige informatie in de scène slechts voor de spreker zichtbaar was en niet voor de luisteraar. Resultaten lieten zien dat sprekers die de alleen voor hen zichtbare contrastieve informatie lekten, een specifieke prosodie (toonhoogte) gebruikten. De studie toonde aan dat sprekers een vorm van prosodische markering gebruiken zelfs als niet alle informatie is gedeeld met de luisteraar.

De derde studie (Hoofdstuk 4) verkende in welke mate contrastieve intonatie aanpassing tussen sprekers signaleert. Daartoe werden perceptie-experimenten uitgevoerd waarin aan proefpersonen gemanipuleerde dialogen van sprekerparen werden gepresenteerd. In de ene helft van de stimuli gebruikten sprekers contrastieve intonatie op een coherente manier. In de andere helft van de stimuli kopieerden sprekers elkaars intonatiepatronen in contrastieve contexten. De resultaten lieten zien dat sprekerparen met hetzelfde coherente gebruik van contrastieve intonatie als betere aanpassers werden waargenomen dan sprekerparen die hun intonatiepatronen kopieerden.

De vierde studie (Hoofdstuk 5) onderzocht de mate waarin sprekers van het Nederlands en het Italiaans zich aanpassen aan atypische prosodie in contrastieve contexten. Experimenten met sprekers van het Nederlands en Italiaans werden uitgevoerd. Eén helft van de sprekers interacteerde met een (Nederlandse of Italiaanse) experimentele complotteur die prosodie gebruikte in overeenstemming met de regels voor prosodische markering van informatiestatus in de betreffende taal (typische prosodie). Voor de andere helft van de sprekers gebruikte de experimentele complotteur prosodie die niet overeenstemde met deze regels (atypische prosodie). Resultaten lieten zien dat sprekers van het Nederlands, dat verplaatsing van toonhoogteaccenten op een flexibele manier toestaat, zich meer aanpasten aan atypische prosodie dan sprekers van het Italiaans, dat toonhoogteaccenten op vaste locaties gebruikt.

Op basis van alle studies in dit proefschrift, kunnen we concluderen dat de manier waarop sprekers informatiestructuur prosodisch markeren in hoge mate wordt bepaald door de interactie met de gesprekspartner. Sprekers passen namelijk hun prosodie aan afhankelijk van hun assumpties over de kennis van zowel hun luisteraar als zichzelf (Hoofdstuk 2 en 3). Daarnaast vinden we een relatie tussen prosodische aanpassing en de mate waarin prosodie functioneel wordt gebruikt. Met name de mate van waargenomen (Hoofdstuk 4) of geproduceerde (Hoofdstuk 5) aanpassing hangt af van de mate waarin prosodie bijdraagt aan de taalkundige betekenis van uitingen.

CURRICULUM VITAE

Constantijn Kaland was born on May 24th 1985 in Vlissingen, The Netherlands. He went to primary schools De Lichtstraal in Westkapelle and De Wegwijzer in Serooskerke until 1997. He obtained his Gymnasium diploma from Christelijke Scholengemeenschap Walcheren in Middelburg in 2004. After that, he studied at Leiden University where he obtained a bachelor degree in linguistics and a certificate in journalism and media studies in 2007. He received an Erasmus-scholarship in 2008 to follow master courses in linguistics at the Albert-Ludwigs University in Freiburg (Germany). In 2009 he obtained a master degree in linguistics with a specialization in phonetics, phonology and psycholinguistics at Leiden University. He was employed as a PhD student from 2009 until 2013 at Tilburg University. This PhD thesis is the result of the research carried out during that period. In 2013 he continued working at Tilburg University as a researcher in an international project on prosodic and gestural entrainment (PAGE), funded by the German Volkswagenstiftung.

TiCC Ph.D. Series

1. Pashiera Barkhuysen. *Audiovisual Prosody in Interaction.* Promotores: M.G.J. Swerts, E.J. Krahmer. Tilburg, 3 October 2008.
2. Ben Torben-Nielsen. *Dendritic Morphology: Function Shapes Structure.* Promotores: H.J. van den Herik, E.O. Postma. Co-promotor: K.P. Tuyls. Tilburg, 3 December 2008.
3. Hans Stol. *A Framework for Evidence-based Policy Making Using IT.* Promotor: H.J. van den Herik. Tilburg, 21 January 2009.
4. Jeroen Geertzen. *Dialogue Act Recognition and Prediction.* Promotor: H. Bunt. Co-promotor: J.M.B. Terken. Tilburg, 11 February 2009.
5. Sander Canisius. *Structured Prediction for Natural Language Processing.* Promotores: A.P.J. van den Bosch, W. Daelemans. Tilburg, 13 February 2009.
6. Fritz Reul. *New Architectures in Computer Chess.* Promotor: H.J. van den Herik. Co-promotor: J.W.H.M. Uiterwijk. Tilburg, 17 June 2009.
7. Laurens van der Maaten. *Feature Extraction from Visual Data.* Promotores: E.O. Postma, H.J. van den Herik. Co-promotor: A.G. Lange. Tilburg, 23 June 2009 (cum laude).
8. Stephan Raaijmakers. *Multinomial Language Learning.* Promotores: W. Daelemans, A.P.J. van den Bosch. Tilburg, 1 December 2009.
9. Igor Berezhnoy. *Digital Analysis of Paintings.* Promotores: E.O. Postma, H.J. van den Herik. Tilburg, 7 December 2009.
10. Toine Bogers. *Recommender Systems for Social Bookmarking.* Promotor: A.P.J. van den Bosch. Tilburg, 8 December 2009.
11. Sander Bakkes. *Rapid Adaptation of Video Game AI.* Promotor: H.J. van den Herik. Co-promotor: P. Spronck. Tilburg, 3 March 2010.
12. Maria Mos. *Complex Lexical Items.* Promotor: A.P.J. van den Bosch. Co-promotores: A. Vermeer, A. Backus. Tilburg, 12 May 2010 (in collaboration with the Department of Language and Culture Studies).
13. Marieke van Erp. *Accessing Natural History. Discoveries in data cleaning, structuring, and retrieval*. Promotor: A.P.J. van den Bosch. Co-promotor: P.K. Lendvai. Tilburg, 30 June 2010.

14. Edwin Commandeur. *Implicit Causality and Implicit Consequentiality in Language Comprehension*. Promotores: L.G.M. Noordman, W. Vonk. Co-promotor: R. Cozijn. Tilburg, 30 June 2010.
15. Bart Bogaert. *Cloud Content Contention*. Promotores: H.J. van den Herik, E.O. Postma. Tilburg, 30 March 2011.
16. Xiaoyu Mao. *Airport under Control.* Promotores: H.J. van den Herik, E.O. Postma. Co-promotores: N. Roos, A. Salden. Tilburg, 25 May 2011.
17. Olga Petukhova. *Multidimensional Dialogue Modelling*. Promotor: H. Bunt. Tilburg, 1 September 2011.
18. Lisette Mol. *Language in the Hands.* Promotores: E.J. Krahmer, A.A. Maes, M.G.J. Swerts. Tilburg, 7 November 2011 (cum laude).
19. Herman Stehouwer. *Statistical Language Models for Alternative Sequence Selection*. Promotores: A.P.J. van den Bosch, H.J. van den Herik. Co-promotor: M.M. van Zaanen. Tilburg, 7 December 2011.
20. Terry Kakeeto-Aelen. *Relationship Marketing for SMEs in Uganda*. Promotores: J. Chr. van Dalen, H.J. van den Herik. Co-promotor: B.A. Van de Walle. Tilburg, 1 February 2012.
21. Suleman Shahid. *Fun & Face: Exploring non-verbal expressions of emotion during playful interactions*. Promotores: E.J. Krahmer, M.G.J. Swerts. Tilburg, 25 May 2012.
22. Thijs Vis. *Intelligence, Politie en Veiligheidsdienst: Verenigbare Grootheden?* Promotores: T.A. de Roos, H.J. van den Herik, A.C.M. Spapens. Tilburg, 6 June 2012 (in collaboration with the Tilburg School of Law).
23. Nancy Pascall. *Engendering Technology Empowering Women.* Promotores: H.J. van den Herik, M. Diocaretz. Tilburg, 19 November 2012.
24. Agus Gunawan. *Information Access for SMEs in Indonesia*. Promotor: H.J. van den Herik. Co-promotores: M. Wahdan, B.A. Van de Walle. Tilburg, 19 December 2012.
25. Giel van Lankveld. *Quantifying Individual Player Differences.* Promotores: H.J. van den Herik, A.R. Arntz. Co-promotor: P. Spronck. Tilburg, 27 February 2013.

26. Sander Wubben. *Text-to-text Generation Using Monolingual Machine Translation.* Promotores: E.J. Krahmer, A.P.J. van den Bosch, H. Bunt. Tilburg, 5 June 2013.
27. Jeroen Janssens. *Outlier Selection and One-Class Classification*. Promotores: E.O. Postma, H.J. van den Herik. Tilburg, 11 June 2013.
28. Martijn Balsters. *Expression and Perception of Emotions: The Case of Depression, Sadness and Fear*. Promotores: E.J. Krahmer, M.G.J. Swerts, A.J.J.M. Vingerhoets. Tilburg, 25 June 2013.
29. Lisanne van Weelden. *Metaphor in Good Shape.* Promotor: A.A. Maes. Co-promotor: J. Schilperoord. Tilburg, 28 June 2013.
30. Ruud Koolen. *"Need I say More? On Overspecification in Definite Reference."* Promotores: E.J. Krahmer, M.G.J. Swerts. Tilburg, 20 September 2013.
31. J. Douglas Mastin. *Exploring Infant Engagement. Language Socialization and Vocabulary. Development: A Study of Rural and Urban Communities in Mozambique.* Promotor: A.A. Maes. Co-promotor: P.A. Vogt. Tilburg, 11 October 2013.
32. Philip C. Jackson. Jr. *Toward Human-Level Artificial Intelligence - Representation and Computation of Meaning in Natural Language.* Promotores: H.C. Bunt, W.P.M. Daelemans. Tilburg, 22 April 2014.
33. Jorrig Vogels. *Referential choices in language production: The Role of Accessibility.* Promotores: A.A. Maes, E.J. Krahmer. Tilburg, 23 April 2014 (cum laude).
34. Peter de Kock. *Anticipating Criminal Behaviour.* Promotores: H.J. van den Herik, J.C. Scholtes. Co-promotor: P. Spronck. Tilburg, 10 September 2014.
35. Constantijn Kaland. *Prosodic marking of semantic contrasts: Do speakers adapt to addresees?* Promotores: M.G.J. Swerts, E.J. Krahmer. Tilburg, 1 October 2014.