Umeå University, Department of Philosophy and Linguistics
PHONUM 9 (2003), *185-188*
Available online at http://www.ling.umu.se/fonetik2003/

Hesitation and interrogative Swedish intonation

David House

Centre for Speech Technology, Department of Speech, Music and Hearing, KTH

This paper presents the results of a perception experiment designed to investigate the contribution of a pre-focal hesitation pause, delayed focal peak and a raised F0 range for signaling interrogative mode in Swedish. The results support the hypothesis that a pause strengthens the interpretation of a focal peak delay as signaling question intonation. The results are discussed in terms of a framework of biological codes for universal meanings of intonation proposed by Gussenhoven.

1. Introduction

In Swedish, question intonation has been traditionally 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 the time alignment of the focal accent rise has not generally been associated with question intonation. In a recent study, however, it was demonstrated that a raised fundamental frequency (F0) combined with a rightwards focal peak displacement is an effective means of signaling question intonation in Swedish when the focal accent is in final position (House, 2002). Furthermore, the results indicated a trading relationship such that a raised F0 had the same perceptual effect as a peak delay of 50 to 75 ms.

This concept of "delayed peak" in which the peak comes very late in the associated syllable or even in the following syllable and results in differences in intonational meaning has received considerable research interest (see Ladd, 1996). The framework of biological codes for universal meanings of intonation, proposed by Gussenhoven (2002) provides an elegant theoretical explanation for how delayed peak can function as the same signal as raised F0. Gussenhoven proposes a frequency code, an effort code and a production code. The frequency code implies that a raised F0 is a marker of submissiveness or non-assertiveness and hence question intonation. The effort code implies that articulation effort is increased to highlight important focal information producing a higher F0. The production code associates high pitch with phrase beginnings (new topics) and low pitch with phrase endings. In this account, higher peaks take longer to reach than lower ones and thus come later in the syllable. Therefore listeners will associate a late peak with a higher pitch.

It is not uncommon in speech to find filled or silent pauses prior to a focal accent (Strangert, 1991). In terms of the three codes discussed above, a pause can be a correlate of the effort code where a build-up of effort is accompanied by a pause prior to an emphatically focused, semantically important content word. However, a pause can also be a hesitation pause which can be part of what we would propose as a fourth code, namely a cognitive code where cognitive loading results in a dysfluency. In this paper a perception experiment will be presented which tests three different cues to question intonation: raised F0, focal peak delay and a pre-focal filled hesitation pause.

2. Method

The test sentence, *Hon kan tänka sig åka bil*, "She can consider going by car," was synthesized using an experimental version of the Infovox 330 diphone Swedish male MBROLA voice implemented as a plug-in to the WaveSurfer speech tool (Sjölander & Beskow, 2000). The sentence was synthesized with the final syllable bearing a focal accent peak. Two sets of four pitch-manipulated stimuli were created by systematically shifting the focal accent peak through the vowel in steps of 50ms. A low-pitch set of stimuli with the accent peaks at 120Hz and a high-pitch set with the accent peaks at 140Hz were used. Two additional sets of stimuli were created by introducing a filled pause into each stimulus by lengthening the pre-focal [a] by 150 ms. The manipulated portions of the stimuli are presented schematically in Figure 1.

21 subjects participated in the experiment. All were native speakers of Swedish with the central Swedish (Stockholm) dialect predominating. Most of the subjects were students at KTH and participated in the experiment as part of a course requirement.

The experiment was conducted using an interactive computer-based program implementing a visual sort and rate method (VISOR) (Granqvist, 1996). In the program, the stimuli correspond to icons on the computer screen. The subject clicks on the icons to listen to the stimuli and moves the icons along a visual scale for sorting, rating and/or ranking. Subjects were instructed to listen to the stimuli and given the task of deciding if the speaker intended to make a statement or ask a question. Subjects were asked to place each icon in the appropriate horizontal field corresponding to statement/question and then to place the icons in position vertically corresponding to their sense of confidence as to their category decision. The results thus reflect the subjects' identification of the stimuli in terms of statement/question mode and ranking of the responses in terms of confidence.

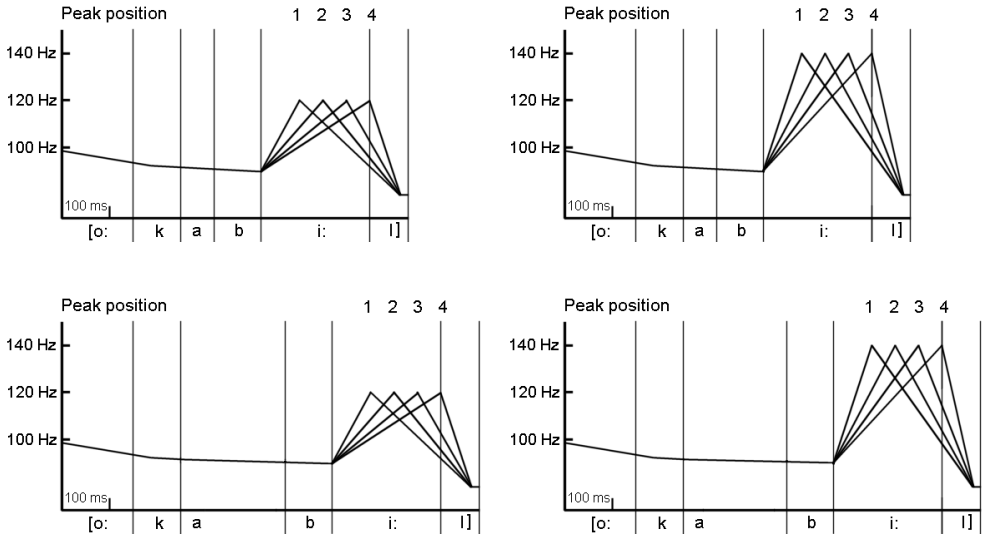

Figure 1. Schematic presentations of the stimuli used in the perception test. The upper two panels represent the low and high pitch configuration with no pause. The lower two panels represent the filled pause configuration.

3. Results

The results are presented in Figure 2 for each peak position and each of the four different conditions, i.e. high and low F0 with and without the filled pause. The results generally indicate that all three variables contribute to the perception of interrogative mode and show considerable interrelationships between the variables. Statement identification was greatest (100%) in the condition of low F0, early peak and no pause, while question identification was greatest (86%) in the condition of high F0, late peak and a filled pause. The effect of the pause is apparent in all conditions moving the responses toward question identification by about 20% except for high F0 at the final peak position which approaches a ceiling effect.

The influence of peak delay on question identification is seen in all four conditions and is greatest in the high F0 condition. Results for peak position are significant in all conditions except the low F0 with pause condition: single factor ANOVA for low F0/no pause $F(3,80)=3.26$, $p<0.05$; for low F0 with pause $F(3,80)=1.08$, $p=0.363$; for high F0/no pause $F(3,80)=16.53$, $p<0.001$; and for high F0 with pause $F(3,80)=4.00$, $p<0.05$.

The high F0 had little influence on question identification when the peak was early. For the late peak positions, however, high F0 was instrumental in eliciting question identification. The confidence response scores plotted in Figure 2 generally follow the identification scores and also reflect the general bias toward statement identification.

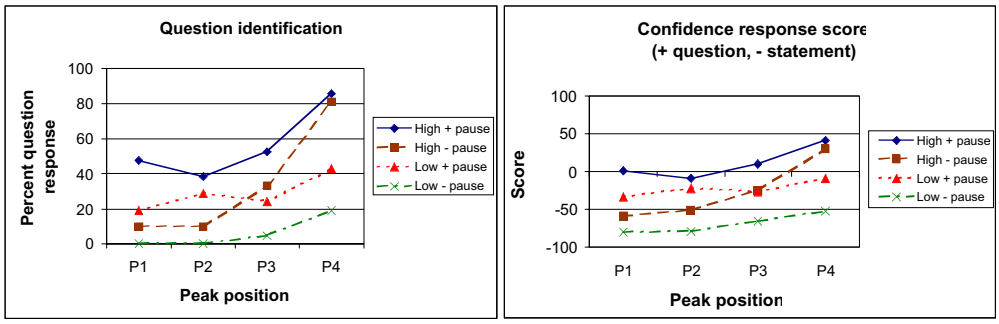

Figure 2. Results of the perception test showing percent question responses (left) and mean confidence response scores for each stimulus (right).

4. Discussion

It is evident from the results that although the combination of a higher F0 and a delayed peak on a final focal accent can alone cue interrogative mode, a pre-focal, filled pause can also contribute to the percept of question intonation. Since pauses can occur before focus as an additional signal of emphasis (Strangert 1991) we could have expected the pause to strengthen the statement percept for the early peak positions and the question percept for the late peak positions. However, this was not the case as the pause moved all responses toward the question percept. This indicates that in the context of the experiment, the filled pause was perceived as hesitation conveying uncertainty and non-assertiveness. In terms of the biological codes discussed above, hesitation has the same function as the frequency code, i.e. a high F0 indicates submissiveness and non-assertiveness. Although the function may be the same, the origins and mechanisms of these two interrogative signals are quite different. While the frequency code originates from anatomical variations in size of different speakers'

speech organs (e.g. child, female and male), hesitation originates from behavioral differences. It is reasonable to conjecture that the perceptual mechanisms for the hesitation code are on a higher cognitive level than are those for the frequency code.

Gussenhoven's proposal presented in the introduction above for mechanisms of substitute variables whereby a peak delay can substitute for a raised peak would also be an example of a mechanism shift to a higher cognitive level involving behavior. According to this argument, listeners use their knowledge that a higher peak takes longer to reach than a lower one and therefore speakers and listeners can incorporate this into a kind of cognitive pitch code. The results presented in the current experiment are not inconsistent with such an interpretation if it is constrained to equating a perceived higher pitch with increased question responses. There are, however, some complications to this argument. Higher pitch is not only a result of the frequency code but also of the effort code which implies that articulation effort is increased to highlight important focal information producing a higher F0. In this experiment higher pitch also signals focal accent. Thus there is a conflict between the phonetic coding of question intonation and focal accent. This conflict can be resolved by the production code which associates high pitch with phrase beginnings (new topics) and low pitch with phrase endings. Thus phrase-final high pitch is the marked case which signals continuation. Here we would argue that interrogative mode often requires a continuation marker which signals an expected response on the same topic. The rightwards shift of the focal peak is not merely a substitute for high pitch in the frequency code but is actually a part of the production code. The three codes therefore work together to differentiate between declarative and interrogative focal accent with peak delay playing an important role.

In conclusion, the results presented here indicate a strong interrelationship between pitch height and peak timing in a final focal accent for signaling echo question intonation in Swedish. Furthermore, a filled, pre-focal hesitation pause contributes to the interrogative percept. It is hypothesized that this mechanism functions on a higher cognitive level where hesitation is perceived as cognitive loading and interpreted as uncertainty.

5. Acknowledgements

This research was in part supported by VINNOVA (The Swedish Agency for Innovation Systems), and by VR (The Swedish Research Council).

6. References

Granqvist, S. (1996) Enhancements to the visual analogue scale, VAS, for listening tests, *Speech, Music and Hearing. TMH QPSR 4/1996*, 61-65, KTH Stockholm.

Gussenhoven, C. (2002) Intonation and interpretation: phonetics and phonology. In *Proceedings of the Speech Prosody 2002 Conference*, pp. 47-57. Aix-en-Provence.

Gårding, E. (1979) Sentence Intonation in Swedish, *Phonetica* 36, 207-215.

House, D. (2002) Intonational and visual cues in the perception of interrogative mode in Swedish, In *Proceedings of ICSLP 2002*, pp. 1957-1960. Denver, Colorado.

Ladd, D. R. (1996) *Intonational phonology.* Cambridge: Cambridge University Press.

Sjölander, K. & Beskow, J. (2000) WaveSurfer - an open source speech tool. In *Proceedings of ICSLP 2000*, vol. 4, pp. 464-467, Beijing, China.

Strangert, E. (1991) Phonetic characteristics of professional news reading. *Papers from the fifth national phonetics conference, PERILUS XIII*, pp. 39-42, Stockholm University.