The interaction of pitch range and temporal alignment in the perception of interrogative mode in Swedish

David House
Centre for Speech Technology, Department of Speech, Music and Hearing, KTH, Stockholm, Sweden
davidh@speech.kth.se

The temporal alignment of fundamental frequency excursions related to segmental and syllable boundaries is a topic of interest involving the perception and integration of short-domain cues (e.g. burst energy) and long-domain cues (relative pitch levels). In certain languages (e.g. Neapolitan Italian, D’Imperio & House 1997) the time alignment of a final accent has been shown to play a decisive role in the perception of interrogative mode. In Swedish, question intonation has been primarily described as marked by a raised topline and a widened F0 range on the focal accent (Gårding 1979). An optional terminal rise has been described, but time alignment of the focal accent rise has not been associated with question intonation. The purpose of this study is to investigate if a late peak in a final focal accent is sufficient to signal interrogative mode in Swedish when it is not expressed by syntactic or lexical means and also to test the perceptual interaction of temporal alignment and pitch range.

The test sentence, *Hon vill bara flyga*, “She only wants to fly,” was synthesized using an experimental version of the Infovox 330 diphone Swedish male voice with the final syllable bearing a focal accent peak. Final vowel duration was set to 150 ms. Two sets of six pitch-manipulated stimuli were created by systematically shifting the focal accent peak through the vowel in steps of 25 ms. In one set of stimuli the accent peaks were set at 130 Hz consistent with the F0 range of the entire sentence. In the second set of stimuli, the accent peaks were set at 160 Hz comprising a widened F0 range on the focal accent.

Results of the perception test indicate that both a widened F0 range on the focal accent and time alignment properties of the rise and peak make important contributions to the interrogative percept. For the low-pitch stimuli, the first four peak positions signaled declarative mode while the final two signaled interrogative mode. For the high-pitch stimuli, all peak positions except the first signaled interrogative mode. The response curve for the low-pitch stimuli set was sharper at the category boundary than for the high-pitch set. These results demonstrate sensitivity to time alignment on the order of 25 ms, and a perceptual interaction between pitch range and temporal alignment. In terms of a trading relationship the effect of a widened F0 range is equivalent to a 75 ms peak delay which results in a salient rise during the spectrally stable portion of the vowel in the low F0 range.