Answers to the Perception of Thematic Contrast and Questions regarding the Perception of Thematic "Non-Contrast"

Bettina Braun

Institute of Phonetics
Saarland University, Saarbrücken, Germany
`bebr@coli.uni-sb.de`

Abstract

In a previous production study [1] we explored the prosodic marking of thematic material in contrastive and non-contrastive contexts in German. While both conditions resulted in a prenuclear rise, we found that themes in contrastive context exhibited a significantly longer stressed vowel, together with a higher and later peak. Interestingly, speakers varied as to whether they used peak alignment, peak height or both for signalling thematic contrast. This might indicate that there is a continuum in contrast-marking rather than distinct accent categories.

In this paper we shall describe several perception studies that investigate which of the factors are important to make an utterance appropriate in contrastive and non-contrastive contexts. More specifically, we explored duration and (extent and temporal alignment of) f_0-movement in German prenuclear accents.

Results show that subjects have clear perceptual preferences in contrastive contexts which disappear in non-contrastive contexts. We therefore conclude that contrastive contexts impose strong constraints on intonational form whereas non-contrastive contexts seem to be rather under-specified which gives room for alternative interpretation. Obviously, this conclusion is problematic for semantic theories. If themes in both contrastive and non-contrastive contexts may be intonationally marked, the assumed direct link between intonational marking and contrastive interpretation becomes questionable.

1. Introduction

Most studies that investigate the phonetic properties of contrast aim at rhematic (newness) accents (e.g. [2, 3]) while research on thematic (given) accents is rather rare (e.g. [4]). This is surprising, given the growing interest of semanticists in contrastive themes (cf. [5, 6, 7]), especially with respect to their force to evoke implicatures.

While the semantics for contrastive themes[1] is well-formalised, the prosodic "categories" employed are not yet sufficiently explored. According to Steedman, e.g., thematic material may only be intonationally marked (= accented), if it "contrasts with a different established or accommodatable theme" [5, p. 656]. In an earlier production study [1], however, themes in both contrastive and non-contrastive contexts were intonationally marked. The only difference was the *extent* of theme marking that manifested itself in various prosodic variables. Such acoustic differences between themes in contrastive and non-contrastive contexts could be interpreted as meaningless phonetic variation or as meaningful phonological differences, depending on how they are perceived.

---

[1] We assume German themes to be sentence-initial.

We therefore investigated how different theme marking (and the accompanying rheme marking) influences perception. Our main aim is to explore the phonetic factors that contribute most to the perception of contrast and whose absence is most likely to induce a "non-contrastive" percept and/or a "neutral interpretation".

In sec. 2 we describe the perception of peak height and peak alignment in contrastive contexts. In sec. 3, the focus is shifted to non-contrastive contexts. There, we first motivate a modification of the stimuli before describing the actual experiment. In this main experiment, the perceptual influence of peak height and segmental duration is explored. Because the use of modified stimuli makes a direct comparison to the first perception experiment impossible, we summarise a control study that tests the perception of the new stimuli in contrastive context.

2. Themes in contrastive contexts

In a first experiment, subjects were asked to judge the appropriateness of resynthesised stimuli in contrastive contexts. As contrast "triggers", we used semantic parallelism, such as *'Marlene loves the sea, but Anna likes to go to the mountains'*. In order to reduce the complexity of the differences found in the production study [1], we concentrated on the manipulation of the f_0-movement in the prenuclear rise.

2.1. Method

Four versions of a rise were resynthesised in five different utterances using Praat [8], resulting in one version presumed to be non-contrastive (nc) and three versions presumed to be contrastive in varying degrees (see fig. 1).
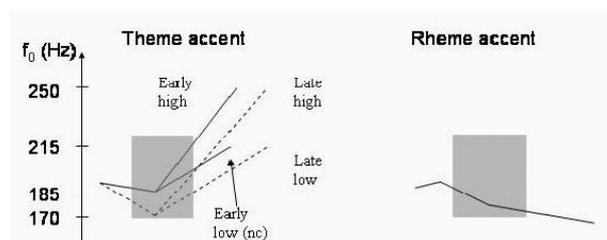

Figure 1: *Schematic figure of the four resynthesised thematic rises of a female voice. Early-low is presumed to be non-contrastive, all other versions are presumed to be contrastive. The shaded areas mark the stressed syllables.*

The choice of these four versions was guided by idealisations of the production data. While the thematic rises were manipulated, the rhematic part was kept constant as an !H+L* (au-

tosegmental annotation, cf. [10]; following [17], we also refer to this accent as *early accent*). This was the rheme accent type most commonly found in the production data. Since flat hat patterns (cf. [6, 9]) were rare in the production data, we resynthesised a short dip before the !H+L* (180Hz in the syllable preceeding the !H-syllable which was set at 200Hz).

16 linguistically naïve German subjects heard all 20 stimuli in random order. They were told to rate the appropriateness of the melody of the test utterances in the given context of a parallel construction using a seven-point scale. Also, they were explicitly told to ignore the actual content of the sentences and any unnaturalness resulting from resynthesis. Sentences are similar to the following one (where "Marlene" was manipulated):

Visual presentation:	Anna liebt das Meer, aber . . .
	'*Anna loves the sea but . . .*'
Auditory presentation:	Marlene fährt gern in die Berge.
	'*Marlene likes to go to the mountains.*'

The first half of the parallel construction was presented visually, followed by three dots to indicate the continuation. After clicking a loud-speaker symbol on the screen, the subjects heard a resynthesised continuation to the utterance. Five training examples were presented to familiarise subjects with the speech stimuli and the task.

We are well aware of the criticism of unnaturalness and the uncontrollable effects of visual presentation but we believe that parallel constructions are syntactically and semantically sufficiently marked to constrain the degrees of interpretational and intonational freedom. In not resynthesising the first part of the parallel construction, we avoid the as yet rather unexplored territory of inter-clause peak-height relations.

2.2. Hypotheses

We predicted that the non-contrastive version (*early-low*) would be rated significantly worse than all contrastive versions in the given contrastive context. Furthermore, it was hypothesised that a stronger marking of contrast (by both higher *and* later peak) would give higher appropriateness judgements than just one contrastive "feature" alone (higher *or* later peak). That is, we assumed the following preference ratings[2]:

> *late-high* > *early-high* = *late-low* > *early-low*

2.3. Results

Normalised ratings were analysed using the non-parametric Friedman test and a Wilcoxon signed ranks test for post-hoc analysis. Surprisingly, we found two groups of subjects whose rating differed significantly $F(3,234)=5.09$, $p=0.02$ (ANOVA with rise-type as within-subjects factor and group as between-subjects factor). The rating behaviour of one group (4 subj.) was not consistent; so they are not discussed further here. Preference ratings of the larger group (12 subjects) differed significantly ($\chi^2=28$, df=3, $p=0.000$), so we concentrate on them. We found a significant preference for all contrastive versions over the non-contrastive stimuli which supports our main hypothesis. However, the hypothesised preference grouping (preference for *late-high* over both *early-high* and *late-low*) was not observed:

> *early-high* = *late-high* > *late-low* > *early-low*

---

[2]Throughout the article, ">" stands for "significantly preferred over ($p<0.05$)" while "=" means "no significant difference".

2.4. Discussion

Versions with **high** peaks formed one group that was rated best. In these versions, alignment differences played no role for appropriateness, which might be due to a ceiling effect. *Late-low* peak versions were judged worse than both high peak versions but still differ significantly from the non-contrastive one (*early-low*). It appears that the peak alignment (or the accompanying lower trough) is crucial only for the interpretation of low peaks.

The primacy of peak height over peak alignment found here is in line with other studies, e.g. Bartels and Kingston who summarise that "peak height was found to be by far the most salient cue to contrastiveness." [3, p. 1]. It is worth noticing, however, that the importance of the scaling of accents on perception has been largely ignored in autosegmental tonal categories [10, 11] and is acknowledged only in perceptually-oriented annotation schemes, like the British School (e.g. [12]).

The fact that subjects rated both high versions (*early-high* and *late-high*) equally good allows for speculations about the impact of overall f_0-excursion on perception. Overall f_0-excursion is greater in the *late-high* versions but these are not rated better. Both high versions, however, have a steeper slope of the rise compared to the versions with low peaks which might be the reason for the observed preference pattern. [3] also suggested that the slope may be important for perception.

The results suggest a kind of trading relation between peak height and peak alignment which mirrors the variance found in production (there, 42% of the contrastive–noncontrastive pairs showed both peak height and peak alignment differences, 15% only peak alignment differences and 29% only peak height differences). Surprisingly, the combination of higher and later peaks which was most common in production was not preferred most in perception.

3. Themes in non-contrastive contexts

In this first experiment we asked subjects to judge the appropriateness of a test utterance in a contrastive context (parallel construction). We found that listeners significantly preferred the presumed contrastive versions over the presumed non-contrastive ones. In this section we discuss the perception in non-contrastive context where the opposite preference pattern might be expected. It became necessary to perform some informal pilot studies (sec. 3.1) that motivate the refinement of the stimuli used. Sec. 3.2 describes the actual perception experiment in non-contrastive contexts. In sec. 3.3 we replicate the first experiment with the modified stimuli to be able to compare the results. These are described and discussed in 3.5 and 3.6.

3.1. Pilot study

On the basis of some informal experiments we concluded that the duration of the stressed syllable is an important contributing factor in the perception of "non-contrast", an issue that we haven't accounted for so far. Consequently, in the following pilot experiment we manipulated the variables peak height and duration rather than peak height and alignment. Note that what we treated as alignment effects in the production study could actually have been due to differences in duration.

For the manipulation of f_0-movement we used those versions that yielded best and worst appropriateness values in the study with contrastive context (*early-high* and *early-low* in fig. 1). The duration of the whole preverbal constituent (prosodic word, including articles or prepositions) of the existing stimuli was shortened to 80% according to the results of

the production study. Since slope seemed to be important for perception (which was also emphasised by [3]), it was held constant across duration conditions. Similarly, the alignment of the f_0-minimum (L) was kept constant, according to various studies that found a consistent alignment of L with the segmental structure (e.g. [13, 14]). This resulted in a slightly later peak in conditions with shorter constituents[3].

To establish a non-contrastive context, we decided to use focus-inducing questions. These were recorded in a way that the constituent that would end up as the pre-verbal (thematic) material of the answer (= target-utterance) was intonationally marked as given in the corresponding question (by deaccenting it). Following [5], thematic parts are labelled with $[\theta \ldots]$, the rhematic, new parts with $[\rho \ldots]$ (capital letters mark the nuclear pitch accent, the focus-exponent).

Question:	Was $[\rho$MACHT] $[\theta$ Marlene] denn gerne?
	'What is $[\theta$ Marlene] $[\rho$INTERESTED] in?'
Answer:	$[\theta$ Marlene] $[\rho$ fährt gern in die BERGE].
	'$[\theta$ Marlene] likes $[\rho$ to go to the MOUNTAINS].'

Most of the eight subjects were highly irritated by the rhematic !H+L* on the focus-exponent (here: "Berge") although this was the accent most often found in the production data. Furthermore, Kohler described the semantics of such early rhematic accents as denoting "established facts" [17] which would be a valid interpretation in the examples used (but one that subjects rejected for some reason). They complained that the "main accent" was not "strong enough" and that the utterance was "wrongly accented". There was no significant preference for any of the resynthesised versions which might be caused by the "inappropriate" rheme accent or by an insufficient context.

3.2. Method

Based on previous experience the stimuli were resynthesised with *medial* rheme accents (terminology by [17], corresponding to autosegmental H*) to make the rhematic part more appropriate to the question. The nuclear syllable with set at 205Hz, the preceding syllable at 180Hz. As described above, stimuli varied in both f_0-movement and segmental duration. Furthermore, we started to use magnitude estimation instead of the seven-point scale because magnitude estimation has proved a highly reliable and highly replicable method for linguistic judgements (e.g. [18]).

To strengthen the intended non-contrastive interpretation, the pre-context was extended by visually presenting two or three sentences with the same topic as the theme of the answer (= target utterance). It was suggested to subjects that this could be interpreted as a dialogue with the precontext said by the same speaker who subsequently provides the answer.

Visual Context: I really don't know what I could buy as a birthday present for Marlene. Do you have an idea?

To familiarise subjects with the magnitude estimation task they were first asked to judge the length of eight lines. Afterwards they were told that this method was now being used for judging the appropriateness of speech data. Mores specifically, their task was to judge the appropriateness of the *intonation* of the answer (= test utterance) with respect to the question.

---

[3]Note that this is not in accordance with the results of [15] who found a stable anchor for H as well which makes slope an epiphenomenon. Perceptual evaluation of the resynthesised stimuli, however, convinced us not to vary the slope.

3.3. Control study: medial rheme and contrastive context

To allow direct comparison of the influence of duration and f_0-movement on preference ratings in different contexts we replicated the above experiment in contrastive contexts (visual presentation of parallel construction). We used the same material (manipulation of both duration and f_0, medial rhematic accent) and method as in non-contrastive contexts. There were two orders of presentation: one starting with the non-contrastive context and one starting with the contrastive context (23 subjects altogether). With this control study we further aim to discover whether a different sort of rheme accent influences perception in contrastive contexts as well. In the following sections, the results for the contrastive context and the non-contrastive context are presented and discussed in parallel.

3.4. Hypotheses

Non-contrastive context: According to the results of the production study we hypothesised that the less marked themes should be more appropriate in non-contrastive contexts. As both *low-long* and *high-short* contain just one cue to prominence we assumed that they would be grouped together:

$$low\text{-}short > low\text{-}long = high\text{-}short > high\text{-}long$$

Contrastive context: For the contrastive control study we expected the opposite pattern, again assuming that f_0-movement and duration are additive factors:

$$high\text{-}long > high\text{-}short = low\text{-}long > low\text{-}short$$

3.5. Results

Magnitude estimation values were logarithmised ($\log_{10}$) and normalised. In contrast to the ratings in the previous studies these are scalar data which can be analysed using parametric tests. To investigate whether the order of presentation (contrastive first vs. non-contrastive first) had an influence on the rating behaviour in the two contexts, we calculated two ANOVAs with rise-type as within-subject factor (with the four conditions *low-short, low-long, high-short* and *high-long*) and presentation-order as between-subject factor, one for contrastive and one for non-contrastive contexts. In non-contrastive contexts, there was a significant interaction between presentation-order and rise ($F(3,276)=3.57$, $p=0.015$). Therefore only the results of the group that first performed the non-contrastive task were analysed (15 subjects). In the contrastive contexts, there was no significant interaction between order and rise, so ratings were pooled (23 subjects).

Non-contrastive context: In non-contrastive contexts, there was no consistent preference across subjects, similar to the results of the pilot study reported in sec. 3.1.

Contrastive context: First, the use of a medial rheme accent (as opposed to the early rheme accent in the first experiment (sec. 2)) had no influence on subjects' ability to perform the task. There was a significant difference in preference for the four versions in contrastive contexts ($F(3,282)=6.41$, $p=0.000$)[4]. Subjects showed a clear preference for versions with a high prenuclear peak over versions with a lower one (as expected). However, contrary to expectation, duration and f_0-movement did not turn out to be additive factors:

$$high\text{-}long = high\text{-}short > low\text{-}long = low\text{-}short$$

---

[4]The same preferences and significances are found in smaller data sets (e.g. with the first 15 subjects ($F(3,165)=4.55$, $p=0.000$)).

3.6. Discussion

A 'wrong' kind of rheme accent impaired subjects' ability to rate the stimuli in non-contrastive context. In contrastive contexts, the kind of rheme accent did not matter. This difference might be partly explained by the *visual* presentation of the parallel constructions which allows more intonational freedom than the rigid question in non-contrastive contexts. It is interesting to compare this to the production data where we also found a significant difference in the distribution of rheme-accents (x^2=9.64, df=1, p<0.01) but in reverse order: in contrastive contexts, there were significantly more early rheme accents (x^2=17.5, df=1, p <0.01), while medial and early rheme accents were equally distributed in non-contrastive contexts.

In contrastive contexts, the acceptability was solely based on f_0-movement, not on duration. This is no surprise given that f_0-excursion is not influenced by different speaking rates [13]. That is, segmental duration can be shortened which need not affect the extent of f_0-movement. Peak height (and slope) were most important for the perception of thematic contrast, but these factors cannot be described in autosegmental terms.

It is surprising that subjects found all versions equally appropriate in the neutral, non-contrastive contexts. [19] similarly found that in clearly marked neutral (declarative) contexts, marked forms (questions) were accepted but seen as carrying additional meaning (e.g. emphasis). Since the opposite does not apply, non-contrastive contexts can also be seen as prosodically under-specified so that they can give room for alternative functions of prosody (e.g. impatience, strong opinion).

A further explanation is that subjects "repair" the context to accommodate for the perceived contrast [5, 20]. In a preliminary study we tried to investigate whether our results are mainly due to accommodation or to alternative functions of prosody. We therefore asked a small group of subjects to choose between the two versions *long-high* and *short-low* in the non-contrastive contexts described above. They were told to explain *why* they preferred a certain version. In the direct comparison, there was a preference for the *short-low* versions, which was mainly justified in information-structural terms. The reasons for choosing the presumed inappropriate *long-high*-versions, on the other hand, were mostly described as "better intelligible", "less bored", "more enthusiastic". These preliminary results cannot be explained by accommodation of contrast but rather by the prosodic freedom of non-contrastive contexts.

4. Conclusions

Peak height was a stronger cue to thematic contrast than both peak alignment and segmental duration. Alignment differences only played a role when the prenuclear peak was low (*late-low > early-low*). As opposed to peak alignment, segmental duration could not compensate in the same way for the missing f_0-movement to achieve a contrastive interpretation (*low-long = low-short*).

The contrastive context allowed for freedom concerning the rheme accent marking but subjects had clear preferences for the theme accent. On the other hand, the non-contrastive contexts demanded a medial rheme accent but theme marking was irrelevant. We plan to investigate further whether there are restrictions to the compositionality of accents (unlike proposed in [5, 21]). Restricted compositionality is not that uncommon, as it even occurs in the case of preheads of different pitch height and nuclear accent types in Dutch [22].

While there were clear preferences in contrastive contexts, all versions were equally appropriate in non-contrastive contexts. From that and a preliminary pilot test, we concluded that non-contrastive contexts are prosodically under-specified so that intonation is free to be used to convey paralinguistic meaning apart from contrast or information structure.

5. References

[1] Braun, B.; Ladd, D.R., 2003. Prosodic correlates of contrastive and non-contrastive themes in German. *Proc. Eurospeech.* Geneva.

[2] Bolinger, D.L. 1961. Contrastive accent and contrastive stress. *Language*, 37, 83-96.

[3] Bartels, C.; Kingston, J., 1994. Salient pitch cues in the perception of contrastive focus. Bosch, P., van der Sandt, R. (eds.) *Focus and Natural Language Processing* vol. 1. Intonaton and Syntax, p 1-10.

[4] Mehlhorn, G., 2001. "Produktion und Perzeption von Hutkonturen im Deutschen", *Linguistische Arbeitsberichte*, 77, 31-57.

[5] Steedman, M. 2000. Information structure and the syntax-phonology interface. *Linguistic Inquiry*, 31(4), 649-689.

[6] Büring, D., 1997. *The Meaning of Topic and Focus – the 59th Street Bridge Accent*. London: Routledge.

[7] van Hoof, H., 2003. The rise in the rise-fall contour: does it evoke a contrastive topic or a contrastive focus. *Linguistics*, 41(3), 515-563.

[8] http://www.fon.hum.uva.nl/praat

[9] Cohen, A.; 't Hart, J., 1967. "On the anatomy of intonation", Lingua, 19, 177-192.

[10] Pierrehumbert, J., 1980. *The Phonology and Phonetics of English Intonation*. Cambridge MA: PhD Thesis.

[11] Beckman, M.; Ayers, G., *Guidelines for ToBI Labelling*. Ohio State University: Technical report.

[12] Crystal, D. 1969. *Prosodic Systems and Intonation in English*. Cambridge: Cambridge University Press.

[13] Caspers, J.; van Heuven, V.J., 1993. Effects of time pressure on the phonetic realization of the Dutch accent-lending pitch rise and fall. *Phonetica*, 50, 161-171.

[14] Prieto, P.; van Santen, J.; Hirschberg, J., 1995. Tonal alignment patterns in Spanish. *J. of Phon.*, 23, 429-451.

[15] Arvaniti, A.; Ladd, D.R.; Mennen, I., 1998. Stability of tonal alignment: The case of Greek prenuclear accents. *J. of Phonetics*, 26, 3-25.

[17] Kohler, K.J., 1991. Terminal intonation patterns in single-accent utterances of German: Phonetics, phonology and semantics. *Arbeitsberichte des Instituts für Phonetik der Universität Kiel (AIPUK)*, 25, 115-185.

[18] Bard, E.; Robertson, D.; Sorace, A., 1996. Magnitude estimation of linguistic acceptability. *Language*, 72(1), 32-68.

[19] Andreeva, B., Barry, W.J., 1999. Intonation von *Checks* in der Sofia Varietät des Bulgarischen. *Research Report Phonetics, Univ. of the Saarland (PHONUS)*, 4, 1-14.

[20] Thomason, R.H., 1990. Accommodation, meaning, and implicature: Interdisciplinary foundations for pragmatics, In *Intentions in Communication*, P. Cohen; J. Morgan; M. Pollack (eds.). Cambridge MA: MIT Press, 325-363.

[21] Pierrehumbert, J., Hirschberg, J., 1990. The meaning of intonation in the interpretation of discourse, In *Intentions in Communication*. P. Cohen; J. Morgan; M. Pollack (eds.). Cambridge MA: MIT Press, 271-311.

[22] Grabe, E.; Gussenhoven, C.; Haan, J.; Marsi, E.; Post, B., 1997. Preaccentual pitch and speaker attitude in Dutch. *Language and Speech*, 41(1), 63-85.