Accounting for the listener: Comparing the production of contrastive intonation in typically-developing speakers and speakers with autism

Constantijn Kaland,[a)] Marc Swerts, and Emiel Krahmer
Tilburg Centre for Cognition and Communication, P.O. Box 90153, NL-5000 LE Tilburg, The Netherlands

(Received 4 July 2012; revised 8 July 2013; accepted 9 July 2013)

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 paper 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.
[http://dx.doi.org/10.1121/1.4816544]

PACS number(s): 43.70.Dn [CGC] Pages: 2182–2196

I. 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, 1984). 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 toward the contrastive information. In the scenario in the preceding text, 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. Subsequently, there is a specific section on 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.

[a)]Author to whom correspondence should be addressed. Electronic mail: c.c.l.kaland@uvt.nl

 0001-4966/2013/134(3)/2182/15/$30.00

A. 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 and Keysar, 1996; Bard *et al.*, 2000; Pickering and Garrod, 2004). Other studies argued that speakers do take the listener's perspective into account at the phonetic level and higher (Clark and Murphy, 1982) as soon as that information is available (Brennan and Hanna, 2009; Galati and 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 and 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 of 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 and 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 and 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 and 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 and Brennan, 2010). The one-bit model proposed by Galati and 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 and Trueswell, 2003; Kraljic and 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 in the preceding text. 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.

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 in the preceding text, 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 in the preceding text, 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 and Hirschberg, 1990), whereas others favored a less clear distinction between these categories of information status in English (Schwarzschild, 1999; Féry and 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 and 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. 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 in the preceding text) 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 in the preceding text) 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 (1984) 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 and 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 1 day after the pattern was heard and concerns the memorization of both the contrastive and the alternative information.

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. Further, 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.

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; the discourse 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 and 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 and 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 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; 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.

5. Disordered prosody in autism: Function versus form

As suggested in the literature on autism discussed in the preceding text, 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 and 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 and 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.

B. Research goals

Given 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 in the preceding text, 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 and 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 article is structured as follows: Section II describes a production experiment with typically developing speakers and speakers with autism; Section III describes the results of perception experiments using elicited NPs of typically developing speakers and speakers with autism; the article ends with a discussion of the results and concluding remarks.

II. PRODUCTION

A. 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

TABLE I. 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

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 I). 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 Sec. II A 2). 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.

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 Fig. 1) and small cards, each the size of a grid-square, displaying colored 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 6×4 grids with rows numbered from 1 to 4 and columns marked by each character of the word "bingo!" (Fig. 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 that 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!" after 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. Participants

Twenty typically developing participants (TYP) acted as speaker in the production experiment (17 women, 3 men,

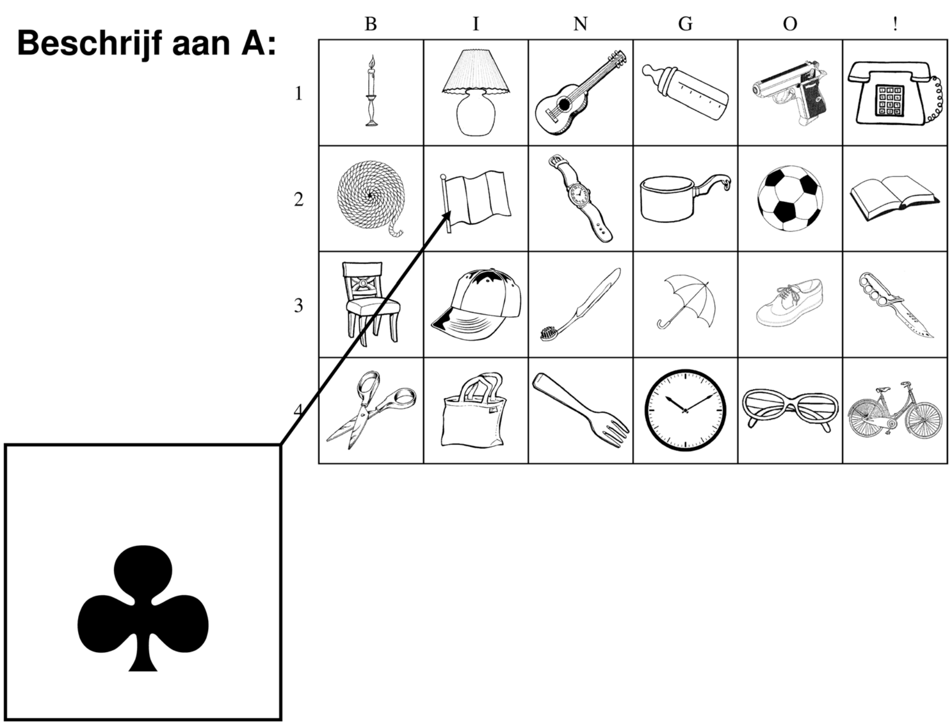

FIG. 1. Example of the speaker's screen, showing in Dutch *Beschrijf aan A* (describe to A), the target figure (bottom left) and A's bingo card. A typical instruction would be: "Put the red clover on the flag."

$M_{age} = 21.8$ yr, age range: 18–29 yr). 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, $M_{age} = 28.9$ yr, age range: 18–51 yr). 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. Eleven 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.

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 (Fig. 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 (Fig. 1). The screen's layout 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 (Fig. 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.

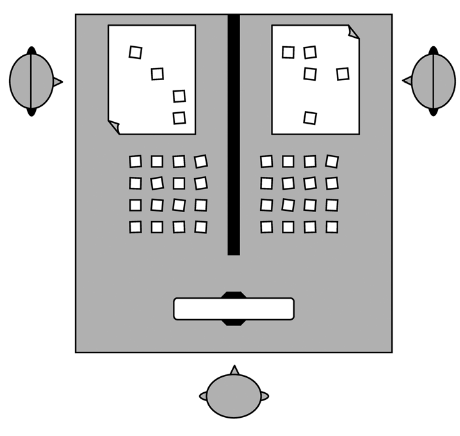

FIG. 2. Birdseye view of the experimental setup showing the speaker facing the screen (bottom) and the listeners, at opposite sides of a partition, facing their bingo cards and figures (top).

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 and 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 (Sec. III).

For pitch measures, F0 maxima in Hertz on the stressed syllable of the adjective and the noun were measured in PRAAT (Boersma and 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 Sec. II A 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 and Weenink, 2011). This method closely resembles the one used by Nilsonne *et al.* (1988) that 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 ms in larger stretches of speech. As the current study focuses on NPs that are per definition short, we obtained a standard deviation directly from PRAAT (Boersma and Weenink, 2011). This standard deviation is based on mean F0 measurements every 10 ms and is therefore arguably more suitable for an analysis of NPs. To abstract over speakers' gender differences standard deviations were measured in ERB (Glasberg and 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.

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 (two levels: Same, different) and focus (two levels: Adjective, noun) as within-subject factors and with development (two levels: Typical, HFA) as between subject factor.

Further, 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 (two levels: Same, different) and focus (two 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 (two levels: Same, different) and focus (two levels: Adjective, noun) as within-subject factors and diagnosis (two 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 (two levels: Same, different) and focus (two 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.

B. Results

Results are discussed following the order of statistical tests summarized in Table II.

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. Further, typically developing speakers produced NPs with a significantly larger pitch range ($M = 0.47$) than HFA speakers ($M = 0.36$).

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

TABLE II. 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 and HFA		
Listener	n.s.	n.s.
Focus	n.s.	n.s.
Listener*focus	$F(1,38) = 5.29, p < 0.05, \eta_p^2 = 0.12$	n.s.
Development	$F(1,38) = 10.01, p < 0.01, \eta_p^2 = 0.21$	$F(1,38) = 4.76, p < 0.05, \eta_p^2 = 0.11$
TYP		
Listener	n.s.	n.s.
Focus	n.s.	n.s.
Listener*focus	$F(1,19) = 7.21, p < 0.05, \eta_p^2 = 0.28$	n.s.
Speaker	n.s.	$F(1,19) = 6.74, p < 0.001, \eta_p^2 = 0.087$
Speaker*focus	$F(1,19) = 13.41, p < 0.001, \eta_p^2 = 0.93$	$F(1,19) = 3.01, p < 0.05, \eta_p^2 = 0.75$
HFA		
Listener	n.s.	n.s.
Focus	$F(1,19) = 23.32, p < 0.001, \eta_p^2 = 0.55$	n.s.
Listener*focus	$F(1,19) = 9.65, p < 0.01, \eta_p^2 = 0.34$	n.s.
Diagnosis	$F(1,18) = 3.16, p = 0.092, \eta_p^2 = 0.15$	$F(1,18) = 4.07, p = 0.059, \eta_p^2 = 0.18$
Speaker	n.s.	$F(1,19) = 21.16, p < 0.001, \eta_p^2 = 0.97$
Speaker*listener	n.s.	$F(1,19) = 2.68, p < 0.05, \eta_p^2 = 0.73$
Speaker*focus	$F(1,19) = 3.88, p < 0.01, \eta_p^2 = 0.80$	n.s.

TABLE III. 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.

			F0 maximum *M* (SD)			Pitch range *M* (SD)
Development	Listener	Focus	Adjective	Noun	Difference	NP
Typical	Same	Adjective	272.16 (60.78)	232.30 (47.45)	39.86 (50.62)	0.48 (0.17)
		Noun	248.58 (59.69)	252.70 (52.71)	4.12 (59.81)	0.46 (0.16)
	Different	Adjective	276.36 (61.48)	241.75 (49.76)	34.61 (45.40)	0.48 (0.16)
		Noun	249.47 (53.75)	258.40 (55.88)	8.93 (52.71)	0.45 (0.15)
HFA	Same	Adjective	208.84 (55.83)	193.77 (59.74)	15.07 (32.70)	0.36 (0.17)
		Noun	196.96 (53.61)	200.20 (52.29)	3.24 (27.22)	0.33 (0.17)
	Different	Adjective	214.10 (58.73)	194.62 (56.67)	19.48 (32.52)	0.36 (0.18)
		Noun	203.18 (55.57)	199.36 (54.19)	−3.83 (26.51)	0.37 (0.19)

and that these differences depend on whether the adjective or the noun is in focus.

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 ($M = -0.29$). 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 III) 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 toward 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 or their interaction. The factor diagnosis revealed that speakers with Asperger produced larger pitch ranges ($M = 0.46$) than speakers with PDD-NOS ($M = 0.31$); this showed a trend. The factor speaker had a significant effect on pitch range and showed an interaction effect with listener; this 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.

C. 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 would be 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 Sec. III C 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 Sec. III, which also gives a more elaborate discussion of all the prosodic analyses.

III. PERCEPTION

A. 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 Sec. II, 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 and Swerts, 2001; Ladd, 2008).

1. Prominence

NPs collected in the production experiments ($N_{\mathrm{TYP}} = 480$, $N_{\mathrm{HFA}} = 480$) 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). Pearson's correlation coefficients as computed for the adjective and noun ratings indicated that the experts' ratings were consistent [$r_{\mathrm{TYP}}(478)$ range = 0.62–0.72, $p < 0.001$] and [$r_{\mathrm{HFA}}(478)$ range = 0.59–0.69, $p < 0.001$].

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 ($N_{\mathrm{TYP}} = 101$, $N_{\mathrm{HFA}} = 45$).[3]

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) × 2 (listener: Same/different) × 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 that 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 min, and the results were collected on a web server.

Thirty different participants completed the judgment task (22 women, 8 men, $M_{\mathrm{age}} = 22.6$ yr, age range: 18–60 yr). 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.

3. Statistics

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

Concerning the speech dynamicity scores RM-ANOVAs were performed with development (two levels: Typical, HFA), listener (two levels: Same, different) and focus (two levels: Adjective, noun) as within-subject factors.

Further, 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 (two levels: Same, different) and focus (two 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 (two levels: Same, different) and focus (two levels: Adjective, noun) as within-subject factors and diagnosis (two levels: Asperger, PDD-NOS) as between subject factor. Concerning speech dynamicity, the data of the speakers with autism were analyzed using an RM-ANOVA with diagnosis (two levels: Asperger, PDD-NOS), listener (two levels: Same, different), and focus (two 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 (two levels: Same, different) and focus (two 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.

B. Results

Results are discussed following the order of statistical tests summarized in Table IV.

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

	Prominence difference	Speech dynamicity
TYP and HFA		
Listener	$F(1,38) = 25.75$, $p < 0.001$, $\eta_p^2 = 0.40$	n.s.
Focus	$F(1,38) = 26.52$, $p < 0.001$, $\eta_p^2 = 0.41$	n.s.
Listener*focus	n.s.	n.s.
Development	$F(1,38) = 2.96$, $p = 0.093$, $\eta_p^2 = 0.07$	$F(1,29) = 7.34$, $p < 0.05$, $\eta_p^2 = 0.20$
TYP		
Listener	$F(1,19) = 16.48$, $p < 0.001$, $\eta_p^2 = 0.46$	n.s.
Focus	$F(1,19) = 11.81$, $p < 0.01$, $\eta_p^2 = 0.38$	n.s.
Listener*focus	n.s.	$F(1,29) = 8.72$, $p < 0.01$, $\eta_p^2 = 0.23$
Speaker	n.s.	$F(1,19) = 15.35$, $p < 0.05$, $\eta_p^2 = 0.99$
Speaker*focus	$F(1,19) = 4.41$, $p < 0.01$, $\eta_p^2 = 0.82$	n.s.
HFA		
Listener	$F(1,19) = 10.05$, $p < 0.01$, $\eta_p^2 = 0.35$	$F(1,29) = 4.07$, $p = 0.053$, $\eta_p^2 = 0.12$
Focus	$F(1,19) = 14.72$, $p < 0.01$, $\eta_p^2 = 0.44$	n.s.
Listener*focus	$F(1,19) = 5.11$, $p < 0.05$, $\eta_p^2 = 0.21$	n.s.
Diagnosis	n.s.	$F(1,29) = 5.36$, $p < 0.05$, $\eta_p^2 = 0.16$
Speaker	n.s.	$F(1,19) = 24.70$, $p < 0.01$, $\eta_p^2 = 0.99$
speaker*focus	$F(1,19) = 10.15$, $p < 0.001$, $\eta_p^2 = 0.91$	n.s.

1. Both development groups taken together (TYP and HFA)

Data for both development groups taken together (Table V) 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 ($M = 0.84$). 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 [$r_{\text{adjective}}(478) = 0.25$, $p < 0.01$ and $r_{\text{noun}}(478) = 0.10$, $p < 0.05$] and speakers with autism [$r_{\text{adjective}}(478) = 0.18$, $p < 0.01$ and $r_{\text{noun}}(478) = 0.12$, $p < 0.01$]. Further, a correlation was found between pitch range and speech dynamicity for both typically developing speakers [$r(78) = 0.33$, $p < 0.01$] and speakers with autism [$r(78) = 0.60$, $p < 0.001$].

2. Typically developing speakers (TYP)

For the prominence difference scores, no negative means were found (Table V), 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 ($M = 1.95$). Further, 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 V. 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. Concerning 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.

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 V). 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. Concerning 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 ($M = 0.36$), see Table V. Further, the factors

TABLE V. 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 (0.31)
		Noun	2.42 (1.96)	4.30 (1.83)	1.88 (3.63)	3.17 (0.32)
	Different	Adjective	4.79 (1.43)	1.67 (1.63)	3.13 (2.84)	3.18 (0.35)
		Noun	2.94 (1.84)	3.71 (1.92)	0.76 (3.57)	3.25 (0.33)
HFA	Same	Adjective	5.03 (1.42)	1.76 (1.40)	3.27 (2.61)	3.08 (0.35)
		Noun	2.96 (2.06)	3.98 (1.78)	1.03 (3.62)	3.14 (0.30)
	Different	Adjective	4.86 (1.48)	1.90 (1.49)	2.96 (2.67)	3.18 (0.36)
		Noun	3.66 (1.98)	3.36 (1.90)	−0.30 (3.55)	3.16 (0.32)

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.

C. Discussion

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 that 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. Further, 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 speaker 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 appeared to produce smaller prominence differences between accented and deaccented words than 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. 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, although 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. There is additional evidence that the results in this study are not simply an artifact of the methodology used, however. 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 that were formally distinct in terms of variation in pitch range and speech dynamicity.

IV. GENERAL DISCUSSION AND CONCLUSION

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 and Murphy, 1982; Galati and Brennan, 2010). The outcomes also refine previous work on contrastive intonation (Chafe, 1976; Pechmann, 1984; Braun and Tagliapietra, 2010). That is, speakers produced this pattern by blending both their own and their listener's perspective. In particular, the contrastive intonation pattern that speakers produced to a different listener was not identical to the pattern they produced to the same listener in the current study. 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. It can be concluded that 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 was 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 and Shaw, 2011).

Our study thus shows that it is important to distinguish prosodic form from prosodic function. Prosodic differences between typically developing speakers and speakers with autism found in the current study relate to form rather than function. Nevertheless, 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 additional prosodic 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. Further, 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 and 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 we cannot conclude on the basis of the present data for Dutch 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.

ACKNOWLEDGMENTS

The authors thank Sonny de Nijs of Zintri Zorggroep for help with the recruitment of and the carrying out of the experiment involving participants with autism and Marieke Hoetjes for help with the prominence ratings. E.K. thanks The Netherlands Organisation for Scientific Research (NWO) for Grant No. 277-70-007.

[1]In the current study, DSM-IV (American Psychiatric Association, 2000) was followed and distinguished between Asperger and PDD-NOS. During the reviewing process, we found out that this distinction is no longer maintained in DSM-V, which is to appear.

[2]To investigate the effect of the high boundary tone, a multivariate analysis of variance with F0 maxima of the noun as a dependent variable and listener (two levels: Same, different), focus (adjective, noun), and boundary tone (two 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 ($M_{\mathrm{TYP}} = 256.82$) compared to when the adjective was in focus ($M_{\mathrm{TYP}} = 236.88$): [$F(1,472) = 11.45$, $p < 0.01$, $\eta_p^2 = 0.02$]. 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 ($M_{\mathrm{HFA}} = 224.23$) compared to when the noun ended with a low boundary tone ($M_{\mathrm{HFA}} = 195.15$): [$F(1,472) = 8.05$, $p < .01$, $\eta_p^2 = 0.02$]. 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 (two levels: Same, different), focus (adjective, noun), and boundary tone (two 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 [$F(1,472) = 150.42$, $p < 0.001$, $\eta_p^2 = 0.24$] and for the speakers with autism [$F(1,472) = 40.83$, $p < 0.001$, $\eta_p^2 = 0.08$] in that prominence scores were higher when the noun was in focus ($M_{\mathrm{TYP}} = 4.25$, $M_{\mathrm{HFA}} = 4.31$) compared to when the adjective was in focus ($M_{\mathrm{TYP}} = 1.93$, $M_{\mathrm{HFA}} = 2.45$). Boundary tone had an effect for the typically developing speakers [$F(1,472) = 38.06$, $p < 0.001$, $\eta_p^2 = 0.08$] and for the speakers with autism [$F(1,472) = 30.00$, $p < 0.001$, $\eta_p^2 = 0.06$] in that those were higher when the noun ended with a high boundary tone ($M_{\mathrm{TYP}} = 3.67$, $M_{\mathrm{HFA}} = 4.81$) compared to when the noun ended with a low boundary tone ($M_{\mathrm{TYP}} = 2.51$, $M_{\mathrm{HFA}} = 2.59$). 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. Further, F0 differences are not always smaller when speakers address a different listener. This can be seen in the case where the noun is focused (Tables III and V). 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 Tables III and V) 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.

American Psychiatric Association (**2000**). *Diagnostic and Statistical Manual of Mental Disorders*, 4th ed. (APA, Washington, DC), 943 pp.
Baltaxe, C. (**1984**). "Use of contrastive stress in normal, aphasic, and autistic children," J. Speech Hear. Res. **27**, 97–105.
Baltaxe, C. A. M., and Guthrie, D. (**1987**). "The use of primary sentence stress by normal, aphasic and autistic children," J. Autism Dev. Disord. **17**, 255–271.
Baltaxe, C. A. M., and Simmons, J. Q. (**1985**). "Prosodic development in normal and autistic children," in *Communication Problems in Autism*, edited by E. Schopler and G. B. Mesibov (Plenum, New York), pp. 95–125.

Bard, E. G., Anderson, A. H., Sotillo, C. F., Aylett, M., Doherty-Sneddon, G., and Newlands, A. (**2000**). “Controlling the intelligibility of referring expressions in dialogue,” J. Mem. Lang. **42**, 1–22.

Bard, E. G., and Aylett, M. P. (**2000**). “Accessibility, duration, and modelling the listener in spoken dialogue,” in *Proceedings of the Fourth Workshop on the Semantics and Pragmatics of Dialogue (Götalog 2000)*, Göteborg, Sweden, pp. 1–8.

Baron-Cohen, S. (**1995**). *Mindblindness: An Essay on Autism and Theory of Mind* (MIT Press, Cambridge), 171 pp.

Baron-Cohen, S. (**2001**). “Theory of mind and autism: A review,” Int. Rev. Ment. Retard. **23**, 169–184.

Boersma, P., and Weenink, D. (**2011**). “PRAAT: Doing phonetics by computer (version 5.2.19),” http://www.praat.org (Last viewed March 16, 2011).

Bonneh, Y. S., Levanon, Y., Dean-Pardo, O., Lossos, L., and Adini, Y. (**2011**). “Abnormal speech spectrum and increased pitch variability in young autistic children,” Front. Hum. Neurosci. **4**, 237.

Braun, B., and Tagliapietra, L. (**2010**). “The role of contrastive intonation contours in the retrieval of contextual alternatives,” Lang. Cogn. Proc. **25**, 1024–1043.

Breckenridge, J. (**1977**). “The declination effect,” J. Acoust. Soc. Am. **61**(S1), S90.

Brennan, S. E., and Hanna, J. E. (**2009**). “Partner-specific adaptation in dialog,” Top. Cogn. Sci. **1**(2), 274–291.

Brown, G., Currie, K., and Kenworthy, J. (**1980**). *Questions of Intonation* (Croom Helm, London), 206 pp.

Calhoun, S. (**2009**). “What makes a word contrastive? Prosodic, semantic and pragmatic perspectives,” in *Where Prosody Meets Pragmatics: Research at the Interface (Studies in Pragmatics 8)*, edited by D. Barth-Weingarten, N. Dehé, and A. Wichmann (Emerald, Bingley), pp. 53–78.

Chafe, W. (**1976**). “Givenness, contrastiveness, definiteness, subjects, topics, and point of view,” in *Subject and Topic*, edited by C. N. Li (Academic, New York), pp. 25–55.

Chevallier, C., Noveck, I., Happé, F., and 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., and Murphy, G. L. (**1982**). “Audience design in meaning and reference,” in *Language and Comprehension*, edited by J.-F. Le Ny and W. Kintsch (North-Holland, New York), pp. 287–299.

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

Diehl, J., Watson, D., Bennetto, L., McDonough, J., and Gunlogson, C. (**2009**). “An acoustic analysis of prosody in high-functioning autism,” Appl. Psycholing. **30**, 385–404.

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

Fine, J., Bartolucci, G., Ginsberg, G., and Szatmari, P. (**1991**). “The use of intonation to communicate in pervasive developmental disorders,” J. Child Psychol. Psychiatry **32**, 771–782.

Fraundorf, S. H., Watson, D. G., and Benjamin, A. S. (**2010**). “Recognition memory reveals just how CONTRASTIVE contrastive accenting really is,” J. Mem. Lang. **63**, 367–386.

Galati, A., and Brennan, S. E. (**2010**). “Attenuating information in spoken communication: For the speaker, or for the addressee?” J. Mem. Lang. **62**, 35–51.

Glasberg, B. R., and Moore, B. C. J. (**1990**). “Derivation of auditory filter shapes from notched-noise data,” Hear. Res. **47**, 103–138.

Gregory, M. L., Jurafsky, D., and Healy, A. F. (**2001**). “The role of the hearer in durational shortening,” in *Proceedings of the 7th Annual Conference on Architectures and Mechanisms for Language Processing (AMLaP 2001)*, Saarbrücken, Germany, p. 1.

Gussenhoven, C., Repp, B. H., Rietveld, A., Rump, H. H., and Terken, J. (**1997**). “The perceptual prominence of fundamental frequency peaks,” J. Acoust. Soc. Am. **102**, 3009–3022.

Horton, W., and Keysar, B. (**1996**). “When do speakers take into account common ground?” Cognition **59**, 91–117.

Keysar, B., Barr, D. J., and Horton, W. S. (**1998**). “The egocentric basis of language use: Insights from a processing approach,” Curr. Dir. Psychol. Sci. **7**, 46–50.

Kochanski, G., Grabe, E., Coleman, J., and Rosner, B. (**2005**). “Loudness predicts prominence: Fundamental frequency lends little,” J. Acoust. Soc. Am. **118**, 1038–1054.

Krahmer, E., and Swerts, M. (**2001**). “On the alleged existence of contrastive accents,” Speech Comm. **34**, 391–405.

Kraljic, T., and Brennan, S. E. (**2005**). “Prosodic disambiguation of syntactic structure: For the speaker or for the addressee?” Cogn. Psych. **50**, 194–231.

Ladd, D. R. (**2008**). *Intonational Phonology*, 2nd ed. (Cambridge University Press, Cambridge), 349 pp.

Levelt, W. J. M. (**1989**). *Speaking: From Intention to Articulation* (MIT Press, Cambridge, MA), 566 pp.

Lewis, D. (**1979**). “Scorekeeping in a language game,” J. Philos. Logic **8**, 339–359.

McCaleb, P., and Prizant, B. M. (**1985**). “Encoding of new versus old information by autistic children,” J. Speech Hear. Disord. **50**, 230–240.

McCann, J., and Peppe, S. (**2003**). “Prosody in autism spectrum disorders: A critical review,” Int. J. Lang. Commun. Disord. **38**(4), 325–350.

Nadig, A., and Shaw, H. (**2011**). “Acoustic and perceptual measurement of expressive prosody in high-functioning autism: Increased pitch range and what it means to listeners,” J. Autism Dev. Disord. **42**, 499–511.

Nilsonne, Å., Sundberg, J., Ternström, S., and Askenfelt, A. (**1988**). “Measuring the rate of change of voice fundamental frequency in fluent speech during mental depression,” J. Acoust. Soc. Am. **83**, 716–728.

Pechmann, Th. (**1984**). “Überspezifizierung und Betonung in referentieller Kommunikation” (“Overspecification and intonation in referential communication”), Ph.D. dissertation, Universität Mannheim, Mahnnheim, Germany.

Peppé, S., McCann, J., Gibbon, F., O’Hare, A., and Rutherford, M. (**2007**). “Receptive and expressive prosodic ability in children with high-functioning autism,” J. Speech Lang. Hear. Res. **50**, 1015–1028.

Pickering, M. J., and Garrod, S. (**2004**). “Toward a mechanistic psychology of dialogue,” Behav. Brain Sci. **27**, 169–226.

Pierrehumbert, J., and Hirschberg, J. (**1990**). “The meaning of intonational contours in the interpretation of discourse,” in *Intentions in Communication*, edited by P. Cohen, J. Morgan, and M. Pollack (MIT Press, Cambridge, MA), pp. 271–312.

Rooth, M. (**1992**). “A theory of focus interpretation,” Nat. Lang. Semantics **1**, 75–116.

Rooth, M. (**1996**). “On the interface principles for intonational focus,” in *Proceedings of SALT VI*, edited by T. Galloway and J. Spence (Cornell University, Ithaca, NY), pp. 202–226.

Rutherford, M. D., Baron-Cohen, S., and Wheelwright, S. (**2002**). “Reading the mind in the voice: A study with normal adults and adults with Asperger’s Syndrome and high functioning autism,” J. Autism Dev. Disord. **32**, 189–194.

Schafer, A. J., Speer, S. R., Warren, P., and White, S. D. (**2000**). “Intonational disambiguation in sentence production and comprehension,” J. Psycholing. Res. **29**, 169–182.

Scherer, K. R. (**1986**). “Vocal affect expression: A review and a model for future research,” Psych. Bull. **99**, 143–165.

Schwarzschild, R. (**1999**). “GIVENness, Avoid F, and other constraints on the placement of focus,” Natural Lang. Semantics **7**, 141–177.

Selkirk, E. (**2008**). “Contrastive focus, givenness and the unmarked status of ‘discourse-new,’ ” Acta Linguist. Hung. **55**, 331–346.

Shriberg, L. D., Paul, R., McSweeny, J. L., Klin, A., Cohen, D. J., and Volkmar, F. R. (**2001**). “Speech and prosody characteristics of adolescents and adults with high-functioning autism and Asperger’s Syndrome,” J. Speech, Lang., Hear. Res. **44**, 1097–1115.

Snedeker, J., and Trueswell, J. (**2003**). “Using prosody to avoid ambiguity: Effects of speaker awareness and referential context,” J. Mem. Lang. **48**, 103–130.

Swerts, M. (**2007**). “Contrast and accent in Dutch and Romanian,” J. Phonetics **35**, 380–397.

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

Veenker, T. J. G. (**2003**). “WWSTIM: A CGI script for presenting web-based questionnaires and experiments,” *version 1.4.4*, Utrecht University, The Netherlands.

Von Benda, U. (**1983**). “Zur Auditiven Beurteilung der Intonation Autistischer Kinder: Ein Hörexperiment” (“On the auditory assessment of autistic children’s intonation: A listening experiment”), in *Mündliche Kommunikation, Störungen und Therapie*, edited by D.-W. Allhoff (Scriptor, Frankfurt am Main, Germany), Vol. 10, pp. 63–70.

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

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