Fine phonetic detail in prosody. Cross-language differences need not inhibit communication

Chapter · January 2012

DOI: 10.1515/9783110301465.259

2 authors, including:

Bistra Andreeva
Universität des Saarlandes
87 PUBLICATIONS **552** CITATIONS

SEE PROFILE

Fine phonetic detail in prosody. Cross-language differences need not inhibit communication.

Bistra Andreeva and William Barry
Universität des Saarlandes

0. Abstract

Two experiments are directed towards the 'function' and 'communication' themes of this volume. They examine the auditory consequences and the communicational effect of systematic differences that have been found between Bulgarian and German in the production of phrasal prominence. In experiment one, five unmanipulated versions of a German sentence differing in focus, and therefore in the degree of prominence of two critical words implicated in the focus conditions, were presented in pairs to Bulgarian and German subjects. Their task was to judge the degree to which the prominence of the critical words differed. In a second experiment, the questions used to elicit the different focus realizations of the sentence were paired with each of the focus versions. Subjects were asked to judge how well the answer matched the question. In the meta-linguistic first task, evidence was found to support the hypothesis that the greater use of signal intensity in Bulgarian prominence production is paralleled by greater sensitivity to intensity differences in perception. In the functional second task, there was no indication that the differences in production and indeed the greater sensitivity to signal intensity have any communicational consequences.

Accentuation, perception, cross-language differences, communicational importance

1. Introduction

Important as the recognition of the systemic nature of language structure has been for the development of linguistics in the 20th century (de Saussure 1916, Trubetzkoy 1939), the fateful seed of separation sown by Trubetzkoy has had unintended but enduring consequences for speech and language research. The argument that the study of form and function

of language, i.e. phonology, should be studied separately from the study of the substance, the acoustics and articulation of speech, i.e. phonetics, may have allowed scholars to focus on less complex scenarios of communicative reality and, presumably, considerably advance our understanding of the separate pictures. However, it has been at the expense of the wider vista. In over-simplified terms, the result has been, until this decade, to a large extent, the study of the phonetics of speech in relation to phonological constructs rather than a study of speech communication where the direct phonetic contribution to the communicative process is considered in its own right both within and outside the particular phonological currency of the period (cf. Kohler 2007a,b for a more thorough-going discussion).

The seminal paper by Hawkins and Smith (2001) reflected the unease of many at the primacy of phonological concepts, in particular the phoneme, but it also highlighted problems of the separation of formal descriptive levels and the compartmentalization of processing modules. It brought together, discussed and re-interpreted a vast amount of disparate work within the Polysp model – "a polysystemic, phonetically-rich approach to speech understanding". This may have triggered some refocusing of attention in phonetics to the importance of fine phonetic detail for the speech communication process, and it has certainly provided an alternative theoretical umbrella for speech communication work outside mainstream formal phonology.

There have, of course, always been islands of recognition for work which has focused on the direct importance of phonetic detail for communication.

Phonetic studies of 'talk in interaction' (Auer 2006, Couper-Kuhlen and Selting 2006, Local, 2005, 2007) show how durational, amplitude and voice-quality modifications operate on top of the phonologically structured utterances to help steer the interactional process.

The empirical basis for the validity of exemplar theory in phonetic perception (Goldinger et al. 1991; Goldinger 1997, 2000; Johnson 1997) is the long accepted fact that we do not only understand an utterance but recognize the speaker who uttered it. The theory postulates that we store mental records not only of the segmental and prosodic structural relations of the languages we speak but also of the individual colouring of the properties carrying those relations by the speakers we are familiar with. At a pre-scientific level, many of us could presumably agree that those properties attain a quasi-categorical status, since we are able to say that X (the voice of an unknown person we hear) has a voice quality similar to Y (a person we know).

Nolan (1992) found that the incomplete alveolar gesture of ostensibly assimilated (i.e. not palatographically registered) word-final alveolars preceding word-initial /k/ (the road collapsed / the rogue collapsed) still projected sufficient transition information in the preceding vowel to distinguish it from the /-g#k-/ case.

Heid and Hawkins (2000) and Hawkins and Heinrich (2009) have shown that the preparatory articulatory setting for an upcoming /r/ can be present much earlier than in the preceding vowel, and that lexical access for the /r/-word is easier because of the consequent /r/-resonances.

In speech synthesis research Hawkins and Slater (1994) demonstrated the importance of longer-domain effects for processing synthetic speech, and our own work (Barry et al. 2001) demonstrated that concatenative (diphone) synthesis is immediately more natural when diphone-selection is controlled for the previous and the following vowel context as well as for the two segments involved in the diphone.

The nub of Hawkins and Smith's (2001) argument is that distributed phonetic properties, fine phonetic detail reflecting the segmental structure at some level of production, form a 'perceptually coherent' pattern within the language experience of the listener. This, together with situational information, can support successful "speech" communication when nearly all the properties of the underlying segmental structure are missing, as their example of [ə̃ə̃ə̃] for "I don't know" illustrates.

The overarching message from the studies mentioned is that the sequentially correct segmental skeleton of an utterance is not exclusively important for utterance comprehension (and is certainly not the only contributor to communicative ease). Residual information remains and with the support of a prosodic structure (supporting 'perceptual coherence') reflecting the information structure of the intended utterance the semantic content that would have been carried by the segments is reconstituted.

As Hawkins and Smith's (2001) extreme "I don't know" example illustrates, the leeway for segmental distortion is very considerable. The leeway that is allowed for prosodic distortion, on the other hand, has not been considered and it is certainly not assumed that prosodic ambiguity is compensated for by articulatory accuracy, although there has long been evidence that durational manipulation of segments can effect focus interpretation (Huggins 1972). Also, the line of research on segmental intonation by Niebuhr (2008, 2009) shows that there is segmental variation which is able to contribute to the pitch course of utterances. Linguistic and meta-linguistic observation, however, paint a conflicting picture of how important prosodic accuracy is for communication.

Over the decades, intonation studies have found ample evidence of timing constraints for tonal accent contours relative to segmental structure (cf. e.g. Bruce 1977, Arvaniti, Ladd and Mennen 1998, Atterer and Ladd 2004, Dilley, Ladd and Schepman 2005) and these can have direct communicative consequences (Kohler 2005). Kohler (1987) demonstrated the categorical switch from one prosodic-semantic category to another that a small shift in the start of a falling nuclear tone induces and in Niebuhr (2006) and Niebuhr and Pfitzinger (2010) it was shown that the timing of accompanying changes in intensity are also critical. Niebuhr (2007) showed the importance of steepness of the rise and fall and of the breadth of a tonal peak in distinguishing between an early, medial or late tonal accent. It is also accepted that small shifts in the alignment of tonal peaks and troughs reliably distinguish between dialects (e.g. Kohler 2007). On the other hand, dialects can also vary in the way they signal information structure, sometimes to the point of completely contradictory signals, without noticeable loss of mutual understanding. A well-known example is the tendency of some English dialects to continue to make repeated words (i.e. known information) prominent while others de-accent them. Everyday observation of radio and television announcers using contextually incorrect accentuation and intonation contours (apparently without complaint from listeners) also indicates that inaccuracies occur – and are registered and compensated by the listener, presumably using the same contextual information that made the error conspicuous.

In summary then, prosodic structure has been shown to operate within an intricate system of tonal and energy dynamics and of timing in relation to the segmental skeleton that carries it. Yet, just like its mirror image, the segmental bones floating on the prosodic waves of energy and tone, it is prone to distortion, apparently without undermining the communicative process. So how important is all the detail?

The goal of this study is to consider the effect of phonetic detail in phrasal accentuation for the perception of information-structure (focus) across languages. It seeks to determine whether the degree of inter-language variation that has been observed in the use of duration, intensity and f_0 in the production of focused and non-focused lexical items has perceptual consequences for the communicationally important interpretation of focus.

The perceptual correlate of placing a linguistic 'accent' on a word in an utterance in order to bring it into focus is the increased prominence of the stressed syllable of the word. Conversely, a de-accented word has reduced prominence. Of the four acoustic parameters that are universally available to induce an increase or decrease in perceived prominence –

duration, intensity, f_0 and spectral energy distribution – it has been shown that different languages use them to differing degrees, i.e. they are differently weighted from language to language (Andreeva et al. 2007, Barry et al. 2007, Koreman et al. 2008, Koreman et al. 2009, Andreeva et al. 2010). Recent work has shown that languages attach different degrees of importance to duration and f_0 in functional prosodic judgments. Cumming (2010) has shown that the two parameters interact in marking group boundaries, and that Swiss German attach a lesser degree of importance to increased syllabic duration and greater importance to a rising f_0 contour than (Swiss) French native speakers when deciding whether a sequence of five elements (digits and or letters) is divided into two groups (e.g. AB+ CDE or ABC+DE).

Such differences are potentially a source of interference in an L2 situation. If a language makes little use of one parameter that is used dominantly in another, obvious production difficulties arise and it is assumed that perceptual processing becomes more difficult. In some cases it could even result in misunderstanding of the intended focus condition (though whether it would result in mis-communication or in contextual correction and the recognition of 'foreign-ness' in the speaker is presumably a matter of listener awareness). In the case of rhythmic grouping a misinterpretation of the intended grouping would be unlikely because both French and German use some degree of final lengthening and final rise to mark the group. It was only in the experimental condition created by Cumming, where conflicting cues were located on the second and third element, forcing a choice between the two cues to perform the grouping task, that the differing preference was revealed. The case of phrasal prominence examined in this study is rather different. The different weighting of cues could theoretically lead to a focused item in one language being interpreted as less prominent, perhaps even as not focused, by a speaker of another. Then indeed communication could be impaired beyond a mere slowing down of processing.

An immediate objection to such suggestions of interference-led breakdowns in communication is that the differences in weighting are statistical not categorical on the production side, and on the perception side they are identified by experimental means, almost exclusively by parameter manipulation. Since in natural production the prominence-giving parameters are, for the greater part, correlated, and it is mainly the steepness of the slopes (e.g. intensity increasing more than syllable duration in one language and less in another) which differs between languages, the inherent redundancy of the multiple interacting cues should prevent any misinterpretations. The experiments reported here were carried out to test this objection, using naturally produced stimuli. Thus

the parameter weightings used were part of the natural variation found in prompted production, neither enhanced nor reduced by manipulation.

Two tasks were chosen, one a meta-judgment of greater or lesser prominence, the other a communication-linked task dependent on the ability to make an implicit judgment on the information structure of an utterance. German and Bulgarian were chosen for structural reasons and because analyses of their prominence-giving production patterns had revealed differences.

2. Production differences and hypotheses for perception

With increasing levels of accentuation (from de-accented to nuclear accented), Bulgarian (BG) and German (D) show different degrees of change in the three main accent-bearing parameters – f_0, intensity and duration (cf. Andreeva et al. 2007, Barry et al. 2007, Koreman et al. 2008, Koreman et al. 2009, Andreeva et al. 2010a, Andreeva et al. 2010b). German employs increased duration significantly more than Bulgarian, which shows less durational change with increased prominence. Conversely, Bulgarian increases syllable intensity with accentuation significantly more than German. f_0, on the other hand is clearly important for both Bulgarian and German speakers. These production differences between focus conditions, resulting in different accentuation levels, calculated over all speakers and all sentences, are shown in table 1. Values are normalized relative to the average for the corresponding measurement units over the sentence for critical word 1 (CW1), which is early in the sentence but not initial and critical word 2 (CW2), which is late but not sentence-final.(see table 4):

duration difference in % from accented to de-accented		
	BG (CW1/CW2)	D (CW1/CW2)
rime duration	13,5 / 12,7	24,8 / 18,6
intensity difference in dB from de-accented to accented		
	4,3 / 8,6	2,7 / 6,8
f_0 difference in semi-tones from accented to deaccented		
	5.3 / 6.1	4.8 / 5.5

Table 1 Normalized duration, intensity and f_0 difference between nuclear accented and (de-)accented conditions over the first critical word (CW1) and the second critical word (CW2) averaged over all speakers and all sentences.

A complementary way to view these differences is by showing the differing production ranking of the parameters within each language (see table 2):

Bulgarian:	f_0 > intensity > duration
German:	f_0 > duration > intensity

Table 2 Production ranking

The stimuli used in this study reflect the above production differences between the two languages. Table 3 gives the range of normalized values over all speakers for the German sentence (used in both experiments) and the Bulgarian sentence (presented to Bulgarian listeners in experiment 2). The measures for calculating f_0 change for the first and second critical words (CW1 and CW2), vowel intensity in dB for the stressed syllable and rime duration in the stressed syllable, are given in appendix 1. Values for the Bulgarian sentence used in experiment 2 are also given.

Mean duration difference in % from accented to de-accented		
	BG (CW1/CW2)	D (CW1/CW2)
rime duration	19,36 / 0,74	30,01 / 15,2
Mean intensity difference in dB from de-accented to accented		
	4,0 / 9,6	3,0 / 4,9
Mean f_0 difference in semi-tones from accented to de-accented		
	4.9 / 6.9	3.5 / 4.1

Table 3 Normalized duration, intensity and f_0 difference between nuclear accented and (de-)accented conditions over the first critical word (CW1) and the second critical word (CW2) averaged over all speakers for the experimental sentence

Given the assumption, which is not only intuitively plausible but has considerable empirical support[1], that different patterns in production result in different perceptual sensitivities, it is to be expected that speakers of BG and D attach different perceptual weighting to changes in the relevant parameters. An immediate hypothesis from the observed

1 Long established evidence is the accordance between the production and perception of VOT in the voiced-voiceless distinction (Lisker and Abramson 1964, 1967). Recent evidence in the prosodic domain is the differing use in production and perception of duration and f_0 for phrasal demarcation in Swiss German and French (Cumming 2010).

production patterns is, therefore, that speakers of one language will judge the degree of prominence in the other language differently from a native speaker. More specifically: If a parameter that is exploited strongly in language A changes less with an increase in accentuation in language B, language A listeners will judge the prominence difference of a language B stimulus pair to be weaker than language B listeners do. Likewise, if a parameter that is exploited to a *lesser* extent in language A changes *more* with an increase in accentuation in language B, the language A listeners will again judge the prominence difference of a language B stimulus to be weaker than language B listeners do because they are less sensitive to the larger change.

This cross-language interference may have communicational implications. Information structure is signaled prosodically in many languages, with increased prominence identifying focal elements, often with accompanying de-accentuation of non-focal elements. A functionally oriented hypothesis would therefore be that differences in parameter weighting can result in mis-recognition of information structure. A minor mis-recognition within the question-answer pairings of experiment 2 could, for example, be the acceptance of a contrastive focus as a perfect match to a question eliciting a non-contrastive narrow focus. The communicative implications of such a case in a natural situation are arguably negligible, since the focal element remains the same and the contrastive function is contextually as well as prosodically determined. Confusion between narrow and broad focus, which is prosodically less strongly differentiated, could be communicationally more serious (Wells and Local 1983, Cruttenden 2006).

3. Experiments and methods

Two experiments were carried out to test the hypotheses stated in section 2. In experiment 1, subjects judged greater or less prominence; experiment 2 was a communication-linked task dependent on the ability to judge the information structure of an utterance.

3.2. Experiment 1

Speech Material

The German sentence used in the experiments was "Der Mann fuhr den Wagen vor" (The man brought the car round). The two 'critical words'

(CW) under scrutiny were CW1 "Mann" (man) and CW2 "Wagen" (car). The five questions presented to the speakers in the original production experiments (Andreeva et al. 2007) are shown in table 4. These questions triggered a broad-focus response, a non-contrastive and a contrastive narrow-focus response for CW1 and a non-contrastive and a contrastive narrow-focus response for CW2.

Question	**Focus condition**
Was passierte? (*What happened?*)	broad focus
Wer fuhr den Wagen vor? (***Who*** *brought the car round?*)	non-contr. early focus on CW1
Was fuhr der Mann vor? (***What*** *did the man bring round?*	non-contr. late focus on CW2
Die **Dame** fuhr den Wagen vor? (*The* ***lady*** *brought the car round?*)	contr. early focus on CW1
Der Mann fuhr die **Klagen** vor? (*The man brought the* ***charges*** *round?*)	contr. late focus on CW2

Table 4. German questions used for eliciting utterances with broad and (non-contrastive and contrastive) narrow focus questions.

The difference values for the parameters under the stimulus-pairing conditions which are the acoustic basis for the prominence-difference judgments are given in tables 5 a-c, together with comparative values for corresponding Bulgarian pairings.

For experiment 1 each focus condition was paired with itself (identical pairing) and with each of the other four conditions in both orders of presentation, giving 45 pairs for each CW. A part of each version of the sentence was masked by low-pass filtering to help the subjects concentrate on the relevant CW. "Der Mann" was masked when "Wagen" was to be judged; "fuhr den Wagen vor" was masked when "Mann" was to be judged.

The five identical pairings were used to ascertain the basic ability of the subjects to judge relative prominence. Comparable behaviour across the two subject groups was a necessary condition for analyzing the other stimulus groups. Five repetitions of each stimulus pair in each order of presentation were offered to the groups, giving 450 stimulus pairs in total. Five practice pairs, which were not evaluated, were included for familiarization with the task at the beginning.

	Diff. in f_0 change (semi-tones)		Diff. in intensity (dB)		Diff. in duration (%)	
CW1		D	BG	D	BG	D
contr. late – contr. early	6.0	8.6	11.2	4.8	8.5	46.4
contr. late – non-contr. early	9.3	5.9	7.6	1.5		35.0
non-contr.late – contr. early	3.0	8.2	12.1	3.8	0.9	60.2
non-contr.late – non-contr.early	6.3	5.5	8.5	0.5	0.9	47.7
CW2						
contr. late – contr. early	9.7	4.9	21.1	6.7	0.0	10.1
contr. late – non-contr. early	9.7	4.9	12.1	6.7	3.8	24.2
non-contr. late – contr. early	11.6	4.9	11.8	6.0	5.8	9.5
non-contr. late – non-contr. early	11.6	4.9	14.4	4.0	1.9	23.5

Table 5a Parameter differences between pairings of maximally different stimuli

	Diff. in f_0 change (semi-tones)		Diff. in intensity (dB)		Diff. in duration (%)	
CW1	BG	D	BG		BG	D
broad – contr. early		2.3	9.9	3.2	2.6	25.8
broad – non-contr. early	6.1	0.4	6.3	0.1	2.6	16.0
broad – contr. late	3.2	6.3	1.3	1.6	11.3	16.4
broad – non-contr. late	0.2	5.9	2.2	0.6	1.7	27.3
CW2						
broad – contr. early	11.8	4.4	5.4	4.0	3.8	10.5
broad – non-contr. early	11.8	4.4	8.0	2.0	0.0	2.0
broad – contr. late	1.7	0.5	6.7	2.7	3.8	21.7
broad – non-contr. late	0.2	0.5	6.4	2.0	3.8	10.5

Table 5b Parameter differences between pairings of contrasts with broad

	Diff. in f_0 change (semi-tones)		Diff. in intensity (dB)		Diff. in duration (%)	
CW1	BG	D	BG	D	BG	D
non-contr. early – contr. early	3.3	2.7	3.6	3.3	0.0	8.4
non-contr. late – contr. late	3.0	0.4	0.9	1.0	9.4	9.3
CW2						
non-contr. early – contr. early	0.0	0.0	3.6	3.3	3.8	12.8
non-contr. late – contr. late	1.9	0.0	0.3	0.7	5.8	0.5

Table 5c Parameter differences between pairings of stimuli with minimal differences

Experimental task and subjects

In experiment 1, 30 Bulgarian and 30 German subjects were presented with German sentence pairs which differed only in the focus conditions under which the sentences were produced. The task was to judge to what *degree* a critical word in one version differed in prominence from the same word in the other version. The subjects registered their judgments of the prominence difference on a graphic interface (see figure 1) by moving a slide upwards or downwards on an uncalibrated scale (with concealed values of ±100). The neutral (zero difference) position was marked in the centre of the scale and moving the slide upwards signified stronger prominence in CW of the first sentence of the pair. Moving it downwards signified a more prominent CW in the second sentence.

Response data pre-processing

The uncalibrated scale used in experiment 1 allowed each subject to set his/her own level of response. Since there are always some subjects who make fuller use than others of the scale's potential range, simply averaging subjects repeat responses can lead to a distortion of differences between groups. Therefore z-values were calculated for each subject to normalize for individual scaling differences.

Figure 1: Graphic interface for Experiment 1.

Consistent responses resulted in negative response values for one presentation order which were similar in magnitude to the positive values for the opposite presentation order of the same stimulus pair. Since it was difference in response magnitude which was of interest, negative means were multiplied by -1 (after averaging over repetitions).

Since, in experiment 1, differences between the two listener groups might be expected to vary depending on the stimulus-pairs being judged (see for example the different degrees of acoustic contrast between the critical words of the pairs given in table 5 a-c), a descriptive variable was defined to split the data into four stimulus groups: i) pairs with identical stimuli, ii) pairs with near-identical stimuli with the same narrow focus location (early or late) but differing in level (non-contrastive vs. contrastive) iii) stimulus pairs comprising the broad focus realization with a narrow focus version (early or late, non-contrastive or contrastive), iv) maximally different stimulus pairs comprising early narrow-focus sentences with late narrow-focus ones. In general, the degree of perceived prominence difference should increase from i) to iv) for both listener groups but, depending on the degree of acoustic contrast and the parameter bearing that contrast, there could be differences between the D and BG subjects for some stimulus pairs.

3.2. Experiment 2

Speech Material

The recording of each of the questions used to elicit the different focus conditions of the test sentence "Der Mann fuhr den Wagen vor" was paired with a realization of each of the answers. This resulted in 5

question-answer (QA) pairs with the original matching answer and four pairs each of every other combination (e.g. Q broad + A non-contrastive early, contrastive early, non-contrastive late, contrastive late), a total of 25 stimulus pairs. Again, 5 repetitions were offered, giving 125 stimuli, presented in randomized order. The subjects carried out the task interactively, listening over headphones in a quiet room, and were allowed to repeat the stimulus by clicking on the 'repeat' button (see figure 2). The number of repeats and the reaction time (time from first presentation until one of the response buttons is pressed) were recorded with the response itself.

Experimental task and subjects

Two groups of 30 German and 30 Bulgarian subjects were tested. Some subjects had taken part in experiment 1, but this was not seen as an impediment because the tasks in the two experiments were completely different: a meta-linguistic task vs. a functional task. Both groups judged German question-answer pairs (task groups D and BG_D). The D and BG_D conditions are central to the hypotheses to be tested, but in experiment 2, the Bulgarian subjects also judged Bulgarian question-answer pairs (BG_BG task-group). This served as an indicator of whether prosodic marking is comparably important for signaling information structure in Bulgarian and German. The subjects were required to judge the goodness of the QA match on a 1 to 5 scale: 1 = perfect match; 2 = acceptable but not perfect; 3 = don't know; 4 = not a good match; 5 = unacceptable.

Response data pre-processing

The combinations of questions and answers are grouped into four categories:

i) i) the question is paired with the actual focus condition it elicited (expected response 1 = "perfect match");
   ii) the question is paired with the focus condition elcitied but with a different level of contrast: contrastive Q + non-contrastive A or non-contrastive Q + contrastive A.(expected response 2 = "near match").

ii) iii) the question eliciting a broad-focus answer is paired with a narrow-focus answer, or a question eliciting a narrow-focus

answer is paired with the broad-focus answer (expected response 3 = "uncertain") and

iii) iv) a question eliciting an early narrow-focus answer paired with a late narrow-focus answer, or a question eliciting a late narrow-focus answer paired with an early narrow-focus answer. Contrastive and non-contrastive pairings are pooled (expected response 4 or 5 = "bad match").

Figure 2: Graphic interface for Experiment 2.

3. Results

Experiment 1: Prominence-difference judgments

The z-values of the individual means for each stimulus pairing were used in a univariate ANOVA with the two subject groups (BG and D) and the stimulus groups as factors. Overall, the subject groups were not significantly different (F [1, 2992] = 1.251, p = 0,263), while the stimulus groups, as is to be expected, were highly significant (F [3, 2992] = 826.153, $p < 0.001$). There was also a significant interaction between the subject groups and the stimulus groups.

Post hoc, the stimulus groups were shown to increase in perceived prominence-difference from:

iv) the 'matching' stimulus pairs with an average z-value for prominence difference of 0.1070, to

v) the 'near-match' stimulus pairs with the same *narrow* focus location (CW1 or CW2) but differing in level (non-contrastive vs. contrastive) with an average z-value for prominence difference was 0.3642, to
vi) the stimulus pairs comprising the *broad* focus realization with a *narrow* focus version (CW1 or CW2, non-contrastive or contrastive), with an average z-value for prominence difference was 0.777, to
vii) the 'bad match' stimulus pairs comprising *early* narrow focus sentences with *late* narrow focus ones. The average z-value difference judgment was 1.1186.

These four groups were significantly different from one another, which is not in itself surprising. Together with the subject-stimulus-group interaction, however, it provides the basis for considering the BG and D subjects' judgments for the stimulus pairs that make up each of the stimulus groups. Separate univariate ANOVAs were carried out for the four stimulus groups with subject groups (BG and D) and stimulus pairs as independent factors and absolute mean z-values as dependent measure. The results were as follows:

BG and D show almost identical behaviour ($F [1, 598] = 0.002$, $p = 0.962$) for the 'matching' pairs and were not significantly different for the 'near-match' pairs ($F [1, 478] = 0.447$, $p = 0.504$) nor for the pairs containing broad-focus stimuli ($F [1, 958] = 1.871$, $p = 0.172$). They differed significantly, however for the maximally disparate 'bad match' stimulus pairs ($F [1, 958] = 8.667$, $p = 0.003$).

Post hoc tests for differences between the stimuli in this stimulus group showed that both BG and D listeners distinguished CW1 stimulus pairs containing non-contrastive early focus from the pairs with contrastive early focus (the two homogeneous subsets in table 6a-b). Subset 1 has significantly lower prominence difference score than subset 2 (z-value ranges 0.57-0.72 vs. 1.30-1.42 for the BG subjects, 0.58-0.69 vs. 1.15-1.30 for the D subjects (see table 6a-b). The significant difference between the BG and D subjects for these maximally different stimulus pairs is clearly due to the higher prominence-differences perceived by the BG subjects for subset 2, the stimulus pairs not containing non-contrastive early focus stimuli. The values are 12.5% higher on average compared to a difference of 2.15% for subset 1.

Against the background of the Bulgarian and German production patterns (table 1) and the parameter differences between pairings of maximally different stimuli shown in table 5a we can attempt an explanation of the observed divergence in prominence perception between Bulgarian and German subjects. Given the comparable

exploitation of f0 in the production of accentuation, the 5.5/5.9 semi-tone difference between the CW1 'non-contrastive early' condition and the CW1 '(non-)contrastive late' conditions should be heard as equally prominent by both the D and the BG group. The 0.5/1.5 dB difference should contribute little to Bulgarian prominence impression in view of the much greater average dB differences found in the Bulgarian production data. Similarly, the large difference in duration (47.7/35.0%) should not signal prominence for the Bulgarian to the extent it does for the German listeners, given that Bulgarian changes in rime duration with prominence are on average rather low and very irregular, (sometimes even negative), suggesting relative irrelevance for prominence perception. In summary, then, the intensity and durational properties should militate against the Bulgarian subjects hearing the differences between these pairs as strongly as the German subjects. However, this is not the case since, as already mentioned, the Bulgarian average z-value for these stimuli is (non-significantly) 2.15% higher than the German.

For the remaining CW1 judgments ('contrastive early' vs. '(non-) contrastive late'), greater tonal differences (8.2/8.6 ST) should signal greater prominence differences for both groups; 60.2/46.4% durational differences should be important for the German subjects, and 3.8/4.8 dB differences in intensity are within the range found for Bulgarian and should therefore contribute to the perceived prominence difference for the Bulgarian subject group. So here the 12.5% higher average Bulgarian prominence-difference is also rather unexpected. Parity would be easier to explain, with the large duration differences triggering the large perceived prominence-difference for the German subjects while the intensity differences are the trigger for the Bulgarian subjects.

An explanation which presents itself, but which requires independent experimental validation with manipulated stimuli systematically decoupling intensity and duration, is that the Bulgarian subjects are in fact responding to the greatly increased duration of the German narrow-focus stimuli not in terms of duration but as a psycho-acoustic reaction to the integration of duration and signal energy which has been shown to operate for durations up to 250-300 ms (Moore 2003). This means that both German and Bulgarian subjects react to the durational difference, though for different reasons, whereas only the Bulgarian subjects react to the increase in intensity, with the effect that they perceive the same signal differences as greater prominence differences.

A smaller (6.6%) but significant difference between the Bulgarian and German subjects is found for the CW2 maximum-difference stimulus pairs, which all belong to a single homogeneous group. Carrying these CW2 judgments there are smaller tonal differences than for the CW1 pairs

(4.9 ST vs. 8.2/8.6 ST for CW1) – important for both groups but hardly responsible by themselves for the large perceived-prominence differences.

a) BG			
		Subset	
Stimulus pairs	N	1	2
non-contr. late CW1 – non-contr. early CW1	30	.5735	
non-contr. early CW1– non-contr. late CW1	30	.6426	
contr. late CW1 – non-contr. early CW1	30	.6995	
non-contr. early CW1 – contr. late CW1	30	.7228	
contr. early CW1 – non-contr. late CW1	30		1.3046
contr. early CW1 – contr. late CW1	30		1.3453
contr. late CW1 – contr. early CW1	30		1.4016
non-contr. late CW1 – contr. early CW1	30		1.4172

b) D			
		Subset	
Stimulus pairs		1	2
non-contr. late CW1 – non-contr. early CW1	30	.5769	
contr. late CW1 – non-contr. early CW1	30	.6295	
non-contr. early CW1 – contr. late CW1	30	.6842	
non-contr. early CW1 – non-contr. late CW1	30	.6922	
contr. late CW1 – contr. early CW1	30		1.1521
non-contr. late CW1 – contr. early CW1	30		1.1867
contr. early CW1 – contr. late CW1	30		1.2181
contr. early CW1 – non-contr. late CW1	30		1.3019

Table 6 a and b. Average z-values and homogeneous subsets for maximally disparate CW1 stimulus pairs for Bulgarian (BG) and German (D) subjects

The intensity differences (4.0-6.7 dB) are strong, but the durational changes (9.5-24.2%) are only moderate. Again, the rationale of greater sensitivity to intensity differences, perhaps supported by the energy-duration integration can explain the observed higher average Bulgarian prominence-difference.

In summary, in the meta-linguistic task of judging differences in degree of prominence, the Bulgarian and German subject groups behave consistently similarly, except where there are sufficiently large intensity differences coupled with duration differences. The Bulgarian subjects then perceive significantly greater prominence differences that the German subjects, which, we have argued, reflects their greater sensitivity to signal energy. The second experiment was designed to test whether such different sensitivities carry over into communicative functions.

Experiment 2: Question-answer matching

A multivariate ANOVA with the three task groups (Bulgarian subjects who judged German QA pairs: BG_D, Bulgarian subjects who judged Bulgarian QA pairs; BG_BG, and German subjects who judged German QA pairs: D) and three question-answer combination categories (QA categories i + ii, iii and iv) as factors was carried out with the subjects mean acceptability judgments for each question-answer combination as dependent measures. Both factors differed significantly (subject group: F [2, 11241] = 58.9, $p < 0.001$; QA category: F [2, 11241] = 6427.8, $p < 0.001$). Post hoc test showed that all three task groups differed significantly in their acceptability judgments. With an overall average of 2.84 (vs. 2.72 for BG_BG and 2.58 for BG_D), the German subjects were closest to the central response category of 3 and were also most clearly symmetrical in their acceptance of matching QA combinations (categories i and ii) and their non-acceptance of disparate QA combinations (category iv): Bulgarian overall averages are considerably lower than the category iii) averages in table 7).

Separate tests for each task group confirmed that all three clearly separated the QA categories. Table 7 gives the averages for each group and category, showing the greater spread and greater symmetry of the mean D responses across the 1-5 range.

		QA cat. i & ii	QA cat. iii	QA cat. iv
BG	Mean accept.	1.60	2.65	4.06
BG_D	Mean accept.	1.46	2.42	4.01
D	Mean accept.	1.21	2.83	4.68

Table 7 Group average acceptability scores for the QA categories.

The fact that 40% of the Bulgarian subjects (BG_BG and BG_D together) failed to find the disparate category iv) combinations unacceptable, i.e. gave values < 4.0, is the most striking difference between the German and Bulgarian subjects. This is not plausibly attributable in any way to different perceptions of *degrees* of prominence, since a) experiment 1 showed strong parallels between the German and L2-Bulgarian subjects' judgment of prominence differences and b) the answer stimuli contained de-accented and strongly accented CWs produced naturally in contrast within the same sentence, a categorical difference in both languages. These would not disappear, even if the differences in degree of perceived-prominence hypothesized for experiment 1 became manifest within the communicative-functional task of experiment 2. In the L2 task, i.e. for the BG_D group, it might be convincingly mooted that a lack of confidence in their ability to judge the match between question and answer would affect the subjects asymmetrically: They might well be less able to recognize, or less prepared to judge a combination as a *bad* match than to recognize a *good* match. However, this explanation cannot be applied to the BG_BG group, who also failed in a high proportion of cases to mark the contrasting focus combinations as unacceptable.

These results suggest a different processing mode by the Bulgarian listeners which is independent of the L1 or L2 task. It could well be the case that Bulgarian makes relatively less use of prosody than alternatives means of marking narrow focus than German, so that incorrect prominence patterns in an answer to particular focus-oriented questions are, as far as the prominence pattern is concerned, less negatively marked. Linguistic resources relevant for the information structure of Bulgarian utterances (Avgustinova 1997) involve:

word order, remarkably flexible and discourse conditioned, as in all Slavic languages;

morphological category of *definiteness*, unusual in the Slavic language family;

cliticisation, as in any language having a system of weak and strong forms of personal pronouns: entities that are fully recoverable from the context can be structured as insignificant for the current communication purpose by using a clitic, an informationally inert element occurring in the utterance for reasons of (morpho-)syntactic wellformedness;

clitic replication of nominal material, specific to Bulgarian;

intonation, fairly malleable, as in languages like English and unlike, e.g., Czech.

Avgustinova models the information structure in Bulgarian utterances as an interplay of three factors: the lexeme-specific obliqueness hierarchy of grammatical relations, the actually observable constituent order, and the contingent clitic replication. While this study does also list prosodic factors under the general term "intonation", it is clear that there must be considerably more expectation of syntactic signalling for the Bulgarian than for the German listener.

Subject -group		perfect match (1)	Relati-vely good (2)	I don't know (3)	Relati-vely bad (4)	unaccept -able (5)
BG_B	Cat. i)	922 (68.3%)	246 (18.2%)	25 (1.9%)	115 (8.5%)	42 (3.1%)
	Cat. ii)	306 (25.5%)	387 (32.2%)	84 (7.0%)	272 (22.7%)	151 (12.6%)
	Cat. iii)	79 (6.6%)	88 (7.3%)	49 (4.1%)	447 (37.3%)	537 (44.7%)
BG_D	Cat. i)	1020 (75.6%)	198 (14.7%)	29 (2.1%)	49 (3.6%)	54 (4.0%)
	Cat. ii)	354 (29.5%)	425 (35.4%)	74 (6.2%)	252 (21.0%)	45 (3.8%)
	Cat. iii)	97 (8.1%)	111 (9.3%)	36 (3.0%)	392 (32.7%)	564 (47.0%)
D	Cat. i)	1126 (83.4%)	194 (14.4%)	10 (0.7%)	16 (1.2%)	4 (0.3%)
	Cat. ii)	138 (11.5%)	483 (40.3%)	111 (9.3%)	378 (31.5%)	90 (7.5%)
	Cat. iii)	2 (0.2%)	9 (0.8%)	18 (1.5%)	315 (26.2%)	856 (71.3%)

Table 8 Absolute and percentage distribution of judgments for the three QA categories for the BG, BG_D and D subjects

In the search for a possible prominence-based reason for the difference in acceptability responses, the subject-group responses to the QA categories and the different QA pairs within those categories were examined. Table 8 gives the absolute and percentage distribution of judgments for the QA categories (categories i) 'perfect match' and ii) 'near match' pooled). The following patterns are immediately apparent:

1. None of the listener groups appear keen to use response 3 ('don't know'), and all of them resort to it more frequently for the category ii pairs (stimuli with a broad-focus eliciting question or a broad-focus answer). This may be seen as an implicit statement that the subjects feel they are performing a meaningful task.

Combining the 1 + 2 and the 4 + 5 responses as acceptance and non-acceptance judgments, there is a marked linear increase in the acceptance of category i) pairs (matched pairs): BG, 86.5% < BG_D, 90.2% < D, 97.8% and a different pattern in the non-acceptance of category iii) pairs (conflicting focus pairs): BG, 82.0% > BG_D, 79.7% < D, 97.5%. This suggests that the BG_D subjects are more confident in the recognition of (L2) matching QA focus conditions than their L1-task compatriots.

2. The BG_D subjects are much more ready to accept category ii) pairs (64.9%) and less prepared to reject them (24.8%) than the D subjects (41.8% vs. 39%) or even the BG subjects (57.7% vs. 35.3%). This category is therefore the most promising candidate to examine for the possible influence of differing perceptual weighting in the responses. Table 9 lists the BG_D and D responses for the category iii) pairs.

QA combination	Subj-group	response 1+2	response 3	response 4+5	Diver-gence
Q-broad + A-contr_early	BG_D	91	8	51	– 15
	D	100	5	45	
Q-broad + A-contr_late	BG_D	105	7	38	+ 5
	D	98	16	36	
Q-broad + A-noncontr_early	BG_D	114	8	28	+15
	D	104	13	33	
Q-broad + A-noncontr_late	BG_D	86	16	48	– 19
	D	97	13	40	
Q-contr_early + A-broad	BG_D	112	7	31	**+ 70**
	D	70	21	59	
Q-contr_late + A-broad	BG_D	100	10	40	**(+29)**
	D	85	9	56	
Q-noncontr_early + A-broad	BG_D	84	9	57	**+ 107**
	D	26	18	106	
Q-noncontr_late + A-broad	BG_D	87	9	54	**+ 82**
	D	41	16	93	

Table 9 BG_D and D responses for the category iii) pairs (responses 1+2 and 4+5 pooled) The positive divergence sums (see text) are marked in bold.

The 'divergence sum' in the final column of table 9 indicates the degree to which the BG_D subjects' acceptance of a category iii) QA pair exceeds that of the D subjects. The number of BG_D 'good matches' (1 & 2 responses) that exceeds the number of D 'good matches' plus the number of D 'non-acceptances' (4 & 5 responses) that exceeds the BG_D 'non-acceptances' is registered as a reflection of the Bulgarian listeners greater tolerance. When the question is one that is aimed at eliciting a broad-focus answer, the divergence value is rather low and even has a negative number in two of the four cases, indicating that, overall, the German subjects were no more critical of the information-structural clash between Q and A than the Bulgarian subjects. The explanation that suggests itself for this pattern is the non-directive nature of the broad-focus question, which allows the respondent to associate an acceptable context with the particular narrow-focus answer presented in the QA pair.

This freedom to imagine a fitting context does not apply to the same extent when the question calls for a narrow-focus answer and the German subjects' responses confirm the unacceptability of the combinations, especially the broad-focus answers following the 'non-contrastive early' and 'non-contrastive late' questions. The rather high level of acceptability indicated in the BG_D response pattern for these combinations calls for an explanation and, given the subjects' performance for other QA combinations and the parallels discussed above between BG and BG_D responses, a lack of L2 competence is not a convincing claim.

With regard to the prosodic structure of expected answers to questions eliciting narrow-focus answers in comparison to the broad-focus answer given we can consult the parameter table in the appendix. (We summarize the comparison in table 10). As table 10 shows, the values for the CWs in the broad-focus answer lie mainly, as expected, between the de-accented and the accented versions. The ranges of tonal movement on the broad-focus tone accents are very close to the values for the (non-)contrastive accents and, in the case of the non-contrastive early accent, the broad-focus value in fact exceeds it. However, the tone-accent categories are not the same, as the tonal-movement descriptions show. What is more, the same broad-focus realization, which has both CWs accented, is accepted where versions with one accented and one de-accented CW (i.e with none or only negligible tonal movement) would be normal. The intensity values are almost identical for CW1 and CW2 in the broad-focus version and although they have sufficient energy to support the perception of an accented syllable, their prominence is unconvincing for de-accented tokens, given that the mean difference between de-accented and accented syllables in the Bulgarian production data 4.3/8.6dB. (see table 1). The duration of CW1 in the broad-focus

realization lies between the values for the de-accented and (non-) contrastive accented realizations, while the duration of CW2 is slightly less than that of the de-accented versions. In summary, none of the properties support the perception of a de-accented element except CW2 duration, but its effect there would be counteracted by the tonal accent and, to some extent, by the intensity of the /va:/ syllable.

Der	**Mann** (CW1)			fuhr den	**Wa-**(gen) (CW2)			vor
Focus	Inton	dB	S-Dur.		Inton	dB	S-Dur.	
contr. early	Rise fall 8.6ST	83.4	316ms		de-acc 0 ST	76.1	204ms	
n-contr. early	Fall 5.9ST	80.1	268ms		de-acc 0.5ST	78.1	207ms	
broad	**Rise 6.3ST**	**80.2**	**242ms**		**Early fall 4.4ST**	**80.1**	**197ms**	
n-contr. late	De-acc. 0.4ST	79.6	187ms		rise fall 4.9ST	82.1	244ms	
contr. late	De-acc. 0 ST	78.6	214ms		rise fall 4.9ST	82.8	247ms	

Table 10 Signal properties of the broad-focus answer presented with narrow-focus questions and of the narrow focus answers for which it was frequently accepted.

The acceptance of the broad-focus realization as an answer to the narrow-focus questions would, it seems, only be plausible as a non-random response if the listener's attention is strongly directed to the accented word defined as 'in focus' by the question, and the accentuation of the other CW is not registered. The acceptance of broad focus occurs to some extent with the German subjects and even more strongly with the Bulgarian subjects for the two contrastive questions (contrastive early and contrastive late) than for the non-contrastive questions, supporting the conclusion derived from the above discussion, that the effect is not prosodically induced.

4. Discussion and conclusions

The finding of this study can be summarized as showing the difference between auditory phonetic sensitivity and recognition of communicative function. The consistent differences in the exploitation of duration and intensity found previously in the *production* of different degrees of phrasal prominence do result in Bulgarian learners of German *perceiving* the degree of prominence in German sentences somewhat differently from German subjects. Scrutiny of the acoustic differences behind those stimulus pairs which were judged significantly differently by the Bulgarian and German subjects, compared to those that were judged similarly, suggests that it is a greater sensitivity to intensity which led the Bulgarian subjects to hear greater prominence differences.

Interestingly, greater durational differences, to which German subjects should be more sensitive, given their greater exploitation of duration for prominence in production, did not result in their hearing greater prominence differences than the Bulgarian subjects. It is suggested that the psycho-acoustic integration of signal intensity over time, within the syllabic time spans involved, enabled the Bulgarian subjects to perceive the prominence difference to a similar degree to the German subjects because they based their judgments on the summed energy rather than the duration of the event. This requires further investigation with stimuli in which intensity and duration are decoupled, since durational sensitivity in perception linked to differences in production has been found for other languages (cf. Cumming 2010 on Swiss German and (Swiss) French). However, the functional linguistic task involved in Cumming' study that case was phrasal grouping, which may well be triggered by duration alone, without any participation of the summed energy.

In the second experiment, a communication-oriented task which tested the subjects' acceptance of the Question-Answer pairing with matching and variously clashing focus conditions between the eliciting question and the answer following it, none of the significant differences between the German and the Bulgarian subjects were attributable to differences in degree of perceived prominence. There was a general difference between the German subjects and both (L1 and L2) Bulgarian subject groups in the readiness to fully accept the matching or fully reject the non-matching QA pairs. German subjects acceptance of matching question-answer sequences and non-acceptance of clashing were closer to response values 1 and 5, respectively, with lower standard deviations. But the distinct classification of the three QA categories was maintained in all three subject groups.

This general lack of differences between the subject groups with regard to the focus (mis-) match, may possibly be explained by the more or less categorical nature of the decision, at least for the QA pairs in which their was no broad-focus element (question or answer) involved. Whatever the production weighting of an early (CW1) or late (CW2) prominence, it is unlikely to be confused with the de-accented version of the same word. Thus, while the perceived prominence might not reach the degree that the listener expects, it would not diverge enough to trigger the opposing category and thus prompt a different 'acceptability' judgment.

In the case of broad-focus answers following a narrow-focus question or questions to elicit broad-focus answers paired with narrow-focus answers, the intermediate 'appropriateness' values could result either from a consistent ranking of the QA match as neither good nor bad (i.e. a predominance of '3' judgments) or from an uncertainty resulting in a wide range of judgments. An examination of response distribution for the individual QA pairs showed a general tendency for responses to group around the 1 and 2 or the 4 and 5 response categories and avoid a 3 response. In most cases the proportion of subjects 'accepting' and 'rejecting' was similar across the three task groups. The exceptions were the non-contrastive (early- and late-focus) questions followed by the broad-focus answer, where the BG_D group was significantly more accepting than the D (or the BG) group. However, there was no evidence that this divergence from the otherwise consistent group behaviour was attributable to differential sensitivity to prominence-lending acoustic properties.

The conclusion to be drawn from these experiments is that production differences between languages may have perceptual-processing consequences in a non-functional discrimination task, but that such differences do not necessarily carry over to a more communication-linked functional task. The results also illustrate the well-known fact that the closer one gets in experimentation to the level of normal communication-linked processing, the more likely it is that subjects make systematic use of factors that have nothing to do with the question being addressed experimentally. Although only a conjectural explanation, it appeared that the effect of narrow-focused question formulation induced an attentional state in a considerable number of listeners (German as well as L1 and L2 Bulgarian subjects) which enabled them to hear and accept the prominence of the 'focused' word and ignore the more or less equal prominence of the non-focused word. Interestingly this was manifested more strongly in the BG_D subject group (L2 task) since it was apparent for the weaker non-contrastive focus conditions and the divergence from the D subjects' distribution pattern was very strong. But in the stronger,

contrastive condition, although the acceptance level for the BG_D group was higher, the divergence from the German response pattern diminished because the D subjects apparently also succumbed to the effect. It is once more a case, as so often in speech communication, a case of 'trying to make sense of the input'. This basic need cannot be kept out of experimentation[2] but, as this study shows, it is important to combine and compare substance- orientated and communicationally relevant tasks.

Acknowledgments

The work reported here was carried out with the support of the German Research Council (Deutsche Forschungsgemeinschaft – DFG), grant BA 737/10-1.

Literature

Andreeva, B., W.J. Barry and I. Steiner, 2007. Producing Phrasal Prominence in German. *Proceedings International Congress of Phonetic Sciences*, Saarbruecken, 1209-1212.

Andreeva, B., S. Dimitrova and W.J. Barry (2010). Prosodic transfer in L2 speech: Evidence from phrasal prominence and rhythm in English, Bulgarian and German,Discourses of Globalization. *Selected Papers of the 13th International Conference of the Bulgarian Society for British Studies*, November 7-9, 2008, Sofia. St. Kliment Ohridski University Press, 2010, 9-19.

Arvaniti, A., D.R. Ladd and I. Mennen, 1998. Stability of tonal alignment: the case of Greek prenuclear accents. *Journal of Phonetics* 26, 3-25.

Atterer, M. and D.R. Ladd, 2004. On the phonetics and phonology of "segmental anchoring" of F0: evidence from German. *Journal of Phonetics* 32, 177-197.

Auer, P. (2006). On the prosody and syntax of turn-continuations. Couper-Kuhlen, E. and Selting, M. (eds.), *Prosody in Conversation. Interactional Studies*. Cambridge: Cambridge University Press, 57-100.

T. Avgustinova, 1997. Word Order and Clitics in Bulgarian. PhD dissertation. Universität des Saarlandes, (Published in *Saarbrücken Dissertations in Computational Linguistics and Language Technology*, Vol.5, 1998).

Barry, W., Nielsen, C., Andersen, O. (2001). Must diphone synthesis be so unnatural? *Proc. Eurospeech 2001 Scandinavia*, Aalborg

Bruce, G., 1977. *Swedish word accents in sentence perspective*. Lund: Gleerup.

Bühler, K. (1934). *Sprachtheorie*, Jena: Fischer Verlag. 2nd edition 1965. English translation *Theory of Language*, Amsterdam/Philadelphia: Benjamins.

2 An amusing anecdotal experience by the second author underlines this need: A listener-group for a test of synthesizer quality comprised students from a department of theological studies. Many of their 'misperceptions' resulted in words with religious connotations.

Couper-Kuhlen, E. and Selting, M. (2006). Towards an interactional perspective on prosody and a prosodic perspective on interaction. Couper-Kuhlen, E. and Selting, M. (eds.), *Prosody in Conversation. Interactional Studies*, Cambridge, Cambridge University Press, 11-56.
Cruttenden, A. (2006). The deaccenting of old information: A cognitive universal? G. Bernini and M.L. Schwarz (eds.) *Pragmatic Organisation of Discourse in the Languages of Europe*. Berlin: de Gruyter. 1-33.
Cumming, Ruth E. (2010). *Speech Rhythm. The language specific integration of pitch and duration*. PhD. Dissertation, University of Cambridge, UK.
De Saussure, F. (1916). *Cours de Linguistique Générale*.
Dilley L., D.R. Ladd and A. Schepman, 2005. Alignment of L and H in bitonal pitch accents: testing two hypotheses. *Journal of Phonetics* 33, 115-119.
Goldinger, S. D., Pisoni, D.B. and Logan, J.S. (1991). On the nature of talker variability effects on serial recall of spoken word lists. *Journal of Experimental Psychology: Learning, Memory and Cognition* 17:152-162.
Goldinger, S. D. (1997). Words and voices: Perception and production in an episodic lexicon. K. J.W. Johnson and Mullennix (eds) *Talker Variability in Speech Processing*, San Diego: Academic Press, 33-66.
Goldinger, S.D. (2000). The role of perceptual episodes in lexical processing. Anne Cutler, James M. McQueen and Rian Zondervan (eds.), *Proceedings of the Workshop on Spoken Word Access Processes (SWAP),* Nijmegen, Max-Planck Institute for Psycholinguistics, 155-158.
Hawkins, S. and Slater, A. (1994). Spread of CV and V-to-V coarticulation in British English: Implications for the intelligibility of synthetic speech. *Proceedings of the 1994 International Conference on Spoken Language Processing,* Yokohama: 57-60.
Hawkins S, Smith R (2001). Polysp: a polysystemic, phonetically-rich approach to speech understanding. *Italian Journal of Linguistics – Rivista di Linguistica* 13: 99-188.
Heid, S. and Hawkins, S. (2000). An acoustical study of long domain /r/ and /l/ coarticulation. *Proceedings 5th seminar on speech production*. 77-80.
Heinrich, Antje / Hawkins, Sarah (2009). Effect of r-resonance information on intelligibility. *INTERSPEECH-2009*, 804-807.
Huggins, A. William F. (1972a). Just noticeable differences for segment duration in natural speech". *Journal of the Acoustical Society of America* 51:1270-1278.
Johnson, Keith (1997). Speech perception without speaker normalization: An exemplar model. K. J.W. Johnson and Mullennix (eds) *Talker Variability in Speech Processing*, San Diego: Academic Press, 145-165.
Kohler, K.J.(1987). Categorical pitch perception. *Proc.11th ICPhS, Tallinn, vol. 5*, 331-333.
Kohler, K.J. (2005). Timing and communicative functions of pitch contours. *Phonetica* 62, 88-105.
Kohler, K.J. (2007a). Beyond Laboratory Phonology: The Phonetics of Speech Communication. Solé, Maria Josep, Beddor, Patrice Speeter, and Ohala, Manjari (eds): *Experimental Approaches to Phonology*. Oxford: Oxford University Press, 41-53.
Kohler, K.J. (2007b). Two Anniveraries. 75 Years of International Congresses of Phonetic Sciences and 50 Years of Phonetica. Some epistemological reflexions on the theory and methodology of Phonetic Science. *Phonetica* 64, 73-79.
Kohler, K.J. (2007c). Heterogeneous exponents of homogeneous prososodic categories – The case of dialectal variability of pitch contour synchronization with

articulation. *Paper in AG7 "Standard Prosody or Prosody of Linguistic Standards? Prosodic Variation and Grammar Writing". 29. Jahrestagung der Deutschen Gesellschaft für Sprachwissenschaft*. Siegen, 28. Februar 2007.
Koreman, J., B. Andreeva and W.J. Barry (2008). Accentuation cues in French and German, P. A. Barbosa, S. Madureira, and C. Reis (eds.), *Proceedings of the 4th International Conferences on Speech Prosody*, Campinas: Editora RG/CNPq, 613-616.
Koreman, J., B. Andreeva, W.J. Barry, W. Van Dommelen, R.-O. Sikveland, (2009). Cross-language differences in the production of phrasal prominence in Norwegian and German, Martti Vainio, Reijo Aulanko, and Olli Aaltonen (eds.), *Nordic Prosody, Proceedings of the Xth Conference*, Helsinki 2008, Frankfurt: Peter Lang, 139-150.
Lisker, L. and Abramson, A. S. (1964). A cross-language study of voicing in initial stops -Acoustical measurements. *Word – Journal of the International Linguistic Association,* 20(3), 384-422.
Lisker, L. and Abramson, A. S. (1967). Some effects of context on voice onset time in English stops. *Language and Speech,* 10(1), 1-28.
Local, J. (2005). On the interactional and phonetic design of collaborative completions. W. Hardcastle and J. Beck (Eds.), *A Figure of Speech: a Festschrift for John Laver*, 263-282. New Jersey: Lawrence Erlbaum.
Local, J. (2007). Phonetics detail and the organisation of talk-in-interaction. *Pro-ceedings of 16th International Congress of Phonetic Sciences*, Saarbrücken 2007, 1-10.
Moore, B. C. J. (2003). An Introduction to the Psychology of Hearing. Academic Press.
Niebuhr, O. (2006). The role of the accented-vowel onset in the perception of German early and medial peaks. *Proceedings of the 3rd International Conference of Speech Prosody*, Dresden, Germany, 109-112.
Niebuhr, O. (2007). Categorical perception in intonation: a matter of signal dynamics? *Proceedings of the Interspeech 2007*, Antwerp, Belgium, 109-112.
Niebuhr, O. (2007). The signalling of German rising-falling intonation categories – The interplay of synchronization, shape, and height. *Phonetica* 64, 174-193.
Niebuhr, O. (2008). Coding of intonational meanings beyond F0: Evidence from utterance-final /t/ aspiration in German. *Journal of the Acoustic Society of* America, Volume 124, Issue 2, 1252-1263.
Niebuhr, O. (2009). Intonation segments and segmental intonations. Proceedings of the 10th Interspeech conference, Brighton, UK.
Niebuhr, O.and Pfitzinger, H.R. (2010). On pitch-accent identification. The role of syllable duration and intensity. *Proc. of the 5th Int. Conf. on Speech Prosody*, Chicago.
Nolan, F. (1992). The descriptive role of segments: evidence from assimilation. G. Docherty and D.R. Ladd (eds.), *Laboratory Phonology* 2*,* 261-280. Cambridge: CUP.
Trubetzkoy, N. S. (1937). *Grundzüge der Phonologie*. Travaux du Cercle Linguistique der Prague 7. (1971) 5th edition Göttingen.
Wells, W. and Local, J. (1983). Deaccenting and the structure of English intonation. *Linguistics* 21, 701-715.

Appendix: Parameter values of the Stimuli in Exp 1 (non-manipulated stimuli)

	f_0 -range		V-Intensity (dB)		Rime-Duration (ms)	
CW1	*BG da(ma)*	*D Mann*	*BG da(ma)*	*D Mann*	*BG da(ma)*	*D Mann*
Broad	118-160 (5.3ST) (L*+H))	104-150 (6.3ST) (L+H*)	75.7	80.2	118	163
Ncontr early	250-1209 (11.4ST) (H*+L)	145-103 (5.9ST) (H*+L)	82	80.1	115	189
Ncontr late	102-137 (5.1 ST) (L+H)	105-103 (0.4ST) (deacc)	73.5	79.6	116	128
Contr early	258-161 (8.1ST) (H*+L)	103-170 (8.6ST) (L+H*)	85.6	83.4	115	205
Contr late	115-129 (2.1ST) (L+H)	107-107 (0ST) (deacc)	74.4	78.6	106	140
CW2	*BG ka(ka)*	*D Wa(gen)*	*BG ka(ka)*	*D Wa(gen)*	*BG ka(ka)*	*D Wa(gen)*
Broad	202-102 (11.8ST) (H+L*)	137-106 (4.4ST) (H+L*)	75	80.1	108	152
Ncontr early	120-creak (0 ST) (deacc)	100-98 (0.5ST) (deacc)	67	78.1	108	149
Ncontr late	256-131 (11.6 ST (H*+L)	111-147 (4.9ST) (L+H*)	81.4	82.1	110	184
Contr early	136-creak (0ST) (deacc)	98-98 (0ST) (deacc)	69.6	76.1	104	168
Contr late	250-143 (9.7ST) (H*+L)	113-150 (4.9ST) (L+H*)	81.7	82.8	104	185