THE INTERACTION OF FUNDAMENTAL FREQUENCY AND INTENSITY IN THE PERCEPTION OF INTONATION

K. J. Kohler

Institut für Phonetik und digitale Sprachverarbeitung
Kiel, Germany

ABSTRACT

The temporal alignments of three terminal FO peaks (early, medial, late) with stressed syllables, the parallelism of FO and intensity timing in these patterns, and the importance of intensity in pitch accent signalling are discussed for German.

1. FO PEAK POSITIONS IN TERMINAL INTONATION

In (4,5), I have shown that terminal intonation contours in German can have three different, specific meaning related types of FO peak positions around one and the same stressed vowel: (1) the peak may be early, before the stressed vowel, which only gets an FO fall (early peak), (2) the peak may be in the centre of the stressed vowel, which therefore has an FO rise and an FO fall (medial peak), (3) the peak may follow a stretch of low FO in the stressed vowel and therefore not occur until the second half or even the beginning of a subsequent unstressed syllable (late peak), which means that the FO rise dominates the stressed vowel and the FO fall is not always realised in it.

The early peak differs categorically from medial and late ones by only having a falling FO during the stressed vowel, thus accentuating the lower pitch range directly, but the utterance sounds postponed and thus the high FO level extended (Figs. 1a, b).

Although the positioning of FO peaks contributes to the perception of stressed syllables, this FO feature is not the only factor. Durations of vowels and postvocalic consonants are also important cues, particularly inside hat patterns, where the FO movements are minimal. Similarly, in a hat pattern uniting two abutting stressed syllables, as in 'Der Ring glänzt.' (The ring glitters.), with a late peak rise on the first and an early peak fall on the second, the segment durations in the second stressed syllable as well as the FO timings are important for it to be perceived as stressed and thus differentiated from a single stress with late peak on the first syllable only (Figs. 1b, c).

In early and medial peaks, the FO fall, which is not possible at the beginning of a hat, and non-early ones can only be signalled initially, if in the initial position of a hat the FO fall is shifted further and further into the stressed vowel from an early via a medial to a late position, this shift lacks the change-over from fall to rise, because the preceding syllables are not lower in FO. Similarly, if in the initial position of a hat the FO rise is shifted further and further to the left from a late via a medial to an early position, this shift lacks the change-over from rise to fall because the subsequent syllables do not have a dip in FO. In both cases we get continua of fall and rise timings, respectively, and the concomitant perception is equally continuous. Because of this, the early peak is the most natural FO pattern at the end of a hat. It also accentuates the contrast between the low FO in the stressed vowel and the high FO level preceding it, thus adding to stress perception, which is weakened if the FO fall is postponed and thus the high FO level extended (Figs. 1a, b).

2. FO AND INTENSITY TIMING

The precise FO timing of terminal peak contours not only depends on the peak type but also on the segmental structure of the stressed syllable. In final stressed peaks, the left-hand base point occurs at the beginning of the first consonant preceding the stressed vowel, the peak point at a time after vowel onset determined by the quantity and quality of the vowel, and the right-hand base point some 150 ms after the peak point. In early peaks, the peak point is postponed where medial peaks have their left-hand peak base; the right-hand base point occurs at the end of a lax (short) or about the same as a tense (long) stressed vowel. In late peaks, the left-hand base point is positioned where medial peaks have their peak point, the stretch from the syllable beginning being low and descending slightly; the rise to the peak point then occurs within about 100 ms, after which we get a descent to the right-hand base point in another approx. 100 ms. The point, the stretch from the syllable beginning being low and descending slightly; the rise to the peak point then occurs within about 100 ms, after which we get a descent to the right-hand base point in another approx. 100 ms. The point, the stretch from the syllable beginning being low and descending slightly; the rise to the peak point then occurs within about 100 ms, after which we get a descent to the right-hand base point in another approx. 100 ms. The
husky at the end and overloaded in the middle because F0 and intensity are in opposite directions in these two places.

The loss of the particular characteristics of a peak pattern is illustrated by the synthesis of late peaks in an utterance-final word structure "stressed vowel + voiceless consonant + syllable nasal" as in 'Er ist ja geritten.' (He has been riding.). A voiceless consonant after a late-peak stressed vowel interrupts the F0 course; it can only be successfully reconstructed by a listener if, in addition to an indication of a fast F0 rise speed (of ca 0.5 Hz/ms), the onset of voicing following the voiceless consonant receives the F0 peak and if the F0 descent from this value to the terminal low level can be clearly perceived. This means that the source amplitude must be high enough to guarantee sufficient intensity in the final nasal for the high falling F0 contour to be audible. If a natural medial peak speech signal with its low final intensity in the above utterance is taken for LPC resynthesis with a late peak, positioned at the nasal onset, the percept lacks the significant attributes of the late peak, because the intensity of the final nasal is too low and the F0 contour, therefore, not perceivable. Contrariwise, in a RULSYS TTS formant synthesis-by-rule of the above sentence [1], a reduction of the voice source A0 from 20 dB to 12 dB and of the nasal source from 30 dB to 10 dB in the final \( /n/ \) within a late peak (fig. 2) results in a loss of the perceptual late peak feature.

3. THE IMPORTANCE OF INTENSITY IN ACCENT SIGNALLING

The question now arises as to whether it is possible to change stress perception simply by varying intensity. Obvious instances for testing this hypothesis are utterances that are ambiguous with regard to containing one or two stresses. When a late F0 rise is immediately followed by a medial F0 fall without an intervening F0 dip in two abutting stressed syllables, (fig. 1a) the second stress is weakened. If intensity alone can change stress perception, then it should be possible in a case like this to produce a switch in focus to initial sentence stress simply by reducing the intensity in the second accent and by simultaneously raising it in the first.

This has been interactively tested by changing the A0 values accordingly in the RULSYS TTS synthesis-by-rule. The result has been negative: the focussing and consequently the number of stresses, does not change; it is more the loudness relations that are affected. This is further support to the long-established finding that intensity has a low signalling value for stress compared with F0 and duration [2].

4. REFERENCES