

Perceiving question intonation: the role of pre-focal pause and delayed focal peak

David House

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

KTH, Stockholm, Sweden

E-mail: davidh@speech.kth.se

ABSTRACT

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 are consistent with a previous study which demonstrated that a delayed peak combined with a raised F0 range can effectively signal interrogative mode in Swedish echo questions. In addition, the present study supports the hypothesis that the presence of a pre-focal hesitation 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. An additional biological code is proposed on a behavioral level where cognitive loading results in a hesitation pause or dysfluency which is perceived as non-assertiveness.

1. INTRODUCTION

The signaling of interrogative mode in speech through intonation is a topic which has long attracted interest from intonation researchers. Question intonation, however, has remained somewhat elusive in both descriptive phonetics and intonation models. Not only does question intonation vary in different languages but also different types of questions (e.g. wh, yes/no or echo questions) result in different kinds of question intonation [1]. The most commonly described tonal characteristic for questions is high final pitch and overall higher pitch [2]. In some languages, however, e.g. Neapolitan Italian [3], a late 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 [4]. An optional terminal rise has been described, but the time alignment of the focal accent rise has not generally been associated with question intonation. Instead, a rightward shift of the focal accent peak has been associated with lending prominence to given domain-specific information in a dialogue context [5]. In a recent study, however, House [6], 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. Perception results confirmed the importance of timing where an early peak was perceived as a statement while a late peak was perceived as a question in a manner similar to that shown in Neapolitan Italian. Furthermore, there was a trading relationship between peak height and peak displacement so 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 [1]). The framework of biological codes for universal meanings of intonation, proposed by Gussenhoven [7] provides an elegant theoretical explanation for how delayed peak can function as the same signal as raised F0. Gussenhoven proposes three codes or biological metaphors: 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 [8]. 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 [9]. However, a pause can also be a hesitation pause [10] 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. The experiment was also designed to test the previous results on peak delay using an accent I word (the previous work used an accent II word) and further to test the hypothesis that the presence of a hesitation pause strengthens the interpretation of the focal peak delay as signaling question intonation in Swedish. Question intonation will also be discussed in the light of the theoretical framework presented above, and implications

for general mechanisms of tonal perception will be explored.

2 METHOD

2.1 Stimuli

The test sentence, *Hon kan tänka sig åka bil*, “She can consider going by car,” was synthesized using an experimental version of the Invox 330 diphone Swedish male MBROLA voice implemented as a plug-in to the WaveSurfer speech tool [11]. The test 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. The first peak was located 80ms into the stressed vowel in order to maintain the initial rising contour important for the identity of a focal accent I. The last peak in the stimuli continuum was located at the end of the stressed [i:] vowel. In the low-pitch set of stimuli the accent peaks were set at 120Hz consistent with the F0 range of the entire sentence. In the high-pitch set of stimuli, the accent peaks were set at 140Hz comprising a widened F0 range on the focal accent. 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. The peak position numbers correspond to the timing location of the peaks in both the low-pitch and high-pitch set with and without the pause.

2.2. Subjects

21 subjects participated in the experiment. 19 subjects were students at KTH and two were recruited outside KTH. All were native speakers of Swedish with the central Swedish (Stockholm) dialect predominating. None of the subjects reported any hearing loss. The students participated in the experiment as part of a course requirement.

2.3. Test procedure and task

The experiment was conducted using an interactive computer-based program implementing a visual sort and rate method (VISOR) [12]. 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. No specific instructions were given in terms of vertical scale. 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.

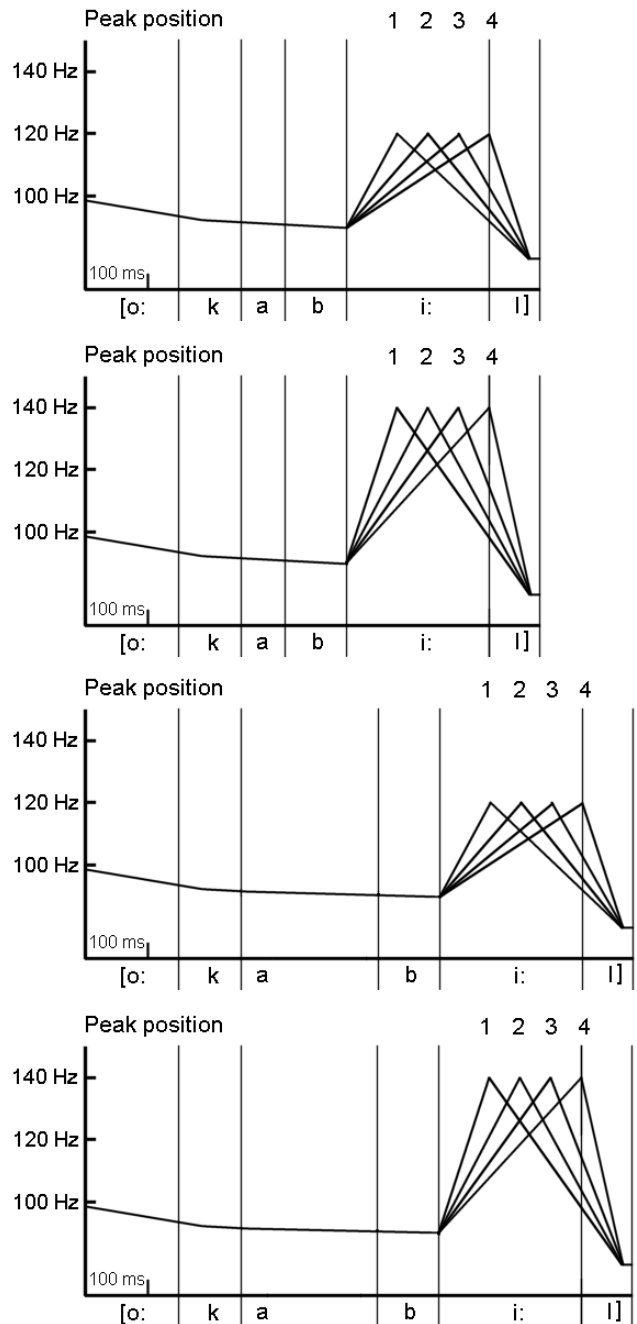


Figure 1: Schematic presentations of the stimuli used in the perception test. The upper two panels represent the low and high pitch configuration with no pause. The lower two panels represent the filled pause configuration.

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 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 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.

Confidence response scores are plotted in Figure 2 for each peak position and each condition. These scores generally follow the identification scores with the low confidence scores for questions reflecting the general bias toward statement identification.

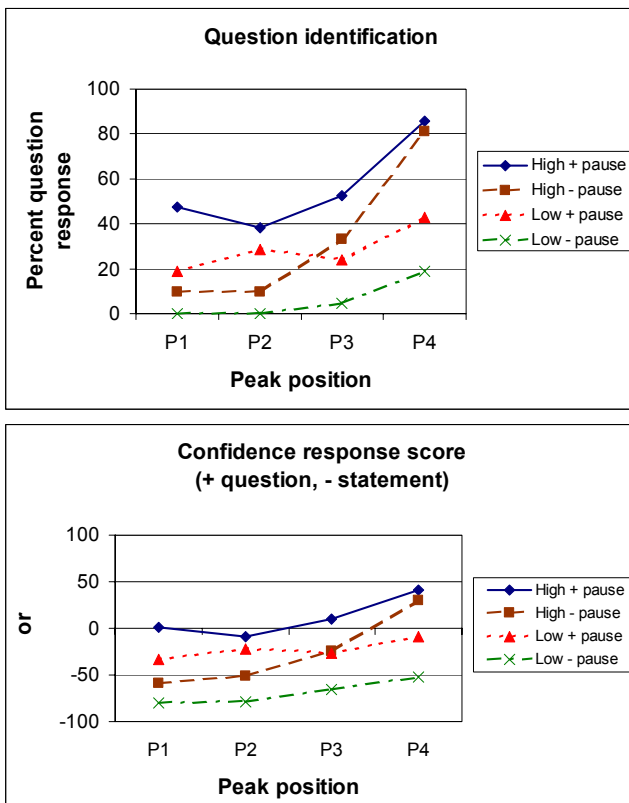


Figure 2: Results of the perception test showing percent question responses (upper panel) and mean confidence response scores for each stimulus (lower panel).

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

must be seen as primary cues to interrogative intonation, 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 [8] 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. Another reason to expect this could be based on a psychoacoustic argument where the pause gives the listener more time to process peak timing [13]. However, this was not the case as the pause moved all responses toward the question percept, effectively introducing ambiguity to stimuli with statement intonation (early low peak) and strengthening the question responses for stimuli with question intonation (late high peaks). 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, the hesitation code has the same function as the frequency code, whereby a high F0 indicates submissiveness and non-assertiveness and thereby signals question intonation [7] [14]. 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 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.

What we see here may be evidence that perception is not limited to equating a delayed peak with a higher pitch. There is also psychoacoustic evidence that late rises are not always perceived as high pitch [15]. The rise itself may be an extra cue which is needed in certain instances being part of the production code. In the experiment presented here, both a high pitch and a rise are needed to unambiguously signal interrogative mode. These results can be interpreted as evidence that perception of pitch height and pitch rise are not altogether equivalent but rather can function in a complementary fashion as a simultaneous signal of two codes. Ladd (1996) argues that a delayed peak should be interpreted as a low pitch target in the accented vowel (p. 104). In the context of this paper there is an argument for the interpretation of delayed peak as involving the perception of an additional rise.

5. CONCLUSIONS

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 hesitation pause contributes to the interrogative percept. It is hypothesized that this mechanism functions on a higher cognitive level where pausing is perceived as cognitive loading and interpreted as uncertainty. It will be interesting to compare production results from ongoing studies of pauses in different communicative situations to the perceptual results presented here [16]. Characteristics of pre-focal pauses are of particular interest as they may reveal differences between hesitation codes for non-assertiveness and effort codes for assertiveness.

ACKNOWLEDGEMENTS

This research was carried out at the Centre for Speech Technology, a competence centre at KTH, supported by VINNOVA (The Swedish Agency for Innovation Systems), KTH and participating Swedish companies and organizations. This work has also been supported in part by the Swedish Research Council (VR).

REFERENCES

- [1] D.R. Ladd, *Intonation phonology*. Cambridge: Cambridge University Press, 1996.
- [2] D. Hirst and A. Di Cristo, "A survey of intonation systems," In D. Hirst and A. Di Cristo (eds.) *Intonation Systems*, pp. 1-45, Cambridge: Cambridge University Press 1998.
- [3] M. D'Imperio and D. House, "Perception of questions and statements in Neapolitan Italian," In *Proceedings of Eurospeech 97*, pp. 251-254, Rhodes, Greece, 1997.
- [4] E. Gårding, "Sentence Intonation in Swedish," *Phonetica* 36, pp. 207-215, 1979.
- [5] M. Horne, P. Hansson, G. Bruce, J. Frid and A. Jönsson, "Accentuation of domain-related information in Swedish dialogues," In *Proceedings of ESCA International Workshop on Dialogue and Prosody*, pp. 71-76. Veldhoven, The Netherlands, 1999.
- [6] D. House, "Intonational and visual cues in the perception of interrogative mode in Swedish," In *Proceedings of ICSLP 2002*, pp. 1957-1960. Denver, Colorado, 2002.
- [7] C. Gussenhoven "Intonation and interpretation: phonetics and phonology," In B. Bel and I. Marlien (eds.), *Proceedings of the Speech Prosody 2002 Conference*, pp. 47-57. Aix-en-Provence, 2002.
- [8] E. Strangert, "Phonetic characteristics of professional news reading," *Papers from the fifth national phonetics conference, PERILUS XIII*, pp. 39-42, Stockholm University, 1991.
- [9] L. Ferrer, E. Shriberg and A. Stolcke, "Is the speaker done yet? Faster and more accurate end-of utterance detection using prosody," In *Proceedings of ICSLP 2002*, pp. 2061-2064. Denver, Colorado, 2002.
- [10] E. Shriberg, "Phonetic consequences of speech disfluency," In *Proceedings of the 14th International Congress on Phonetic Sciences*, pp. 619-622, San Francisco, 1999.
- [11] J. Beskow and K. Sjölander, "WaveSurfer - a public domain speech tool," In *Proceedings of ICSLP 2000*, vol. 4, pp. 464-467, Beijing, China, 2000.
- [12] S. Granqvist, "Enhancements to the visual analogue scale, VAS, for listening tests," *Speech, Music and Hearing. TMH QPSR 4/1996*: pp. 61-65, KTH Stockholm. 1996.
- [13] D. House, "The influence of silence on perceiving the preceding tonal contour." In *Proceedings of the 13th International Congress of Phonetic Sciences*, pp. 122-125, Stockholm, Sweden, 1995.
- [14] J.J. Ohala, "Cross-language use of pitch: an ethological view," *Phonetica* 40, pp. 1-18, 1983.
- [15] D. House, "Perception of pitch and tonal timing: implications for mechanisms of tonogenesis," In *Proceedings of the 14th International Congress of Phonetic Sciences*, pp. 1823-1826. San Francisco, 1999.
- [16] R. Carlson, B. Granström, M. Heldner, D. House, B. Megyesi, E. Strangert and M. Swerts, "Boundaries and groupings - the structuring of speech in different communicative situations: a description of the GROG project," In *Proceeding of Fonetik 2002, TMH-QPSR 44*, vol. 1, pp. 65-68, 2002.