Perceived Prominence Reflected by Imitations of Words with and without F0 Continuity

Hansjörg Mixdorff[1], *Angelika Hönemann*[1], *Oliver Niebuhr*[2] *and Christoph Draxler*[3]

[1] Department of Computer Science and Media, Beuth University Berlin, Germany
[2] Department of General Linguistics, ISFAS, Christian-Albrecht-University of Kiel, Germany
[3] Institute of Phonetics und Speech Processing, Ludwig-Maximilian-University Munich, Germany

{mixdorff|ahoenemann}@beuth-hochschule.de, niebuhr@linguistik.uni-kiel.de,
draxler@phonetik.uni-muenchen.de

Abstract

This paper continues our work on the perception of prominence as a function of *F0* continuity. In an earlier study the first author had shown that *F0* intervals occurring at lexically stressed syllables – and measured using the amplitude of Fujisaki model accent commands – strongly contribute to the perceived prominence of that syllable. More recent work explored how *F0* continuity influenced prominence ratings of single word utterances. The outcome indicated that listeners made use of the physically available *F0* information and therefore words containing gaps in the contour were perceived as less prominent. It was also shown that subjects were able to interpolate missing parts as long as the *F0* peak was still present. The current study explores whether subjects compensate the lack of prominence in words containing *F0* gaps by asking them to produce a word with the same accent strength as that of a spoken word stimulus, the spoken word being either the same or different from the one they are asked to utter. We evaluated word durations, *F0* intervals and intensities of the responses as correlates of prominence and found that listeners indeed seem to adjust depending on the kind of stimulus they have heard.

Index Terms: prominence, perception, Fujisaki model, *F0*

1. Introduction

The information structure of an utterance is reflected in the relative saliency of its lexical constituents. At the acoustic level we observe that accented syllables serve as anchoring points of this structure. They are emphasized or toned down by phonetic means. The perceptual correlate of this process is the so-called prominence [1]. Various segmental and supra-segmental factors have been shown to affect prominence, cf. [1][2][3], such as F0 excursions, segment durations, intensity as well as vowel type and syllable coda structure. In an earlier study [4], the first author and his co-worker investigated the relationship between perceived syllable prominence and the *F0* contour in terms of the parameters of the Fujisaki model [5]. The model was used to parameterize a subcorpus of the Bonn Prosodic Database [6]. Analysis showed that prominences labeled on a scale from 0-31 strongly correlated with the excursion of *F0* movements, but only when it was anchored to accented syllables. This indicates that the prominence judgment is partly guided by linguistic considerations. Evidence in support of this assumption has been presented for many languages, including German ([7],[8], [9]), which is the language of the present study.

While we are well aware that other factors such as syllable duration [4] influence prominence perception, we concentrate in the current work on the contribution of *F0*. As mentioned above the Fujisaki model decomposes natural *F0* contours with a defined value for each speech frame, irrespective of segmental structure. This entails that it smoothly interpolates or extrapolates *F0* gaps owing to unvoiced sounds. However, from a communicative point of view, the implicit claim of using the same underlying prosodic gesture for voiced and unvoiced sound sections is that listeners are *also* able to interpolate or extrapolate *F0* gaps. Recent evidence from a tonal scaling study [10] seemed inconsistent with this assumption.

Subjects were presented with short resynthesized utterances and asked to rate the tonal height of accent-related *F0* rises. The rises led to a peak that was either present or absent due to an unvoiced stop consonant. Tonal height ratings were made and analyzed relative to reference utterances in which the *F0* rise was replaced by a flat *F0* stretch, yielding a constant tonal height. The findings of [10] suggested that the subjective continuity of pitch contours in speech is due to the fact that the auditory system simply ignores rather than fills *F0* gaps.

In [11] we therefore examined the implications of these findings for the perception of word prominence. We investigated how gaps in the *F0* contour due to unvoiced consonants affect prominence perception, given that such gaps can either be filled or blinded out by listeners. For this purpose we created a stimulus set of real disyllabic words which differed in the quantity of the vowel of the accented syllable nucleus and the types of subsequent intervocalic consonant(s). Participants rated pairs of these stimuli in a 2AFC discrimination task for accent strength, that is, they decided which word in a pair sounded more strongly accented. Results included, inter alia, that stimuli with unvoiced gaps in the *F0* contour are indeed perceived as less prominent. The prominence reduction was smaller for monotonous stimuli than for stimuli with *F0* excursions across the accented syllable. Moreover, in combination with *F0* excursions, it also mattered whether *F0* had to be interpolated or extrapolated, and whether or not the gap included a fricative sound. The results supported both the filling-in and blinding-out of *F0* gaps, which fits in well with earlier experiments on the production and perception of pitch.

The current experiment examines whether speakers compensate for the inherent difference in prominence when they produce words with or without continuous *F0* contours. To this end we asked participants of a production experiment to reply to an acoustic stimulus of an isolated word and speak either the same or a different word with the same accent strength. Our hypothesis are: (1) if speakers indeed compensate for missing F0 parts and their prominence contribution when they reproduce a given accent strength, then they will realize higher *F0* targets in low-sonority than in high-sonority words in which *F0* is more/ all F0 is present; (2) replicating our previous findings, acoustic stimuli in which *F0* is either continuous *or* can be interpolated or filled-in will make speakers produce higher *F0* targets.

10.21437/SpeechProsody.2014-49

2. Stimuli and Experiment Design

The acoustic stimuli were taken from the set employed in [11]. They are shown in Table 1 with their critical segments set in bold in the SAMPA column. As can be seen, the critical segment extends from the beginning of the open vowel to the end of the intervocalic consonant(s).

Table 1. *Five target words, SAMPA transcription, English translation, type and mean energy of the critical segment.*

Word	SAMPA	English	critical segment	mean energy (dB)
Rahmen	[R**a:m**@n]	frame	long vowel (LV), voiced (vcd) nasal	74.23
Rasen	[R**a:z**@n]	lawn	LV, vcd fricative	72.10
Raten	[R**a:t**@n]	guess	LV,voiceless (vcl) plosive	68.10
Ras(s)ten	[R**ast**@n]	rest	short vowel (SV), vcl fricative+plosive	66.19
Ratten	[R**at:**@n]	rats	SV, long vcl plosive	50.68

For acoustic uniformity, the stimuli had been created using the *MBROLA* concatenative speech synthesizer driving the German male voice *de8* [12]. The base stimulus had a monotonous *F0* at 100Hz. The long vowel [a:] was adjusted to 244ms and the central consonant portion to 126ms. In the short-vowel words the V/C portions were 142ms and 228ms. These critical segment durations and the word durations, respectively, were kept constant for all stimuli, in order to avoid that durational differences influenced the prominence judgments.

Using the *FujiParaEditor* [13] and Praat PSOLA resynthesis [14] we created further stimuli by adding *F0* peak contours to the monotonous stimuli. The contour basis was laid by a phrase component, constant for all stimuli. One accent component with duration of 200ms was superimposed on the base contour. Unlike in [11] we decided to employ only medial-peak stimuli in the current study. Their *F0* maxima were aligned close to the accented-vowel offset in the long-vowel words and in the following coda in the short-vowel words, in line with previous findings [15] and observations in citations forms.

Figure 1 displays the stimuli Rahmen, Raten and Ras(s)ten with *Aa*=0.6, *Aa* being the amplitude of the underlying Fujisaki model accent command which corresponds to the interval of the *F0* excursion associated with each word (see box-shaped accent command at the bottom of Figure 1. The F0-peak range was varied in three steps, represented by three different accent command amplitudes (*Aa*): 0.4 (interval of approximately 3 semitones), 0.6 (about 3 semitones higher) and 0.8 (about 6 semitones higher). Hence, including the monotonous condition, we generated four different acoustic versions of every word and hence a total of 100 acoustic stimulus/text pairs.

The experiment was conducted using WikiSpeech [16], a framework developed at Ludwig-Maximilian-University Munich for web-based perceptual testing and speech data collection. Participants were asked to enroll on the WikiSpeech-Website and accept the download of the Speech Recorder audio recording tool. When participants executed the program on their computers, the task was explained to them on the start-up screen. Every trial consisted of the automatic playback of the acoustic stimulus, that is, one of the synthetic stimuli, followed by the display of the word to be produced written as text. A traffic light indicated when to speak. The duration of the recording slot was fixed at five seconds. After having recorded the subject's response to the given synthetic stimulus, the experiment immediately continued with the next stimulus. Each of the acoustic words was either paired with its text equivalent or that of one of the other target words. Subjects were asked to pronounce the text word with the same accent strength as the acoustic stimulus.

Prior to the experiment the subjects listened to two examples in which of an acoustic stimulus was followed by the reply of a dummy subject. Subsequently a training session started which four pairs of the same or different words had to reproduced by the subjects with the same accent strength.

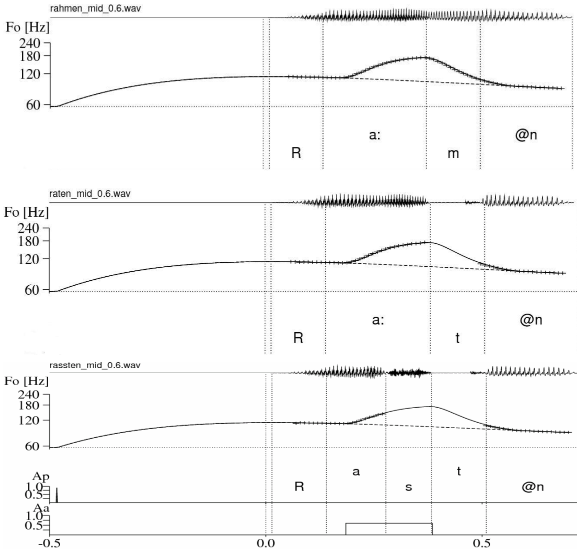

Figure 1. Examples of stimulus words Rahmen, Raten and Ras(s)ten with Aa=0.6. Panels display waveform (top), F0 contour (+++extracted, —Fujisaki model-based, middle), and underlying phrase/accent commands (bottom).

3. Acoustic Analysis of Stimuli

We recorded 19 native speakers of German (10 females, 9 males, 22-46 years old), most of them students at Beuth University Berlin or Kiel University. Since the experiment was web-based, we had no control over the recording equipment and environments. Although we had requested the use of headsets, many participants apparently used built-in microphones picking up environmental noise. Thus, the quality of recordings varied considerably with a wide range of gain settings, background noise and audible audio compression effects. But as long as the recording quality was sufficient to reliably extract prosodic parameters, and F0 in particular, we included data which would not have qualified for other acoustic analyses.

We first checked the audio files for the correct intended word, then admitted data sets with more than 70% correct word replies to further analysis, excluding, of course, all erroneous tokens within these sets. As a consequence, two participants were entirely excluded (1 female, 1 male), leaving a total of 17 data sets or 1,700 recordings.

Of the remaining sets 1,561 contained the correct word (92%). Only three of the subjects were a 100% successful. Errors chiefly concerned the words [Rast@n] and [Rat@n]. This was because, [Rast@n] (*"to rest"*) and [Ra:st@n] (past tense of *"to speed"*) have the same spelling <Rasten> in German. Since we had anticipated this interference we had spelled [Rast@n] in the non-standard way 'Rassten' to indicate the short vowel. However, the presence of 'Rasen' (infinitive of “to speed”, in addition to *“lawn”*) in the data set possibly triggered errors, especially when a long-vowel word had to be reacted to. Eleven of the participants produced [Rast@n] wrongly, in a total of 79 trials. [Rat@n] exhibited 38 two-way confusions with [Ra:t@n]. In most cases the vowel quantity of the acoustic stimulus matched the one produced erroneously. In contrast, participants reacted to the words [Ra:m@n] and [Ra:z@n] without difficulties. A few cases of errors also concerned empty audio files (9), stuttered words (8) or repetition of the acoustic stimulus word (5) when this was not requested.

After having selected the analyzable sets, their recordings were down-sampled to 16 kHz, and the first 0.85 seconds removed as they often contained audible traces of the acoustic stimulus. These traces disturbed the subsequently conducted forced alignment with the WEVOSYS LINGWAVES aligner [17]. Automatic phone segmentations were checked and corrected manually in the *PRAAT* TextGrid Editor [14].

We calculated word, syllable and phone durations in *PRAAT* based on the segmentations. *F0* values were extracted at intervals of 10 ms with *F0* floors and ceilings for male (50-300Hz) and female participants (120-400Hz). All *F0* contours were then subjected to Fujisaki model parameter extraction [18]. Results were checked and if necessary corrected in the *FujiParaEditor* [19]. In this way, we obtained a smooth, interpolated model *F0* contour even when the natural *F0* contour was interrupted by unvoiced segments. Hence the accent command amplitude *Aa* is a measure of the underlying *F0* gesture magnitude.

Intensity contours were extracted in *PRAAT* with default settings, and mean intensities in dB, as well as maxima employing parabolic interpolation were determined for each phone.

In order to analyze our data with respect to the research question, it was sufficient use only the following four Fujisaki model parameters as dependent variables: *Aa* (accent command amplitude, as a measure of magnitude of *F0* excursion at the accented syllable), *Fb* (base frequency of the *F0* pattern), *Ap* (phrase command magnitude, a measure of the magnitude of *F0* reset before phrase onset), and *T1rel* (the timing of the accent command onset relative to the onset of [a:] or [a:] in the first, accented syllable). Each word exhibited one phrase and one accent command, like the acoustic stimuli in Figure 1. *F0* con-tours of several reactions to monotonous stimuli were absolutely flat, so that neither a phrase nor an accent command was extracted.

The four dependent variables provided by the Fujisaki model were complemented by five dependent derived from the segmentation and measurement procedures conducted with PRAAT. These additional five dependent variables were *duration syllable 1*, *duration syllable 2*, as well as *vowel duration*, *mean vowel intensity*, and *maximum vowel intensity*. The latter three concerned the accented vowel [a:] or [a].

It should be stressed prior to the results presentation that we did not aim to perceptually evaluate whether participants had actually succeeded in producing words of equal prominence, but explore the patterns in their reactions to the acoustic stimuli depending on the magnitude of *F0* excursion and type of word. Our only assumption is that subjects reacted to the stimuli as requested and aimed at producing equal prominences.

4. Results of Analysis

Our 4+5=9 dependent variables were analyzed statistically with a three-way multivariate ANOVA based on the fixed factors *Word Heard*, *Word Realized*, and *F0 Range Heard*. The latter factor included the four *Aa* levels of the acoustic stimuli, i.e. 0.0 (monotonous), as well as 0.4, 0.6, and 0.8 (henceforth “*F0* peak conditions”). The other two factors had five levels each, corresponding to the disyllabic target words Rahmen, Rasen, Raten, Ras(s)ten (henceforth Rasten), and Ratten.

All three fixed factors had significant effects on many dependent variables. Our results section can only summarize a subset of these findings; and since the monotonous conditions differ considerably from the *F0* peak conditions, we will deal with the two conditions separately, starting with the *F0* peak conditions.

The most important finding in the *F0* peak conditions was that the *Aa* levels in the speakers’ productions were highly significantly affected by *F0 Range Heard* (F[3,1296]= 7.582, p<0.001). More specifically, the *Aa* levels produced by the participants clearly mirrored those of our acoustic stimuli, independently of the target word (cf. Figure 2, center). The *Aa* level 0.4 was on average even exactly reproduced, whereas the higher *Aa* levels 0.6 and 0.8 of the acoustic stimuli were slightly underestimated and on average produced as 0.5 and 0.6.

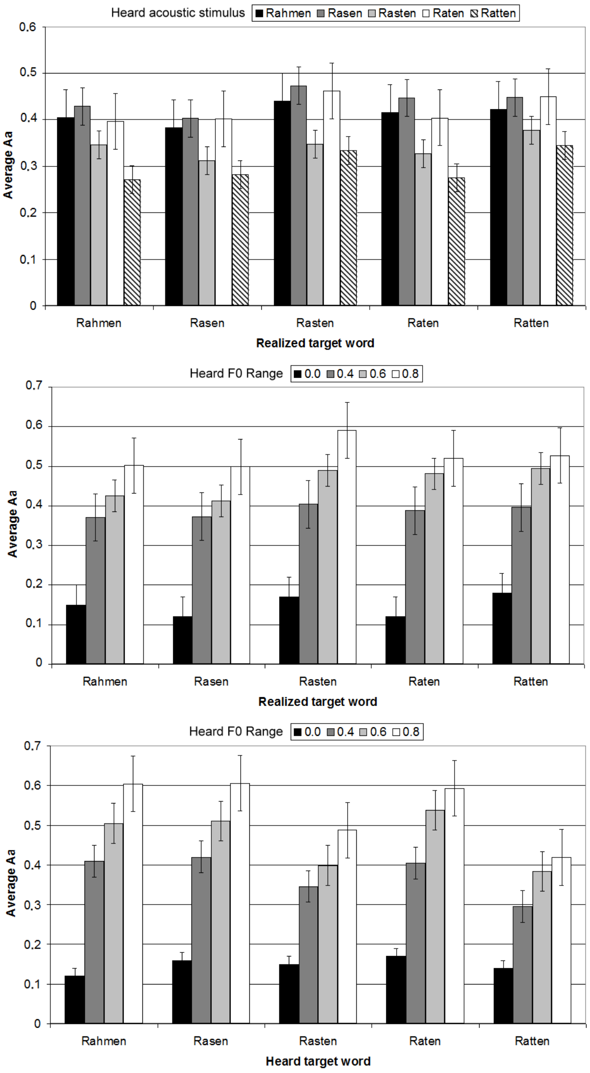

Figure 2: *Effects of* Word Heard *on average Aa levels in* Word Realized *(top), as well as combined effects of* F0 Range Heard *and* Word Realized *(center) or* Word Heard *(bottom) on the produced average Aa levels.*

Next to this general parallelism between heard and actually produced *Aa* levels, the *Aa* levels in the speakers’ productions showed target word-specific differences. This is reflected in the MANOVA in a significant interaction between *Word Heard* and *F0 Range Heard* (F[12,1296]=3.813, p<0.001). However, there was also a separate significant main effect of *Word Heard* (F[4,1296]= 35.522, p<0.001). Breaking down this main effect in terms of multiple post-hoc comparisons with Sidak correction revealed that our speakers produced higher *Aa* levels across all target words when the *F0* peaks they heard came from the more sonorous stimuli Rahmen, Rasen, and Raten rather than from the less sonorous stimuli Rasten and Ratten. The *Aa* level difference was on average 0.1 (p<0.001 in all cases, 278 ≤ n ≤ 286, cf. Figure 2, bottom). Moreover, the latter two words Rasten and Ratten yielded an additional weak *Aa* difference (p<0.05) with the Rasten stimuli triggering significantly higher *A*a level productions (of about 0.05) than the Ratten stimuli.

Exactly the opposite pattern led to a separate significant main effect of *Word Realized* (F[4,1296]= 6.829, p<0.001). We analyzed this effect by means of multiple post-hoc comparisons with Sidak correction (p<0.05 in all cases, 224 ≤ n ≤ 305, cf. Figure 2, top) and found that the *F0* peaks in the less sonorous target words Rasten and Ratten were produced with an *Aa* on average about 0.7 points higher than the more sonorous target words Rahmen, Rasen, and Raten (p<0.05 in all cases, cf. Figure 2, top and center). So, while successive desonorization and/or devoicing caused a reduction of the *F0* range at the level of perception, it triggered an extension of the *F0* range at the level of production. It is not a usual finding in the area of segment-related microprosodic perturbations that production and perception findings go in opposite directions. The present outcome suggests the existence of a compensatory strategy in *F0* production.

The speakers' productions after monotonous stimuli differed considerably from those after the *F0* peak stimuli. The nature of these differences suggests that the speakers used a stylized, singing speech mode when producing target words after monotonous stimuli. For example, *Fb* (i.e. the base *F0*) was significantly higher and *Ap* significantly smaller after monotonous than after all other stimuli, which is among others reflected in a main effect of *F0 Range Heard* (F[3,1296]= 20.752, p<0.001; F[12,1296]= 103.065, p<0.001). Furthermore, we found with respect to *F0 Range Heard* that target word productions after monotonous stimuli were softer in terms of a lower *maximum vowel intensity* (F[3,1296]= 2.253, p<0.05) and longer due to increased durations in both the first and especially the second syllable (*duration syllable 1*: F[3,1296]= 2.169, p<0.05; *duration syllable 2*: F[3,1296]= 5.095, p<0.001).

Besides these major findings, *Word Heard* resulted in an interesting further effect, which could be characterized as "transfer" or "echo effect" of the acoustic stimuli on the produced target words. For example, after hearing the short-vowel stimuli Rasten and Ratten, speakers produced their target words with shorter first syllables and vowels, independently of the target word or the quantity of its accented vowel (F[4,1296]= 10.965, p<0.001; F[4,1296]= 17.097, p<0.001). So, even Rahmen was produced with shorter first syllables and vowels when preceded by a stimulus like Rasten or Ratten. Finally, a significant main effect of *Word Realized* on *T1rel* (F[3,1296]= 5.309, p<0.001) replicated known effects of syllable structure and/or consonant type on *F0* peak timing [20], for example, in the form of an earlier timing relative to the vowel onset with increasing desonorization of the produced target words.

Last but not least, we have no indications from randomly composed sub-samples for speaker- or gender-specific effects on the crucial *Aa* variable. However, other duration and intensity variables show differences, pointing to individual F0-peak timings (cf. [24]), speaking rates or loudness levels, the latter of which could also be due to the different recording equipments.

5. Discussion and Conclusions

This paper presented results from an imitation study comparing reactions to word stimuli with either the same or a different word. Effects of *F0* on prominence perception are usually investigated with continuously voiced speech material. But what happens in more natural speech conditions, i.e. when parts of prominence-related *F0* movements are missing due to interruptions by voiceless sound segments? Do speakers and/or listeners compensate for the missing *F0* sections? The answers to these questions provided by the present findings are consistent with our hypotheses (1) and (2).

That is, in accord with previous findings on English [10], our speakers' responses to the stimuli suggest that *F0* gaps are ignored rather than filled so that word intonations in which the peak maximum and/or adjacent high *F0* section are missing sounded lower and were hence imitated with lower *Aa* on the same or other words. As in our previous study, voiceless fricatives, here the Rasten/Ratten distinction, seem to be an exception. Speakers produced higher *Aa* as a reaction to Rasten stimuli than to Ratten stimuli, which suggests that they also heard more of the prominence-related *F0* peak in the Rasten stimuli, although the *F0* gap was physically equally long as in Ratten. That is, it seems that voiceless fricatives lend themselves better to a perceptual fill-in of *F0* gaps. This matches well with the notion of "segmental intonation", developed with reference to *F0* adjusted fricative productions by the third author [21].

However, even though *F0* sections masked by voiceless segments can be restored by listeners under certain circumstances, our present findings also suggest the existence of a compensatory mechanism. Speakers did not use the same *Aa* level across all target words when they imitated the prominence level of a given stimulus. Rather, they adjusted the *Aa* level of the realized target word such that they used higher *Aa* levels for words with *F0* interruptions, hence compensating for their inherently lower prominence. As far as we know, such a compensatory mechanism is observed for the first time here, although it is known for a long time that microprosodic variation is compensated for in speech production and/or perception. For example, listeners compensate for intrinsic *F0* variation [22] or intrinsic duration variation [23]. Such compensatory mechanisms are typically only partial; and a comparison between the *Aa* levels in perception and production/imitation suggests that the same also applies to our findings.

Finally, given the general parallelism between heard and realized *Aa* levels, the clear distinction between short and long vowels, the observed "transfer/echo effect" (short/long vowels in the stimuli resulted in generally shorter/longer syllable productions), and the replication of known interactions between syllable structure and F0 peak timing, we can be confident that our speakers generated valid data and did a good job in reproducing the prosodic prominence patterns of acoustic stimuli on the same of different target words. Nevertheless, follow-up studies should aim at replicating the present findings with a larger number of speakers. and probably without the monotonous *F0* stimuli, as it is unclear how their very special character biased the prominence imitations in the *F0* peak stimulus conditions. Future work should also take up the observed "echo effects", which could point to phonetic accommodation or the way in which acoustic properties are mapped onto perceptual measures, which is another promising field for imitation tasks.

6. Acknowledgements

Special thanks go to all students at Beuth University Berlin and Kiel University participating in this experiment.

7. References

[1] Fry, D.B., “Experiments in the perception of stress”, Language and Speech 1, 126-152, 1958.

[2] Gay, T., “Physiological and acoustic correlates of perceived stress. Language and Speech 21, 347-353, 1978.

[3] Koreman, J., Van Dommelen, W., Sikveland, R., Andreeva, B., Barry, W.J., ”Cross-language differences in the production of phrasal prominence in Norwegian and German”, in M. Vainio, R. Aulanko, O. Aaltonen (eds.): Nordic Prosody X. Frankfurt: Peter Lang, 139-150, 2009.

[4] Mixdorff, H., Widera, C., “Perceived Prominence in Terms of a Linguistically Motivated Quantitative Intonation Model”, Proc. Eurospeech 2001, Aalborg, Denmark, 403-406, 2001.

[5] Fujisaki, H., Hirose, K., “Analysis of voice fundamental frequency contours for declarative sentences of Japanese”, Journal of the Acoustical Society of Japan 5, 233-241, 1984.

[6] Heuft, B., “Eine prominenzbasierte Methode zur Prosodieanalyse und –synthese”, in W. Hess and W. Lenders (eds.): Computer Studies in Language and Speech, Vol. 2, Peter Lang, Frankfurt am Main, 1999.

[7] Niebuhr, O., “Interpretation of pitch patterns and its effects on accentual prominence in German”, Proc. 3rd Tone and Intonation in Europe Conference (TIE3), Lisbon, Portugal, 2008.

[8] Niebuhr, O., “F0-based rhythm effects on the perception of local syllable prominence”, Phonetica 66, 95-112, 2009.

[9] Kleber, F., Niebuhr, O., “Semantic-context effects on lexical stress and syllable prominence”, Proc. 5th Speech Prosody, Chicago, USA, 1-4, 2010.

[10] Barnes, J., Brugos, A., Veilleux, N., Shattuck-Hufnagel, S., “Voiceless Intervals and Perceptual Completion in F0 contours: Evidence from scaling perception in American English”, Proc. 16th ICPhS, Hong Kong, China, 108-111, 2011.

[11] Mixdorff, H., Niebuhr, O., “The Influence of F0 Contour Continuity on Prominence Perception”, Proc. Interspeech 2013, Lyon, France, 230-234, 2013

[12] Dutoit, T., Pagel, V., Pierret, N., Bataille, F., van der Vreken, O., “The MBROLA Project: Towards a Set of High-Quality Speech Synthesizers Free of Use for Non-Commercial Purposes”, Proc. ICSLP, Philadelphia, USA, 1393-1396, 1996.

[13] Mixdorff, H., “FujiParaEditor”, http://public.beuth-hochschule.de/~mixdorff/thesis/fujisaki.html, 2009.

[14] Boersma, P., “Praat, a system for doing phonetics by computer”, Glot International 5, 341-345, 2001.

[15] Niebuhr, O., Ambrazaitis, G.I., “Alignment of medial and late peaks in German spontaneous speech”, Proc. 3rd Speech Prosody, Dresden, Germany, 161-164, 2006.

[16] Draxler, Chr., Jänsch, K., “WikiSpeech - A Content Management System for Speech Databases”, Proc. Interspeech 2008, 1646-1649, Brisbane, 2008.

[17] http://www.wevosys.com/products/lingwaves/lingwaves.html

[18] Mixdorff H., “A novel approach to the fully automatic extraction of Fujisaki model parameters”, Proc. ICASSP 2000, vol. 3, 1281-1284, Istanbul Turkey, 2000.

[19] Mixdorff, H., “FujiParaEditor”, http://public.beuth-hochschule.de/~mixdorff/thesis/fujisaki.html, 2009.

[20] Wichmann, A., House, J. Rietveld, T., “Discourse constraints on F0 peak timing in English”, in A. Botinis (ed.): Intonation. Dordrecht/Norwell: Kluwer Academic Publishers. 163-182, 2000.

[21] Niebuhr, O., “At the edge of intonation – The interplay of utterance-final F0 movements and voiceless fricative sounds”, Phonetica 69, 7-27, 2012.

[22] Niebuhr, O., “Intrinsic pitch in opening and closing diphthongs of German”, Proceedings of the 2nd international conference of speech prosody, Nara, Japan, 733-736, 2004.

[23] Gussenhoven, C., “Explaining two correlations between vowel quality and tone: the duration connection2”, Proceedings of the 2nd international conference of speech prosody, Nara, Japan, 179-182, 2004.

[24] Niebuhr, O., D’Imperio, M., Gili Fivela, B., Cangemi, F., “Are there "shapers" and "aligners"? Individual differences in signaling pitch accent category”, Proc. 17th ICPhS, Hong Kong, China, 120-123, 2011.