Segment Duration and Vowel Quality in German Lexical Stress Perception

Klaus J. Kohler

University of Kiel, Kiel, Germany
kjk AT ipds DOT uni-kiel DOT de

Abstract

A perception experiment based on systematic variation of di-syllabic vowel duration and vowel quality, and of intervocalic fricative duration in the German minimal lexical stress pair '*Kaffee* "coffee" vs. *Ca'fé* (locality) in a low f0 tail or a high f0 plateau shows highly significant effects of the 3 variables as cues to lexical stress. The combined vowel quality and fricative variables outweigh vowel duration. The effect of the prosodic frame is marginal. It is concluded that acoustic stress variables do not form a generally valid cue hierarchy.

Index Terms: lexical stress, accent, perceptual hierarchy

1. Research question

The cues to the perception of lexical stress in Germanic languages have been studied very extensively since Fry [1,2,3] set up a descending hierarchy for f0 change, syllable duration, spectral expansion in vowels and energy for cue strength in English (for a survey see [4]). As regards the influence of f0 change, these analyses have often conflated the coding of lexical stress as a feature of the phonology of the word and the coding of sentence accent as it manifests itself on the stressed syllable of the accented word, see [5] for a discussion. When the word is not pitch-accented f0 loses its primacy in the cuing of lexical stress, and duration is generally found to be the most influential lexical stress cue. Spectral expansion and acoustic energy are generally considered weak cues, and the former has received the least attention in experimental analysis.

In a recent study of spectral and durational cue strength in Dutch [6], van Heuven and de Jonge excluded f0 change from the test word by putting it in the post-accentual low-pitch tail. By systematically varying vowel duration and spectral composition (expansion/reduction) in a dual-parameter 7x7 design of a single minimal stress pair *canon* /ˈkaːnɔn/ 'canon, round song' ~ *kanon* /kaˈnɔn/ 'cannon', they confirmed that temporal organisation is a strong cue to stress perception when the target word is without a pitch accent, while spectral expansion/reduction is very weak. Its effect was noticeable only when temporal structure was ambiguous. The authors conclude that spectral expansion/reduction is indeed the weakest of the four traditional stress cues, at least in the West Germanic stress-accent languages Dutch and English.

This paper applies the Dutch research question to a similarly contrastive stressed – unstressed and unstressed – stressed German word pair of otherwise identical segmental types – *Kaffee* /'kafe:/ "coffee" and *Café* /ka'fe:/ (the locality), and adapts it to a new experimental paradigm. The two words are again placed in a sentence-prosodic frame that excludes accentual f0 change from the test items. As in the Dutch experiment, the pattern is a low tail, but it is supplemented by a hat-pattern, which levels f0 on a high plateau. The words were pronounced by the author in the two prosodic patterns in the same sentence slot: *Wir treffen uns "regelmäßig beim ˌKaffee/Caˌfé dort an der ˌEcke*. ("We "regularly meet for ˌcoffee/at the ˌcafé on the ˌcorner.") with focal accent on *regelmäßig* for the low-tail condition, and *Wir treffen uns 'regelmäßig beim 'Kaffee/Ca'fé dort an der 'Ecke*. ("We 'regularly meet for 'coffee/at the 'café on the 'corner.") with evenly distributed accents across a high f0 plateau spanning a hat pattern over *'regelmäßig beim 'Kaffee/Ca'fé dort an der* followed by the nuclear fall on *'Ecke*. Remaining differences in f0 and acoustic energy across the two syllables in the natural productions of the two words in each of the prosodic frames were smoothed under auditory control.

The vowels in the stressed syllables were longer and of more peripheral quality than the corresponding vowels in the unstressed syllables. The single intervocalic consonant was longer and stronger in the unstressed-stressed than in the stressed- unstressed syllable sequence.

The investigation tests the following hypotheses by analysing initial vs. final-stress judgements of the generated stimuli as an index of the certainty of the responses.

Hypothesis 1
As German disyllables with initial or final stress in the same phonological segment sequence are produced with different vowel duration structures across the two stress patterns, a step-wise complementary disyllabic change of vowel duration within the range of a naturally produced stress pair effects a change in lexical stress perception.

Hypothesis 2
As different vowel qualities are produced in stressed and unstressed positions of such stress pairs, they have a further effect on lexical stress perception. Thus the complementary disyllabic vowel duration changes along the same scale on each token of such a stress pair create different stress perception profiles.

Hypothesis 3
As the duration of an intervocalic consonant in such stress pairs differs, depending on whether the unstressed or the stressed vowel follows, this also has an effect on lexical stress perception. Thus the complementary disyllabic vowel duration changes in tokens with a long or a short consonant create different stress perception profiles.

Hypothesis 4
The effects of dyadic vowel duration pattern, dyadic vowel quality pattern and intervocalic consonant duration are cumulative, i.e. across the vowel duration scale, "initial-stress quality" + short consonant will yield the highest number of initial stress responses, "final-stress quality" + long consonant the highest number of final stress responses; "initial-stress quality" + long consonant will reduce the number of initial stress responses and "final-stress quality" + short consonant will reduce the number of final stress responses.

Hypothesis 5
Although focal prosody on the test word pair is excluded by insertion in low-tails and high-plateaux, the two different pro-sodic frames also have an effect on lexical stress perception, because they trigger different vowel qualities in production.

2. Method

2.1. Stimulus generation

Table 1 lists the durations and mid-vowel formant frequencies F1, F2, F3 of the first-syllable /a/ and the second-syllable /e:/, as well as the duration of the intervocalic fricative /f/ of *Kaffee/Café* in both prosodic frames.

Table 1. *Durations (in ms) and formant frequencies F1, F2, F3 (in Hz, mid-vowel position) of the vowels /a/, /e:/, and durations of the fricative /f/ in "Kaffee" and "Café" of the low-tail and high-plateau frames.*

		low		high	
		Kaffee	Café	Kaffee	Café
		ms / Hz	ms / Hz	ms / Hz	ms / Hz
/a/	dur	110	88	124	96
	F1	545	482	510	495
	F2	1504	1682	1427	1528
	F3	3121	3316	2974	3110
/e:/	dur	98	112	126	135
	F1	383	352	409	357
	F2	1795	1920	1955	1941
	F3	3251	3336	3270	3335
/f/	dur	76	104	72	100

The /k/ release + aspiration has very similar durations in all 4 words (*Kaffee* low 53 ms, *Café* low 54 ms, *Kaffee* high 62 ms, *Café* high 65 ms), but their spectra are different, so it is the whole syllable /ka/ that differs across both words and both sentence frames.

The words *Kaffee* and *Café* were excerpted from the frames and 100 ms of silence added at the right end to guarantee proper f0 and energy analysis in praat. As the vowels in *Café* had equal energy maxima in both vowels, the lower energy in the second vowel of *Kaffee* was raised by 4 dB in the tail and by 2 dB in the high plateau frame in CoolEdit to equalize. The f0 courses were levelled to auditory pitch equality in the two syllables by using Manipulation in praat, leaving microprosodic fluctuations to guarantee natural sound quality.

Disyllabic vowel duration was then varied in a complementary fashion in praat to create 5 steps from equal to long1-short1, long2-short2 and short1-long1, short2-long2 for the word pair in both prosodic patterns, creating 4 duration series, one from original *Kaffee* and one from original *Café* in each of the two prosodic frames. Table 2 gives the resulting vowel duration scales for the word pair in either prosodic frame. The 5 * 4 f0 and duration manipulations of the word excerpts from the original utterances were then psola resynthesized in praat.

Table 2. *The complementary first-second vowel durations (equal, long1-short1, long2-short2 and short1-long1, short2-long2 for both "Kaffee" and "Café" in either the low-tail or the high-plateau frame.*

	EQU	L1S	L2S	S1L	S2L
l-tail	100-100	110-90	120-80	90-110	80-120
h-plat	115-115	125-105	135-95	105-125	95-135

In each of the 20 resynthesized wav files, the /f/ section was removed and replaced by the /f/ section from both the original *Kaffee* and the original *Café*, in the low and the high frame, respectively, thus generating 2 x 5 x 4 = 40 wav files. The 20 low-pitched and the 20 high-pitched word files were then spliced into the low-tail and high-plateau, respectively. The sentence frame was in either case taken from the original utterance containing *Kaffee*. The resulting stimuli sounded fluent and completely natural without any technical artefact.

This experimental design can test whether there are different profiles of initial and final stress perception for the vowel duration series imposed on the two disyllables, which, due to their distinct original dynamic patterns, differ in their vocalic qualities and in the duration and intensity of the intervocalic fricative. Thus the aim of the paper is to quantify this interdependence between segmental duration and vowel quality, and to assess its adjustment to two different prosodic patterns, which both eliminate macroprosodic f0 movements across the disyllables, but leave microprosodic ones in to guarantee a natural sound quality.

2.2. Test files

The stimuli were indexed according to 4 parameters:

- **F0 Frame** *high*, *low* – **Fh**, **Fl**.
- **Vowel quality dyad** *Word Kaffee, Café* – **Wk**, **Wc**.
- **Vowel duration dyad** *equal, long1-short1, long2-short2, short1-long1, short2-long2* – **EQU**, **L1S**, **L2S**, **S1**L, **S2L**.
- **Fricative Consonant** *long, short* – **Cl**, **Cs**.

In the stimulus name, the parameter labels are given in the above order, separated by _, e.g. Fh_Wc_EQU_Cl. The vowel duration patterns L2S, L1S, EQU, S1L, S2L form a scale of descending first vowel durations, and a concomitantly ascending scale in the second vowel, going through equal.

Two test files were created for two perception tests with the Kiel reaction measuring instrumentation for recording responses and reaction times. All the generated stimuli were duplicated 5 times. The 5x20 high-plateau stimuli were randomized in test Fh, the 5x20 low-tail stimuli in test Fl. The length of the answering window was 4 sec. The sequence of test Fh and test Fl was changed for different subject groups to get a balanced test paradigm.

2.3. Subjects and tests

16 students of linguistics and languages (3 male, 13 female, 1 male 66 yrs, 1 female 33 yrs, the rest 21-26yrs) did the two listening tests in 4 groups of (a) 5, (b) 4, (c) 5 and (d) 2, respectively. Each group did both tests in one session, groups (a) and (d) in the order Fh – Fl, groups (b) and (c) in the reverse order. The subjects were given simultaneous written and oral instructions. These instructions provided illustrations of word differentiation by different stress placement. The word pair *Kaffee* vs. *Café* was introduced as a minimal pair with 1st-syllable and 2nd-syllable stress, respectively, to differentiate the drink from the locality in the northern variety of German. Since it could not be ruled out altogether that subjects might also use final stress for the drink, subjects were told that they should decide whether they heard the word in the utterance with stress on the first or the second syllable. They were to give their answers by pressing either button 1 or button 2 of the response boxes in front of them, and to react as quickly as possible following perception. They were told that there may be uncertain cases, but that they should still come to a quick decision. The instructions were followed by a practice run of

responding to 10 stimuli, 5 from Fl, then 5 from Fh. The test session took 15 minutes for each test and about 10 minutes for the instructions, i.e. a total of approximately 45 minutes.

3. Results

There were 17 responses missing in the high-plateau and 6 in the low-tail frame, in each case out of 16*100, i.e. ≤ 1%, which means that the data are reliable. Since the misses are spread randomly across stimuli and subjects, they are treated as non-1 responses.

Figure 1 breaks the response data down into the two test frames Fh and Fl, and the two vowel quality dyads Wc and Wk in each, along the 5-point vowel duration scale, separating the fricative durations Cl and Cs at each point. The graph shows that the two vowel quality dyads form distinct sets in both frames, point by point along the scale for corresponding fricative durations.

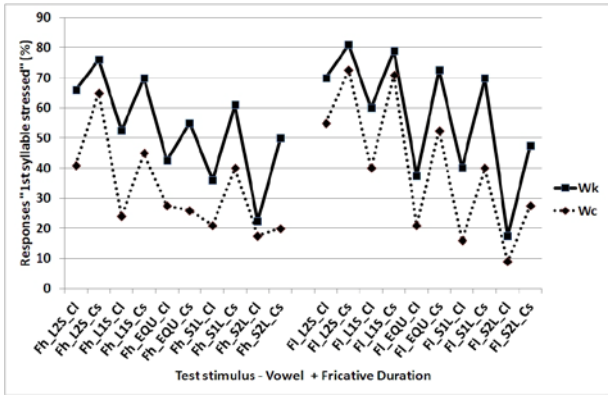

Figure 1: *Relative frequencies of "1st syllable stressed" responses to each of the 20 test stimuli in the high-plateau (Fh, left). and the low-tail (Fl, right). Plain/dotted lines connect responses to Wk/Wc for Cl and Cs along the 5-point duration scale L2S to S2L. Proportions out of 80 responses in each case.*

Figure 2 breaks the data down further into the 4 combinations of vowel quality and fricative duration Wk_Cs, Wk_Cl, Wc_Cs, Wc_Cl. From this graph we can deduce:

(1.1) In both frames, the stimuli Wk_Cs (original *Kaffee* with short /f/) show decreasing initial-stress responses along the duration scale from L2S to S2L, but the relative frequency for S2L stays around 50%, i.e. there is no change of lexical stress category from L2S to S2L.

(1.2) For all the corresponding stimuli with the other vowel dyad quality (original *Café* with short /f/: Wc_Cs), there are fewer initial-stress judgements in both frames, now resulting in a lexical stress category change from L2S to S2L.

(1.3) A similar response profile across the duration scale occurs when short /f/ of Wk_Cs is replaced by long /f / Cl.

(2.1) In both frames, the responses to stimuli Wc_Cl (original *Café* with long /f/) along the duration scale stay below or around 50% 'initial stress', i.e. there is no category change in the opposite direction from S2L to L2S.

(2.2) For all the corresponding stimuli with the other vowel dyad quality (original *Kaffee* with long /f/: Wk_Cl), there are more initial-stress judgements in both frames, now resulting in a lexical stress category change from S2L to L2S.

(2.3) A similar response profile across the duration scale occurs when long /f/ of Wc_Cl is replaced by short /f / Cs.

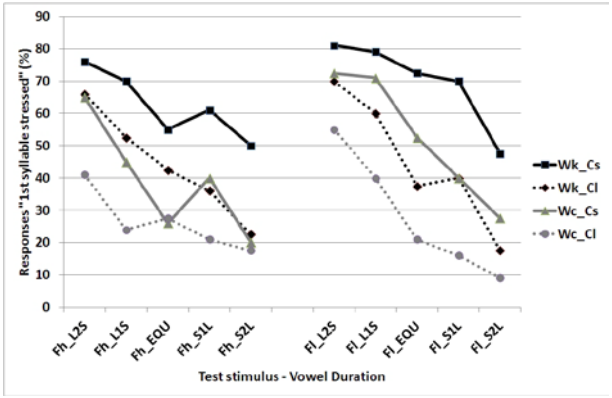

Figure 2: *Relative frequencies of "1st syllable stressed" responses to each of the 20 test stimuli in the high-plateau (Fh, left). and the low-tail (Fl, right), along the 5-point duration scale L2S to S2L in the 4 sets Wk_Cs, Wk_Cl, Wc_Cs, Wc_Cl. Proportions out of 80 responses in each case.*

These data make it very clear that complementary vowel duration patterns are not sufficient to differentiate the German minimal stress pair *'Kaffee/Ca'fé* in levelled f0. Vowel quality in the two syllables and the intervocalic fricative do not just determine judgement at vowel durations that are indecisive as to one or the other category, but even at the polar ends of the duration scale. Their cue value is thus on a par with duration.

The frame has a minor effect on stress perception. The greatest differences between the two sets relate to the entire response profiles for Wc_Cs and Wc_Cl, and to an increase in initial-stress responses for the two durations at the left end of the scale in all 4 vowel-quality – fricative pairs of the low-tail.

To test the significance of these observations, a Repeated Measures ANOVA was applied to the response data with the 4 factors Frame (2 levels high/low, Fh/Fl), Word (2 levels (original) *Café/Kaffee*, Wc/Wk), Vowel Duration (5 levels, L2S – S2L), and Fricative (2 levels long/schort, Cl/Cs). The within-subject factors Word, Vowel Duration and Fricative yield highly significant main effects: Word $F(1,15)=36.290$, $p<0.001$, $\eta p^2=0.708$; Vowel Duration $F(4,60)=48.312$, $p<0.001$, $\eta p^2=0.763$; Fricative $F(1,15)=55.886$, $p<0.001$, $\eta p^2=0.788$. These significances are linked to considerable and very similar effect sizes in terms of partial eta-squared, i.e. dyadic vowel quality and duration of intervocalic /f/ have equally significant effects as dyadic vowel duration. Frame misses significance: $F(1,15)=4.350$, $p=0.054$, $\eta p^2=0.225$, and also has a much lower effect size. Of the interactions only 3 reach significance with even smaller effect sizes: Frame * Vowel_Duration $F(4,60)=2.719$, $p=0.038$, $\eta p^2=0.153$; Frame * Vowel_Duration * Fricative $F(4,60)=2.950$, $p=0.027$, $\eta p^2= 0.164$; Word * Vowel_Duration * Fricative $F(4,60)=2.975$, $p=0.026$, $\eta p^2=0.165$.

The results of the inferential statistics buttress the descriptive observations. Vowel quality and intervocalic fricative duration have equally strong effects as vowel duration on initial or final stress perception in disyllables when f0 is smoothed on a high plateau or in a low tail, and these different f0 frames have no independent effect. Furthermore, the effects of vowel quality and fricative duration on the response profiles of the vowel duration scale are cumulative, with Wk_Cs – Wk_Cl – Wc_Cs – Wc_Cl yielding a descending order for initial and an ascending order for final stress.

4. Discussion

This paper has investigated the contribution to stress perception of vowel quality and intervocalic fricative duration besides complementary vowel duration changes in the German disyllables *Kaffee* and *Café*, excluding critical f0 changes on the test words by placing them either in a low tail or a in high-plateau hat pattern of a sentence frame. The analysis of the response data has confirmed Hypotheses 1-3 as vowel duration, vowel quality and fricative duration have proved to be strong significant main effects in shaping the stress response profiles in German *Kaffee* vs. *Café* along the duration scale.

All 4 combinations of vowel quality and fricative duration create different profiles, but there is no category change in the vowel quality and fricative combination Wc_Cl, which has the strongest cue for final stress and does not change to initial stress even at the appropriate vowel durations, nor in Wk_Cs, which has the strongest cue for initial stress and does not change to final stress even at the appropriate vowel durations. In the weaker combinations Wc_Cs and Wk_Cl, where vowel quality and fricative duration provide opposing cues to initial or final stress, the category change occurs. This means that the vowel quality and fricative duration values are cumulative to create a basic cue power for initial or final stress, onto which vowel duration is grafted This applies to both prosodic frames. Hypothesis 4 has thus also been confirmed.

Hypothesis 5 has not been confirmed, as there are only marginal effects of such high or low levelled prosodic frames on stress perception.

5. Conclusions

As f0 on test items is levelled in a low tail or a high-plateau hat pattern vowel quality can become an important cue in lexical stress perception, further enhanced by the strength/ duration of an intervocalic consonant. The combined cue force of these two parameters can outweigh the effect of vowel duration. van Heuven and de Jonge [6] also showed the effect of vowel quality for the Dutch minimal initial-final stress pair *canon* and *kanon*, but the effect is much smaller.

This is probably due to the different experimental design and to the lack of a proper minimal vowel quantity pair, since *canon* has a long stressed vowel, *kanon* a short unstressed one in the first syllable. */ˈkanɔn/ with a short stressed vowel in the first syllable would be a possible phonotactic structure but does not exist as a word. Therefore, changing duration along a 7-point scale that spans the long stressed and the short unstressed vowel in the Dutch word pair needs a wider range than the stressed-unstressed vowels in the German word pair, and a longer first-syllable vowel for *canon* identification. The manipulation of the first-syllable vowel covered a range of 120ms, that of the second syllable 40ms (at a 85/15 ratio for vowel/final /n/). This compares with just 40ms in the vowels of both syllables of the German pair. Furthermore, the duration manipulation goes through */ˈkanɔn/ which does not have a straightforward association with *canon* in a word identification test and consequently with initial stress. The most reduced spectral steps do not go together with the longest duration steps in the production of /ˈkaːnɔn/, and listeners may therefore find it difficult to allocate them to either "kanon" or "canon". Similarly, the most expanded spectral steps do not go together with the shortest duration steps in production, and may point to */ˈkanɔn/, which is not a possible response. Thus duration is bound to have greater cue strength over vowel quality. The stimulus synthesis method may also have interfered with the naturalness of the sound output.

In a similar experiment to the German one, [7] manipulated duration in 7 steps across the two syllables of Catalan *mama* [ˈmamə] and *mamà* [məˈma], ranging from 87-45ms to 45-87ms, i.e. vowel durations are closer to the German than to the Dutch data, and range over 40ms as in the former. Vowel quality has a strong effect: [ˈmamə] cannot be changed to final stress by short-long vowels, and vice versa for [məˈma]. The effect is even stronger than the authors realise because their instruction to listeners to respond only on hearing *mamà* must have increased false alarms and created a bias towards final stress. The results thus parallel the ones for German.

This is a field that needs further cross-linguistic investigation with more words, including different vowel and consonant types. Other test examples for German are *Das ist bestimmt für den August nicht typisch.* "That is certainly not typical for Augustus / the month of August." or *Bei uns hat immer die 'Mama/Ma'ma / der 'Papa/Pa'pa das letzte Wort.* "In our house, mum/dad always has the final word.", with different stylistic forms for mum/dad, The same stimulus generation with the variables Vowel Quality and Vowel Duration in the two syllables, Intervocalic Consonant, Low-tail or High-plateau frames, and the same test design would be applied.

From the limited data we have we can already conclude that investigations into the exponents of lexical stress and into their effects on stress perception need to open a broad analysis window that not only spans at least a stressed and an unstressed syllable, but includes all segmental and prosodic variables in it. There is no general hierarchy of the variables traditionally adduced to signal lexical stress. F0, vowel duration, vowel quality, consonant duration, acoustic energy operate together at all times to transmit lexical stress information. They do this not just in the stressed syllable but across at least a stressed/unstressed or unstressed/stressed syllable sequence, and the weight of each individual parameter is adapted to the contextual and situational demands of speech communication. Thus, every new prosodic embedding of syllable sequences defines the hierarchy afresh, and artificial separation of variables in experimental designs should always pay attention to naturalness and communicative appropriateness of the stimuli.

6. References

[1] Fry, D.B., "Duration and intensity as physical correlates of linguistic stress", J. Acoust. Soc. Am. *27*, 765-768 , 1955.

[2] Fry, D.B., "Experiments in the perception of stress", Lang. Speech 1, 126-152, 1958.

[3] Fry, D.B., "The dependence of stress judgments on vowel formant structure", in Zwirner, Bethge, Proceedings of the 6th International Congress of Phonetic Sciences, 306-311, Karger, Basel, 1965.

[4] Heuven, V.J. van; Sluijter, A.M.C., "Notes on the phonetics of word prosody", in Goedemans, van der Hulst, Visch, Stress patterns of the world, Part 1: Background, HIL Publications (vol. 2), pp. 233-269, Holland Academic Graphics, The Hague, 1996.

[5] Sluijter, A.M.C.; Heuven, V.J. van, "Effects of focus distribution, pitch accent and lexical stress on the temporal organisation of syllables in Dutch". Phonetica 52, 71-89,1996.

[6] Heuven, V.J. van, Jonge, M. de, "Spectral and temporal reduction as stress cues in Dutch", Phonetica 68, 1-13, 2011.

[7] Ortega-Llebaria, M., Vanrell, M.M., Prieto, P., "Catalan speakers' perception of word stress in unaccented contexts", J. Acoust. Soc. Am. 127, 462-471, 2011.