EVIDENCE FOR TONAL IDENTITY FROM PEAK SCALING UNDER PITCH SPAN VARIATION

Martine Grice, Stefan Baumann & Nils Jagdfeld

I*f*L Phonetik, University of Cologne
{martine.grice; stefan.baumann; nils.jagdfeld}@uni-koeln.de

ABSTRACT

In a reading task we investigate the scaling of pitch accents in neutral and lively speech in German. We first show that lively speech tends to increase the pitch span, raising the F0 targets for H tones but little affecting those for L tones. We then investigate the scaling of a tone whose identity is controversial: the second tone, X, of an early peak accent (H+X*). This pitch accent is employed on inferentially accessible referents and has been analysed as H+!H* as well as H+L*. Our finding that the F0 target for X is clearly raised in lively speech favours its analysis as a downstepped high tone in a H+!H* pitch accent.

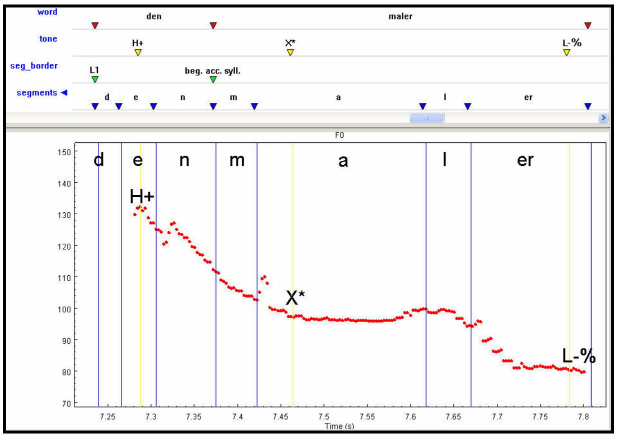
Keywords: Early peak, leading tone, F0 scaling, downstep, information status, intonational phonology

1. INTRODUCTION

[1] and [2] found that in German pitch accents with the H peak on the pretonic syllable (early peak accents) are appropriate in declarative utterances on a referent which is inferentially accessible. For example, if a speaker has just mentioned a *convent*, the word *nun* in a sentence like *He was waiting for the nun* is not 'out of the blue' for the hearer; instead, the presence of a nun can be inferred from the scenario, which, in this case, is the convent.

Within the autosegmental-metrical framework, the phonological representation of the early peak accents employed in this context is still unclear. An example F0 trace is given in Fig.1. In the example, the early peak is clearly on the pretonic syllable, *den*, and is labelled H+. The final low pitch towards the end of the phrase is taken to be an edge tone, and is labelled L-% (a sequence of the intermediate phrase boundary tone L- and the Intonation Phrase boundary tone L%). The open question is whether the apparent F0 target around the middle of the speaker's range on the tonic syllable *MA-*, labelled X*, should be classified as a downstepped H tone or a (somewhat raised) L tone. That is, is the accent best analysed as H+!H* or H+L*?

Figure 1: Example of early peak accent on *den MAler* ('the painter') with F0 trace and labels; from: *Er wartete auf den Maler* ('He was waiting for the painter').



For American English, the early peak accent has been analysed as H+L* [10], with the specification that L*, when part of this bitonal pitch accent, is scaled higher than monotonal L*. Thus, although the starred tone was observed to be phonetically mid, the early peak accent was analysed as underlyingly H+L*. In the ToBI annotation scheme [3], which is somewhat less abstract in terms of scaling, this same mid tone is transcribed as a downstepped H tone within the H+!H* accent. This ToBI label is not based on a phonological reanalysis but is merely introduced for ease of learning.

The German ToBI system [5] distinguishes between H+L* and H+!H*, although this distinction has been disputed [11]. We do not explore this issue here, but rather restrict our study to a well-defined context in which one type of early peak accent is appropriate – on referents which are inferentially accessible – and aim to shed light on the underlying analysis of the starred tone in that accent. To do this we investigate its scaling across different pitch range conditions.

It is known that pitch range and pitch span can be affected by a number of paralinguistic, attitudinal and emotive factors (see [9]). For example, [6] show in a perception experiment that as stimuli sound more surprised, H* tones are higher, whereas L* tones are lower. Translated into production, surprised utterances will use an expanded pitch range – also referred to as an increased *pitch span* [7] – where the topline (affecting the scaling of the H tones) is raised, and the baseline (affecting the scaling of L tones) is lowered.

Eliciting comparable speech material with different pitch spans can thus inform us about the tonal identity of X* in early peak accents. As the pitch span increases, X would be raised if it were a (downstepped) H and lowered if it were a L.

Our first concern was to find a paralinguistic dimension which was likely to manipulate pitch span without affecting the speaker's choice of pitch accent, so that the same accent type could be compared across neutral and modified pitch ranges. Surprise is problematic, as it can affect the information stuctural interpretation: If an utterance is spoken with surprise, it could be that the content of that utterance has not been correctly inferred from the context. This could lead to a different pitch accent on the focussed word (e.g. L+H*, as in new or contrastive cases), rather than an early peak accent.

Our second concern was to make sure there was an increase in pitch span without a raising of the overall level. From the attributes which [6] suggest as promising for this purpose (expressiveness, liveliness and insistence), informal testing led to a preference for liveliness.

2. METHOD

2.1 Speakers and recordings

Seven native speakers of German, three female and four male, were recruited from the undergraduate and graduate student population of Cologne University. They were aged between 22 and 28. They were asked to read aloud mini-dialogues in which the following types of exchange were embedded (target tones in bold face, H+X* stands for an early peak accent).

1A:Ein Bekannter von mir hat seinen Urlaub in einem Kloster verbracht. Jeden Tag betete er stundenlang in der Kapelle. Er wartete auf **[H+X*] die Nonne [L-%]**.
1B: Auf die **[L*] Nonne**? Warum das denn?
(1A: Someone I know spent his holidays in a convent. Every day he prayed for hours in the chapel. He was waiting for the ***nun****. 1B: For the* ***nun****? Why did he do that?)*

2A: Auf wen hat Klaus gewartet?
2B: Auf die **[L+H*] Nonne**. Sie wollte ihm das Kloster zeigen.
(2A: What was Klaus waiting for? 2B: For the ***nun****. She wanted to show him the convent)*

Exchanges of type 1 and 2 were separated by at least one other, to avoid givenness of the target noun in 2B.

2.2 Labelling

A tonal analysis was carried out according to the categories in GToBI [5] with the exception of early peak accents, which were labelled H+X*. Alignment points for each of the target tones were placed by hand after inspection of the F0 display in the EMU labelling and database system, see Fig.1 for a screen shot. In addition, in the early peak contours, the beginning of the tonic syllable and the pretonic syllable were labelled, so that the location of the peak could be determined in relation to these landmarks. (For the *den Maler* context, only cases where a clear segment boundary could be found were labelled, e.g. decrease in F2 in the nasal spectra from /n/ to /m/.) F0 values were transformed into semitones relative to a reference of 50 Hz [8] to facilitate cross-speaker comparisons.

3. RESULTS AND DISCUSSION

The data from three female and three male speakers were analysed, making a total of 224 tokens. The values of the fourth male speaker were excluded because he frequently employed a (L+)H* pitch accent on the test word in 1A rather than the expected early peak contour.

3.1 Timing of early and medial peaks

To verify the analysis of accents in contexts of type 1A as early peaks, latencies were calculated between the beginning of the accented syllable and the labelled H+ peak of H+X*, as compared with the medial H* peak of L+H*.

Table 1: Mean position of early peak H tone (H+) and medial peak tone (H*) relative to the beginning of the accented syllable in ms. (F=female, M=male)

speaker	(H+) – begin acc. syll.	(H*) – begin acc. syll.	p
F1	-59.5	170.4	***
F2	-40.9	166.4	***
F3	-56.4	183.3	***
M1	-51.6	187.1	***
M2	-65.2	82.8	***
M3	-38.2	171.4	***
overall	-52	160.2	***

Table 1 shows that the H+ tone occurs, on average, 52 ms *before* the beginning of the accented syllable, whereas the H* tone occurs, on average, 160 ms *after* it. A series of t-tests show that the difference in timing between H+ and H* is highly significant for all speakers (p<0.001).

3.2 Pitch span and identity of X*

Fig. 2 shows the mean F0 values in semitones for H and L tones in neutral and lively speech. These are tones whose analysis as H or L is undisputed, both in the case of the starred tones H* (from the context of type 2B) and L* (1B), as well as the leading tone H+ and the boundary tone L-% (both from the early peak context 1A).

Figure 2: F0 values in semitones (reference 50 Hz) for H*, L*, H+ and L-% in neutral (N) and lively (L) renditions for each speaker.

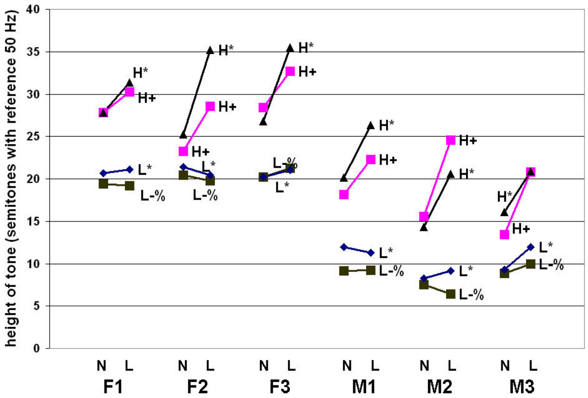


Table 2: Difference in semitones between neutral and lively conditions for each speaker and each tone.

	H+		X*		L-%		H*		L*	
speaker	Δ (st) L–N	p	Δ (st) L–N	p	Δ (st) L–N	p	Δ (st) L–N	p	Δ (st) L–N	p
F1	+2.47	**	+1.34	*	-0.21	0.655	+3.58	***	+0.43	0.342
F2	+5.27	***	+1.07	*	-0.62	**	+9.98	***	-0.91	**
F3	+4.28	***	+3.77	***	+0.98	*	+8.69	***	+0.83	0.055
M1	+4.08	**	+2.22	*	+0.05	0.922	+6.22	***	-0.71	0.071
M2	+8.99	***	+7.18	***	-1.11	*	+6.26	***	+0.85	*
M3	+7.33	***	+2.63	***	+1.12	**	+4.85	***	+2.64	***

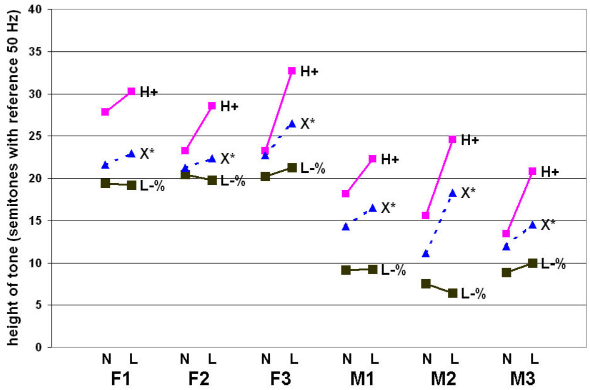
T-tests revealed that both H tones (H+ and H*) are significantly higher in lively than in neutral speech. By contrast, L-% tones were either significantly lower (speakers F2, M2), higher (F3, M3) or not significantly different (F1, M1). Similarly, L* tones were either significantly lower (F2), higher (M2, M3) or not significantly different (F1, F3, M1); see Table 2.

The difference between neutral and lively values for H tones ranged from 2.5 to 10 semitones, whereas in the few cases where L tones were higher in lively speech, the differences were very small, ranging from 0.1 to 1 semitone (except for speaker M3, who had a difference as high as 2.6 semitones for L*).

Having discussed the behaviour of H and L tones in neutral and lively speech, we now turn to the controversial tone, X*. This tone was always significantly higher in lively speech (p<0.001), the differences ranging from 1.1 to 7.2 semitones, with an overall mean of 3 semitones. Unlike the L tones, which were rather variable in their scaling, the raising of its value was consistent, similar to H+ and H*.

Fig.3 shows how X* is scaled in neutral and lively speech in comparison to H+ and L-% of the same phrase. These two tones define the upper and lower bounds of the local pitch range [12].

Figure 3: F0 values in semitones (reference 50 Hz) for H+, X* and L-% occurring within the same phrase, neutral (N) and lively (L) renditions.



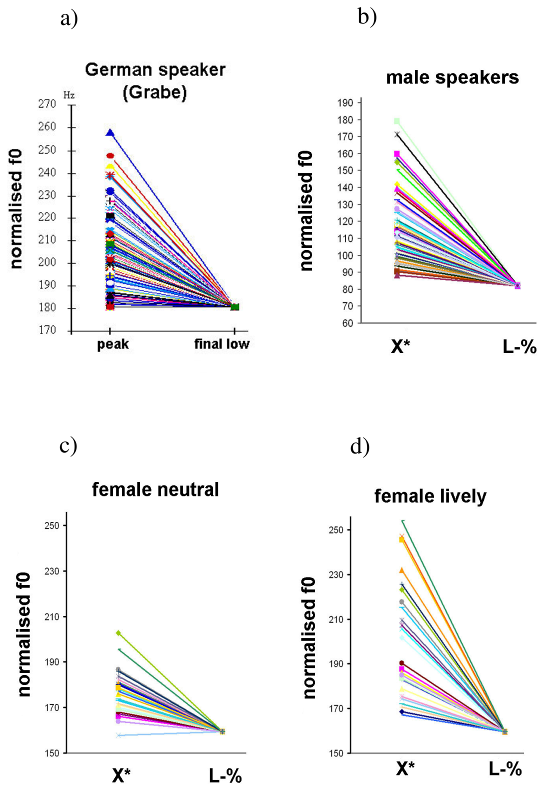
Speaker M3 not only increased the span but also raised his pitch level, as did F3, albeit by a small margin. The remaining four speakers increased their pitch span without raising the pitch level. These speakers are ideal subjects for investigating tonal identity. In their speech, the difference between neutral and lively values for X* constitute a clear increase, in some cases even parallel to the H+ values, which provides a strong argument for considering it as a H tone.

3.3 Variation in scaling of downstepped tones

Although for four speakers (F3, M1, M2, M3) the values for X* were towards the middle of the range (40%) between H+ and the final low (as would be expected if the tones are downstepped), for the remaining two speakers (F1, F2), the scaling of X* is relatively low in the range (28%). This variation in scaling is what Grabe [4] found in German !H*+L accents (accents without an early peak which are downstepped relative to a previous accent): As opposed to English, where the fall from a final downstepped peak to the final low was relatively large, the German contours exhibited a 'gradiently varying continuum of excursion for final falls' [4:198]. Fig. 4(a) illustrates this continuum in terms of normalised F0 values for the downstepped tone and the final low, and (b) shows that the distribution for our male speakers is strikingly similar. The F0 values were normalised as in [4]: the final low was set to the mean for all speakers for L-%, and the values for X* were calculated in relation to that value. In our data male and female speakers were treated separately.

Fig. 4(c) and (d) show that although there are low values for X in neutral speech, the values are generally higher and more diverse in lively speech. If X had been a L tone, we would have expected fewer high values.

Figure 4: Variation in the scaling of tones in relation to the final low tone: (a) !H* reproduced from [4], (b) X* for male speakers, (c) X* for female speakers in neutral speech and (d) for female speakers in lively speech.



4. CONCLUSIONS

We have shown that liveliness involves an increase in pitch span without an overall raising of the pitch range in all but two of our speakers. For those speakers who increase the span in this way, we have found clear evidence for a raising of the values of X* in the H+X* early peak accent, suggesting an analysis as H+!H* - an early peak with a downstepped starred tone. Although variation in the F0 values for X* (including some low values for two female speakers) might at first call this analysis into question, our results are in line with those from an independent study on downstep in German, where this type of variation was also found [4].

In terms of methodology, we show that scaling within a neutral utterance is not always enough to inform us as to tonal identity. Only if we observe how this tone is scaled when changes are made to the pitch span do we obtain a clear indication as to whether a tone is phonologically high or low.

5. REFERENCES

[1] Baumann, S. 2006. *The Intonation of Givenness - Evidence from German.* Tübingen: Niemeyer.
[2] Baumann, S., Grice, M. 2006. The Intonation of Accessibility. *Journal of Pragmatics* 38 (10), 1636-1657.
[3] Beckman, M., Hirschberg, J., Shattuck-Hufnagel, S. 2005. The Original ToBI System and the Evolution of the ToBI Framework. In: Jun, S. (ed), *Prosodic Typology: The Phonology of Intonation and Phrasing*. Oxford: OUP, 9-54.
[4] Grabe, E. 1998. *Comparative Intonational Phonology: English and German.* MPI Series in Psycholinguistics 7, Wageningen, Ponsen en Looien.
[5] Grice, M., Baumann, S., Benzmüller, R. 2005. German Intonation in Autosegmental-Metrical Phonology. In: Jun, S. (ed), *Prosodic Typology: The Phonology of Intonation and Phrasing*. Oxford: OUP, 55-83.
[6] Gussenhoven, C., Rietveld, T. 2000. The behavior of H* and L* under variations in pitch range in Dutch rising contours. *Language and Speech* 43, 183-203.
[7] Ladd, D.R. 1996. *Intonational Phonology*. Cambridge: Cambridge University Press.
[8] Nolan, F. 2003. Intonational equivalence: an experimental evaluation of pitch scales. In: Sole, M.J., Recasens, D., Romero, J. (eds), *Proc. 15th ICPhS* Barcelona, 771-774.
[9] Patterson, D., Ladd, D.R. 1999. Pitch range modelling: Linguistic dimensions of variation. *Proc. 14th ICPhS* San Francisco, 1169-1172.
[10] Pierrehumbert, J. 1980. The Phonetics and Phonology of English Intonation. PhD thesis, MIT. Bloomington: Indiana University Linguistics Club.
[11] Rathcke, T., Harrington, J. 2006. Is there a distinction between H+L* and H+!H* in Standard German? Evidence from an acoustic and auditory analysis. *Proc. Speech Prosody 2006* Dresden.
[12] van den Berg, R., Gussenhoven, C., Rietveld, T. 1992. Downstep in Dutch: Implications for a model. In: Docherty, G.J., Ladd, D.R. (eds), *Papers in laboratory phonology II*, 335-359.

Acknowledgements: We would like to thank Stefanie Shattuck-Hufnagel for suggesting looking at early peak accents with expanded pitch range.