The Influence of Vowel Quality Features on Peak Alignment

Matthias Jilka [1], *Bernd Möbius* [2]

[1] Department of English Linguistics, University of Stuttgart, Germany
[2] Institute of Natural Language Processing, University of Stuttgart, Germany
jilka@ifla.uni-stuttgart.de, moebius@ims.uni-stuttgart.de

Abstract

This study continues an approach that uses a unit selection corpus in order to investigate aspects of the phonetic realization of tonal categories. The focus lies on the peak position of German H*L pitch accents, specifically on the question of whether it is influenced by vowel quality. It is confirmed that vowel backness does not affect peak alignment at all. The distinction between tense and lax vowels initially promises to be relevant, as the H*L peaks seemingly occur significantly earlier in lax vowels. The effect is however demonstrated to be caused by the far greater number of lax vowels in the closed syllables found in the corpus. Finally, the feature of vowel height is revealed to be a significant factor (peaks are aligned latest in high vowels, earliest in low vowels). Various parameters (e.g., syllable structure, position in the phrase) are examined for interactions, but cannot account for the effect. While vowel height correlates with vowel duration, vowel duration itself does not influence peak position. The only possible explanation found involves peak height, which is intrinsically higher in high vowels, thus it may require more time to reach the peak.

Index Terms: peak alignment, vowel height, unit selection corpus, German

1. Introduction

This study describes an investigation of factors influencing peak alignment for German H*L pitch accents, based on a unit selection corpus, the IMS German Festival synthesis system [1]. The voice database includes a large number of individual instances of the phenomenon under investigation, thus allowing a thorough description of the segmental and prosodic features of the linguistic units in which the H*L accents occur.

While previous investigations [2], [3] concentrated on more conventional influences on peak alignment such as syllable structure and position of the accented syllable in the phrase, the analysis presented here gives an overview of the potential effect of vowel quality features on peak alignment. It examines vowel height (high, mid, low), backness (front, central, back) and the dichotomy between tense and lax vowels. As none of these factors is commonly expected to be relevant, possible correlations with co-occurring parameters are taken into consideration as well.

The primary objective of this study is thus to contribute further insights into the phonetic realization of prosodic categories, and, as the data is provided by a unit selection corpus and context information is available in newly synthesized sentences as well, to improve prosody quality during the synthesis process - either directly via the unit selection itself or by means of subsequent prosodic modification.

2. Corpus

The speech database of the IMS German Festival synthesis system serves as a corpus for the investigation. It mainly consists of sentences 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 and contains approximately 160 minutes of speech (2601 utterances with 17489 words [4]). It was prosodically labeled using the GToBI System [5] (2681 instances of the H*L pitch accent).

3. Procedure

The Festival framework provides a precise description of the segmental and prosodic environment in which H*L peaks occur. The measurement of the peaks is achieved in a straightforward, automatic fashion by locating the F_0 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 coarticulatory effects cannot be excluded.

The general analysis of all labeled H*L accents disregards the fact that timing differences could also be phonological. 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 by [6] for early, medial, or late peaks in German) are not captured.

From the point of view of speech synthesis, this is not an urgent problem 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 newspaper articles. It must be stressed that while this approach has the advantage of allowing for the effective analysis of large amounts of data, it does not provide a controlled environment in which influences from parameters other than the one investigated are excluded. Such possible interactions must therefore be thoroughly checked.

4. Tenseness

A comparison of tense vs. lax vowels in the corpus was carried out first and yielded the result that the peak of an H*L pitch accent occurs later when the accented vowel is tense (mean: 41.50%) than when it is lax (mean: 37.34%). This difference is significant ($F [1, 2668] = 17.684$, $p < 0.001$).

However, this apparent effect of a vowel quality feature on the phonetic realization of a prosodic category can actually be explained by the fact that, in the database, a much greater number of lax vowels (1132) is found in closed syllables than tense vowels (287). Previous research [2] has already shown

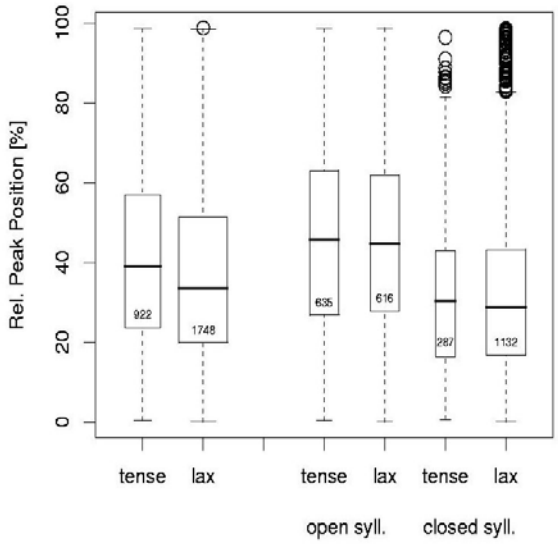

Figure 1: *Peak alignment in tense and lax vowels. The earlier peaks in lax vowels (see left column) are explained by their greater frequency in closed syllables. Figures within the boxes indicate number of instances.*.

that peaks occur significantly earlier in closed syllables than in open ones. Figure 1 illustrates that the difference between tense and lax vowels in open (45.85% vs. 45.88%) and closed syllables (31.87% vs. 32.70%) by itself is not significant, but in fact very small. The open/closed dichotomy must therefore be regarded as the truly distinctive factor. There is no reason to assume that tenseness is relevant to peak position.

5. Backness

Another major parameter of vowel quality is backness. Distinguishing front, central, and back positions, no clear tendency with respect to the position of the H*L peak in the accented syllable was found. The mean point of alignment for front vowels is at 40.35 % of syllable duration, 35.10% for central, and 40.91% for back vowels. There is no discernible effect of backness on peak placement (see Figure 2).

An interpretation of the deviating behavior of central vowels must take into account that the great majority of instances consists of variants of the phoneme /a/ (considered a central vowel in German), as the typical central vowels /@/ or /6/ are unstressed and thus do not receive pitch accents.

Considering the fact that a very similar tendency is found when comparing average peak height for the three degrees of backness (front vowels: 143.1 Hz, central vowels: 138.6 Hz, back vowels: 145.9 Hz), there is an indication that the chararacteristic of /a/ being a low vowel (as opposed to all others in the inventory) is more likely to be responsible for the differences both in peak height and position.

A closer look at the feature of vowel height is therefore the next logical step.

6. Height

The Festival feature functions also allow the classification of vowels in terms of their height as low, mid or high vowels. Like the other vowel quality features investigated in this study, this parameter is typically not assumed to be connected to peak alignment.

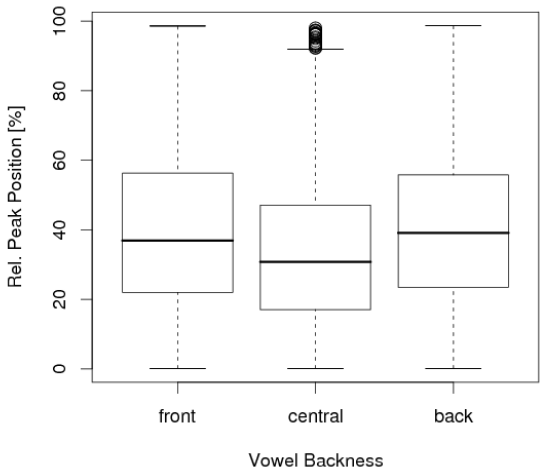

Figure 2*: Boxplot showing relative peak position depending on vowel backness: front (median: 36.9%) central (median: 30.8%), back (median: 39.1%).*

In the context of the IMS German corpus however, a significant tendency can be observed. As depicted in Figure 3 peaks occur earliest in low vowels (mean value 35.16% of syllable duration) and successively later in mid (38.28%) and high vowels (43.40%). A Welch two sample t-test (repeated measures with significance level discounted) shows that in pairwise comparisons the values for high, mid and low vowels are indeed significantly different from each other, the difference between mid and low slightly less so.

high vs. mid	t= 4.4244	p< 0.001
high vs. low	t= 6.7959	p< 0.001
mid vs. low	t= 2.8312	p< 0.005

The clarity of this result is somewhat unexpected and suggests a possible alternative explanation similar to the one presented in our investigation of tenseness in section 4. For this reason a thorough investigation was carried out, focusing on other, more plausible, factors that vowel height might interact with naturally or due to a coincidental unbalanced

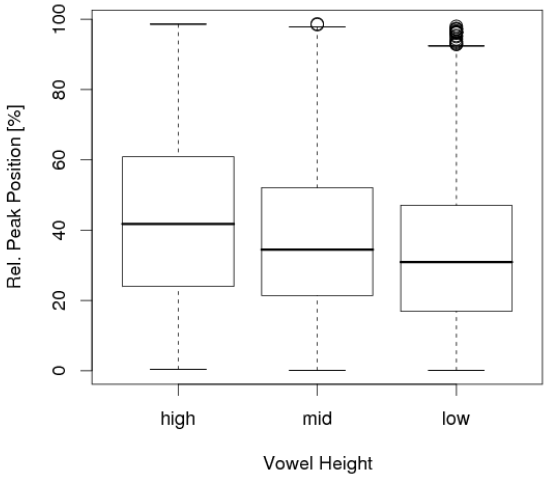

Figure 3: *Boxplot showing relative peak position depending on vowel height: high (median: 41.8%), mid (median: 34.5%), low (median: 30.9%).*

distribution of syllables in the corpus, e.g. relatively more high vowels in accented open syllables.

6.1. Correlations with previously established factors

In previous research [2], [3] various factors had been established as stable influences on the alignment of the peak in H*L pitch accents. While there is no obvious connection to vowel height, they were nevertheless checked as to whether there were any links.

Syllable structure has, for example, been shown to affect the location of the peak via the composition of onset and coda respectively. The peak occurs significantly later in open than in closed syllables. If, for some reason, our corpus contained a substantially higher number of high vowels in open syllables, then this could account for the result. This is, however, not the case as high and low vowels do not have distinctly different shares of open syllables (347 (open) / 429 (closed) = 0.81 for high vowels and 367 (open) / 475 (closed) = 0.77 for low vowels). The open/closed distinction therefore does not seem to have explanatory power.

Similar results are found for factors relating to the position of the pitch accent in the intonation phrase. Peaks in nuclear pitch accents occur significantly later than in non-nuclear pitch accents, especially in phrase-final syllables when they experience an effect of tonal repulsion triggered by the boundary. This effect may also be caused by other tonal targets immediately following the pitch accents. Unsurprisingly, in neither case does the database contain a disproportionately high number of low vowels in these positions that would account for earlier peaks in this vowel type.

6.2. Vowel duration

Vowel height has been shown to be intrinsically connected to vowel duration (see e.g., [7]). This effect can be reproduced in our corpus, also when restricted to vowels with H*L pitch accents (see Figure 4). Vowel duration is shortest in high vowels (mean: 93 ms) and longest in low vowels (mean: 139 ms) with mid vowels in between (mean: 119 ms).

A Welch two sample t-test shows that these differences are significant in all pairwise comparisons:

high vs. mid	t = -11.12	p < 0.001
high vs. low	t = -17.49	p < 0.001
mid vs. low	t = -8.15	p < 0.001

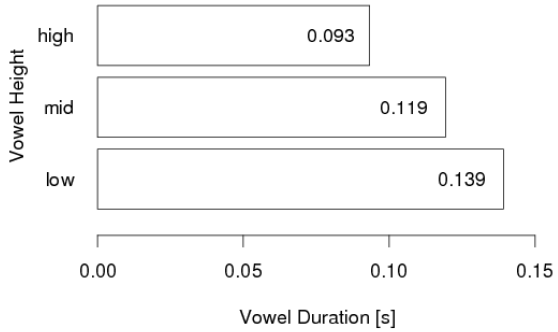

Figure 4: *Barplot showing mean vowel duration for high, mid and low vowels.*

This relationship between vowel height and vowel duration would appear to suggest that since high vowels are both associated with later peaks and shorter duration, vowel duration itself would also correlate with peak alignment. The idea of peaks occurring relatively later in a shorter reference area would include not only vowels but also syllable structure (peaks are latest in open syllables and succesively earlier the more elements there are in the coda [3]). Furthermore, an analysis of variance connecting relative peak position and vowel duration seems to indicate a significant influence of the latter factor ($F [1, 2668] = 168.75$, $p < 0.001$).

However, closer inspection of the distribution of peak position in relation to vowel duration shows no clear tendency of peak positions occurring gradually earlier with increasing vowel duration. As a consequence it cannot be claimed that vowel duration is the true cause behind the apparent influence of vowel height on peak alignment.

6.3. Peak height

The intrinsic pitch of high vowels is known to be higher than for non-high vowels (see again [7]). This also affects the height of the H*L peaks analyzed in our corpus. An analysis of variance shows that vowel height significantly affects peak height ($F [1, 2668] = 22.709$, $p < 0.001$), which is on average 145.5 Hz in high vowels, 142.9 Hz in mid vowels and 138.6 Hz in low vowels (see Figure 5). These values do not seem to be spectacularly different, but another t-test confirms that the difference in peak height is strongly significant between high and low vowels and slightly less so between mid and low vowels. It is marginally significant in the comparison of high and mid vowels ($p = 0.056$), but still shows a marked tendency for high vowels to have higher peaks.

high vs. mid	t = 1.91	p = 0.056
high vs. low	t = 4.86	p < 0.001
mid vs. low	t = 3.13	p < 0.01

6.4. Shape of the slope

If a certain importance of vowel height to peak alignment is acknowledged, then it could be associated with the course of the F_0 contour within the accented syllable. The shape of the slope rising toward the peak might manifest itself differently depending on vowel height.

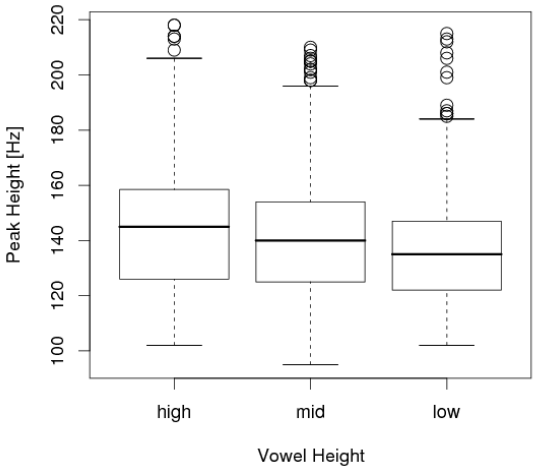

Figure 5: *Boxplot showing peak position depending on vowel height: high (median: 145 Hz), mid (median: 140.5 Hz), low (median: 135 Hz)*

6.4.1. Initial pitch

Taking into account the phenomenon of intrinsic pitch, it would be interesting to determine whether pitch values in the beginning of the syllable are also higher in high vowels. As a consequence, the slope of the rise to the peak would have to be flatter. It turns out, however, that there are no remarkable differences between these initial pitch values for high (mean: 126.6 Hz), mid (126.3 Hz) and low (121.4 Hz) vowels. For this reason, the steepness of an assumed slope - determined as the quotient of the absolute rise from the initial pitch value to the peak and the peak's distance to the syllable start – also does not yield a significant difference between high, mid and low vowels.

6.4.2. Intermediate valleys

Such a direct theoretical measurement from the initial F_0 value to the peak will of course not always give a true indication of the steepness or even generally of the slope. The rise might well start later in the syllable, preceded by a downward dip from the intial value. Such intermediate valleys are found in 791 of the 2670 H*L pitch accents in the corpus. They are distributed fairly evenly across the three levels of vowel height (high vowels: 252 instances, mid vowels: 287 instances, low vowels: 252 instances) and mean pitch values at these locations are also not significantly different (124 Hz for high vowels, 124 Hz for mid vowels, 119 Hz for low vowels). As a consequence there are also no differences in the steepness of the rise from the valleys to the corresponding F_0 peaks.

The F_0 values at the valley locations are naturally lower than at the beginning of the syllable. Nevertheless the slope from valley to F_0 peak is determined to be less steep. The reason for this is that the peaks occur substantially later when they are preceded by a valley in the same syllable (cause and effect must remain somewhat unclear). The distance is on average 31 ms longer in high vowels, 30 ms longer in mid vowels, and 31 ms in low vowels. The different contour shapes for all vowels are illustrated in Figure 6. This suggests that peak position in absolute terms should also be later when a valley is present and indeed the mean distance from the beginning of voicing to the peak is 100 ms with an intermediate valley present and 40 ms without. However, once again, no distinct behavior relating to vowel height could be determined. The occurrence of intermediate valleys is thus unlikely to be linked to vowel height.

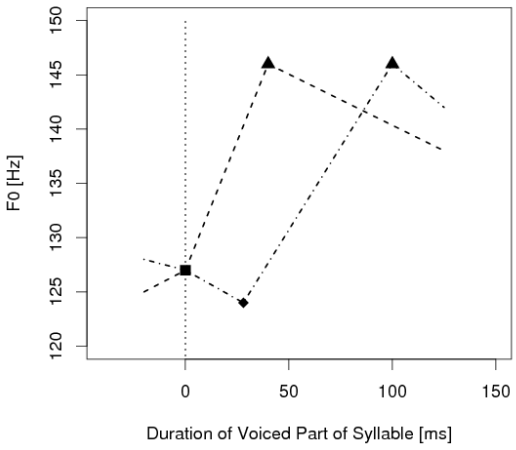

Figure 6: *Stylized contour shapes for rise to H*L peak (◢) with and without intermediate valley.*

7. Conclusions

The present study attempts to provide an analysis of the influence of vowel quality features on peak alignment, specifically in German H*L pitch accents, based on data from a unit selection corpus provided by a single speaker.

The approach allows the effective investigation of large amounts of data, albeit without offering the possibility of immediate control over individual factors in the segmental and prosodic environment of an examined peak.

The results confirm to a large extent the general impression that vowel quality features do not exert influence on peak alignment. This is certainly true for the distinction between tense and lax vowels as well as for vowel backness.

For vowel height, however, the picture is more complex. We did find a stable correlation between vowel height and peak alignment, with peaks occurring late in high vowels and successively earlier in mid and low vowels. This result is of course somewhat unexpected. The investigation of other, seemingly more plausible factors such as syllable structure or position of the accented syllable in the phrase did not yield any convincing interactions. The only parameters related to vowel height, and thus potentially useful in accounting for its influence, are vowel duration and peak height.

As for vowel duration, it has to be stated that no direct influence on peak position was found.

The remaining explanation at this point is that vowel height is indeed related to peak position, possibly via the phenomenon of intrinsic pitch. As different shapes of slopes (with and without intermediate valleys) toward the peak show no significantly different steepness depending on vowel height, the fact that the peak is innately higher in high vowels may simply lead to it taking longer to reach. This is supported by the finding that initial F_0 and steepness of rise are stable while F_0 of the peak and consequently its position are not.

8. Acknowledgements

We greatly appreciate the valuable advice given by David House on investigating the possible causes for the significant influence of vowel height.

9. References

[1] "IMS German Festival Homepage,"[http://www.ims.uni-stuttgart.de/phonetik/synthesis/index.html]

[2] Jilka, M. and Möbius, B. "Towards a Comprehensive Investigation of Factors Relevant to Peak Alignment Using a Unit Selection Corpus". Proc. Interspeech Pittsburgh, 203-206, 2006

[3] Möbius, B. and Jilka, M. "Effects of syllable structure and nuclear pitch accents on peak alignment: A corpus-based analysis". Submitted to ICPhS 2007

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

[5] Mayer, J., Transcription of German Intonation – The Stuttgart System, Technical Report, University of Stuttgart, 1995

[6] 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

[7] Neweklowsky, G. "Spezifische Dauer und spezifische Tonhöhe der Vokale", Phonetica, Vol. 32, 38-60, 1975