On pitch-accent identification – The role of syllable duration and intensity

Oliver Niebuhr[+] *& Hartmut R. Pfitzinger*[*]

[+]Dept. of General & Comparative Linguistics, University of Kiel, Germany
[*]Institute of Phonetics and Digital Speech Processing (IPDS), University of Kiel, Germany
niebuhr AT linguistik.uni-kiel.de; hpt AT ipds.uni-kiel.de

Abstract

The two German pitch accents H+L* and L*+H show different duration and intensity patterns in the triplet of pre-accented, accented, and post-accented syllable. Combining the pattern of H+L* with the F0 peak of L*+H and vice versa lowered the identification of the two pitch accents. The implications of these findings for pitch-accent modelling are discussed.

1. Introduction

Pitch accents are events in the intonation of utterances. Languages like German, English, and Dutch have (partly comparable) inventories of pitch-accent categories that are involved in conveying the argumentation structure (e.g., speakers' attitudes towards the interlocutor, the discourse, or the message, cf. [1,2]) or the information structure (e.g., given vs. new information, broad vs. narrow focus, cf. [3,4]). As demonstrated by numerous perception experiments across languages, the acoustic fundamental frequency (F0) course is crucial for the coding of pitch accents. For example, changing the alignment of rising-falling F0 peaks relative to the accented syllable affects the identification of the pitch-accent category, cf. [2,5, 6,7,8]. Additionally, the perceptual identification of pitch-accent categories is affected by changes in the slopes, the ranges, the durations, and the shapes of the rising and falling movements, cf. [2,9]. Particularly these additional perceptual effects pose a problem for the autosegmental-metrical (AM) framework of intonation (cf. [10,4]) in which pitch accents are represented solely as one or two local turning points in the F0 course. How these turning points are reached, left, and connected should be irrelevant. Therefore, [11] suggested more recently to use the centre-of-gravity concept for representing pitch accents. In this way, complex effects of F0-movement qualities on the identification of pitch-accent categories can be integrated and projected onto the time axis as alignment changes of a single tonal centre of gravity (TCoG).

The TCoG shifts the phonological building blocks of pitch accents from local acoustic F0 values to more holistic and perception-oriented events. This shift can in fact solve many of the problems that relate to perceptually relevant variations in slope, duration, and shape of pitch accent movements. Yet, the TCoG approach falls short, as it cannot to account for acoustic cues to pitch-accent identification that go beyond F0 and concern, for example, intensity.

Starting from German stimulus utterances, it was shown by [5] that the successive shift of a rising-falling F0 peak across the consonant-vowel boundary of an accented syllable triggers an abrupt perceptual change that is linked to two pitch-accent categories which are known as H+L* and H*, cf. [12]. This effect was replicated by [13] in an 11-step peak-shift continuum. In addition, he demonstrated with non-speech (but speech-like) stimuli that the perceptual change does also take place when the F0 peak-shift continuum is presented solely in combination with the intensity course of the original speech stimuli, cf. the grey and black curves in Figure 1.

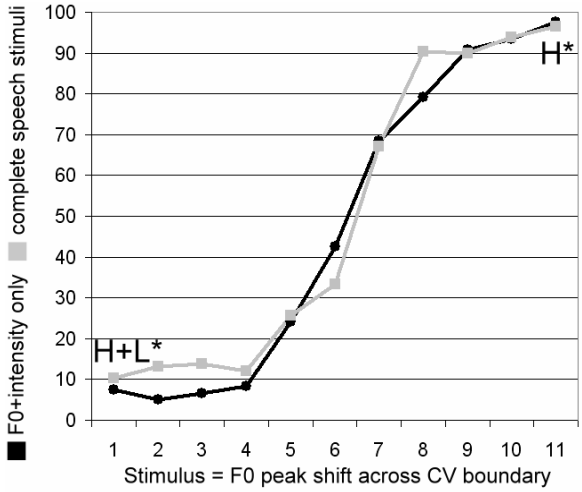

Figure 1: Change in perceived intonation (in percentages), brought about by an F0 peak-shift continuum in stimuli that contain either the complete speech signal (grey curve) or just the intensity course of the speech signal (black curve).

Moreover, by varying the steepness of the intensity increase at the consonant-vowel boundary of the accented syllable, the perceptual change from H+L* to H* identifications in the speech stimuli can be made more or less abrupt ([14]). These findings suggest that it is not the alignment of the F0 peak relative to the consonant or vowel segments, but to their concomitant intensity levels that is involved in pitch-accent identification.

In line with this idea, the production study of [15] revealed pitch-accent specific intensity levels in the triplet of pre-accented, accented, and post-accented syllable. Moreover, the variation in the intensity patterns was linked with a variation in syllable duration. The duration and intensity patterns can be described as contrasts of the pre- or post-accented syllables relative to the accented one. In summary, a reduced duration and intensity contrast was found between the accented syllable and the adjacent syllable that was additionally spanned by the F0 peak of the pitch accent, while at the same time the duration and intensity contrast between the accented syllable and the other adjacent syllable was enhanced. This can be illustrated by the two pitch accents H+L* and L*+H. For H+L* F0 falls into the accented vowel, while parts of the preceding rise are contained in the *pre*-accented syllable. By contrast, the L*+H peak spans the accented and most of the *post*-accented syllable. In view of the diametrically opposed F0-peak alignment at the beginning or the end of the accented syllable H+L* and L*+H are also referred to as *early* or *late* peak, cf. [5]. In the study of [15], the early peak of H+L* led to high duration and intensity levels in the pre-accented syllable. They approximated the ones of the accented syllable. Hence, the contrast between these two syllables in terms of duration and intensity was relatively small. At the same time, the duration and int-

ensity contrast between the accented and the post-accented syllable was very pronounced due to low duration and intensity levels of the post-accented syllable. The late peak of L*+H had an opposite effect on the duration and intensity levels in the pre- and post-accented syllables and hence on their contrasts relative to the accented syllable. Compared with H+L* the contrast between pre-accented and accented syllable was enhanced and the one between accented and post-accented syllable was reduced.

Do these pitch-accent specific duration and intensity levels in the pre- and postaccented syllables and the resulting contrasts to the accented syllable play a role in the identification of the pitch accents? For example, what happens if the F0-peak pattern of pitch accent A is combined with the naturally produced duration and intensity pattern of pitch accent B and vice versa? Do the mismatching duration and intensity patterns shift the identification towards their original pitch-accent category? And which effect have the duration and intensity patterns *per se*, i.e. if they are presented in combination with a flattened F0? The present study explores these questions for German in a perception experiment by using naturally produced F0, duration, and intensity patterns of the diametrically opposed pitch-accent categories H+L* and L*+H.

2. Method

2.1 Stimulus preparation

The perception experiment is based on six stimuli. They were derived from the continuously voiced German utterance *"Eine Malerin"* ('a painter'). It was produced twice by the first author 'ON' with a terminal falling H+L* or L*+H pitch accent on the syllable *"Ma-"*. Measurements in 'praat' [16] showed in line with the findings of [15] that the H+L* and L*+H pitch-accent productions created clearly different duration and intensity levels in the pre- and post-accented syllables (intensity levels were defined as maximum values in the syllable). Divided by the corresponding values of the accented syllable, the H+L* accent yielded duration quotients for the pre- and post-accented syllables of 0.46 and 0.29. By contrast, in connection with the L*+H production the duration quotients were 0.23 and 0.65. The quotients that were calculated for the intensity levels also resulted in diametrically opposed values for the H+L* and L*+H accents. While the pre- and post-accented syllables in the context of H+L* had intensity quotients of 1.03 and 0.87, the values created by L*+H were 0.98 and 1.06. So, compared with H+L*, L*+H again enhanced the duration and intensity contrast between pre-accented and accented syllable and reduced the one between post-accented and accented syllable. Figure 2 illustrates the different contrast patterns along with the corresponding absolute durations and intensities.

2.2 Stimulus manipulation

The F0 contours of the two naturally produced utterances were stylized in 'praat' at five contour points. Points (1) and (5) were the utterance onset and offset; the remaining contour points represented rise onset (2), maximum (3), and fall offset (4) of the F0 peak. While keeping the temporal positions constant, the contour points with the same numbers in the two utterances were given identical F0 values that were intermediate between the originally produced ones (cf. Fig.2). This was to control for potential effects of F0 register or range on the judgements of the subjects. After stylizing and equating the frequency values of the F0 contours the two utterances were resynthesized via the PSOLA algorithm in 'praat'. The resyntheses represented the first two stimuli of the perception experiment. Since they still have the different original duration and intensity (DI) patterns caused by the H+L* and L*+H pitch-accent productions, and since the marginal F0 modifications did not affect the pitch-accent category, the two stimuli are referred to as DI(H+L*)FO(H+L*) and DI(L*+H)FO(L*+H). A further pair of resyntheses was created by exchanging the H+L* and L*+H F0 patterns in the DI(H+L*)FO(H+L*) and DI(L*+H)FO(L*+H) stimuli. Thus, the stimuli are referred to as DI(H+L*)FO(L*+H) and DI(L*+H)FO(H+L*). In exchanging the contours the temporal distances of the five contour points to the vowel onsets of the coinciding syllables were adapted from the original to the new utterance in order to control for the crucial role of alignment in pitch-accent identification, cf. [2, 5,6,7,8]. A final pair of stimuli resulted from substituting the F0 contours in the DI(H+L*)FO(H+L*) and DI(L*+H)FO(L*+H) stimuli by a constant completely flat, but slightly declining F0 course from 128Hz to 118Hz. The stimuli will hence be called DI(H+L*)FO(FLAT) and DI(L*+H)FO(FLAT).

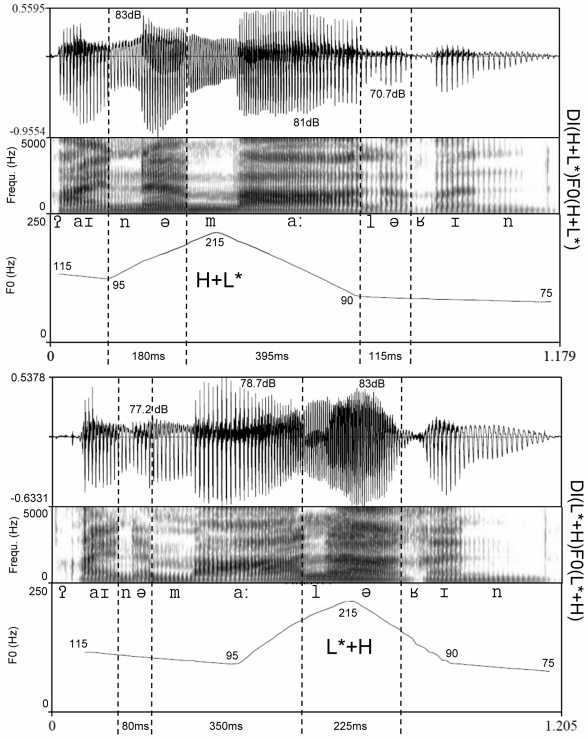

Figure 2: Acoustic analyses (oscillogram, spectrogram, F0 with contour-point values) of the stimuli DI(H+L*)FO(H+L*), *(top)*, and DI(L*+H)FO(L*+H), *(bottom)*. The boundaries of the pre-accented, accented, and post-accented syllables were labelled and their durations and intensity maxima are given.

2.3 Stimulus presentation

The perception experiment consisted of two parts in which stimulus pairs were judged with regard to attributes that aimed at the meanings of the H+L* and L*+H pitch accents. The larger part of the experiment comprised the entire 12 pairs that result when the 4 stimuli DI(H+L*)FO(H+L*),DI(L*+H)FO(L*+H), DI(H+L*)FO(L*+H), and DI(L*+H)FO(H+L*) are assembled in all possible combinations and orders, disregarding pairings of identical stimuli. The 12 pairs were arranged to 15 differently randomized sequences. In each sequence, the stimulus pairs were separated by pauses of 4 seconds, during which the subjects made their judgements. The sequences themselves were delimited by acoustic signals (double bleeps) that served as reference points for the subjects, and that coincided with the beginnings and ends of the prepared answer sheets. Furthermore, the 15 sequences were organized into five subsequent groups of three. In each group the subjects judged the stimulus

pairs with regard to a different attribute. The attributes were (1) *concluding* ('abschließend'), (2) *dominant*, (3) *questioning* ('fragend'), (4) *emotional*, and (5) *artificial* ('künstlich'). Attributes (1)-(2) aimed at the meaning of the H+L* pitch accent; (3)-(4) were dedicated to the meaning of L*+H. The attributes were selected on empirical grounds. That is, previous studies that applied the semantic differential paradigm to German intonation continua (e.g., [17,18,19]) showed consistently that the semantic concepts expressed by (1)-(2) and (3)-(4) are suitable to distinguish H+L* and L*+H. Attribute (5) was primarily added to test whether the two stimuli with mismatches between the F0 patterns on the one and the duration and intensity patterns on the other hand are more frequently perceived as artificial than the two (almost) natural stimuli.

The smaller part of the experiment focused on the stimuli with the flattened F0 contours, i.e. DI(H+L*)FO(FLAT) and DI(L*+H)FO(FLAT). Analogues to the larger part of the experiment, they were arranged to pairs in the orders AB and BA, excluding identical AA and BB pairings. Linking the two resulting stimulus pairs to each of the five attributes yielded a sequence of 10 stimulus pairs. Three of these sequences with different randomizations were created for the smaller part of the experiment. So, across the three sequences, each pair is judged three times for each attribute, as in the case of the larger part of the experiment.

Overall, 21 native speakers of German (13 females, 8 males, average age 33.5 years) participated in the experiment. They were instructed to listen carefully to the utterance pairs and to decide for each pair, which of the two stimuli (i.e. "utterances") matched better with the given attribute. Decisions were to be made by ticking boxes on prepared answer sheets. The boxes were arranged in two parallel columns called 'Eine Malerin (A)' and 'Eine Malerin (B)'. The corresponding attributes were placed in between the columns. After the instruction the subjects judged a practise sequence of 10 stimulus pairs which familiarized them with the procedure and the stimuli. The sequence was arranged differently for each subject with regard to both attributes and stimulus pairs. As for the latter, the practise sequence contained all six stimuli at least once, but none of the pairs twice. Subsequent to the short practice, the 2x5x3=30 stimulus pairs of the smaller experimental part were judged first, followed by the 12x5x3=180 stimulus pairs of the larger part. The entire experiment took 35 minutes.

3. Results

The judgements of the perception experiment were derandomized, counted and submitted to descriptive and inferential statistics. Mean judgements and standard deviations of the four stimuli with F0 peaks are normalized to a percentage scale and displayed in Figure 1. The stimuli with the original F0, duration, and intensity patterns of H+L* and L*+H are given white or black. The adjacent grey bars represent the stimuli with the same pitch accent, but a different duration and intensity pattern. One-way repeated-measures ANOVAs were applied to each of the five attributes separately in order to test for differences between the judgements of the four stimuli. It turned out that the attributes 'concluding' ($F=62.17$, $p<0.001$), 'dominant' ($F=23.33$, $p<0.001$), 'questioning' ($F=119.48$, $p<0.001$), 'emotional' ($F=33.95$, $p<0.001$), and 'artificial' ($F=10.61$,$p<0.001$) were all judged highly significantly different. The perception results for the two stimuli with flat F0 are shown in Figure 4. Here, inferential statistics revealed that the two stimuli were not judged as being different with regard to any of the attributes ('concluding', $F=0.11$, $p=0.741$; 'dominant', $F=1.33$, $p=0.261$; 'questioning', $F=0.41$, $p=0.530$; 'emotional' $F=0.21$, $p=0.649$; 'artificial', $F=0.68$, $p=0.419$).

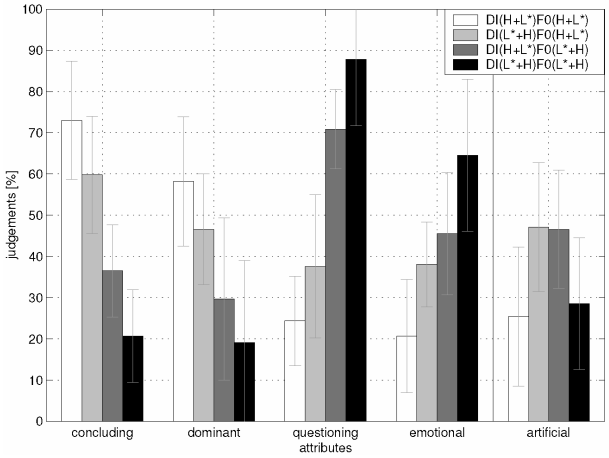

Figure 3: Summary of comparative judgements yielded by the 4 stimuli with the F0 peaks. The percentages show how often the stimulus matched better with the 5 attributes in relation to the other 3 stimuli (n=3x21x3=189).

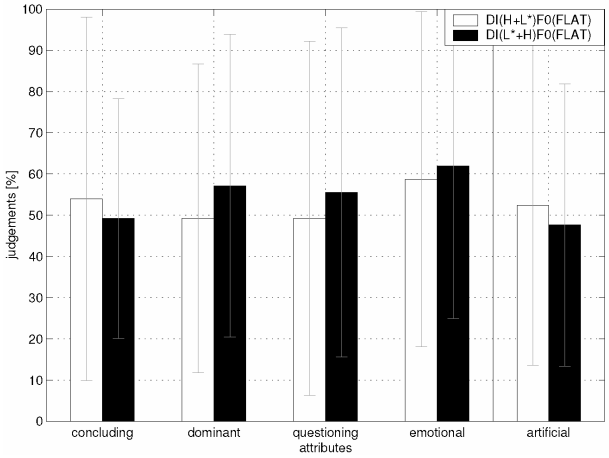

Figure 4: Summary of comparative judgements yielded by the 2 stimuli with flat F0.

4. Discussion

The comparative meaning judgements may be interpreted as indirect identifications of two pitch-accent categories. Pitch accents judged as *concluding* and/or *dominant* were identified as H+L*, whereas *questioning* and *emotional* judgements indicate L*+H identifications. From this point of view, the present experiment showed that those pitch accents were clearly identified as H+L* or L*+H for which the stylized F0 peaks were combined with their original duration and intensity levels in the triplet of pre-accented, accented, and post-accented syllable. However, combining the F0 peaks of H+L* and L*+H with the duration and intensity levels of the other pitch-accent category significantly reduced H+L* and L*+H identifications. Moreover, these stimuli sounded significantly more 'artificial' than the stimuli with the matching F0, duration, and intensity patterns. Thus, in line with [13,14] the present study provided further evidence that pitch-accent identification goes beyond mere F0-related parameters. The pitch-accent specific duration and intensity levels that were observed by [15] in the syllables around the accented one (cf. also Fig.2) do play a role in pitch-accent identification. This perceptual relevance cannot be explained by concepts like the TCoG that are restricted to F0.

As against the TCoG, the theoretical framework developed by [15] for the perception of speech melody can offer an explanation for the present findings. The crucial assumption that underlies this explanation is that a pitch accent consists of two connected gestalt-like patterns. The first pattern refers to pitch. It is constituted by tonal events derived from F0 movements and variations in the spectral energy distribution of speech sounds. The number of tonal events within the gestalt can vary due to the duration of the F0 movements and their perceptual decomposition in the string of speech sounds, cf. [20]. Each tonal event is further associated with a perceptual prominence. The resulting gestalt-like prominence pattern is the second component of a pitch accent. It is, *inter alia*, determined by the syntagmatic duration and intensity contrasts of the syllables that coincide with the tonal events, cf. [21,22]. Thus, variations in the alignment of F0 movements are regarded by [15] as an epiphenomenon, i.e. as an efficient strategy to create particular prominence patterns along with the pitch patterns.

From this point of view, it is obvious that the effects found in the present study are due to mismatches of pitch and prominence patterns. For example, it is argued by [15] that the rising-falling pitch pattern of the L*+H accent is associated with subsiding prominences in the coinciding syllables, i.e. starting from the accented syllable. This was given in the stimulus DI(L*+H)FO(L*+H), cf. Figure 2. The accented syllable in this stimulus gets a high perceptual prominence due to its great contrast with the relatively low duration and intensity levels of the pre-accented syllable (this contrast is in some cases further enhanced by increasing the duration and/or intensity level of the accented syllable itself, cf. [23]). The post-accented syllable has lower duration and intensity levels than the accented one. Yet, they are relatively high compared with the following syllable. These intermediate levels create an intermediate prominence for the post-accented syllable. By contrast, the DI(H+L*)FO(L*+H) stimulus yields rather a swelling than a subsiding prominence pattern around the accented syllable. This pattern then highlights inadequate tonal events within the late aligned F0 peak, which, in turn, caused the decrease in L*+H identifications. According to [15], a swelling prominence pattern is a feature of the H+L* pitch accent. This is supported in the present study by the fact that H+L* was better identified in the DI(H+L*)FO(H+L*) than in the DI(L*+H)FO(H+L*) stimulus.

Moreover, it fits in well with the explanatory framework sketched above that the different duration and intensity patterns in the triplet of pre-accented, accented, and post-accented syllable did only have a systematic effect on the judgements when they were presented in combination with F0-peak patterns. This suggests the duration and intensity patterns did not convey any meanings in themselves. Instead, they are indirectly related to meanings as part of the pitch-accent code. That this part is inseparably bound up with the pitch pattern, as claimed by the bipartite gestalt concept, is reflected in the significant increase in 'artificial' judgements for the stimuli with mismatching F0, duration, and intensity patterns.

In summary, the present study has demonstrated that our notion of 'pitch accent' must be detached from local turning points or movements in the acoustic F0 course and shifted towards more holistic, perception-oriented concepts. These concepts must take into account that the term 'pitch' does actually mean 'melody'. That is, the meaning of a 'pitch accent' is not solely composed of tone, but also of prominence and maybe even timbre. And through all these components, pitch accents are interwoven with the string of speech sounds. Developing such complex pitch-accent representations will be a major future challenge, which also requires getting a better understanding of the meanings conveyed by pitch accents.

5. References

[1] Kohler, K.J. (2008). Prosody in speech interaction – expression of the speaker and appeal to the listener. *Proc. 8th Conference of China and the International Symposium on Phonetic Frontiers, Beijing, China*.

[2] Niebuhr, O. (2007). The Signalling of German Rising-Falling Intonation Categories – Interplay of Synchronization, Shape, and Height. *Phonetica 64*, 174-193.

[3] Baumann, S., K. Hadelich (2003). Accent type and giveness: An experiment with auditory and visual priming. *Proc. 15th ICPhS, Barcelona, Spain*, 1811-1814.

[4] Ladd, D.R. (1996). *Intonational Phonology*. Cambridge: Cambridge University Press.

[5] Kohler, K.J. (1987). Categorical pitch perception. *Proc. 11th ICPhS, Tallinn, Estonia*, 331-333.

[6] Kleber, F. (2006). Form and function of falling pitch contours in English. *Proc. 3rd International Conference of Speech Prosody, Dresden, Germany*, 61-64.

[7] Redi, L. (2003). Categorical effects in the production of pitch contours in English. *Proc. 15th ICPhS, Barcelona, Spain*, 2921-2924.

[8] D'Imperio, M. (2000). *The role of perception in defining tonal targets and their alignment*. PhD thesis, Ohio State University.

[9] Petrone, C. & M. D'Imperio (in press). From tones to tunes: Effects of the prenuclear F0 region in the perception of Neapolitan statements and question. *Proc. Tone and Intonation in Europe 3, Lisbon, Portugal*.

[10] Pierrehumbert, J. B. (1980). *The phonology and phonetics of English intonation*. PhD thesis, MIT.

[11] Barnes, J., N. Veilleux, & A. Brugos. (2008). Alternatives to F0 turning points in American English Intonation. *JASA 124*, 2497.

[12] Grice, M., & S. Baumann (2000). Deutsche Intonation und GToBI. *Linguistische Berichte 181*, 1-33.

[13] Niebuhr, O. (2006). The role of the accented-vowel onset in the perception of German early and medial peaks. *Proc. 3rd International Conference of Speech Prosody, Dresden, Germany*, 109-112.

[14] Niebuhr, O. (2007). Categorical perception in intonation: a matter of signal dynamics? *Proc. Interspeech 2007, Antwerp, Belgium*, 642-645.

[15] Niebuhr, O. (2007). Perzeption und kognitive Verarbeitung der Sprechmelodie. Theoretische Grundlagen und empirische Untersuchungen. NY: Mouton de Gruyter.

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

[17] Dombrowski, E. (2003). Semantic features of accent contours: effects of F0 peak position and F0 time shape. *Proc. 15th ICPhS, Barcelona, Spain*, 1217-1220.

[18] Kohler, K. J. (2005). Timing and communicative functions of pitch contours. *Phonetica 62*, 88-105.

[19] Niebuhr, O. (2008). The coding of intonational meaning beyond F0. *JASA 124*, 1251-1263.

[20] House, D. (1990). Tonal perception in speech. *Travaux de l'institute de linguistique de Lund 24*, 7-163.

[21] Fry, D.B. (1958). Experiments in the perception of stress. *Language and Speech 1*, 126-152.

[22] Gay, T. (1978). Physiological and acoustic correlates of perceived stress. *Language and Speech 21*, 347-353.

[23] Gartenberg, R. & C. Panzlaff-Reuter (1991). Production and perception of F0 peak patterns in German. *AIPUK 25*, 29-113.