Proc. 5th int. Congr. phon. Sci., Münster 1964, pp. 326–331
(S. Karger, Basel/New York 1965).

From the Haskins Laboratories, New York

Intonation Contours Evaluated by American and Swedish Test Subjects

By K. Hadding-Koch* and M. Studdert-Kennedy**,
New York, N.Y.

Both in Swedish and in American English the pitch contours of yes-no questions tend to display a final rise, pitch contours of statements a final fall. But naturalistic observations suggest that earlier portions of the contour are also relevant. Such questions tend to be spoken on a comparatively high pitch in Swedish[2], and with a continuously rising contour in American English[3, 6]. The two languages agree in displaying a statement contour with a moderate rise followed by a fall. The present study simulated Swedish intonation contours, by controlled variation of the fundamental frequency, and examined the distributions of responses to these contours made by Swedish and American subjects. The two groups were expected to agree more in their statement than in their question responses.

The words, "*Fӧr Jane*" [foe'jein] = *for Jane* (duration: 640 msecs), so spoken as to be acceptable as English by Americans, as Swedish by Swedes, were recorded on magnetic tape. The tape was passed through a special synthesizer at Haskins Laboratories, New York, which permits the fundamental frequency of the utterance to be varied without otherwise changing the recorded speech[1]. The output of this synthesizer was recorded on magnetic tape.

Forty-two intonation contours were used. The fundamental frequency values were based on detailed spectrographic analyses of the Swedish speaker's natural speech. All started at 250 cps (sustained over "Fӧr" for 140 msecs), rose to a peak of either 370 cps or 310 cps,

* Also at Institute of Phonetics, University of Lund, Sweden.
** Also at Barnard College, Columbia University, New York.

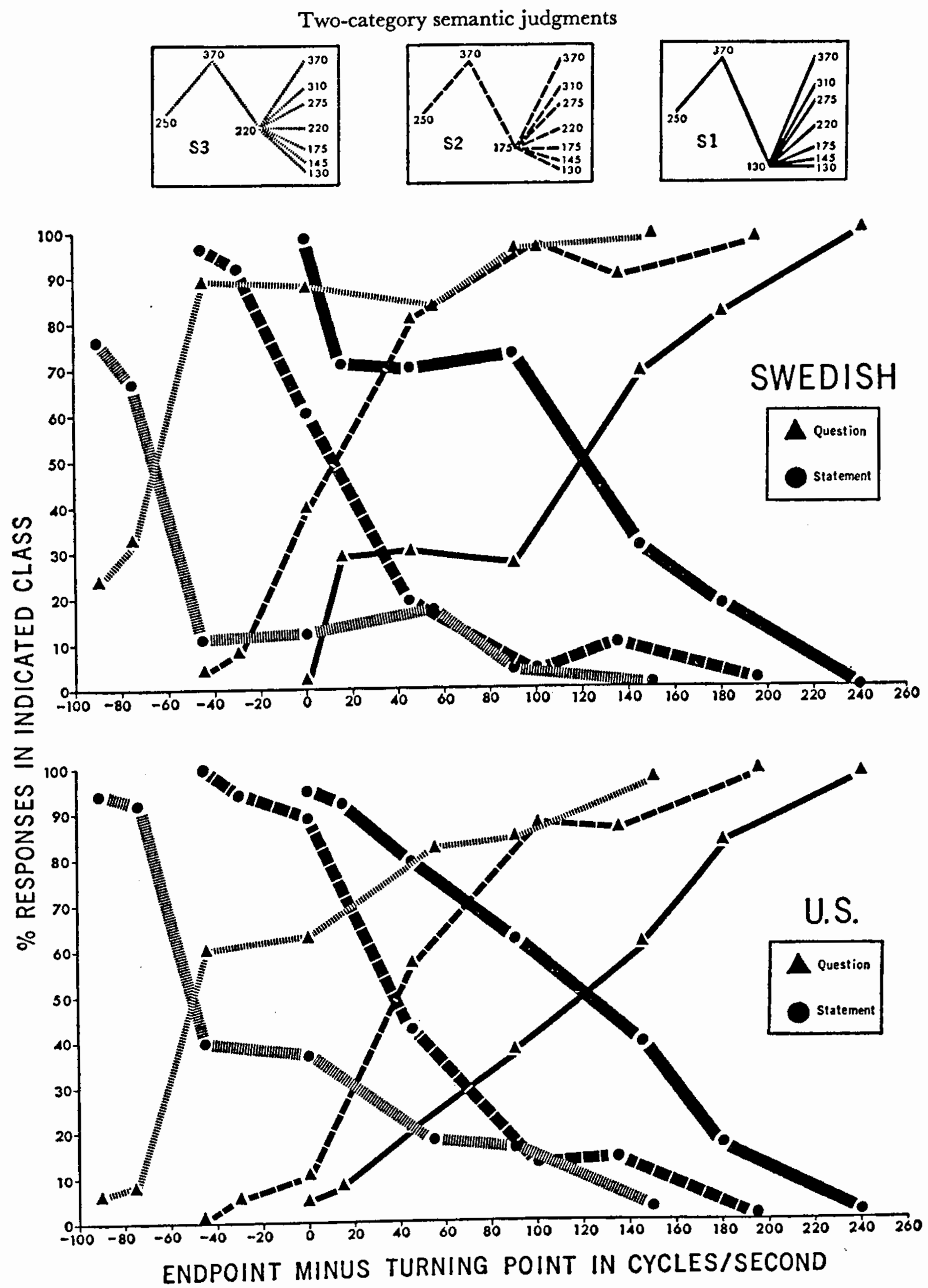

Fig. 1 and *2.* Two-category semantic judgments with peak fundamental frequency at 370 cps: percentage of statement and question responses as a function of the terminal rise (positive) or fall (negative) in cycles/second of fundamental frequency (end point minus turning point). Parameters of the curves are turning point fundamental frequency: 130 cps *(S1)*, 175 cps *(S2)*, and 220 cps *(S3)*. The Swedish data are plotted above, the U.S. data below.

dropped to one of three turning points: 130 cps, 175 cps, or 220 cps (300 msecs), and then proceeded to one of seven end points between 130 cps and 370 cps (200 msecs).

The forty-two stimuli were arranged into five different random orders and presented in an appropriately counterbalanced series to 25 Swedish and 24 American undergraduates. In four separate sessions, subjects were instructed to indicate for each stimulus: (1) whether it would be better characterized as a statement or a question; (2) whether it ended with a rise or a fall; (3) whether it would best be characterized as a question, a statement or a non-communicative utterance, that is, a private "reflection", spoken by the speaker to herself; (4) whether it ended with a rise, a fall or a level pitch. The third category of (3) and (4) was suggested by previous analyses of natural speech in which "reflections" appeared to display more or less level final contours, and was introduced here to see whether it might not collect the responses to stimuli that were found to be ambiguous when only the statement-question categories were used.

In the two-category semantic test, both Swedish and American responses to a stimulus with a given final rise or fall in frequency were found to vary with the frequency values of peak and turning-point: the higher the frequency of either, the more questions were heard. Thus, if we consider the value in cycles of the final rise at which Swedish responses crossed over from predominantly statements to predominantly questions, we find that: with the peak at 310 cps, the crossover is at 160 cycles for a turning point of 130 cps, at 130 cycles for a turning point of 175 cps, at 80 cycles for a turning point of 220 cps; with the peak at 370 cps, the crossovers for the three turning-points are at a rise of 120 cycles, a rise of 12 cycles and a fall of 65 cycles, respectively. From this last figure (crossover from statement to question at a final fall of 65 cycles), it is evident that the effect of the peak and turning point frequencies overrode the effect of the final fall. See Figs 1 and 2.

Turning to the results of the psychophysical sessions in which listeners were asked simply to indicate whether the final movement of the contour was a rise or a fall, we find more overall uncertainty – more disagreement between subjects – but, at the same time, a remarkable degree of agreement with the semantic judgements. That is to say, stimuli heard as questions tended to be heard also as having a final rise (even if, in fact, the contour displayed a fall), while stimuli heard as statements tended to be heard as having a fall

Swedish two-category semantic judgments

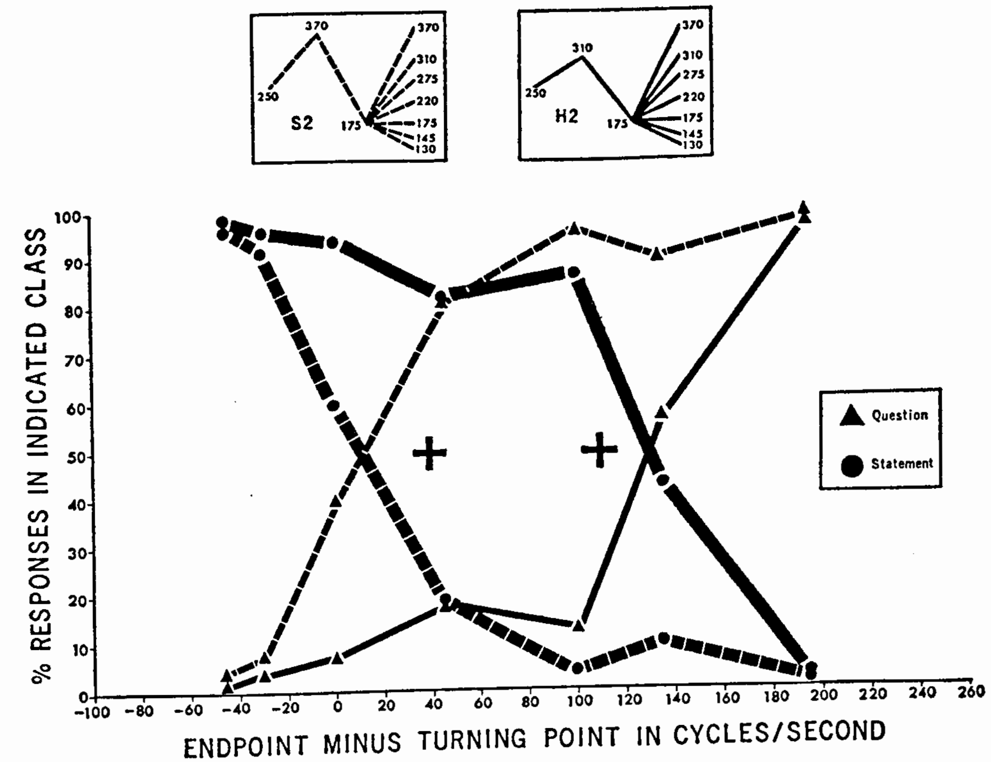

Fig. 3. Two-category semantic judgements of Swedish subjects on the *S2* (peak frequency: 370 cps) and *H2* (peak frequency: 310 cps) series. Percentage of responses in indicated class is plotted against terminal rise or fall in cycles/second of fundamental frequency. The crosses indicate the crossover values for the U.S. subjects on the *S2* (left) and *H2* (right) series.

(even if, in fact, the contour displayed a rise). Thus the discrimination of pitch within these speech signals was markedly less fine than might have been predicted from pure tone pitch discrimination[4].

The results of the two sessions in which three category judgements were called for were somewhat unsatisfactory – although once again there was remarkably good agreement between semantic and psychophysical judgements. For the most part, the third category served only to confuse the picture. The response curves were, on the whole, very similar to those of the two category responses, except that there was less agreement, the third category "nibbling" pieces from the other two. But in one group of stimuli, for which the peak

U.S. two-category semantic and psychophysical judgments.

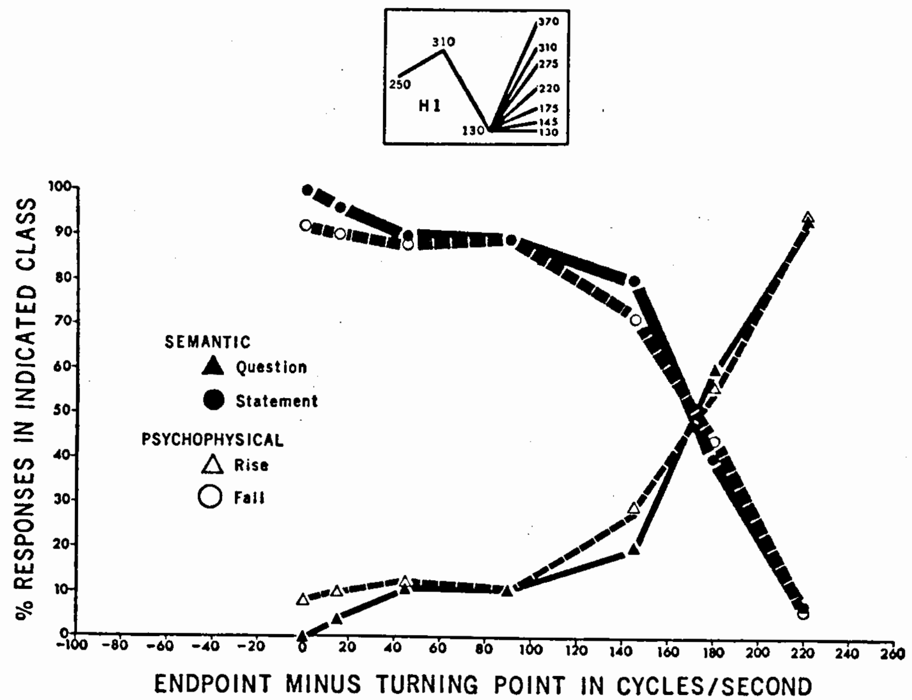

Fig. 4. Two-category semantic (solid line) and psychophysical (hatched line) data for U.S. subjects on the *H1* stimulus series. Percentage of responses in indicated class is plotted against terminal rise or fall in cycles/second of fundamental frequency.

was relatively low (310 cps) and the turning point relatively high (220 cps), the third category did show an appreciable percentage of responses: 58 % and 72 % in the semantic test for the Swedish and American groups respectively, 70 % and 66 % in the psychophysical test. As was expected, these maximum figures were reached on stimuli with a level final contour. See Fig. 4.

To sum up, at least three variables were operating to determine listeners' semantic responses: peak frequency, turning point frequency and final contour. In general, the higher the peak and the turning point, the more likely were both Swedish and American listeners to hear a question, the lower the peak and the turning point, the more likely were they to hear a statement. Only marked rises and falls overrode the effects of peaks and turning points. The third category ("reflection") proved to be of use for a limited group of stimuli

having a level final contour and a relatively low precontour. All these effects were also present, though less marked, in the psychophysical data. For the most part, particularly in the semantic responses, these effects were more marked in the Swedish group than in the American. The American listeners showed greater disagreement among themselves – as was expected, since the contours used were derived from the spectrographic analysis of Swedish speech.

Further experiments, designed to clarify this and other points, are now in progress. Our new stimuli include typical American English, as well as typical Swedish, contours. For purposes of control we are collecting psychophysical data on pure tone contours identical with the fundamental frequency contours of the speech.

Acknowledgement. We should like to acknowledge that the research reported here was supported in part by a grant from the National Science Foundation, Washington, D.C.

References

1. *Borst, J. M.* and *Cooper, F. S.:* Speech research devices based on a channel vocoder (abstract). J. acoust. Soc. Amer. *29:* 777 (1957).
2. *Hadding-Koch, K.:* Acoustico-phonetic studies in the intonation of Southern Swedish (Lund 1961).
3. *Pike, K. L.:* The intonation of American English (Ann Arbor 1945).
4. *Stevens, S. S.* and *Davis, H.:* Hearing, its psychology and physiology (New York 1938).
5. *Uldall, E. T.:* Ambiguity: question or statement? or "Are you asking me or telling me?", Proc. IVth Int. Congr. Phonet. Sci., Helsinki, 1961, pp. 779–783 (The Hague 1962).

Authors' addresses: Docent Dr. Kerstin Hadding-Koch, Fonetiska Institutionen, Kävlingevägen 20, *Lund* (Sweden) and Dr. Michael Studdert-Kennedy, Haskins Laboratories, 305 East 43rd Street, *New York, N.Y.* (USA).

Discussion

Gårding (Lund): If the generated stimuli that have been described are part of experimental work in juncture I do not think some of them are very realistic. A prejunctural fall of fundamental frequency is most often accompanied by a loss of intensity. Four years ago, at Haskins Laboratories, *Arthur Abramson* and I varied the fundamental frequency in some American English intonation contours. One of the experiments was to pit a falling statement contour against a rising question contour. By means of listening tests we found that we needed a considerable drop of frequency to make listeners place the generated contour in the statement category. A contour with a drop of ten cycles for instance was rather put in the question category. I think this is due to the fact that our carrier phrase did not have the intensity curve that normally goes with a falling contour.

Answer *Hadding-Koch:* Yes, I believe it is quite probable at least that we would have got more statement responses if we had used falling intensity with falling intonation. It was not possible to vary the intensity feature with the device used but we have done it in our new sets of stimuli: we have one set with sustained intensity and one with falling intensity. This could be done simultaneously on two identical spectra stored in the Digital Spectrum Manipulator at Haskins Laboratories, New York.