Accentuation as a cue for speaker adaptation

Constantijn Kaland, Emiel Krahmer, Marc Swerts

Tilburg centre for Communication and Cognition (TiCC), The Netherlands
{c.c.l.kaland, e.j.krahmer, m.g.j.swerts}@uvt.nl

Abstract

A spoken interaction can be viewed as a collaborative process between interlocutors. In order to get a smooth exchange of information, speakers tend to adapt their utterances to their addressees. The present study investigates to what extent prosodic accentuation in Dutch may cue whether speakers adapt to each other. It is known that accents in Dutch can be used to highlight whether information is given or new to the discourse. The current study aims to find evidence for the fact that accents as markers of information status can signal successful adaptation between speakers. In particular, the study uses dialogues in which accent distributions are congruent or incongruent with respect to the given-new distinctions across speaking turns to test whether listeners' judgements about partner adaptation are affected. Results indeed show that listeners can attribute such judgements to pitch accent distributions.

Index terms: discourse prosody, contrastive intonation, accentuation, partner adaptation

1. Introduction

Speaking partners share information in a discourse. This process is seen as collaborative in the sense that interlocutors gradually build up a common ground in the course of their interaction [1], [2]. The collaborative aspect of such discourse would appear from the fact that speakers tend to adapt their utterances to each other [3]. This has been shown, for instance, from the way speakers align lexical items. For instance, when one speaker uses the word *novel* to refer to a book, it is likely that a collaborative speaking partner may start using the same word later on in the interaction [4], [5]. Furthermore, speakers tend to lexically signal to each other whether information in the discourse is important or not. For instance, once entities are introduced in discourse speakers can refer to them later on using shorter descriptions such as pronouns [6]. Moreover, speakers use definite articles to refer to entities mainly when their listener shares knowledge of that entity [7]. Similarly, at the level of prosody there is evidence that speakers use analogous strategies. For example, repeated references tend to be shortened for the listener [8].

The current paper focuses on the extent to which variation in pitch accent distribution may signal whether speaking partners adapt to each other. Indeed, especially in Germanic languages, speakers can use accentuation to signal that new information should be added to the discourse model and that given information is already part of it [9]. In that sense, accents make discourse coherent and comprehensible. This is especially clear from what has been termed contrastive intonation patterns, that typically occur when two referents are minimally distinguishable between two successive discourse contexts [10]. For example, a speaker may first refer to a blue ball and later on to a red ball. In Germanic languages such as English and Dutch, *red* in the second utterance will generally be accented and *ball* will be de-accented [11]. The accented member of a contrast set will generally be uttered with a higher pitch compared to the de-accented member.

Various kinds of perception studies have shown that accent patterns are relevant for listeners as well. For example, Terken and Nooteboom [12] show that inappropriately accented information results in longer comprehension times. Nooteboom and Kruyt [13] reported that listeners appreciate sequences of utterances better when accent distributions matched given-new distrinctions. In a study by Dahan et al. [14] listeners are instructed to find the right referent among competitors. They show that contrastive intonation facilitates referent recognition in early stages of language comprehension. Further, Fraundorf et al. [15] find that listeners update the information status of both members of a contrast set (i.e. *red* and *ball*) when pronounced with a contrastive intonation. That is, listeners try to map the accented member to a new discourse entity and the unaccented member to a context inferable entity. Moreover, listeners accommodate for the lack of context when a contrastively accented utterance is presented in isolation [16]. That is, words that can be associated with the contrast set (i.e. *yellow* and *sport*) are recognized faster when a contrastive intonation is present than when it is absent.

The studies just mentioned indicate that listeners appear to be very sensitive to variations in accent distributions. In particular, they show that listeners appear to associate accents with specific contrast relations. Given the strong association between accents and information structure in Dutch, one could predict that dialogues are perceived as being more coherent if accents reflect given-new distinctions across speaking turns. That is, if a speaker signals by the presence or absence of accents that he or she has taken into account what a speaking partner had just said in a prior turn, this may be interpreted as evidence that this speaker has clearly paid attention to that other person's contribution to the conversation. Conversely, a dialogue in which the accents of a speaker are not linked to information given in a partner's previous turn, may be classified as incoherent and as an exchange in which speakers do not adapt to each other.

Thus, the present research investigates whether listeners can use contrastive intonation as a cue for the adaptation process in discourse. Therefore we create short dialogues of speakers uttering contrastively accented noun phrases (henceforth NPs). The order in which the NPs occur is manipulated such that the contrastive intonation is congruent or incongruent with respect to the previous NP produced by a different speaker. For example, *blue BALL* (accented word in capitals) is accented congruently when following after *blue triangle*, whereas *BLUE ball* is accented incongruently when following after *blue triangle*. Both congruently and incongruently accented utterances contribute to discourse equally with respect to their content. However, only congruently accented utterances can act anaphorically. That is, they implicitly show the speaker's knowledge of the previous utterance. Thus, an adaptive speaker can use accent

distribution to show that he or she accounts for what is said in the previous turn. We hypothesize that listeners judge the adaptation process as being more successful when they hear congruently accented dialogues as compared to incongruently accented ones.

2. Method

Our study consists of two parts. First an object naming task is carried out. Utterances collected in this task are analysed acoustically and used to create short dialogues. These dialogues are presented in a perception experiment.

2.1. Data collection

To elicit utterances with a contrastive intonation a cooperative object naming task is carried out. The participants are instructed to construct a specific shape by using geometrical figures in different forms (triangle, square, parallelogram) and colours (red, blue, yellow). The order in which the geometrical figures should be put together is manipulated such that two successive figures are minimally distinguishable. For example, a blue square may be followed by a blue triangle (shape contrast) or a yellow square (colour contrast). In this way shape contrasts elicit accented nouns and colour contrasts elicit accented adjectives. Twelve participants carry out the object naming task together with an experimenter. They alternately put a geometrical figure in place. Participants are instructed to describe their act using the same matrix sentence for each description: *Ik leg de blauwe driehoek hier* (I put the blue triangle here). In this way the grammatical position of the NP referring to the geometrical figure is kept constant. Furthermore, each utterance ends with the word *hier* (here) so that boundary tones do not occur within the target NP. A computer screen shows how the target shape consists of the geometrical figures. In addition, geometrical figures are numbered according to the order in which they have to be constructed (Figure 1). In total 38 shapes are constructed such that each description of a geometrical figure occurs repeatedly in each accent condition. All utterances are recorded as wave-files using a Marantz PMD-600 solid state recorder.

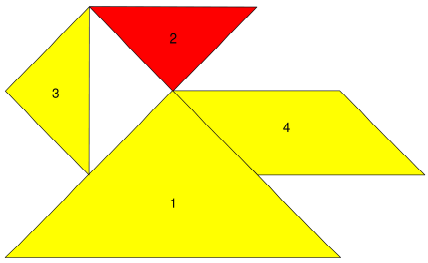

Figure 1: *Example shape consisting of geometrical figures.*

2.2. Acoustic analysis

A total of 288 NPs are selected from the twelve speakers that are recorded during the data collection (see section 2.1) to meet the requirements of the perception experiment (see section 2.3). NPs produced by the experimenter are omitted. On the basis of auditory judgements NPs with the clearest accent on either the adjective or the noun are selected. A software program [17] is used to extract the NPs from the wave-file recordings. As a sanity check, we investigate the effect of accent on the NPs by carrying out an acoustic analysis. That is, for each NP the pitch maximum of the adjective and the noun is measured using Praat [18]. Pitch is measured in ERB [19] for two reasons. First, ERB closely resembles the perception of prominence due to accentuation. Second, its logarithmic scale abstracts largely over gender differences. A repeated measurements analysis of variance is carried out with maximum pitch as dependent variable and with accent (2 levels: accented, de-accented) and word type (2 levels: adjective, noun) as independent variables.

Table 1: Mean pitch maxima (ERB) and standard deviations as a function of accent and word type.

Word type	Accented	De-accented
Adjective	5.21 (*1.17*)	4.40 (*.99*)
Noun	5.32 (*1.33*)	4.66 (*1.09*)

Speakers produce both accented adjectives and accented nouns with a higher pitch than their de-accented counterparts: $F(1,11) = 48.69$, $p < .001$, $\eta_p^2 = .82$ (Table 1). Although nouns exhibit slightly higher pitch maxima than adjectives, there is no significant effect of word type nor an interaction effect of accent and word type. Further, it is known that in Dutch the perception of contrastive intonation depends on both the prominence of the adjective and the noun [11]. Therefore, a pitch maxima difference measure is computed by subtracting the maximum pitch value of the de-accented word from the maximum pitch value of the accented word. The difference measure covers the values of both adjectives and nouns. Thereafter, values are computed per speaker pair used in the perception experiment (section 2.3) such that the value of pair AB is the average pitch maxima difference of speaker A and of speaker B (Table 2). In this way we obtain one value per speaker pair that resembles the clarity of their contrastive intonation. Results indicate that speaker pairs differ in the degree they use pitch to mark the difference between accented and de-accented words. Table 2 shows that pitch maxima differences are the smallest for pair EF and the largest for pair GH.

Table 2: Mean pitch maxima differences and standard deviations per speaker pair.

Speaker pair	*M* (*SD*)
AB	.76 (*1.08*)
CD	.80 (*.62*)
EF	.45 (*.50*)
GH	1.07 (*.98*)
IJ	.53 (*.64*)
KL	.85 (*1.58*)

2.3. Stimuli

Short dialogues are constructed consisting of four NPs (Table 3). NPs are taken alternately from one male and one female speaker (two NPs each). Background noise is added to the dialogue to mask edges at turn shifts and to favour the perception of the dialogue as a whole. NPs are ordered in such a way that accents are either used congruently or incongruently throughout the dialogue. The accentuation

Table 3: Examles of congruently and incongruently accented dialogues per accentuation pattern (accented words in capitals)

Congruity	Pattern	NP1	NP2	NP3	NP4
Congruent	AANN	RED triangle	BLUE triangle	yellow SQUARE	yellow TRIANGLE
	NNAA	yellow SQUARE	yellow TRIANGLE	RED triangle	YELLOW triangle
Incongruent	AANN	BLUE square	YELLOW triangle	red TRIANGLE	yellow TRIANGLE
	NNAA	red TRIANGLE	blue TRIANGLE	YELLOW square	YELLOW triangle

pattern is balanced over the dialogues. That is, in one half of the dialogues the first two NPs of a dialogue have an accented adjective whereas the last two NPs have an accented noun (AANN) and vice versa in the second half of the dialogues (NNAA). In total 72 dialogues are constructed using 6 speaker pairs. Per speaker pair 6 congruently and 6 incongruently accented dialogues are constructed.

2.4. Perception experiment

A web-based perception task presents pairwise one congruently and one incongruently accented dialogue (36 pairs in total). In order to reduce effects of other sound sources participants are instructed to do the experiment in a quiet room or to wear headphones. The entire experiment lasts about 20 minutes. Dialogue pairs occur in a different random order for each participant. Each dialogue pair is presented on an html-page designed using WWStim [20]. Each pair of dialogues is represented by two buttons which participants can click to hear them (Figure 2). Participants are told that speakers in the dialogue constructed a tangram figure together while describing the geometrical figures they used. Instructions to participants note that some speaker pairs collaborated better than others. The participants' task is to choose the dialogue in which speakers account for each other's utterances the best. They are instructed to pay close attention to intonation. Participants are allowed to listen to the dialogues as much as needed. Once they make their choice this is registered digitally and cannot be altered. The data is analysed as proportions where 1 corresponds with a choice for a congruently accented dialogue and 0 corresponds with a choice for an incongruently accented dialogue.

Figure 2: Screenshot of the perception experiment showing a play button for each dialogue and in Dutch the question: *In which dialogue do speakers account for each other the best?*

2.5. Participants

As for the data collection 12 participants acted as speaker (6 males, 6 females, M_{age} = 27.8 years, age range: 20-33). The perception experiment was done by 20 participants (9 males, 11 females, M_{age} = 27.5 years, age range: 20-35). None of the participants acted in both the data collection and the perception experiment. All participants who acted as speaker as well as the ones who completed the perception experiment are native speakers of Dutch.

2.6. Statistical analysis

To investigate participants preference for congruently or incongruently accented dialogues a chi-square test is performed on the proportion of choices. Furthermore, we test whether participants' choices depend on the speaker pair they heard. Therefore, a repeated measurements analysis of variance with mean proportion of choices as dependent variable and speaker pair (6 levels: pairs AB to KL) as independent variable is carried out.

3. Results

Participants have a preference for indicating the congruently accented dialogue as the dialogue in which speakers account for each other's utterances the best. That is, the congruently accented dialogues are chosen in 71.94% of the cases. This rate is statistically above chance: $\chi^2(1, N = 720) = 138.69$, $p < .001$. Furthermore, mean proportions of choices for the congruently accented dialogues show that participants are equally likely to choose that dialogue no matter whether the accentuation pattern was AANN or NNAA: $M_{AANN} = .72$, $M_{NNAA} = .72$. Results further indicate that the factor speaker pair has an effect on participants choices: $F(1,19) = 4.78$, $p < .01$, $\eta_p^2 = .20$ (Table 4). Pairwise comparisons (after Bonferroni correction) show that the pairs EF and GH differ significantly: *MD* (*SE*) = .19 (*.05*), 95% CI = (.02, .37), $p < .05$. Other pairs do not differ significantly from each other.

Table 4: Means and standard deviations of proportions of choices for congruently accented dialogues per speaker pair.

Speaker pair	*M* (*SD*)
AB	.76 (*.05*)
CD	.72 (*.06*)
EF	.63 (*.06*)
GH	.82 (*.04*)
IJ	.78 (*.05*)
KL	.63 (*.06*)

4. Conclusions

The present study shows that listeners can use the congruity of contrastive intonation patterns at a dialogue level to judge the interlocutors' adaptation process. That is, our judges appear to assess dialogue partners in a conversation as being more cooperative if the accent distributions match the given-new distinctions across speaking turns. Thus, accentuation is not

only a cue within discourse to signal its structure. Accentuation also exhibits information about the discourse process itself as a collaborative interaction between speakers.

Nevertheless, there is an effect of speaker pair in that listeners judge some pairs as adapting better than other pairs. Inspection of both Table 2 and Table 4 suggests that this side-effect is not simply a listeners' preference for certain speakers. On the contrary, speaker pairs that show large acoustic differences between accented and de-accented words are generally perceived as better adapters compared to pairs showing small acoustic differences (cf. pair EF and GH). Thus, the effect of speaker pair seems to be grounded in their acoustic characteristics of accentuation. Such a conclusion favours the view that that accentuation plays an active role in the perception of collaborative discourse.

Note that adaptation in the current study is different from a more classical interpretation of alignment. That is, from a social perspective adaptation has been explained as the interlocutors' mimicry of, for example, speech rate [21] or intensity [22]. Note that in the present approach we explicitly focus on the way prosody relates to meaning. That is, speaker A's "BLUE triangle" followed by speaker B's "RED square" would be a perfect mimicking of the contrastive intonation pattern. Nevertheless, the current study shows that for such sequences speakers are judged as not adapting well.

Further, it is crucial to note that listeners in the perception experiment are not addressees. In the perception experiment listeners hear the dialogue passively. That is, they are 'overhearers' [23] and do not actively build common ground with an interlocutor. With respect to this distinction one could think of a variant of the perception experiment in which participants are both interlocutor and judge. For example, they act in a dialogue with a partner that either responds with a congruent or incongruent intonation pattern. Participants may then be expected to judge the adaption process in a more radical way due to their active role in dialogue [23].

The current findings confirm previous work on the use of accents in discourse. That is, accentuation provides cues for how successive utterances are related [24]. The structure that can be derived from accents thus is a useful tool for both speakers and listeners to build new information on top of what has been said before. This study extends these findings by showing that the clearer accents are used to signal discourse structure the better the interlocutors are judged to account for each other. Thus, pitch accents signal both structure and speakers' interactional behaviour in discourse.

It has to be noted that in the current setup dialogues are reduced to combinations of short utterances (NPs). In spontaneous speech there may be more complex discourse structures and other cues to partner adaptation. Although lexical cues are beyond the scope of this study, future work will focus on other potential prosodic markers of partner adaptation. One such cue could be the use of boundary tones. It is known that those tones mark whether the speaker has more to say, asks for response or has finished the utterance [24]. It can therefore be assumed that besides pitch accents boundary tones may signal how well interlocutors account for each other.

5. Acknowledgements

The authors thank Marie Postma and Martijn Goudbeek for helpful discussions about the experimental setup and three anonymous reviewers for comments on a previous version of this paper.

6. References

[1] Clark, H. H., "Using Language", Cambridge: Cambridge University Press, 1996.

[2] Brown-Schmidt, S., "Partner-specific interpretation of maintained referential precedents during interactive dialog", JML 61: 171-190, 2009.

[3] Clark, H. H. and Wilkes-Gibbs, D., "Referring as a collaborative process", Cognition 22: 1-39, 1986.

[4] Pickering, M.J. and Garrod, S., "Toward a mechanistic psychology of dialogue", Behavioral and Brain Sciences 27: 169-225, 2004.

[5] Brennan, S. E. and Clark, H. H., "Conceptual pacts and lexical choice in conversation", Journal of Experimental Psychology: Learning, Memory, and Cognition 22: 1482-1493, 1996.

[6] Chafe, W., "Language and consciousness", Language 50: 111-133, 1974.

[7] Clark, H. H. and Marshall, C. R., "Definite reference and mutual knowledge", In A. K. Joshi, B. Webber and I. Sag (editors), Elements of discourse understanding, 10-63, Cambridge: Cambridge University Press, 1981.

[8] Galati, A. and Brennan, S.E., "Attenuating information in spoken communication: For the speaker, or for the addressee?" JML 61: 35-51, 2010.

[9] Terken, J.M.B. and Hirschberg, J., "Deaccentuation of words representing 'given' information: effects of persistence of grammatical function and surface position", Language and Speech 37(2): 125-145, 1994.

[10] Pechmann, Th., "Überspezifizierung und Betonung in referentieller Kommunikation", Dissertation, Universität Mannheim, 1984.

[11] Krahmer, E. and Swerts, M., "On the alleged existence of contrastive accents", Speech Communication 34: 391-405, 2001.

[12] Terken, J.M.B. and Nooteboom, S.G., "Opposite effects of accentuation and deaccentuation on verification latencies for given and new information", Language and Cognitive Processes 2(3/4): 145-163, 1987.

[13] Nooteboom, S.G. and Kruyt, J.G., "Accents, focus distribution, and the perceived distribution of given and new information", JASA 82(5): 1512-1524, 1987.

[14] Dahan, D., Tanenhaus, M. K. and Chambers, C. G., "Accent and reference resolution in spoken-language comprehension", JML 47: 292-314, 2002.

[15] Fraundorf, S.H., Watson, D.G. and Benjamin, A.S., "Recognition memory reveals just how CONTRASTIVE contrastive accenting really is", JML 63: 367-386, 2010.

[16] Braun, B. and Tagliapietra, L., "The role of contrastive intonation contours in the retrieval of contextual alternatives", Language and Cognitive Processes 25: 1024-1043, 2010.

[17] Audacity Team, "Audacity" (version 1.2.6), retrieved from http://audacity.sourceforge.net. 2006.

[18] Boersma, P. and Weenink, D., "Praat: doing phonetics by computer" (version 5.2.19), http://www.praat.org. 2010.

[19] Glasberg, B. R. and Moore, B. C. J., "Derivation of auditory filter shapes from notched-noise data", Hearing Research 47: 103-138, 1990.

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

[21] Giles, H., Coupland, N., and Coupland, J., "Accommodation Theory: Communication, Context, and Consequence", In H. Giles, N. Coupland, and J. Coupland (editors), Contexts of Accommodation. Developments in Applied Linguistics, Cambridge: Cambridge University Press, 1-68, 1991.

[22] Natale, M., "Converge of mean vocal intensity in dyadic communication as a function of social desirability", Journal of Personality and Social Psychology 32: 790–804, 1975.

[23] Schober, M. F. and Clark, H. H., "Understanding by addressees and overhearers", Cognitive Psychology 21: 211-232, 1989.

[24] Pierrehumbert, J. and Hirschberg, J., "The Meaning of Intonational contours in the Interpretation of Discourse", In P. Cohen, J. Morgan and M. Pollack (editors), Intentions in Communication, Cambridge: MIT Press, 271-311, 1990.