Shaping phrase-final rising intonation in German

Ernst Dombrowski & Oliver Niebuhr

Institute of Phonetics and Digital Speech Processing (IPDS), University of Kiel, Germany
ed AT ipds.uni-kiel.de; niebuhr AT linguistik.uni-kiel.de

Abstract

In German, the course of phrase-final rising intonation can be modelled by an interpolation between three points in the pitch contour, giving a more appropriate description than current approaches to German intonation that only include the alignment of rises and their pitch range. Within this 3-point model, the parameter *range proportion* produces different contour shapes: concave and convex. The perceived meaning of concave and convex rises is studied using the semantic differential technique, starting with the functional distinction of *activating* vs. *restricting contours*. Communicative meaning differences between concave and convex contour shapes are supported by data of 31 listeners.

Index terms: nuclear rise, shape, intonational meaning.

1. Introduction

The simplest way to characterize phrase-final rising intonation in acoustic terms is by two interpolated points, the onset and the offset of the F0 rise. This very basic model can describe different alignments of rising contours, i.e. temporal shifts of the two points relative to the segmental string, and it can account for variation in the total F0 range, caused by either one or both edges of the rise. By adding a further point somewhere in between the F0 onset and offset, the model becomes also capable of describing variation in the overall shape, i.e. in the internal dynamics of the rising movement. Current approaches to intonation such as the Kiel Intonation Model, KIM [1], or the autosegmental-metrical (AM) framework [2,3] disregard this additional dimension, since the communicative meaning expressed by the rise is assumed to be only a matter of alignment and total range. However, recent evidence from different languages suggests that shape does play a role, either as a further cue to communicative meaning along with alignment and range, or even as a separate carrier of communicative meaning, cf. [4,5]. Moreover, the communicative relevance of intonational shapes seems to go beyond phrase-final rises and does also concern falls related to prenuclear pitch accents, cf. [6,7].

Phrase-final rises that start with a constantly aligned nuclear pitch accent were studied for German by [5]. They derive the measure 'range proportion' (*rprop*) from a 3-point model, in which the joint between rise onset and offset is located at the accented-syllable offset. This point is judged to be the end of the first section of the rising movement. *rprop* is the quotient between the frequency range from rise onset to joint and the total range (in semitones, st). It can take values between 0 and 1; low values indicate a *concave* and high values a *convex* rise (seen from above, cf. Fig.1). Applying the 3-point model to the Kiel Corpus of German Spontaneous Speech (cf. [8]), [5] find that convex and concave rises are highly reliable predictors of two discourse environments, which, in turn, are correlated with question and continuation. While concave rises occur primarily turn-finally, turn-internal rises more often have a convex shape. The total range does not contribute substantially to the prediction of the discourse environments, contrary to claims in the literature, cf. [5]. Moreover, both kinds of rises in [5] have their onset immediately before the accented-vowel. In consequence, the two rises cannot be projected onto different alignment categories. In KIM they all belong to the early-valley category [9] and in the AM framework they could be represented by L+H* with high edge tones [3] – because of the rising movement towards the accented vowel.

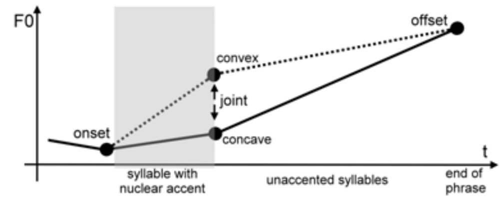


Figure 1: 3-point model of German phrase-final rises by [5].

As the discourse-based shape prediction does not refer to variation in overall range, and supported by additional examples from the literature and from the *Kiel Corpus* (cf. [10,11]), [5] argue that the shape difference plays a more general role in communication going beyond the distinction between turn-yielding vs. turn-holding or question vs. continuation. They suggest a basic communicative meaning opposition concerning the addressee-related structuring of the discourse: *activation* vs. *restriction*. That is, concave rises activate and convex rises restrict the dialogue partner. Depending on the semantic-pragmatic context, this can result in different concrete interpretations. For example, within statements like *Dann ging ich zum <u>Su</u>permarkt* ('Then, I went to the <u>su</u>permarket'; nuclear accent underlined) the shape difference can signal the dialogue partner 'confirm that you understood before I go on' (concave) or 'don't interrupt me, just listen' (convex). In the case of questions like *Sind Sie An<u>ge</u>la?* ('Are you <u>An</u>gela?') the meanings conveyed by concave and convex shapes may be paraphrased as 'give me an answer and then tell me a bit more about you' (concave) or 'I just want the name and nothing else' (convex).

The present study represents one of a larger set of perception experiments that primarily aimed at testing the differences in the meaning of concave and convex rises as postulated by [5]. It is a secondary aim of this study to test the claim of the 3-point model of [5] that the meaningful shape difference of the overall rise may be represented by a joint at the end of the accented syllable, independent of the number of subsequent unaccented syllables until the end of the phrase (Fig.1). The rising movements were resynthesized as a cross-combination of two 2-step conditions: (1) a *concave* vs. *convex* shape condition with *rprop* values of 0.2 vs. 0.8, based on a joint at the accented-syllable offset, and (2) a 2 vs. 3 or 4 syllable condition, including the pitch-accented syllable (*disyllabic* vs. *polysyllabic* nucleus). For judging the stimuli, the semantic-differential paradigm was used, which proved to be successful in detecting and outlining intonational meaning differences, cf. [6,7,12,13,14,15], and which is hence particularly suitable for the present study. Two *hypotheses* were tested:

(A) The *rprop* difference yields two clearly separate meaning profiles across the scales of the semantic differential, particularly on the activating-restricting scale; (B) The meaning profile should not be affected by the number of syllables covered by the phrase-final rise.

2. Method

The stimuli of the perception experiment are based on 6 syntactically marked question utterances (with verbs in clause-initial position) that were produced naturally by a trained male speaker (the author ED) with a rising F0 contour. The rise started around the onset of the only and hence nuclear pitch-accent syllable and continued until the end of the utterance. The accented syllable and almost all other segments covered by the rise were phonologically and phonetically voiced. This reduces microprosodic perturbations and ensures at the same time that the macroprosodic F0 course is well represented in the signal. The speaker also aimed at creating rises with moderate *rprop* values that allowed resynthesizing stimuli with more extreme *rprop* values without lowering the stimulus quality. Finally, in order to control for meaningful prosodic variation, the utterances were produced with comparable speaking rate, voice quality, loudness, F0 range, and rise-onset alignment in the end rhyme of the pre-accented syllable. The following utterances were used (accented syllables underlined):

(1a) *Bist du im Urlaub?* (‘Are you on holiday?’)
(1b) *Sind Sie der Eigentümer?* (‘Are you the owner?’)
(2a) *Liegt das bei Lübeck?* (‘Is that near to Lübeck?’)
(2b) *Liegt das in Niedersachsen?* (‘Is that in Lower Saxony?’)
(3a) *Sind Sie Angela?* (‘Are you Angela?’)
(3b) *Sind Sie Angelika?* (‘Are you Angelika?’)

The 3x2 utterances represent a cross-combination of two variables. First, (1), (2), and (3) differ in the semantics of the accented word in question. While condition (1) asks for a *situation/fact*, (2) and (3) concern *names* or *places*, respectively. Moreover, the utterance pairs (a) and (b), within each semantic condition, differ in the number of syllables following the accented one. While (a) represents the *disyllabic condition*, i.e. the F0 rise extended just over the accented and one subsequent syllable, (b) forms the *polysyllabic condition*. In the latter, the F0 rise was more than twice as long, since the accented syllable is followed by 2 or 3 unaccented ones. In summary, the 3x2 utterances allowed determining the extent to which the assumed meaning difference caused by the *rprop* variation is affected by different semantic and durational contexts.

The *rprop* variation itself was done by means of F0 manipulation with *PSOLA* in *praat* [cf. 16]. For each utterance, the final-rise section of the overall F0 contour was stylized at three contour points: rise onset, accented-syllable offset, and rise offset. Then, the frequency value of the middle contour point was raised or lowered in order to create a *0.2 and a 0.8 rprop condition* (cf. Fig.1 and 2). Resynthesizing these two conditions yielded 2x(3x2)=12 stimulus utterances.

The communicative meaning contrast expected for the two *rprop* conditions was measured in a semantic differential. For this, the communicative concept of *activation vs. restriction* [5,10] was represented in 12 semantic scales. These scales accounted for all three basic dimensions of semantic differentials, i.e. evaluation, potency, and activitiy[15]. The scales were:

(1) *aktivierend – hemmend* (‘activating – restricting’)
(2) *ruhig – erregt* (‘calm – upset’)
(3) *emotional – nicht emotional* (‘emotional - non-emotional’)
(4) *freundlich – unfreundlich* (‘friendly – unfriendly’)
(5) *gibt Spielraum – schränkt ein* (‘gives room - narrows options’ for designing the response)
(6) *interessiert – uninteressiert* (‘interested – not interested’)
(7) *einladend – abweisend* (‘inviting – rejecting’)
(8) *nicht herausfordernd – herausfordernd* (‘not challenging – challenging’)
(9) *nicht dominant – dominant* (‘not dominant – dominant’)
(10) *offen – geschlossen* (‘open – closed’)
(11) *fragend – nicht fragend* (‘questioning – non-questioning’)
(12) *spontane Äußerung – Routineäußerung* (‘spontaneous utterance – routine utterance’)

For ten of the twelve scales, the left alternative is associated with *activation*, the right one with *restriction*. For ‘emotional – non-emotional’ and ‘calm – upset’ the direction of predictions was not specified.

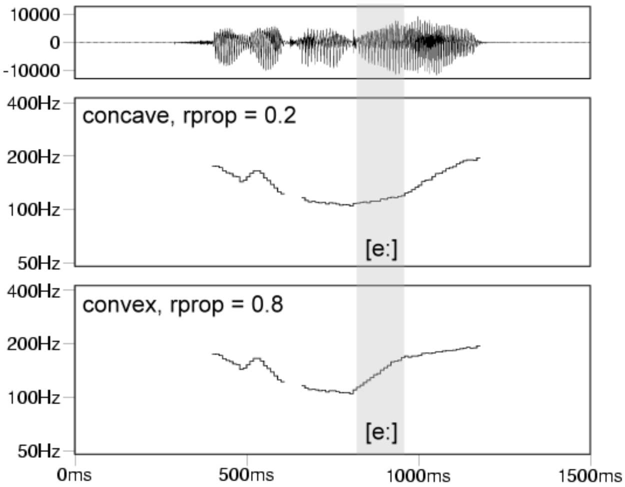


Figure 2: The stimulus utterance “Sind Sie Angela?” (‘Are you Angela’) with a concave (rprop 0.2) and a convex (rprop 0.8) phrase-final rise. Oscillogram, F0 contours. The accented vowel is indicated in grey.

The semantic scales had 7 steps (i.e. 3-2-1-0-1-2-3), which included the option of neither-nor judgements by responding with ‘0’. Listeners had to decide, if one of the terms of a scale matched better with a stimulus utterance, and if so, whether it matched slightly, well, or perfect.

Since each stimulus was combined with each semantic scale, the perception experiment comprised a total of 12x12=144 stimuli. Each stimulus presented was judged on a single dimension at a time. The twelve stimulus utterances were randomized within the responses to each scale. Additionally, scales were randomized. Scales were presented with different polarity changing the left and the right pole of the scales, e.g. from ‘friendly – unfriendly’ to ‘unfriendly – friendly’. Thus, effects of serial position were excluded.

31 naïve Northern Standard German listeners, 11 male and 20 female (average age 24,2 years), participated in the experiment. All participants were undergraduates at Kiel University.

The perception experiment was carried out in a sound-treated room of the Institute of Phonetics and Digital Speech Processing (IPDS) at the University of Kiel. The stimuli were played from a computer and presented to the listeners via loudspeakers. Overall, the experiment took about 30 minutes.

3. Results

Differences between the averaged semantic profiles in the experimental conditions were tested with a four-factor repeated measures MANOVA. The factors were (1) *range proportion* (*rprop* 0.2 vs. 0.8), (2) *syllable count* (2 syllables vs. 3 or 4 syllables in the nuclear stretch of the utterance), (3) *utterance pairs* (3 semantic conditions), and *listener's gender* (control variable). To get an interpretation of the multivariate results with respect to the twelve semantic scales, discriminant analyses and univariate tests were calculated.

(1) The factor *range proportion* is clearly significant ($F=78.277$, $df_1=12$, $df_2=18$, $p=.000$, $\eta^2=.981$): Phrase-final rises with a low and with a high range proportion differ in their meaning profiles – as predicted in *hypothesis A*. Differences are found on almost all semantic scales in the expected direction (cf. Fig. 3). Mean differences of at least about 1 scale point can count as having practical importance in research with the semantic differential. For those semantic scales that are central to the concept of *activation vs. restriction* (as 'activating – restricting' and 'spontaneous utterance – routine utterance', [cf. 5,10]) differences are larger than 1 scale point. However, some differences are considerably smaller. Thus, the average scale-point difference amounts to 0.9.

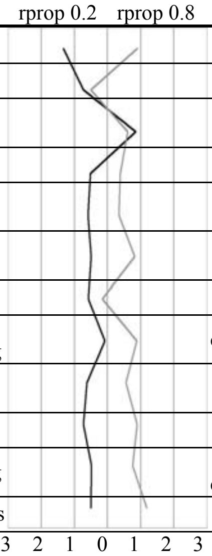
The factors *syllable count* and *utterance pairs* reveal different meaning profiles as well ($F=33.685$, $df_1=24$, $df_2=6$, $p=.000$, $\eta^2=.993$ and $F=15.154$, $df_1=12$, $df_2=18$, $p=.000$, $\eta^2=.910$). However, here the scale-mean differences are only 0.6 and 0.7 scale points on average. The gender of the listeners does not affect semantic judgements at all.

(2) Beyond its significant main effect, *range proportion* shows interactions with the *syllable count* of the nuclear contour and with the lexical semantics of the utterances, i.e., the factor *utterance pairs* ($F=7.906$, $df_1=12$, $df_2=18$, $p=.000$, $\eta^2=.841$ and $F=4.431$, $df_1=24$, $df_2=6$, $p=.036$, $\eta^2=.947$; for the triple interaction $F=8.184$, $df_1=24$, $df_2=6$, $p=.007$, $\eta^2=.970$). That is, the lexical content and inferred communicative meaning of the stimulus utterances as well as the syllable count in the nuclear contours have influence on which semantic scales are used to differentiate the concave and the convex shaping of rising contours. Thus, the assumption of *hypothesis B* that the effect of the contour manipulation is uninfluenced by the number of syllables covered by the rising movement does not hold. Moreover, the lexical and inferred communicative contexts lead to different interpretations of the convex-concave contrast in rising patterns. Here, also effects of unintended and hence uncontrolled prosodic variation between the stimulus utterances may be included.

For the modelling of rising intonation (cf. Fig.1), the effect of the syllable count in the nuclear accent group is of particular interest: The pitch contour spreads over the syllable chain, and the turning point in the three-point stylization has to be positioned at the structurally right place to produce the semantic differences expected between concave and convex patterns. The effect of the *rprop* manipulation on the perceived meaning of the contours is shown here for contours with a different syllable count (cf. Fig. 3). With regard to the disyllabic and polysyllabic nucleus conditions additional multivariate F-test were calculated to test the effect of the contour manipulation (*rprop* 0.2 vs. 0.8) separately. Both yield significant results ($F=32.909$, $df_1=12$, $df_2=19$, $p=.000$, $\eta^2=.954$ and $F=25.804$, $df_1=12$, $df_2=19$, $p=.000$, $\eta^2=0.942$).

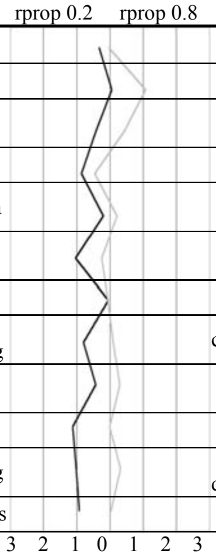
Related univariate tests on the level of the individual semantic scales become significant ($\alpha \leq 0.05$) in 9 of 12 cases in both conditions. However, differences between the scale values of the two contour shapes are larger in the disyllabic than in the polysyllabic condition (1.1 compared with 0.7 scale points on average). In addition, they apply to different scales:

(a) disyllabic nucleus



		weights	loadings	p
activating	restricting	.623	.359	.000
calm	upset	.357	.327	.267
emotional	non-emotional	-.281	.105	.148
friendly	unfriedly	.312	-.053	.000
gives room	narrows options	.142	.269	.000
interested	not interested	.301	-.257	.000
inviting	rejecting	.005	.317	.052
not challenging	challenging	.225	-.043	.000
not dominant	dominant	.295	.549	.000
open	closed	.278	.232	.000
questioning	not questioning	.203	.210	.000
spontaneous	routine	.405	.296	.000

(b) polysyllabic nucleus



		weights	loadings	P
activating	restricting	-.080	.368	.294
calm	upset	.339	.329	.000
emotional	non-emotional	.342	.133	.000
friendly	unfriedly	.225	.089	.001
gives room	narrows options	.107	.284	.059
interested	not interested	.163	.270	.000
inviting	rejecting	-.436	-.327	.947
not challenging	challenging	.336	.004	.000
not dominant	dominant	.225	.478	.001
open	closed	.436	.237	.000
questioning	not questioning	.601	.236	.000
spontaneous	routine	.257	.286	.001

Figure 3: Semantic profiles of concave (rprop 0.2, black) and convex (rprop 0.8,grey) nuclear rises, (a) results for contours comprising the nuclear syllable and one following syllable ("disyllabic"), (b) results for contours with 2 or 3 syllables following the nuclear syllable ("polysyllabic"). Standardized discriminant weights, loadings on the discriminant factor, and significances (df=1;29) are added.

In the disyllabic nuclear contours, the semantic distinction is mainly based on the scales 'activating' vs. 'restricting' and 'spontaneous' vs. 'routine utterance', i.e., expressions which are central for the concept of *activation vs. restriction*, and furthermore the scales 'questioning' vs. 'non-questioning' and 'interested' vs. 'not interested'. In the discriminant analysis concerning the utterances with a disyllabic nucleus the scale 'activating' vs. 'restricting' contributes most to the assignment of the semantic profiles to concave or convex contours (with a standardized discriminant weight of 0.623; cf. Fig. 3).

In the polysyllabic nuclear contours, classification is mainly based on the scales 'open' vs. 'closed', 'questioning' vs. 'non-questioning', and 'calm' vs.' upset'. For 'activating' vs.

'restricting' there is no significant result in this case. Instead, the scale 'open' vs. 'closed' reaches the largest standardized discriminant weight (0.601). Although the two profiles of the polysyllabic condition are significantly different the scale means for *rprop* of 0.8 are near 0 in many cases. This indicates that the profile difference principally depends on the means for *rprop* of 0.2 which differ from 0.

The proportion of correct predictions of the contour class (concave or convex) based on the semantic profiles amounts to 92.5% for the disyllabic nuclear contours and 82.8% for the 3- or 4-syllable contours. Separate discriminant analyses for all six utterances used in the study have prediction rates of 82.3% (*Eigentümer*) up to 100% (*Angela*, *Angelika*).

4. Discussion

In the present study, German phrase-final rises with constant alignment and F0-range but different contour shape are analysed in a semantic judgement task. Contour shapes are *concave* vs. *convex*, which is expressed by the parameter *range proportion*. The meaning profiles resulting for the two contour shapes prove to be different. Significant differences are found on most of the included semantic scales and across all six stimulus utterances, effect sizes are large, and the prediction rates in the discriminant analyses are substantial and sometimes next to perfect. Therefore, data support the claim of the previous corpus study [5,10] that there are two communicative types of rising movements, based on a *rprop* contrast. The characterization of the assumed meaning distinction with the terms *activation vs. restriction* is appropriate.

However, the general line of the findings is not expressed in all stimulus utterances in the same way. When the utterance-final rises comprise more than two syllables (i.e., if there is a long tail after the accented syllable) the semantic differences between the two *rprop* steps are on the whole smaller and the prediction rate of the two contour types based on the semantic profiles decreases. At the same time, the remaining meaning contrast is based on a different selection of scales. Above all, the scale 'activating' vs. 'restricting' only shows reaction to the contour manipulation in the *disyllabic* nuclear patterns.

The different results in convex and concave rises can be explained in at least two ways: First, in polysyllabic (i.e. more than disyllabic) rises the two shapes of the rising movement may be detected less easily by listeners. Second, the present version of the 3-point pitch model may meet the main characteristics of concave as opposed to convex contour shapes less well in the polysyllabic condition. Both explanations allow to conclude that the pitch manipulation may affect different signal systems in the disyllabic vs. polysyllabic conditions: There may be one system that particularly responds to the scales 'activating' vs. 'restricting', 'spontaneous' vs. 'routine utterance', and 'not dominant' vs. 'dominant' and that requires a clearly perceivable rising interval towards the accented syllable – so that the amount of the phrase-final, perhaps gliding pitch movement can be assessed as well. This system would refer to *rprop* variation in the narrow sense and may be salient in the disyllabic or already in a monosyllabic condition. A further, less subtly diversified signal system may refer to the general impression of a more or less extended pitch movement within and after the accented syllable, i.e. range information without a clear specification of the step towards the joint of the concave or convex rise. Range information of this type (and thus the second signal system) may be salient when no clear *rprop* information can be extracted by the listener, which may be the case in the present polysyllabic condition. So, it must be questioned whether the *rprop* manipulation was made at the correct position in the 3- and 4-syllabic nuclear accents. This would mean that the underlying contour model should be revised in order to adjust it to the rhythmic structure in the syllable chain and thus get the perceived meaning of concave and convex rises unaffected by their syllable count.

All in all, the *rprop* parameter can be used to model an important aspect of formal and functional variation in rising intonation. The 3-point description characterizes the quality of the pitch rise beyond its range and its alignment relative to the accented vowel. In this, the range proportion controls the relation between the pitch information of the first part of the nuclear rise (roughly up to the joint of the model, cf. Fig.1) and the pitch information of its second part. In this way, it determines the shape of the nuclear contour as a whole. This allows to relate variation in range proportion to the notion of *intonemes* [17] which can be seen as an attempt to define intonation in communicative terms. To which degree and in which sense a phrase-final rise represents a *contact intoneme* (C↑) or a *non-terminal intoneme* (N↑) may be expressed in its range proportion [18].

5. References

[1] Kohler, K., "Prosody in speech synthesis: the interplay between basic research and TTS application". JPhon 19, 121-138, 1991.

[2] Pierrehumbert, J., "The phonetics of English intonation". Bloomington: IULC, 1980.

[3] Grice, M. and Baumann, S., "Deutsche Intonation und GToBI", Linguist. Ber. *191*: 267-298, 2002.

[4] Asu, E.L., "Rising intonation in Estonian: an analysis of map task dialogues and spontaneous conversations". Phonetic Symposium 2006, Helsinki, Finnland, 1-9 , 2006.

[5] Dombrowski, E. and Niebuhr, O., "Acoustic patterns and communicative functions of phrase-final F0 rises in German: Activating and restricting contours", Phonetica 62: 176-195, 2005.

[6] Petrone, C. and D'Imperio, M., "From tones to tunes: The contribution of the prenuclear region in the identification of intonation contours in Italian". Paper presented at the 3rd Conference on Tone and Intonation, Lisbon, Portugal, 2008.

[7] Petrone, C. and Niebuhr, O., "The role of the prenuclear F0 region in the identification of German questions and statements. Paper presented at the 4th PaPI conference, Las Palmas, Spain, http://www.linguistik.uni-kiel.de/Niebuhr_index.html, 2009.

[8] Kohler, K.J., "Labelled data bank of spoken standard spoken German. The Kiel Corpus of Read/Spontaneous Speech. Proc. 4th ICSLP, Philadelphia, USA, 1938-1941, 1996.

[9] Niebuhr, O. and Kohler, K.J., "Perception and cognitive processing of tonal alignment in German". Proc. of the International Symposium TAL, Beijing, China, 155-158, 2004.

[10] Dombrowski, E. and Niebuhr, O., "Phrase-final rises in German: some examples". URL http://www.ipds.uni-kiel.de/kjk/forschung/lautmuster.en.html, 2005, accessed on 10 Nov 2009.

[11] Kohler, K., „A model of German intonation". AIPUK 25, Christian-Albrechts-Universität, Kiel, 295-360, 1991.

[12] Dombrowski, E., "Semantic features of accent contours: Effects of F0 peak position and F0 time shape", Proc. 15th ICPhS, Barcelona, Spain, 1217-1220, 2003.

[13] Kohler, K., "Timing and communicative functions of pitch contours", Phonetica 62: 88-105, 2005.

[14] Ambrazaitis, G., Prosodic signalling of (un)expected information in South Swedish – An interactive manipulation experiment. Proc. 3rd International Conference on Speech Prosody, Dresden, Germany, 911-914, 2006.

[15] Osgood, C., Suci, G. and Tannenbaum, "The measurement of meaning", Urbana, IL: University of Illinois Press, 1957.

[16] Boersma, P., "Praat – a system for doing phonetics by computer". Glot International 5, 341-345, 2001.

[17] Stock, E and Zacharias, C., "Deutsche Satzintonation", Leipzig: VEB Verlag Enzyklopädie, 1982

[18] Mixdorf , J. and Pfitzinger, H., "Analysing fundamental frequency contours and local speech rate in map task dialogs, Speech communication 46: 310-325, 2005.