Final Rises in Spontaneous Swedish Computer-Directed Questions: Incidence and Function

David House

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

Abstract

Phrase-final intonation was analysed in a subcorpus of Swedish computer-directed question utterances with the objective of investigating the extent to which final rises occur in spontaneous questions, and also to see if such rises might have pragmatic functions over and beyond the signalling of interrogative mode. Final rises occurred in 22 percent of the utterances. Final rises occurred mostly in conjunction with final focal accent. Children exhibited the largest percentage of final rises (32%), with women second (27%) and men lowest (17%). These results are discussed in terms of Swedish question intonation and the pragmatic social function of rises in a biological account of intonation.

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 two recent perception studies, however, House [6] and [7], demonstrated that a raised fundamental frequency (F0) combined with a rightwards focal peak displacement is an effective means of signaling question intonation in Swedish echo questions when the focal accent is in final position. Perception results confirmed the importance of timing where an early peak followed by a final falling contour was perceived as a statement while a late peak resulting in a rise through the final syllable was perceived as a question. This applied to test material where focal accent was carried on words with accent 1 and words with accent 2. 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.

It is clear that listeners perceive a final rise as an interrogative signal in Swedish even though a rise has not been described as a necessary component of interrogative intonation. These perception study results call for a study of speech production data to see to what extent final rises are actually represented in spontaneous speech.

There has moreover been recent interest in the automatic analysis of phrase final tones and short utterances with the objective of categorizing and extracting dialogue acts such as agreement, acknowledgement, backchannels, turntaking and speaker attitude (see e.g. [8], [9], [10] and [11]). Such information is highly useful in building and improving spoken dialogue systems using animated agents and spoken human-system interaction. The availability of an annotated corpus of spontaneous computer-directed speech presented very suitable speech material for analyzing phrase-final intonation in questions directed to an animated agent. We may conjecture that the optionality of a final rise is represented in the data and that an optional rise may in fact signal more than question intonation in Swedish. There may also be a relationship between a final rise and a final focal accent. This paper examines the extent to which final rises may occur in questions in computer-directed spontaneous speech and discusses the function these rises may have in signalling dialogue acts and speaker attitude over and beyond an information question.

2. Material and analysis

The utterances used in this study were taken from the August database recorded in 1998 and 1999 [12]. The August spoken-dialogue system was a multimodal system using a talking head as an animated agent. The agent was modeled on the Swedish author August Strindberg, and the system had several simple domains including facts about Strindberg, Stockholm, KTH and local information such as the locations of restaurants in Stockholm. The system was built into a kiosk and placed in public in central Stockholm for a period of six months. The speech recordings resulted in a database consisting of 10,058 utterances. The utterances were transcribed orthographically and labeled for speaker characteristics and utterance types by Bell and Gustafson [13] and [14] (see also Gustafson [15] and Bell [16] for recent reviews of this work). The total number of speakers represented in the database was 2685 of which 50% were men, 26% women and 24% were children. Bell and Gustafson [13] further analyzed these utterances according to the presumed intentions of the speakers. Two major categories were used: socializing or information seeking. In other words, was the user mainly interested in a social interaction with the agent, or was the user genuinely trying to retrieve information?

To obtain material for the present study, utterances containing the question words *vad* "what" and *var* "where" were extracted from the corpus. A total of 334 utterances contained the word "what" and 161 contained the word "where". For this analysis, 100 questions containing "what" and 100 questions containing "where" were randomly extracted from the subcorpus. The 200 utterances were analyzed auditively and visually using the WaveSurfer speech analysis package [17].

Each utterance was marked for presence or absence of final rise, presence or absence of final focal accent, and whether the utterance was produced by a child, woman or man. Waveforms, spectrograms and F0 traces of two typical utterances are shown in figures 1 and 2. The utterance in figure 1 displays a characteristic final rise while the utterance in figure 2 displays a focal accent ending in a final low.

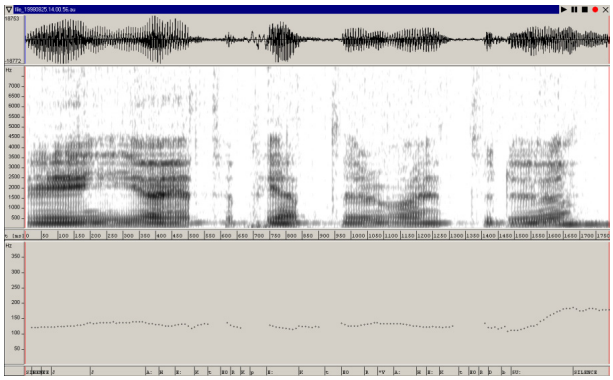

Figure 1: *Waveform, spectrogram and F0 contour of the utterance "Hej, jag heter Peter vad heter du?" (Hi, my name is Peter what is your name?) showing final rise.*

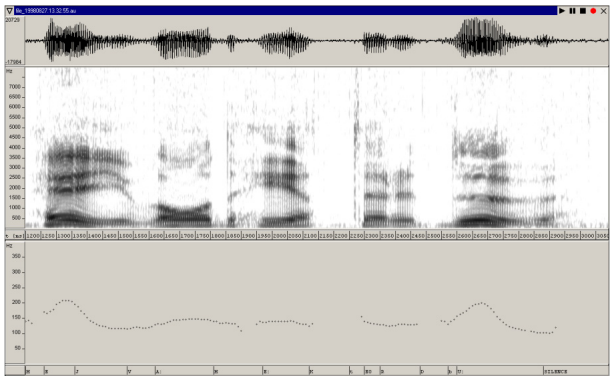

Figure 2: *Waveform, spectrogram and F0 contour of the utterance "Hej, vad heter du?" (Hi, what is your name?) showing final low.*

3. Results

3.1. Final rises

22% of the questions in this material ended in a phrase final rise. Proportionally, children produced more questions with final rises (32%), with women second (27%) and men the lowest percentage of questions with a final rise (17%). This breakdown is presented for both question words in figure 3.

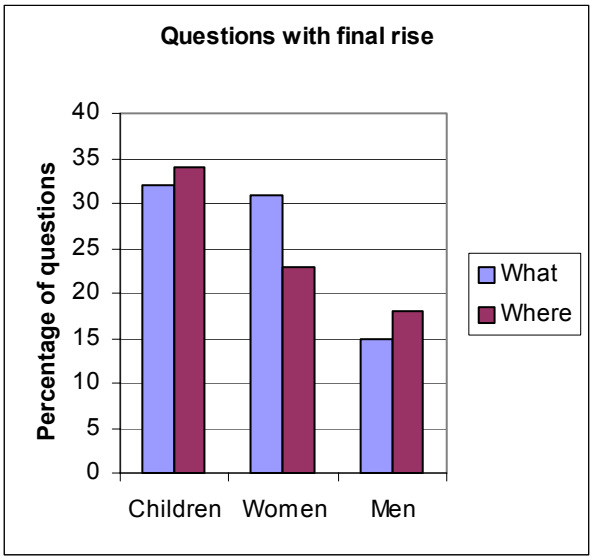

Figure 3: *Percentage of questions with a final rise distributed by speaker type and question word.*

3.2. Distribution by speaker type and question word

Of the 200 questions analyzed for this study, 17% were produced by children, 26% by women, and 57% by men. This compares to the distribution of the entire corpus which was 24% children, 26% women and 50% men [13]. Children asked a proportionally larger number of "what" questions (22) than "where" questions (12), while men asked a proportionally larger number of "where" questions (62) than "what" questions (52). This is displayed in figure 4.

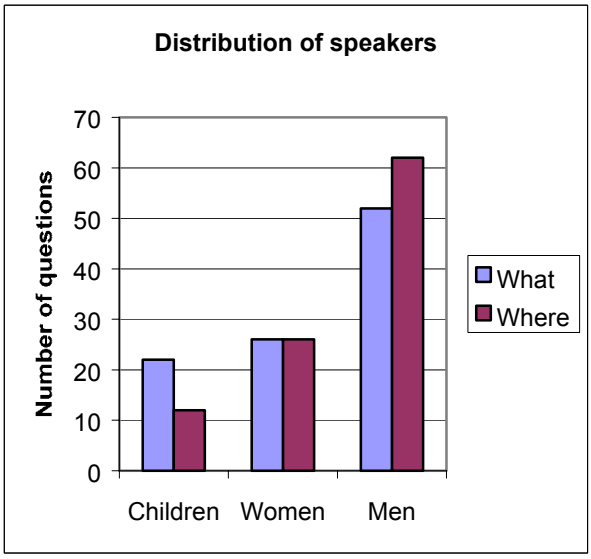

Figure 4: *Distribution of the total number of questions by speaker type and question word.*

"What" questions were generally more oriented to social interaction such as "What is your name?" "What are you doing?" and "What time is it?". "Where" questions were of a more strictly information retrieval character pertaining to the location of streets and restaurants and public facilities.

3.3. Final focal accent

77% of the questions ending in a final rise also ended in a focal accent. Of the questions ending in a final low, 48% ended in a focal accent. The distribution for each question word is presented in figure 5.

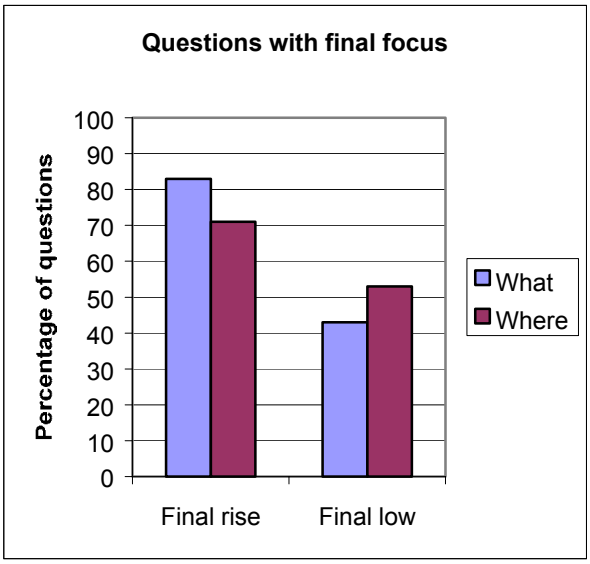

Figure 5: *Distribution of questions with final focus.*

3.4. Speaker dialect

Most of the speakers represented in the material had dialects from central Sweden. There were a few speakers of dialects of West Sweden where a final rise is characteristic of focal accent. There were also a few non-native speakers of Swedish.

4. Discussion

4.1. Final rise as interrogative marker

The results of this study are consistent with the traditional description of interrogative intonation in Swedish [4], where a final rise is not a necessary component for questions. However, as an optional interrogative marker, a final rise does occur in a considerable number of spontaneous questions and therefore should be considered as an important component of interrogative intonation when describing and modeling Swedish intonation.

In this material, final rises occur mostly in conjunction with final focal accent. The rise can be seen as a replacement accent (i.e. an interrogative focal accent) or a deformation of the focal accent where the focal peak is delayed as in the perception experiments cited above [6] and [7]. The fact that the final rise can also occur on a non-focal accent can be seen as evidence that the rise is an extra intonational factor which either delays the focal peak or surfaces as an extra rise. This extra intonation factor may not only be an optional reinforcment for interrogative mode, it may also signal pragmatic meaning.

4.2. Final rise and pragmatic meaning

There has been much discussion about the role of the phrase final tone, or boundary tone, in intonational meaning (see Ladd [1] for a review). A phrase final high or rise can convey speaker attitude such as uncertainty, or be a signal of dialogue act such as feedback seeking or turngiving (e.g. [10] and [18]).

Using the utterance categories proposed by Bell and Gustafson [13], socializing and information seeking, there is some evidence that the final rise in this material is a signal of socializing while the final low is a signal of information seeking. In the complete August corpus, children produced more socializing utterances than did women and men. In this material, children produced proportionally more questions with final rises than did women and men. Furthermore, children asked proportionally more "what" questions than "where" questions with the "what" questions being more socially oriented. This interpretation is consistent with a final rise being a signal of feedback seeking rather than purely information question intonation.

4.3. Biological codes and intonation

An explanation for the function of a final rise as a socializing marker in question intonation can be seen in the framework of biological codes for universal meanings of intonation, proposed by Gussenhoven [19]. 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 of the frequency code, the final rise as a cue to submissiveness or non-assertiveness can be interpreted as an invitation to socialize and engage in conversation. A final low in a question is more a signal of assertiveness and command, thus a more information-seeking question.

In terms of the effort code, a rise is already associated with focal accent to highlight important information. Therefore, to signal the intention to socialize, the production code needs to be exploited. In the production code, high pitch is normally associated with new topics at phrase beginnings. In the case of the final rise, however, the high comes at the end of the phrase and signals the invitation to continue the social interaction.

4.4. Focal structure of question utterances

The fact that final rises occur mostly in conjunction with final focus is also interesting. In this material the focal accent in final position seems to facilitate the final rise resulting in a focal peak delay. The timing of the focal accent rise has been shown to be more variable than the timing of the word accent [20]. A more detailed study of the focal peak timing in the questions with non-final focal accents would be interesting from the perspective of focal accent variability. However, it is clear that the final rise in these questions contribute to the variability of the timing of the focal accent rise.

It also appears that in many of the instances in this material of final lows with non-final focal accent, final focus is avoided by an optional sentence structure containing a final non-focal word or phrase, such as *Vet du var Kulturhuset ligger?* "Do you know where the Culture Center is located?" In this way, the tendency of a final rise coupled to final focus is avoided and the information-seeking intention of the speaker is emphasized.

5. Conclusions

The results of this study show final rises to occur in about 20% of a small corpus of spontaneous questions directed to an animated agent in a spoken dialogue system. The results confirm the optionality of a final rise in Swedish interrogative intonation. The distribution of the rises in the material also indicated the pragmatic function of the rises as signaling intended social interaction rather than purely information seeking. The study of such phrase-final characteristics can have considerable importance for our understanding of human-computer interaction and for improving the responsiveness of animated agents.

6. 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 a grant from the Swedish Research Council (VR) to the project: "Boundaries and groupings - the structuring of speech in different communicative situations [21]. Special thanks to Linda Bell and Joakim Gustafson for providing the transcribed August database and to Rolf Carlson for extracting the question utterances.

7. References

[1] Ladd, D.R., 1996. *Intonation phonology*. Cambridge: Cambridge University Press.

[2] Hirst, D.; Di Cristo, A., 1998. A survey of intonation systems, In D. Hirst and A. Di Cristo (eds.) *Intonation Systems*. Cambridge: Cambridge University Press. 1-45.

[3] D'Imperio, M.; House, D., 1997. Perception of questions and statements in Neapolitan Italian, In *Proceedings of Eurospeech 97,* 251-254, Rhodes, Greece.

[4] Gårding, E., 1979. Sentence Intonation in Swedish, *Phonetica 36*, 207-215.

[5] Horne, M.; Hansson, P.; Bruce, G.; Frid, J.; Jönsson, A., 1999. Accentuation of domain-related information in Swedish dialogues, *Proceedings of ESCA International Workshop on Dialogue and Prosody*, 71-76. Veldhoven, The Netherlands.

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

[7] House, D., 2003. Perceiving question intonation: the role of pre-focal pause and delayed focal peak. *Proc 15th ICPhS,* Barcelona, 755-758

[8] Ferrer, L.; Shriberg, E; Stolcke, A., 2002. Is the speaker done yet? Faster and more accurate end-of utterance detection using prosody, In *Proceedings of ICSLP 2002*, Denver, Colorado, 2061-2064.

[9] Ishi, C.T.; Mikhtari, P.; Campbell, N., 2003. Perceptually-related acoustic-prosodic features of phrase finals in spontaneous speech. *Proc. Eurospeech 2003*. Geneva. 405-408.

[10] Caspers, J., 2003. On the function of low and high boundary tones in Dutch dialogue. *Proc 15th ICPhS,* Barcelona, 1771-1774.

[11] Bhagat, S.; Carvey, H.; Shriberg, E., 2003. Automatically generated prosodic cues to lexically ambiguous dialog acts in multiparty meetings. *Proc 15th ICPhS*, Barcelona, 2961-1964

[12] Gustafson, J.; Lindberg, N.; Lundeberg, M., 1999. The August spoken dialogue system. *Proceedings of Eurospeech 99,* Budapest, 1151-1154.

[13] Bell, L.; Gustafson, J., 1999. Utterance types in the August System. *Proc of the ESCA Workshop on Interactive Dialogue in Multi-Modal Systems*. 81-84.

[14] Bell, L.; Gustafson, J., 1999. Interaction with an animated agent in a spoken dialogue system, *Proc of Eurospeech '99*, Budapest, 1143-1146.

[15] Gustafson, J., 2002. *Developing multimodal spoken dialogue systems; empirical studies of spoken human-computer interaction,* Doctoral dissertation, Department of Speech, Music and Hearing, KTH, Stockholm.

[16] Bell, L., 2003. *Linguistic adaptations in spoken human-computer dialogues; empirical studies of user behavior,* Doctoral dissertation, Department of Speech, Music and Hearing, KTH, Stockholm.

[17] Sjölander, K.; Beskow, J., 2000. WaveSurfer - a public domain speech tool, *In Proceedings of ICSLP 2000*, vol. 4, 464-467, Beijing, China.

[18] Cerrato, L., 2002. Some characteristics of feedback expressions in Swedish. In *Proceedings of Fonetik 2002*, TMH-QPSR 44, vol. 1, 101-104.

[19] Gussenhoven, C., 2002. Intonation and interpretation: phonetics and phonology, In B. Bel and I. Marlien (eds.), *Proceedings of the Speech Prosody 2002 Conference*, Aix-en-Provence, 47-57.

[20] Bruce, G., 1987. How floating is focal accent? In Gregersen K. and H. Basbøll (Eds.), *Nordic Prosody IV*, Odense: Odense University Press. 41-49.

[21] Carlson, R.; Granström, B.; Heldner, M.; House, D.; Megyesi, B.; Strangert, E.; and Swerts, M., 2002. Boundaries and groupings - the structuring of speech in different communicative situations: a description of the GROG project, In *Proceedings of Fonetik 2002*, TMH-QPSR 44, vol. 1, 65-68.