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 F0 peaks (early, medial, late) with stressed syllables, the parallelism of F0 and intensity timing in these patterns, and the importance of intensity in pitch accent signalling are discussed for German.

1. F0 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 F0 peak positions around one and the same stressed vowel: (1) the peak may be early, before the stressed vowel, which only gets an F0 fall (early peak), (2) the peak may be in the centre of the stressed vowel, which therefore has an F0 rise and an F0 fall (medial peak), (3) the peak may follow a stretch of low F0 in the stressed vowel and therefore not occur until its second half or even the beginning of a subsequent unstressed syllable (late peak), which means that the F0 rise dominates the stressed vowel and the F0 fall is not always realised in it.

The early peak differs categorically from medial and late ones by only having a falling F0 during the stressed vowel, thus accentuating the lower pitch range compared with the other two patterns. This categorical difference in the acoustic manifestation of early vs. non-early peaks is parallelled by a categorical change in perception along a peak position continuum from early to medial and by a continuous one from medial to late [3]. This means that for the signalling of an early versus a non-early peak a simple F0 fall as against the presence of an F0 rise is essential.

It follows from this that in the concatenation of F0 peaks without valleys between them ('hat patterns') [5], early peaks are not possible at the beginning of a hat, and non-early ones can only be signalled initially. If in the final position of a hat the F0 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 F0. Similarly, if in the initial position of a hat the F0 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 F0. 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 F0 pattern at the end of a hat. It also accentuates the contrast between the low F0 in the stressed vowel and the high F0 level preceding it, thus adding to stress perception, which is weakened if the F0 fall is postponed and thus the high F0 level extended (figs. 1a, b).

Although the positioning of F0 peaks contributes to the perception of stressed syllables, this F0 feature is not the only factor. Durations of vowels and postvocalic consonants are also important cues, particularly inside hat patterns, where the F0 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 F0 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 these cases we may ask to what extent intensity contributes to stress perception and whether changing it can alter the interpretation between one and two stresses.

2. F0 AND INTENSITY TIMING

The precise F0 timing of terminal peak contours not only depends on the peak type but also on the segmental structure of the stressed syllable. In medial 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 positioned where medial peaks have their left-hand base point; the right-hand base point occurs at the end of a lax (short) or about the centre of 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 appr. 100 ms. To accommodate these F0 time courses in late peaks the stressed vowels are lengthened after the left base point, more so for lax than for tense vowels, more in final monosyllables than elsewhere. If voiceless consonants intervene between a lax stressed late peak vowel and a following unstressed syllable the target peak value cannot be reached in the stressed vowel itself, but is needed for pattern identification and therefore set at the voice onset of the following unstressed vowel.

In early and medial peaks, the low F0 fall at the end of an utterance is accompanied by a drop in source amplitude, which weakens unstressed vowels and sonorants considerably, often reducing them to creaky voice and to irregular breathy glottal pulses. In late peaks this decline is shifted to the right following the later F0 fall, thus keeping a high source amplitude at the onset of unstressed vowels and syllabic sonorants; on the other hand the low F0 stretch in the stressed vowel before the peak gets its intensity reduced. So there is a natural parallelism in the time courses of F0, source amplitude and sound intensity for the three terminal peak contours. If it is destroyed in synthesis the output sounds either degraded or the peak pattern loses its identity.

The first case occurs, when a natural medial peak speech signal is taken as a point of departure for LPC resynthesis with a late peak in a completely voiced environment, as in 'Sie hat ja gelogen.' (*She has been lying.*): the peak type is signalled correctly, but the utterance sounds husky at the end and overloaded in the middle because F0 and intensity diverge 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 plosive + syllabic nasal" as in 'Er ist ja geritten.' [... 'ɪtn̩] (*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 auditorily monitored. 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 foregoing shows that F0 and source amplitude are linked in production, and that their coupled time courses are expected by listeners. If the coupling is artificially destroyed in synthesis the perception is affected at the levels of voice quality and/or intonation. For pitch accents to be signalled effectively to a listener there has to be sufficient voice intensity in the signal. In the examples discussed so far, an intensity reduction was capable of affecting the identity of a pitch accent, but not its presence, i.e. the stress position remained unaltered.

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

[1] CARLSON, R., GRANSTRÖM, B. & HUNNICUTT, S. (1990), "Multi-language text-to-speech development and applications", in *"Advances in speech, hearing, and language processing"*, Vol. 1 (W.A. AINSWORTH, ed.), London: JAI Press), 269-296.
[2] FRY, D. B. (1958), "Experiments in the perception of stress", *Language and Speech, 1*, 126-152.
[3] KOHLER, K. J. (1987a), "Categorical pitch perception", in *"Proceedings of the XIth international congress of phonetic sciences"*, Vol. 3, pp. 149-152, Tallinn: Academy of Sciences of the Estonian SSR.
[4] KOHLER K. J. (1987b), "The linguistic functions of F0 peaks", in *"Proceedings of the XIth international congress of phonetic sciences"*, Vol. 3, pp. 149-152, Tallinn: Academy of Sciences of the Estonian SSR.
[5] KOHLER, K. J. (1991), "Prosody in speech synthesis", *Journal of Phonetics, 19*.

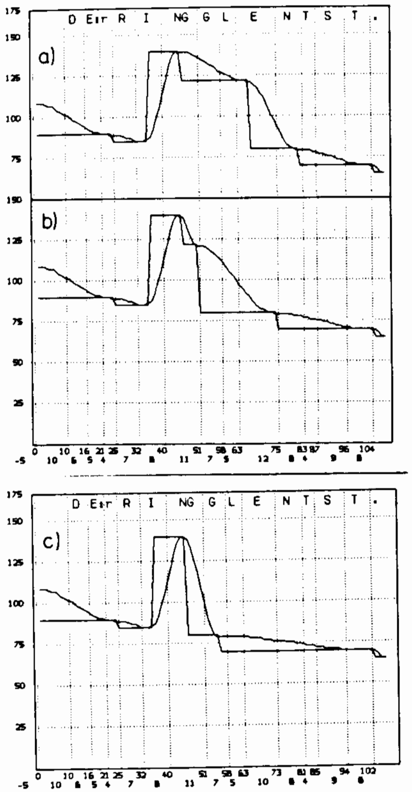

Fig. 1: Phonetic transcription and F0 (squares and cosine interpolations), in the German sentence 'Der Ring glänzt.' (RULSYS TTS); a) two stresses: hat pattern, late rise + medial fall, b) two stresses: hat pattern, late rise + early fall, c) one stress: late peak. Horizontal: cs frames (cumulative and for each segment), vertical: Hz.

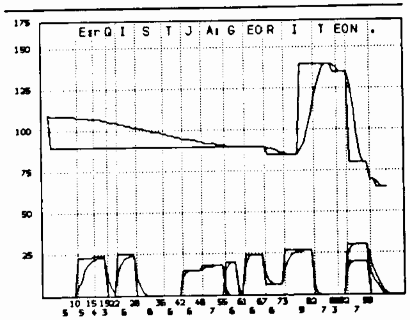

Fig. 2: Phonetic transcription, voice source A0 and nasal source AN (squares and cosine/2nd order interpolations) in the German sentence 'Er ist ja geritten'. with late peak (RULSYS TTS). Horizontal: cs frames (cumulative and for each segment), vertical: Hz for F0, dB for A0, AN.