Comparative Investigation of Peak Alignment in Polish and German Unit Selection Corpora

Grazyna Demenko[1], *Agnieszka Wagner*[1], *Matthias Jilka*[2], *Bernd Möbius*[3]

[1] Dept. of Linguistics, Adam Mickiewicz University, Poznan, Poland
[2] Dept. of English Linguistics, University of Stuttgart, Germany
[3] Institute of Natural Language Processing, University of Stuttgart, Germany
lin@amu.edu.pl, wagner@amu.edu.pl, jilka@ifla.uni-stuttgart.de,
moebius@ims.uni-stuttgart.de

Abstract

This paper presents a comparative study on the temporal alignment of pitch peaks of H*L accents in Polish and German. Speech material used in the study came from the unit selection synthesis corpora of the Polish voice module of the BOSS system and the IMS German Festival TTS system. The major factors investigated were concerned with the influence of syllable structure on the one hand, as well as phrasal and tonal environment on the other hand. For the analysis of Polish falling accents, the effects of accent type, phrase type, and word position were also taken into account. Results show that in both languages, pitch peak placement is consistently affected by onset and coda type and by the tonal context (H or L tonal target preceding or following). Also, the position of the accent in the phrase is found to have a significant influence. Additionally, the results also reveal the difference between the two Polish falling pitch accents (static and dynamic).

1. Introduction

The alignment of pitch peaks is one of the key issues in the generation of natural prosody of synthetic speech. It is generally acknowledged that an F0 peak position on the syllable constitutes a distinctive feature of pitch accents.

In various intonation models the issue of peak alignment is dealt with in different ways. Phonological models distinguish between early and late peaks represented by two bi-tonal pitch accents: L*+H with a low target on the accented syllable followed by a rise or jump up to a peak on the post-accented syllable, and L+H* with a low target on the pre-accented syllable followed by a rise or jump up to a peak on the accented syllable (e.g. [5]).

In phonetic models [18] the alignment of pitch peaks is carefully controlled and determined on the basis of syllable and tonal environment. Contrary to those studies, there is, however, no control of the segmental and prosodic environment of the H*L peaks in the corpora used in this study (for example in [16] only H* peaks in the phrase-final syllable followed by a low phrase accent and a low boundary tone are considered) and the languages examined are Polish and German, not (American) English.

Proper identification of an F0 peak position is essential for the correct approximation and stylization of intonation contours, e.g. the Momel method [6] is based on the detection of target points – F0 maxima and minima – for application of the approximation function. At the moment we are developing a tool for automatic stylization of Polish intonation contours [4]. We hope that the analyses presented in this paper will bring more insight into the factors which influence peak alignment and that their results can also be used to help the automatic detection of pitch peaks for the purpose of intonation stylization.

2. Speech material and annotation

For the purpose of the investigation of peak alignment of Polish falling pitch accents 1150 phrases from the unit selection corpus of the Polish module for the BOSS speech synthesis system were used [2]. They were created by a linguist in order to provide speech samples that include accented syllables in different segmental (e.g. sonorant vs. voiceless obstruent onset and coda) and suprasegmental contexts (e.g. statement vs. exclamation). The phrases are of different length (varying from 1 to 13 words) and include 7136 accented syllables altogether (3031 instances of falling pitch accents).

The speech material was labelled using a tool *Annotation Editor* developed in the Institute of Linguistics AMU and at Poznan University of Technology. Labelling at the segmental level (transcription, segmentation into phonemes, syllables and words) was carried out automatically and with respect to prosodic features – semi-automatically. We distinguished two rising and two falling pitch accents differing with respect to whether the rise/fall is realized on the accented or the post-accented syllable (LH* and L*H for the rising accents, and H* and ΔH for the static and dynamic falling accents respectively), one rising-falling accent (L*HL), one accent realized by F0 interval between pre-accented and accented vowels (LI) and one realized by duration rather than pitch (LD). For the purpose of the current study the two types of falling accents were merged into a single class (H*L)[1].

At the prosodic phrase level information concerning phrase type was provided: statement, exclamation, question, minor intonation phrase (i.e. minor continuation and minor cadence). The speech database labelling made it possible to extract the following features for each accented syllable and include them in the analysis of peak alignment: 1) preceding and following accent type, 2) number of phonemes in the onset and coda, 3) onset and coda type (sonorant vs. voiced obstruent vs. voiceless

[1] An accent can be induced by two different mechanisms, a jump to a new pitch level in the syllable nucleus, and a change within the syllable nucleus. The use of a jump rather than a glide or vice versa is often dependent on the make-up of the syllables over which the accent spreads. If there is only one syllable a glide is more likely to be used. Differences between static and dynamic accent realizations are related to semantic function [3].

obstruent), 4) syllable position in the word, 5) number of syllables in the word, 6) foot position in the phrase (measured as a distance from the phrase end), 7) number of feet in the phrase and 8) prosodic phrase type.

For German the speech database of the IMS German Festival synthesis system [7] served as a corpus for the investigation. It mainly consists of sentences that were selected from a newspaper corpus by means of a greedy algorithm in order to ensure good coverage. The corpus was recorded by a professional male speaker, contains approximately 160 minutes of speech (2601 utterances with 17489 words [13] and was prosodically labeled using the GtoBI(S) System [11] (2681 instances of the H*L pitch accent).

3. Procedure

A *Praat* script was written which enabled to automatically obtain the information necessary for the peak alignment analysis in the Polish corpus. For each file the F0 contour was extracted and smoothed with a median filter with a window of 7 points which is useful for elimination of faulty F0 values and microprosodic effects on the F0 contour (see e.g. [12]). For each syllable the script provided: the F0 value at the syllable start and end, position and height of F0 maximum and minimum, mean F0 over the length of the syllable, F0 standard deviation and syllable duration. All this information was very useful for detection of possible errors in the F0 extraction or prosodic annotation and elimination of faulty data from the analysis.

The Festival synthesis system [1], which was used for the investigation of German, includes the "Festival feature functions" which can be used to describe a multitude of aspects of the segmental, syllabic, and prosodic structure of the utterances in the database. The measurement of the peaks themselves was done in a straightforward, automatic fashion by locating the F0 peak in a syllable labeled with a H*L pitch accent. In the case of H*L, the assumption that the peak is indeed in the same syllable is not problematic, but complications due to microprosody or voiceless regions cannot be avoided.

Throughout the analyses presented in this paper, syllable start was used as a reference point for measuring peak location, which is expressed as percentage of total syllable duration. Alternative reference points, such as start of voicing or start of rhyme, had previously been shown to yield less consistent results [8].

4. Analysis: Polish

4.1. Effects of segmental factors

4.1.1. Onset and coda type

Our analysis confirms previous findings concerning the influence of the onset and coda type - according to the van Santen & Hirschberg classification [16]: -V (voiceless obstruent), +V-S (voiced obstruent), +S (sonorant) - on peak alignment, i.e. that peaks occur earlier in the syllable if there is a sonorant in the onset and later if there is a sonorant in the coda (e.g. [8], [16]). On the basis of our data we found out that peaks occur relatively early in the syllable if there is no onset (mean: 34.22%, median: 23.05%), around the middle of the syllable if it starts with a sonorant (mean: 49.11%, median: 56.37%) or voiced obstruent (mean: 49.53%, median: 56.82%). The F0 peak moves towards the end of the syllable if the onset includes a voiceless obstruent (mean: 59.91%, median: 66.06%). As far as codas are concerned the effect is just the opposite: the peak occurs the earliest in the syllable if it includes a voiceless obstruent in the coda (mean: 39.41%, median: 45.47%), it moves towards the syllable center if there is a voiced obstruent or sonorant in the coda (mean: 43.26%, median: 48.42% and mean: 45.95%, median: 52.48% for the two coda types respectively). Peaks have the latest position if there is no coda (mean: 56.59%, median: 68.12%).

4.1.2. Number of phonemes in the onset and coda

The number of phonemes in the onset and coda also has a significant influence on peak alignment. In general, the more segments the onset includes the later the peak occurs and the less segments the coda consists of the later the peak is aligned in the syllable. The effects discussed here are illustrated in Figures 1 and 2:

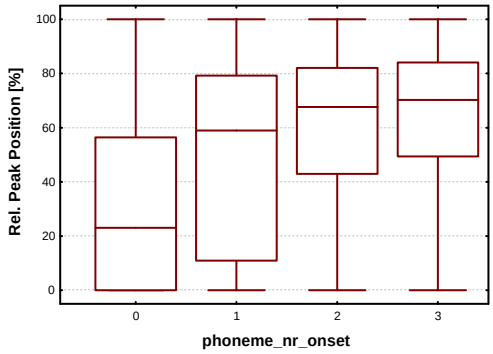

Figure 1: The effect of the phoneme number in the onset on relative peak position: 0 (median: 23.05%), 1 (median: 58.94%), 2 (median: 67.68%) and 3 (median: 70.24%)

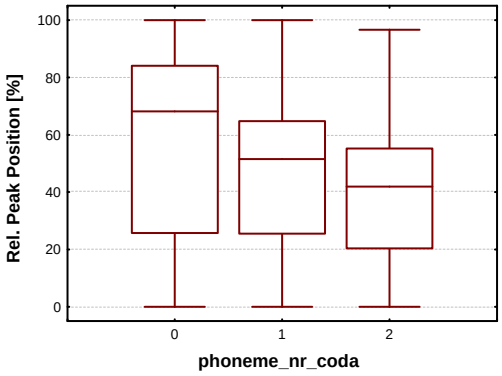

Figure 2: The effect of the phoneme number in the coda on relative peak position: 0 (median: 68.12%), 1 (median: 51.49%), 2 (median 41.93%)

4.2. Effects of suprasegmental factors

Suprasegmental factors investigated in the current study concerned phrasal and tonal environment of the accented syllable and word, and prosodic phrase type.

4.2.1. Syllable position in the word

The position of word accent in Polish is most often the penultimate syllable (see e.g. [14]). In our database there were only three instances of word-final accented syllables and four instances of word-medial accented syllables (in polysyllabic words): they were excluded from the analysis in order to avoid uncertain results. Thus, the effect of syllable position in the word was investigated on the basis of a two-way distinction between word-initial syllables and syllables pre-final in the

word. An insignificant effect of syllable position in the word has been found: F0 peaks occur relatively earlier in word-initial syllables (mean: 53.02%, median: 60.24%) than in pre-final accented syllables (mean: 54.32%, median: 64.43%).

4.2.2. Features of feet

Our results on the influence of syllable number in the feet on peak alignment confirm previous findings. It was shown in [17] that in monosyllabic feet pitch peaks occur somewhere in the middle of the syllable, whereas in polysyllabic feet they are located earlier (i.e. towards the end of the foot-initial syllable). Additionally, in [9] it was observed that this effect concerns only first three syllables in the foot. Therefore, in the current analysis feet consisting of than three syllables and more were merged into one class. The results obtained for our data are illustrated in Figure 3:

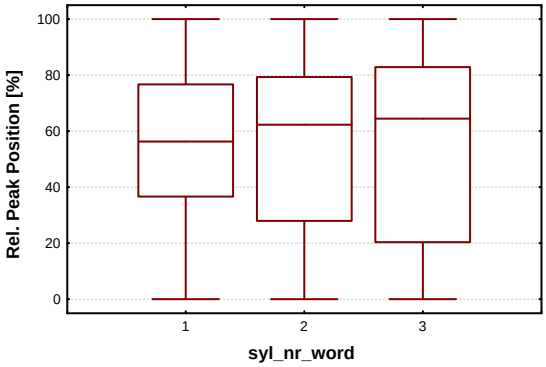

Figure 3: The effect of syllable number in the feet on relative peak position: 1 (median: 56.37%), 2 (median: 62.33%), 3 and more (median: 64.38%)

Considering feet position in the phrase an observation was made that the peak shifts towards the beginning of the syllable as the feet position changes from phrase-initial to phrase-final. In feet at the beginning of major intonation phrases the mean peak position is 76.45% into the syllable (median: 78.94%) and at the beginning of minor intonation phrases: 65.96% (median: 74.25%). The results suggest that intonation phrase type (major vs. minor) also plays a role in peak alignment. Syllables in phrase-medial feet have peaks located relatively later than syllables in feet of a pre-final position in the phrase (mean: 65.45%, medial: 69.94 vs. mean: 61.96%, median: 65.65%). In phrase-final feet F0 peaks occur in the first-third of the accented syllable (mean: 29.94%, median: 16.49%). It can be seen in Figure 4 that the accent type influences relative peak position as well: it seems that even though the mean peak position was the same for the two accent types when all instances of accented syllables were taken into account, the separation of falling accents depending on whether the fall is realized on the accents or post-accented syllable was justified (see section 2). The analysis of variance has proved that peak position is influenced significantly by both falling accent type ($F [4,3031]=86.005$, $p<0.001$) and distance of the accented syllable from phrase boundary ($F [4,3031]=49.34$, $p < 0.001$).

4.2.3. Features of phrases

On the basis of our data an influence of prosodic phrase type on peak alignment was found. We distinguished among three sentence modes: statements, questions and exclamations, and two types of minor prosodic phrases: ending with a cadence and signalling continuation. It has to be explained that in our database questions beginning with a question pronoun received the same phrase type label as statements, because they both.

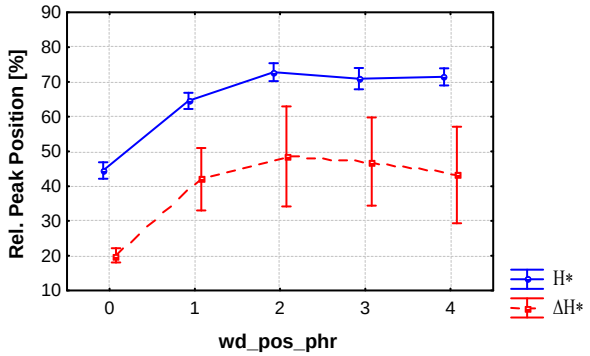

Figure 4: The effect of foot position (regarded as a distance from phrase boundary, i.e. 0 is phrase final) on relative peak position depending on the accent type H* or ΔH*. Mean peak position for phrase final static accented syllables: 43.2% and for dynamic accents: 20.85%.

have falling nuclear melody. Only yes-no questions were marked as representing the interrogative mode

Median pitch peak position in syllables in different types of prosodic phrases is illustrated in Figure 5. It can be seen that pitch peaks are located much earlier in exclamations compared to other phrase types. In statements peaks occur somewhere in the middle of the accented syllable; in questions and minor intonation phrases they are located in the second half of the syllable.

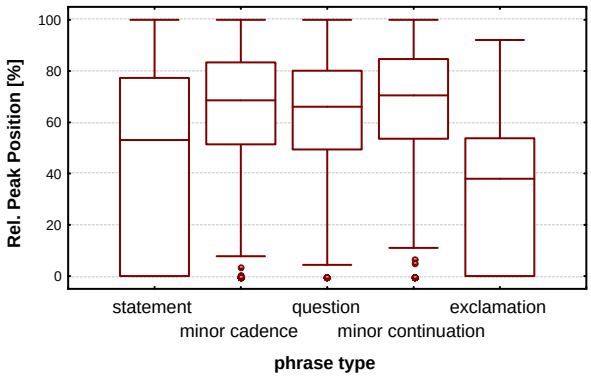

Figure 5: The effect of the phrase type on relative peak position: statement (median: 53.05%), minor cadence (median: 64.24%), question (median: 66.02%), minor continuation (median: 70.5%) and exclamation (median: 36.54%).

Another factor related to phrase structure examined in the study was the number of words in the phrase. This factor appeared to have an opposite effect on peak alignment to the one of syllable number in the word, i.e. with an increasing number of words in the phrase the peak shifts towards the start of the syllable. In single-word phrases the mean peak position is 68.15% (median: 72.71%), in two-word phrases it is 52.67% (median: 60.,89%). But from 3 words up peak location remains fairly constant: the mean position is 53.69% (median: 61.78%) in three-word phrases and 53.55% (median: 62.42%) in phrases consisting of more than three words.

4.2.4. Tonal environment

The influence of neighbouring tonal targets on peak location investigated in [8] concerned the number of syllables from/to the preceding/following pitch accent as well as type of the preceding/following tonal target (H vs. L). In the study of F0 peaks in Polish we examined the influence of the

preceding/following pitch accent type on peak alignment. A general observation is that the type of preceding pitch accent has smaller impact on peak location than the following pitch accent type.

4.3. Comments

Both for segmental and suprasegmental factors the statistical analyses shows large differences between median and mean values. For the interpretation of the statistical significance of the analysed data further research is needed. It is necessary to a) make more precise analyses for each individual factor, b) to examine the interactions between the factors and c) to carry out a multivariate analysis.

5. Analysis: German

5.1. Factors relating to syllable structure

5.1.1. Effects of onset and coda class

Similar to the Polish results (see section 4.1.1) peak placement is significantly influenced by onset and coda type in German as well. These types are again defined according to the van Santen & Hirschberg classification [16]: -V (voiceless obstruent), +V-S (voiced obstruent), +S (sonorant).

With respect to the three onset types (see also Figure 6), the peak is earliest when there is a sonorant in the onset (mean: 32.8% of syllable duration) and latest when a voiceless obstruent forms the onset (mean: 42.0%). Peaks are generally located in-between (mean: 37.0%), if the onset consists of a voiced obstruent.

For the different coda types (see Figure 7) there is a significant difference of peak position between sonorant codas (mean: 41.7% of syllable duration) and codas solely made up of obstruents (mean 28.5% for voiced obstruents; 27.7% for voiceless obstruents). The peak thus occurs clearly later when a sonorant coda is present, whereas there is no significant difference between the two obstruent classes.

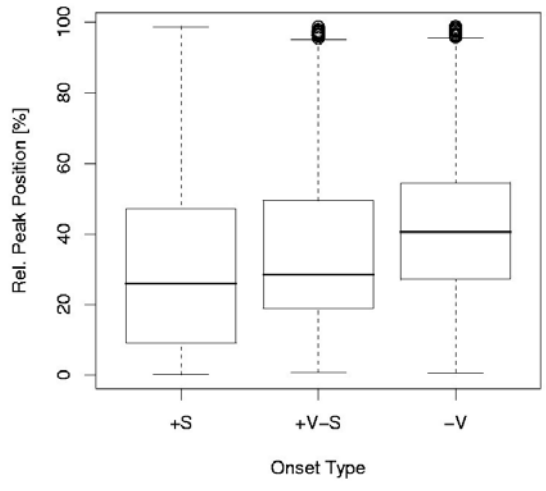

Figure 6: Boxplot showing relative peak position depending on onset type: sonorant (+S, median: 25.95%), voiced obstruent (+V-S, median: 28.4%), voiceless obstruent (-V, median: 40.5%)

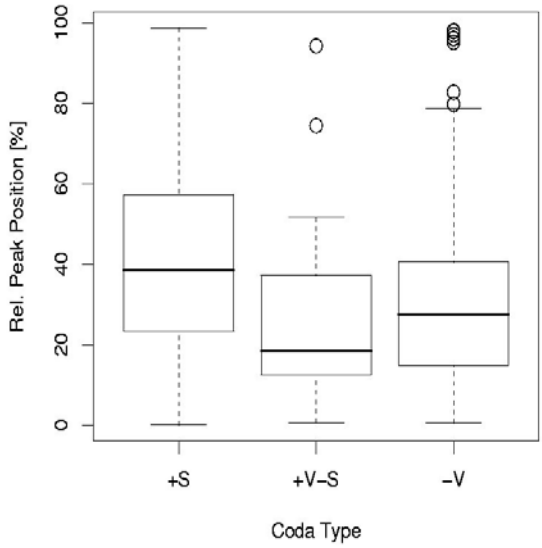

Figure 7: Boxplot showing relative peak position depending on coda type: sonorant (+S, median: 38.4%), voiced obstruent (+V-S, median: 18.45%), voiceless obstruent (-V, median: 27.7%)

If onset type varies but coda type remains unchanged (sonorant), a significant movement of the peak can be observed ($F [2, 2667] = 31.526$, $p < 0.001$), as the peak occurs successively later when the onset is a sonorant (36.23% of syllable duration), a voiced obstruent (39.23%), or a voiceless obstruent (44.97%). Variation of coda type (sonorant vs. voiceless obstruent) has a less distinct effect on the peak locations. Most peaks apparently occur in the same locations. However, the greater frequency of late peaks near the syllable boundary in sonorant codas leads to a significantly later mean value ($F [2, 2667] = 65.005$, $p < 0.001$) for sonorant (36.23%) vs. obstruent codas (25.15%).

5.1.2. Two types of sonorant coda

In the Festival feature functions' classification of syllable structure types, coda type +S covers both closed syllables with actual sonorant coda consonants and open syllables. Differences between the two types must therefore be expected.

Examining the absolute interval between syllable start and peak location, it turns out that there is virtually no difference between open and closed (+S) syllables. For the former, peak location is on average 97.59 ms after the beginning of the syllable compared to 97.24 ms for the latter. Peak position is thus stable in terms of absolute timing in this very specific context. As syllables with actual sonorant codas can be expected to be longer despite possible compensatory effects concerning vowel length (vowels in accented open syllables are unlikely to be short), this has the consequence that, in relation to syllable duration, peaks occur later in open syllables. Indeed, in open syllables the mean value for peak location is 45.87% compared to 37.98 % for syllables with one sonorant coda consonant and 35.08% for syllables with two coda consonants. If the coda consists of only one obstruent, the peak occurs even earlier, at 30.83%. Figure 8 attempts to visualize these results.

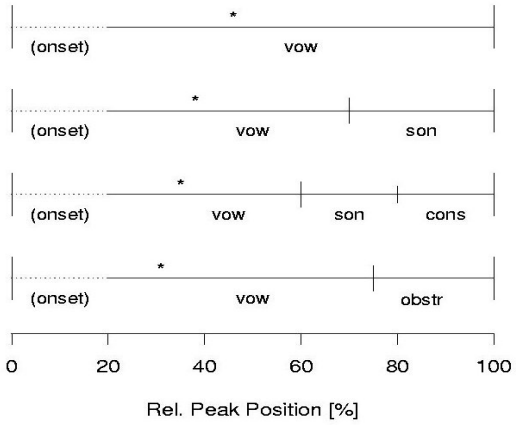

Figure 8: Relative peak position (*) for different coda compositions: open syllable (top row), coda with one sonorant consonant, coda with two consonants, coda with one obstruent consonant (bottom row).

5.2. Influence of the phrasal and tonal environment

The alignment of F_0 peaks is known to be affected by the proximity of other tonal events (pitch accents or phrase boundaries) which may lead to effects of tonal repulsion (e.g., [14]).

5.2.1. Position of the accented syllable in the phrase

A first interesting issue regarding the position of the H*L pitch accent in the intonation phrase is its distance to the preceding or following phrase boundary in number of accents. This amounts to the question whether it is significant if the pitch accent is the first, second, third, next-to-last, last etc. accent of the phrase. As determined by an analysis of variance this distance is shown to be significant ($F [1, 2668] = 475.87$, $p < 0.001$) when looking in the direction of the final phrase boundary. This is not the case with respect to the initial phrase boundary ($F [1, 2668] = 0.2437$, $p = 0.6216$).

Considering this result it is especially interesting to see whether pitch accents in the extreme positions of the phrase, i.e. the first or last (nuclear) accent, behave accordingly.

Indeed we find that peak alignment in the final pitch accent of the intonation occurs significantly earlier ($F [1, 2668] = 21.591$, $p < 0.001$) than in those accents which are not final (mean: 38.5% of syllable duration vs. 53.4% of syllable duration). Peak alignment in the first pitch accent of the phrase is on the other hand not significantly different ($F [1, 2668] = 2.5009$, $p = 0.1139$) from that of the other H*L pitch accents in the phrase.

The early alignment in nuclear pitch accents as opposed to non-final accents raises the question whether this is an effect that is facilitated by those instances that occur in the final syllable of the phrase and are thus being pushed forward by the following boundary tone. The comparison of H*L pitch accents in the phrase-final syllable with all other H*L pitch accents does in fact show a significant difference ($F [1, 2668] = 496.56$, $p < 0.001$). The peaks of final syllable accents are aligned quite early in the syllable (mean: 21.1% of syllable duration vs. 44.0%). The difference remains significant ($F [1, 2668] = 104.98$, $p < 0.001$) also when the nuclear phrase-final pitch accents are compared to nuclear pitch accents that are not in the phrase-final syllable (mean: 21.1% of syllable duration vs. 41.7%). In accordance with these results it is not surprising that similarly to a pitch accent's distance to the next phrase boundary in number of accents, its distance in number of syllables is also significant ($F [1, 2668] = 420.01$, $p < 0.001$), and that, correspondingly, there are no significant results with regard to distance to the preceding phrase boundary ($F [1, 2668] = 0.0194$, $p = 0.8892$).

5.2.2. Influence of neighboring tonal targets

If boundaries can affect peak alignment, then it is reasonable to assume that other neighboring tonal events, i.e., pitch accents, will do so as well. In a comprehensive analysis of the influence of such adjacent tonal events three major questions are of interest, namely whether the adjacent target is high or low, whether it is preceding or following and how far in terms of number of syllables it is away from the examined pitch accent.

High (H) or low (L) targets are defined by the target point closest to the examined pitch accent, a preceding L*H pitch accent would thus be registered as H, a following one as L.

A first analysis shows that peak alignment is not influenced by the type of target preceding it ($F [1, 2461] = 1.643$, $p = 0.2000$). It is, however, of weak significance whether a H or L target follows ($F [1, 2665] = 6.0593$, $p < 0.05$; mean peak position when H target follows: 36.4% of syllable duration vs. mean peak position when L target follows: 39.3% of syllable duration).

Alignment occurs significantly earlier ($F [1, 2665] = 80.584$, $p = 0.001$) when the next tonal target follows immediately in the next syllable (mean peak position: 33.6% of syllable duration vs. 42.2% of syllable duration).

This result is confirmed when comparing the influence of H and L targets at a distance of either one or two syllables from the accented syllable. Here, peak alignment is significantly different for all four possibilities (mean H following after 1 syllable: 26.1%; mean H following after 2 syllables: 39.3%; mean L following after 1 syllable: 34.5% mean L following after 2 syllables: 43.2%). In this case there is thus also a difference depending on whether a high or low target is following (see also Figure 9).

6. Conclusions

Analyses presented in this paper confirmed results of previous studies (e.g. the influence of factors related to syllable and feet structure, tonal environment) and revealed more factors that play significant role in peak alignment such as prosodic phrase structure and type, phoneme number in the onset and coda.

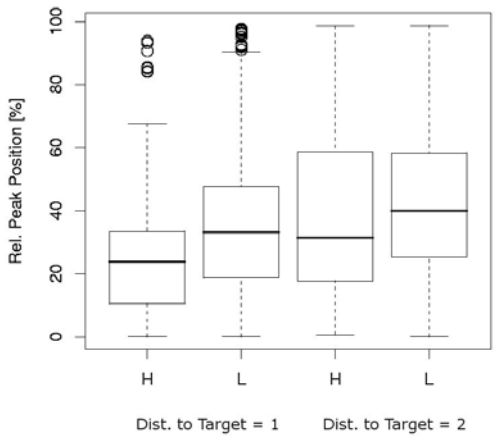

Figure 9: Boxplot showing relative peak position depending on distance and type of following target:H target after 1 syllable (median: 23.8%), L target after 1 syllable (median: 33.2%), H target after 2 syllables

(median: 31.4%), L target after 2 syllables median: 39.9%).

The comparison of peak alignment in Polish and German has shown that a) peaks occur generally later in Polish than in German, b) considering the effects of onset/coda type the same tendency can be observed in the two languages, c) in Polish stronger effect of following tonal target can be observed, d) phrase-final accent has special status in both languages (Polish: final word position, German: nuclear accent).

With respect to speech synthesis it can be added that even very general measurements such as the most frequent peak location in all H*L pitch accents of the corpus may have their use as defaults to fall back on, should more complex rules not apply. In fact, for unit selection the procedure offers the possibility of adapting to the potential prosodic idiosyncrasies of the individual speaker who provides the voice.

The general analysis of all labeled H*L accents must also disregard the fact that timing differences can either be phonetic or phonological in nature. As a consequence, differences in peak alignment that are not caused by the segmental and/or prosodic environment but are actually the expression of a different communicative function (as shown in [10] for early, medial, or late peaks in German) are not captured. From the point of view of speech synthesis, this problem is not too pressing as the prediction of such differences in meaning is not yet possible anyway. Also, this kind of phonological variation is arguably less likely to occur in a corpus that mainly contains readings from newspaper articles. Similarly, the method of detecting peaks may have to be refined, if the current approach is extended to other types of accents which are more likely to have peaks outside of the accented syllable. It may be emphasized again that the approach taken in this study has the advantage of allowing for the effective analysis of large amounts of data. It does, however, not create a controlled environment in which influences from parameters other than the one investigated are excluded. This apparent disadvantage can be dealt with by targeting interactions between specific parameters.

7. Acknowledgements

This research has been realized within the framework of the Project No. R00 035 02 “Technologies of processing and recognizing verbal information in the systems of internal security” financed by the Polish Ministry of Science and Higher Education. It has also been supported by a Alexander von Humboldt Polish Honorary Research Fellowship awarded to the fourth author.

8. References

[1] Black, A. W., Taylor, P. and Caley, R., “The Festival Speech Synthesis System – System documentation”, CSTR Edinburgh, 1999 [http://www.cstr.ed.ac.uk/ projects/festival/manual/].

[2] Breuer, S., Stober, K., Wagner, P., Abresch, J., “Bonn Open Synthesis System BOSS“.[http://www.ikp.uni-bonn.de/boss/].

[3] Cruttenden, A., “ Intonation“, Cambridge University Press, 1986

[4] Demenko, G. and Wagner, A., “The stylization of intonation contours”. Proceedings of Speech Prosody 2006, Dresden 2006

[5] Grice, M., Baumann, S., Benzmüller, R., “German Intontion in Autosegmental-Metrical Phonology”. In Jun, Sun-Ah (ed.) Prosodic Typology: The Phonology of Intonation and Phrasing. Oxford University Press 2005

[6] Hirst, D. and Espesser, R., “Automatic modelling of fundamental frequency using a quadratic spline function”. Travaux de l'Institut de Phonétique d'Aix 15, 71-85

[7] “IMS German Festival homepage,” [http://www.ims.uni-stuttgart.de/phonetik/synthesis/index.html].

[8] Jilka, M. and Möbius, B., “Towards a comprehensive investigation of factors relevant to peak alignment using a unit selection corpus”. Proceedings of Interspeech/ICSLP 2006 (Pittsburgh, PA), 2054-2057.

[9] Klabbers, E. and Van Santen, J., “Clustering of foot-based pitch contours in expressive speech”. ISCA Speech Synthesis Workshop 5, Pittsburgh, PA, 2004

[10] Kohler, K. “Macro and micro F_0 in the synthesis of intonation”. In Kingston, J. and Beckman, M. (eds.), Papers in Laboratory Phonology I. CUP, Cambridge, 115-138, 1990.

[11] Mayer, J., *Transcription of German Intonation – The Stuttgart System*, Technical Report, University of Stuttgart, 1995 [http://www.ims.uni-stuttgart.de/phonetik/ joerg/labman/STGTsystem.html].

[12] Taylor, P., “Analysis and synthesis of intonation using the Tilt model”. J. Acoust. Soc. Am. 107 (3), 1697–1714

[13] Schweitzer, A., Braunschweiler, N., Dogil, G., Möbius, B. 2004. Assessing the Acceptability of the SmartKom Speech Synthesis Voice. *Proceedings of the 5th ISCA Speech Synthesis Workshop*, 1-6.

[14] Silverman, K. and Pierrehumbert, J. “The timing of prenuclear high accents in English”. In Kingston, J. and Beckman, M. (eds). Papers in Laboratory Phonology I. Cambridge: CUP, pp. 72-106, 1990

[15] Steffen-Batóg, M., “Accentual structure of the Polish language”. Polish Scientific Publishers (PWN) 2000.

[16] Van Santen, J. and Hirschberg, J. “Segmental effects on timing and height of pitch contours”. Proceedings of *ICSLP Proc*., 719-722, 1994.

[17] Van Santen, J. and Möbius, B. “A Quantitative Model of F0 Generation and Alignment”. In Botinis, A. (ed.), Intonation – Analysis, Modelling and Technology, Kluwer, Dordrecht, 269-288, 2000.

[18] Van Santen, J., Möbius, B., Venditti, J., Shih, C., “Description of the Bell Labs intonation system”. Proceedings of Third ESCA Workshop on Speech Synthesis,1998.