ARE THERE "SHAPERS" AND "ALIGNERS"? INDIVIDUAL DIFFERENCES IN SIGNALLING PITCH ACCENT CATEGORY

Oliver Niebuhr[a], Mariapaola D'Imperio[b], Barbara Gili Fivela[c], Francesco Cangemi[b]

Dept. of General and Comparative Linguistics, University of Kiel[a]; Aix-Marseiile I & Laboratoire Parole et Langage, CNRS, Aix-en-Provence, France[b]; University of Salento & CRIL, Lecce, Italy[c]
niebuhr@linguistik.uni-kiel.de; mariapaola.dimperio@lpl-aix.fr; barbara.gili@unisalento.it; francesco.cangemi@lpl-aix.fr

ABSTRACT

One of the most studied phonetic dimensions differentiating pitch accent categories is tonal alignment. However, when acoustic alignment data are broken down by individual patterns, one can notice great differences both in the size of the contrast as well as in the dispersion of the temporal values. Moreover, speakers appear to compensate for weaker alignment differences by enhancing shape differences in the contour, such as rise/fall slope. Here we present production data in two languages, German and Italian, for well-established pitch accent contrasts. Our data suggest that the speakers' behavior can be represented on a continuum from most extreme "aligners" to "shapers".

Keywords: Tonal alignment, contour shape, range, pitch accent, F0, German, Italian, intonation.

1. INTRODUCTION

It is well known that the elements of spoken language communication (from vowels and consonants to elements of prosodic structure, such as stress and phrasing) are generally encoded by multiple acoustic parameters. For instance, as regards the distinction between voiceless and voiced stops in English, [1] set up a list of 16 acoustic cues, ranging from VOT to first-formant transition. In most of such lists the parameters were given a hierarchical order, i.e. some parameters are thought to occur more reliably and hence function as more powerful perceptual cues than others [2,3]. While it has always been clear that hierarchies and interactions of cues differ between languages, a growing body of evidence also questions that acoustic cues are used homogeneously within a language, cf. [4] for stop voicing in English. That is, individual speakers of the *same* language seem to employ acoustic parameters in very different ways, which are due to different production strategies.

Individual differences appear to be true also when it comes to prosodic events. For example, [5] emphasize that the realization of phrasal prominence not only differs clearly between Bulgarian and Russian, but also between the speakers of each of the two languages. However, studies on the production of pitch-accent contrasts conducted within the last 10-15 years have mainly focused on the use of a single parameter, i.e. the *temporal alignment* of fundamental frequency (F0) peaks and valleys with the segmental string, see [6] for a review. The widespread focus on tonal alignment had been first motivated by perception experiments in which shifting peaks (or pitch movements) backward or forward within the stressed syllable caused a change in pitch accent category, e.g., [7,8,9,10]. Later, in the realm of production, the hypothesis of *strict segmental anchoring* [11], that is that certain tonal targets would be synchronized with segmental boundaries irrespective of rate changes and/or syllabic structure differences, has further reinforced the alignment-oriented perspective on pitch accents and spurred interest in testing the relevance of various potential segmental references for tonal alignment (e.g., stressed vowel or syllable onset). Note that other relevant tonal alignment measures have been proposed in earlier work [12], such as proportional alignment relative to stressed vowel duration.

Despite a considerable amount of work, it is still highly controversial, which of these measures or "anchors" might be the most relevant one. Against this background, more recent work on a number of different languages has challenged the idea that the alignment of a static tonal target (either the onset or the offset of a pitch movement) might be the only relevant pitch-accent dimension and hence sufficient for understanding the production and perception of pitch accent categories. Specifically, it has been suggested that the *shape* characteristics of rises and falls, whose onsets and offsets share the same relative alignment to the segmental string, can determine pitch accent identification [8,9,13,14]. In some of these studies, the duration of a peak target (within rising-falling configurations) was manipulated, so that a high short plateau would obtain. In other studies, slope, duration and curvature of the rise or the fall were manipulated, keeping peak alignment constant. In [16] a tonal center-of-gravity approach has been proposed in order to explain the perceptual shape effects and their cross-linguistic similarities.

To sum up, since static temporal features of tonal targets are not the only relevant dimension in pitch accent identification, the question arises as to which production strategies might be adopted by speakers when it comes to signaling a contrast between two pitch accent categories. If the multiplicity of pitch-accent cues is, like in the case of segments, reflected in trading relations [2], we expect to find both language and speaker-specific patterns in the use of tonal alignment for distinguishing two pitch accent categories. Specifically, we expect subjects to differ on a continuum from “aligners” to “shapers”. This assumption implies simultaneously that the segmental anchoring of tones might primarily be a statistical phenomenon, in that quite large standard deviations for the mean anchor points are usually found which are due to both inter and intra-individual differences.

2. METHOD

Our cross-linguistic acoustic study was based on well-established pitch accent contrasts. These were H+L* vs. H* in Standard Northern German [7,8], L+H* vs. L*+H in Neapolitan Italian and H* vs. H*+L in Pisa Italian [9,13,14]. The pitch accent categories were elicited in a reading task. The text material of each language variety included 10 short A-B dialogues. Within each dialogue, A’s turn provided a semantic-pragmatic context frame that governed the production of the respective target accent in B’s response. Through this reliable context-based elicitation method (cf. [15]) 5 dialogues aimed at eliciting the first and the other 5 at eliciting the second accent category. Moreover, B’s responses were designed in such a way that the target accents were realized in constant segmental and prosodic environments in all three languages.

Recordings were made in sound-treated rooms at the Universities of Kiel, Naples, and Pisa. The German dialogues were read by 35 speakers; the Italian recordings included 17 Neapolitan and 20 Pisa speakers. All speakers (aged 20-35) were recruited from the student pools of the respective Universities and were roughly balanced for gender. The dialogues were read with an informal speaking style in a face-to-face set up, and repeated three times in a randomized order.

The elicited nuclear pitch accent categories were manually labeled in Praat with regard to (a) the segment boundaries of the accented word and (b) the three F0 landmarks rise onset, peak maximum, fall offset. In the case of plateaux, the peak maximum was marked at the fall onset, cf. [9,13]. The following measures were hence extracted: syllable duration, range and duration of the rise, range and duration of the fall, peak-maximum alignment relative to the segment boundaries of the accented syllable, e.g., vowel and syllable onset. Regarding [15], we also determined the shape index, defined as rise duration divided by fall duration.

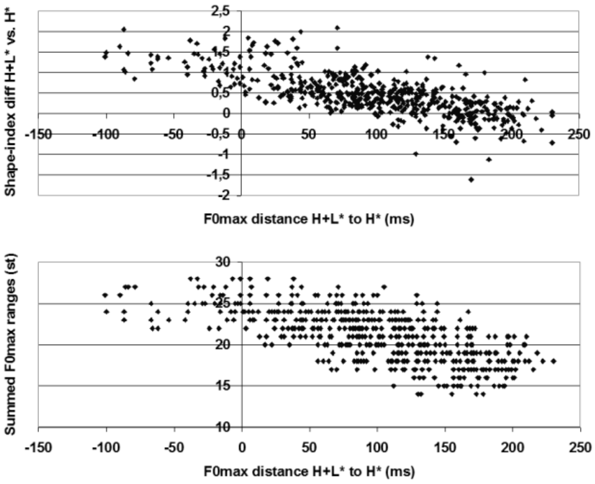
3. RESULTS

3.1. German

When viewed superficially in terms of overall mean values, the results for German simply replicate previous findings on accented syllables with long vowels and voiced codas [7,17]. That is, the F0 peak maximum of H+L* was aligned 29 ms (s=35 ms) before the accented-vowel onset. The H* accent peaks were about 90 ms later, i.e. the peak occurred 68 ms (s=38 ms) after vowel onset, approximately at the vowel midpoint. In addition to this clear alignment difference, H* had almost a symmetrical peak shape with a mean shape index of 0.83 (s=0.25), whereas H+L* was characterized by a slowly rising and fast falling contour shape [10], reflected by an index value of 1.31 (s=0.33). In terms of F0 range, H+L* and H* yielded similar means of 10.5 st (s=1.8 st) or 10.6 st (s=1.7 st).

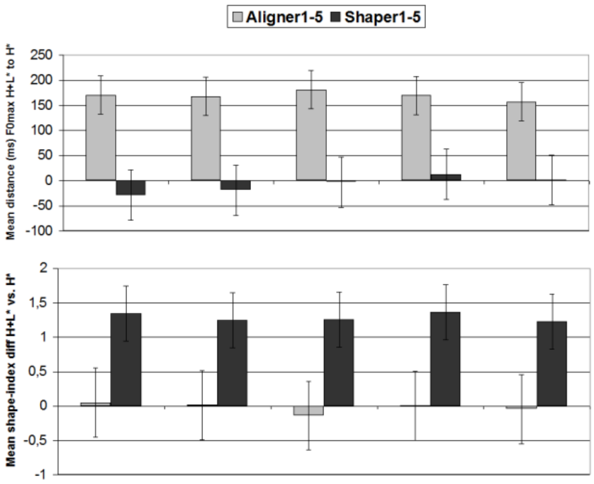
However, when broken down into the individual patterns, this clearly bimodal production picture is blurred. Firstly, the alignment differences of H+L* and H* relative to the accented-vowel onset appear to form an alignment continuum, in which the F0 maxima of the two pitch accents clearly overlap. Fig.1 shows the temporal differences between pairs of H+L* and H* pitch accent peaks (a pair consists of two accents produced by the same speaker in a pair of opposed elicitation contexts). Note that H* peaks can occur more than 100 ms earlier than H+L* peaks. Moreover, the pitch-accent alignment difference is highly negatively correlated ($r=-0.72$, $p<0.001$) with the shape difference for the two accents. That is, the less H+L* and H* differ in alignment, the more they differ in contour shape. The increasing shape differences reflect that both H+L* and H* peaks become increasingly asymmetrical, though in opposite directions. H* peaks rise faster and fall slower, whereas H+L* peaks primarily rise slower and can fall even faster. Fig.1 also shows that there is an additional negative correlation between the alignment differences and the pitch ranges of the two accents (r= -0.67, p<0.001). The summed ranges of the two F0 peaks increase with decreasing (and inverted) alignment differences. This additional correlation is consistent with the previous one, as it was found that an extended peak range enhances shape differences perceptually [8].

Figure 1: F0 peak alignment differences (ms) between H+L* and H* pairs and their negative correlations with shape-index differences (top) and the F0 peak ranges (st) of the accent pairs (bottom). n=525.



Furthermore, Fig.2 uncovers speaker-specific production strategies. In particular, among our 35 German speakers, 5 speakers produced both H+L* and H* around the vowel onset, with no peak alignment difference but with significant shape differences. Another set of 5 speakers used a diametrically opposed production strategy. Their H+L* and H* accents were clearly separable along the alignment dimension, but had very similar peak shapes. T tests that compared the accent productions of the two groups of speakers supported our groupings. The production strategies differ highly significantly between the groups in terms of alignment (t=28.52; *df*=115; p<0.001) and peak shape (t=-27.80; *df*=115; p<0.001). On this basis, we may call the first group of speakers *"shapers"* and the second group of speakers *"aligners"*. The remaining 25 speakers showed an intermediate production behaviour. Though the source of the variability is yet not clear, note that 4 out of the 5 aligners were male and 3 out of the 5 shapers were female speakers. Note also that the produced shape and alignment patterns match very well with findings of the perceptual identification of H+L* and H* in German [8].

Figure 2: Mean alignment differences (ms, top) and mean shape-index differences (bottom) between two groups of 5 speakers, classified as "aligners" (light bars) and shapers (dark bars). Each bar: n=15.

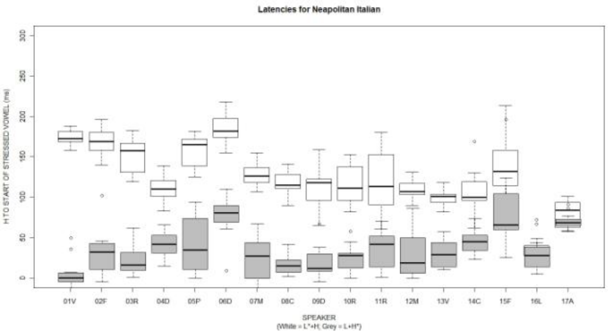


3.2. Italian

First of all, the results of Neapolitan and Pisa Italian show a crucial similarity to those of German: For each of the two language varieties we found considerable inter-speaker differences in the distinction of the pitch-accent contrasts by means of F0 peak (i.e. H) alignment. Moreover, the Neapolitan data parallel the German data also in quantitative terms. Specifically, for Neapolitan Fig.3 shows (from left to right) that more than half of the speakers produced the two accent categories L+H* and L*+H with discrete peak alignment (relative to the accented-vowel onset), whereas a set of 6 speakers made little or no use of the alignment parameter. Preliminary analyses suggest that this subgroup of speakers realized the two accents with greater shape differences than the other group of speakers. Like in German, the overall production pattern seems to include strategies that range from aligners to shapers, with a larger group of speakers who used both types of cues to different extents.

In the case of Pisa Italian, peak alignment had to be normalized relative to the accented-syllable duration, which differed clearly between H* and H+L* (on average 172 ms vs. 244 ms). On this normalized basis, the pitch-accent contrast between H* and H*+L is consistently reflected in discrete peak alignments. However, the mean differences vary across speakers from only 33% to more than 95% (rel. to syllable duration). A further parallel to Neapolitan and German is that the use of alignment is linked with the use of shape differences.

Figure 3: Alignment differences between L+H* (bottom row) and L*+H (top row) shown as separate box plots (ms, relative to vowel onset) for each of the 20 Neapolitan speakers (n=15).



However, unlike in Neapolitan and German, the linkage between the two types of accent parameters is not a trading relation in the sense of 'either-or'. Rather, the relation may be described as 'less-or-more'. That is, those speakers who made the greatest alignment differences additionally distinguished the two pitch accents by means of shape characteristics. Comparing in the Pisa variety the 5 strongest with the 5 weakest aligners in separate ANOVAs for each pitch accent category showed that the strong aligners also created significantly larger, longer and steeper F0 rises for H* ($F_{[1,297]}$=12,066; p=0.001; $F_{[1,297]}$=6,365; p= 0.01; $F_{[1,149]}$=11.210; p=0.001) as well as larger, shorter and steeper F0 falls for H*+L ($F_{[1,149]}$= 4.802; p=0.03; $F_{[1,149]}$=5.766; p=0.01; $F_{[1,149]}$= 4.694; p=0.03). Since the difference between the "alignment-only" and "alignment+shape" strategies concerns both H* and H*+L, it cannot be an artefact of emphasis, as the latter would only be added to the contrastive accent H*+L.

4. DISCUSSION AND CONCLUSIONS

At the level of perception, we already know that tonal alignment is not the only cue to pitch accent identification, and that there are a number of contour shape-related cues. In our study, we provided initial cross-linguistic evidence that this multiplicity of cues is reflected in speaker-specific production strategies with "aligners" and "shapers" at the extreme ends. It will be the task of follow-up studies to investigate, if the production strategies of the Pisan speakers differed from those of German and Neapolitan for functional or structural reasons, and if the production strategies are correlated with speaker attributes. For example, it seems from our data that "shapers" are more likely female than male speakers. Over and above these questions, our study demonstrates that the line of research on the production and perception of intonational cues must be complemented by a closer look at the behavioural patterns of individual subjects.

5. REFERENCES

[1] Lisker, L. 1986. "Voicing in English": A catalogue of acoustic features signaling /b/ versus /p/ in trochees. *Language and Speech* 29, 3-11

[2] Repp, B.H. 1982. Phonetic trading relations and context effects: New experimental evidence for a speech mode of perception. *Psychological Bulletin* 92, 81-110.

[3] Clayards, M., Tanenhaus, M.K., Aslin, R.N., Jacobs, R. A. 2008. Perception of speech reflects optimal use of probabilistic speech cues. *Cognition* 108, 804–809.

[4] Allen, J.S., Miller, J.L., DeSteno, D. 2003. Individual talker differences in VOT. *JASA* 113, 544-552.

[5] Andreeva, B. Barry, W.J. 2007. Cross-language and individual differences in the production of phrasal prominence in Bulgarian and Russian. *Proc. 7th European Conf. on Formal Description of Slavic Languages*, Leipzig.

[6] D'Imperio, M. to appear. Prosodic representations: tonal alignment". *Oxford Handbook of Laboratory Phonology.* (ch. 9). Oxford: OUP Press.

[7] Kohler, K. 1987. Categorical pitch perception. *Proc 11th ICPhS*, Tallinn, 331-333.

[8] Niebuhr, O. 2007. The signalling of German rising-falling intonation categories - The interplay of synchronization, shape, and height. *Phonetica* 64, 174-193.

[9] D'Imperio, M. 2000. *The role of perception in defining tonal targets and their alignment*. PhD thesis, Ohio State University.

[10] Redi, L. 2003. Categorical effects in the production of pitch contours in English. *Proc. 15th ICPhS*, Barcelona, 2921–2924.

[11] Ladd, D.R., Faulkner, R., Faulkner, H., Schepman, A. 1999. Constant „segmental anchoring" of F0 movements under changes in speech rate. *JASA* 106, 1543-1554.

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

[13] Gili Fivela B., D'Imperio, M. 2010. High peaks versus high plateaux in the identification of two pitch accents in Pisa Italian. *Proc. 5th Speech Prosody*, Chicago.

[14] Petrone, C. 2008. *Le rôle de la variabilité phonétique dans la représentation des contours intonatifs et de leur sense. PhD thesis*, Université de Provence, France.

[15] Niebuhr, O. 2010. On the phonetics of intensifying emphasis in German. *Phonetica* 67, 170-198.

[16] Barnes, J., Veilleux, N., Brugos, A., Shattuck-Hufnagel, S. 2010. The effect of global F0 contour shape on the perception of tonal timing contrasts in American English intonation. *Proc. 5th Speech Prosody*, Chicago.

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