STUDIA LINGUISTICA

REVUE DE LINGUISTIQUE
GENERALE ET COMPAREE
XVII 1963

A Study of Semantic and Psychophysical Test Responses to Controlled Variations In Fundamental Frequency.

Kerstin Hadding-Koch
and
Michael Studdert-Kennedy

CWK GLEERUP.LUND
EINAR MUNKSGAARD.COPENHAGUE

Kerstin Hadding-Koch and Michael Studdert-Kennedy

A STUDY OF SEMANTIC AND PSYCHOPHYSICAL TEST RESPONSES TO CONTROLLED VARIATIONS IN FUNDAMENTAL FREQUENCY[1]

In both Swedish and American English the fundamental frequency (F0) of so-called yes-no questions tends to show a final rising contour, while the Fo of statements tends to show a final fall. But earlier portions of the utterance have been observed to be relevant too (see, for instance, Hermann, 1942). Thus, questions tend to be spoken at a comparatively high frequency in Swedish (E. A. Meyer in Hermann, *op. cit.*, Hadding-Koch, 1961) and with a more or less continuously rising contour in American English (Pike, 1945, Uldall, 1962). Statements in both languages are typically spoken with a moderate rise followed by a fall.[2]

In the present investigation, various patterns of fundamental frequency, simulating Swedish intonation contours,[3] were submitted to Swedish and American test subjects. The distributions of the responses obtained were examined.

The words, *För Jane* = Swedish for *For Jane*, were recorded on magnetic tape. They were found to be acceptable as English by American listeners and as Swedish by Swedes. The tape was glued to an acetate loop, on which the contours of a systematically varied series of fundamental frequencies were also painted. The speech was passed through the Vocoder at Haskins Laboratories, with intonation being

[1] This article is a report on an investigation introduced in a paper delivered before the Fifth International Congress of Phonetic Sciences, Münster, in August 1964, a summary of which will appear in the Proceedings of the Congress.
[2] cf., however, p. 75.
[3] In the present study, the term "intonation contour" refers only to a contour of fundamental frequency. Fundamental frequency is usually said to be the strongest single acoustic cue to intonation, although other variables may also play a role as acoustic correlates of intonation, viz., duration and intensity (Bolinger, 1958; Denes, 1959; Denes and Milton-Williams, 1962).

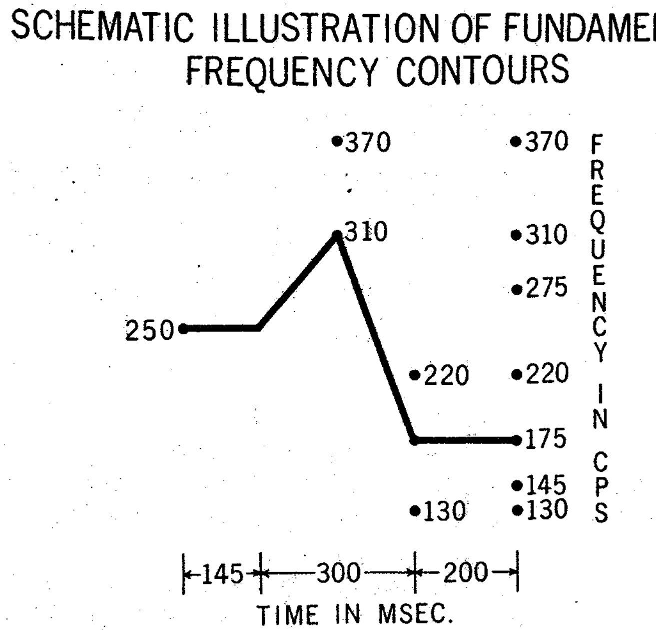

Figure 1. Schematic illustration of fundamental frequency contours, with starting point at 250 cps, two peaks, at 310 and 370 cps, three turning points, at 130, 175 and 220 cps, and seven end points, between 130 and 370 cps.

controlled by the painted contours (see Borst and Cooper, 1957). The output was recorded on magnetic tape.

Forty-two intonation contours were used. They were based on a detailed spectrographic analysis of the Swedish speaker's natural speech (see Hadding-Koch, 1961, p. 85 ff.). Fig. 1 gives a schematic illustration of one of the contours and shows the different points of frequency used. All contours started at 250 cycles per second: this frequency was sustained for 140 msecs (not 145 as indicated in the figure), over the word *För*. They then rose to one of two peaks, one at 310 cps, which we call the high, or *H* peak and the other at 370, which we call the superhigh, or *S* peak. From the peak the contour fell to one of three so-called turning points (cf. Gårding, 1960), numbered 1 to 3 from the lowest (at 130 cps) to the highest (at 220 cps). This rise to the peak and fall to the turning point lasted for 300 msecs. Finally, the contour proceeded to one of seven end points between 130 cps and 370 cps. This final rise, level or fall lasted for 200 msecs. Thus, the total duration of the stimulus was 640 msecs. The rise to the peak was ap-

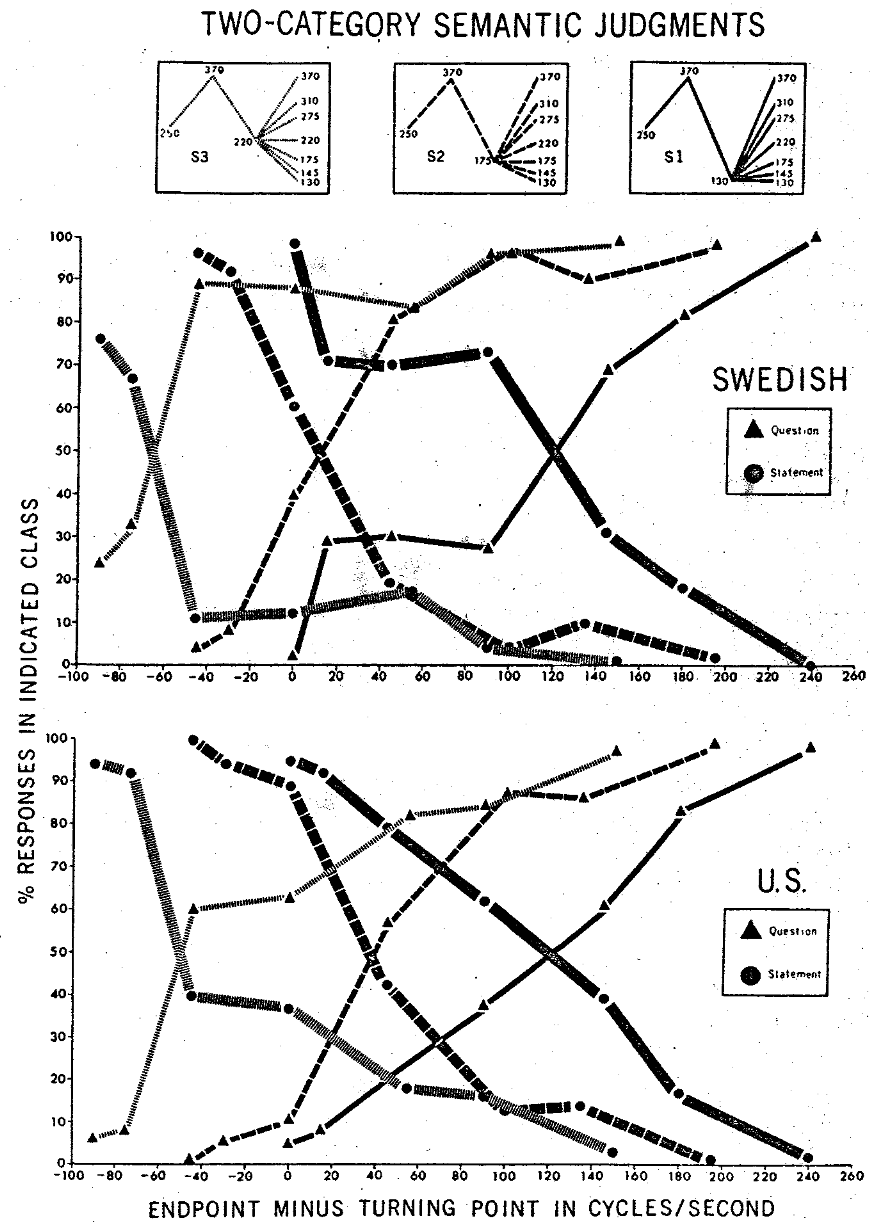

Figure 2. Two-category semantic judgments with peak fundamental frequency at 370 cps: percentage of statement and question responses as a function of the terminal rise (positive) or fall (negative) in cycles/second of fundamental frequency (end point minus turning point). Parameters of the curves are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

proximately half as long as the fall to the turning point. The contours were rounded at peak and turning point, not sharp as in the schematic contour of Fig. 1.

The forty-two stimuli were arranged in five different random orders and presented 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 (two-category semantic judgment); (2) whether it ended with a rise or a fall (two-category psychophysical judgment); (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 (three-category semantic judgment); (4) whether it ended with a rise, a fall or a level pitch (three-category psychophysical judgment). The third category of (3) and (4) was suggested by previous analyses of natural speech, in which "reflections" or "reactions", where the speaker seemed to be speaking more to himself than to his listener, appeared to display more or less level final contours (Hadding-Koch, 1961, p. 121 ff.), and was introduced here to see whether it might not collect the responses to stimuli that were found to be ambiguous when only two categories were used. Cf. also Uldall, 1962.

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. Figure 2 presents the two-category semantic data for the *S* series of contours—contours with a peak fundamental frequency at 370 cps (Swedish data above, American data below). There are three pairs of curves in each graph: a pair for *S1* with a turning point at 130 cps (solid), a pair for *S2* with a turning point at 175 cps (wide hatching), a pair for *S3* with a turning point at 220 cps (narrow hatching). Against the abscissa are plotted the value of the terminal rise or fall in cycles per second of fundamental frequency (end point minus turning point): a negative value indicates a terminal fall, a positive value a terminal rise. Against the ordinate are plotted the percentages of question and statement responses. The questions are graphed with a narrow line and a triangle, the statements with a broad line and a circle.

The effect of the final glide is clear in all three series and for both groups: the higher the final rise, the more likely are subjects to interpret the contour as a question. Moreover, the turning point frequency

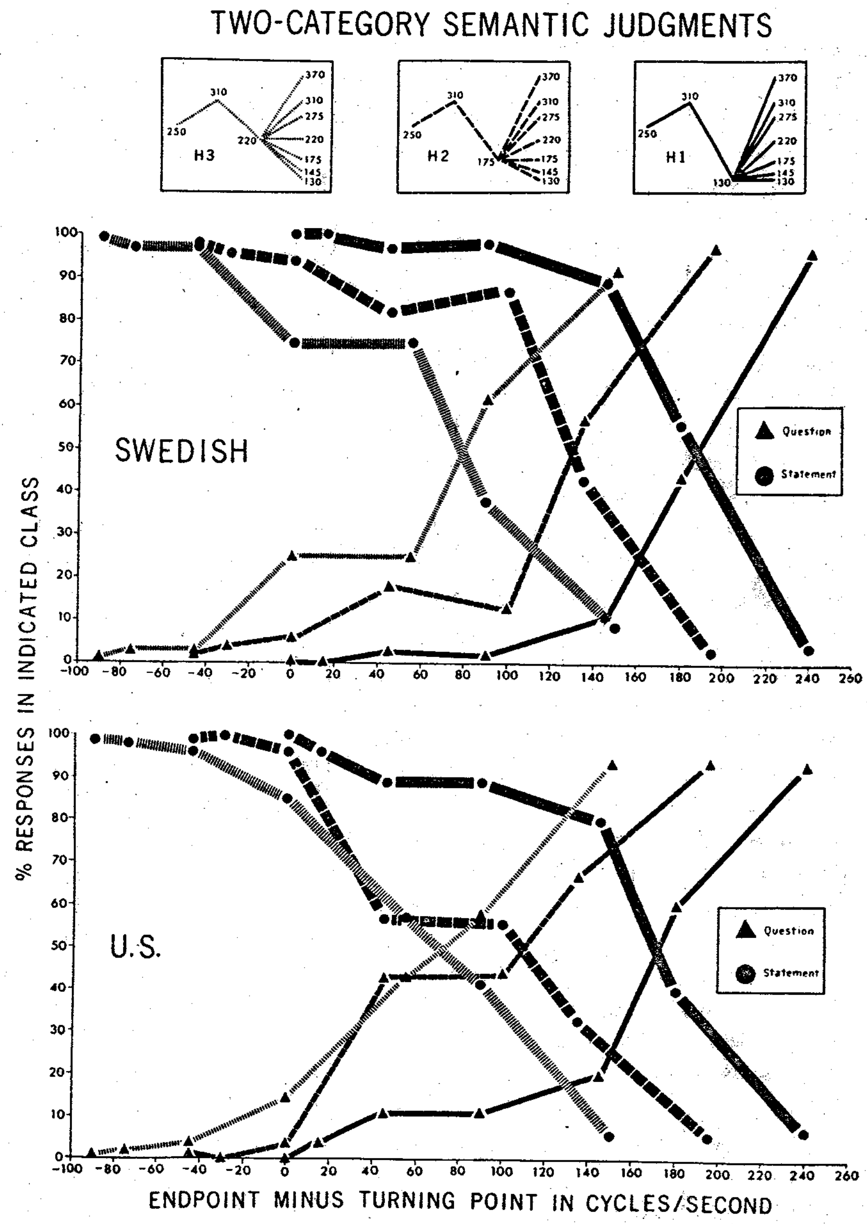

Figure 3. Two-category semantic judgments with peak fundamental frequency at 310 cps: percentage of statement and question responses as a function of the terminal rise (positive) or fall (negative) in cycles/second of fundamental frequency (end point minus turning point). Parameters of the curves are turning point fundamental frequency: 130 cps (*H1*), 175 cps (*H2*), and 220 cps (*H3*). The Swedish data are plotted above, the American data below.

evidently has, a similar effect. As a measure of this we may consider the values of the final rise at which responses cross over from predominantly statements to predominantly questions. For the Swedish subjects the crossover on *S1* is at a rise of 120 cycles, on *S2* at a rise of 12 cycles, on *S3* at a fall of 65 cycles. Thus, the higher the frequency of the turning point before a given rise or fall, the more ready are subjects to hear questions rather than statements. The results are essentially the same for the American group although the effect of the turning point is less evident. Thus, in the *S3* series, the final fall of 90 cycles is heard as a statement 94 per cent of the time as against the Swedish 76 per cent; and the final fall of 45 cycles is heard as a question 60 per cent of the time as against the Swedish 89 per cent.

Similar though somewhat smaller effects of the turning point are evident in the *H* series of contours—those with a peak fundamental frequency of 310 cps instead of 370 cps. However, the chief interest of the *H* series is in the effect which the lowered peak itself had upon the response distributions. The effect of raising the peak is the same as the effect of raising the turning point: the probability of the contour being heard as a question is increased. This is illustrated in Figure 4. For example, a terminal rise of 45 cycles was heard by the Swedish as a *statement* 82 per cent of the time if the peak is at 310 cps, as a *question* 81 per cent of the time when the peak is at 370 cps. The effect was similar in the American data, though less marked.

Turning to the results of the two-category 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—that is, more disagreement between subjects—but, at the same time, a remarkable degree of agreement with the semantic judgments. In Figure 5 the American semantic and psychophysical data are compared for a series on which they agreed particularly well. The semantic data are plotted with a solid line and filled circles and triangles, the psychophysical data are plotted with a hatched line and open circles and triangles. In general, stimuli heard as questions tended to be heard also as having a final rise, while stimuli heard as statements tended to be heard as having a fall (even if, in fact, the contour displayed a rise). The agreement was not always as good as this. However, we may note that, insofar as the psychophysical and semantic judgments agree, subjects appear to have been making use of the *perceived* terminal glide to reach their semantic decisions. We may also note that they perceived the direction of the glide far less accurately than might

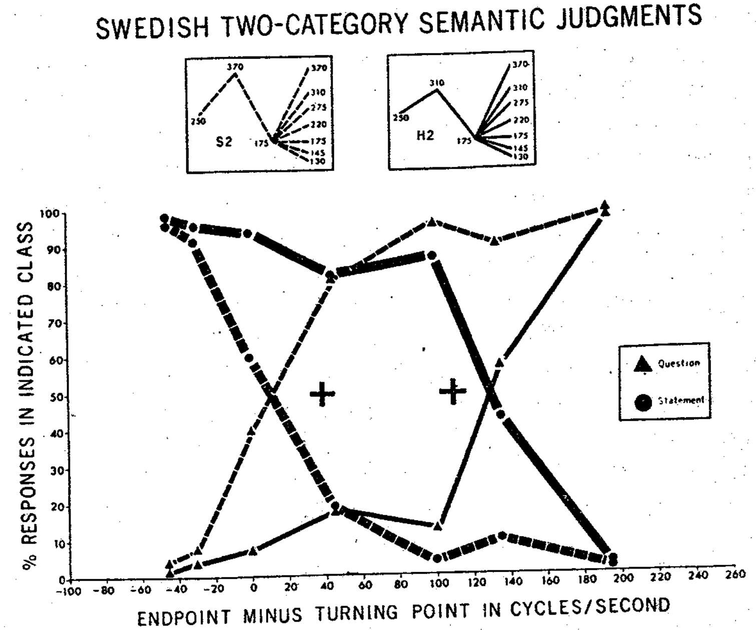

Figure 4. Two category semantic judgments of Swedish subjects on the *S2* (peak frequency: 370 cps) and *H2* (peak frequency: 310 cps) series. Percentage of responses in indicated class is plotted against terminal rise or fall in cycles/second of fundamental frequency. The crosses indicate the crossover values for the American subjects on the S2 (left) and H2 (right) series.

have been predicted from pure tone frequency discrimination (Stevens and Davis, 1938). This loss in accuracy is presumably an effect of the preceding peak and turning point.

The results of the two sessions in which three category judgments were called for showed remarkably good agreement between semantic and psychophysical judgments. The response curves were, on the whole, very similar to those of the two-category responses, except that there was less consistency, the third category "nibbling" pieces from the other two. For the most part the third category served only to confuse the picture. But in one group of stimuli, the third category did show an appreciable percentage of responses. Figure 6 presents the American data from three-category semantic judgments (above) and psychophysical judgments (below) on the *H3* series—a series with a relatively low peak of 310 cps and a relatively high turning point of 220 cps. The question and rise percentages are plotted with triangles; the percentages of the "reflection", or non-communicative utterance, and the level judgments are plotted with squares; the statement and fall percentages are plotted with circles. As was expected, the maximum figures of the third category were reached on stimuli with a level final contour: in the semantic data, 72 per cent of the responses fell into the reflection category when the terminal glide was level, the percentages of statement and question responses increasing on either side of this. For the Swedish subjects the reflection category at the level terminal glide drew 58 per cent of the responses.[1] The corresponding figures in the psychophysical tests were 66 per cent and 70 per cent for the American and Swedish subjects, respectively. The psychophysical data agree well, in fact, with the semantic for stimuli with level or rising terminal glides. For stimuli with terminal falls, listeners were much more uncertain of their psychophysical judgments than of their semantic judgments.

To sum up, at least three variables were operating to determine listeners' responses: peak frequency, turning point frequency and terminal glide. In general, the higher the value of any one of these, the more likely were both Swedish and American listeners to hear questions or terminal rises, the lower the value of any one, the more likely were they to hear statements or terminal falls. From a linguistic viewpoint, the most important finding was that the fundamental frequency values of peak and turning point were powerful determinants of listeners' semantic judgments. At the same time it should be stressed that the apparently simple psychophysical judgments of the direction of the terminal glide were themselves strongly influenced by the parts of the contour that preceded the terminal glide.

As to the performance of the two different language groups, the effect of the peak and turning point frequencies was somewhat greater for the Swedish subjects than for the American. However, since the American listeners also tended to show greater disagreement among

[1] A preliminary three-category semantic test had been run, in which listeners were asked to indicate whether each stimulus would be best characterized as a statement, a question or an exclamation. The exclamation category received a substantial number of responses when the contour was relatively level, but maximal (and very high) figures were reached for the stimuli with the largest final falls. It appeared that these stimuli, which combined a fall with a comparatively sustained intensity, were interpreted as rather strong commands and therefore marked as exclamations. In this case listeners were obviously reacting to an impression of loudness; this was instructive in itself but fell outside our purpose: to provide a "box" for responses to those stimuli that were found to be ambiguous when only two-category responses were allowed.

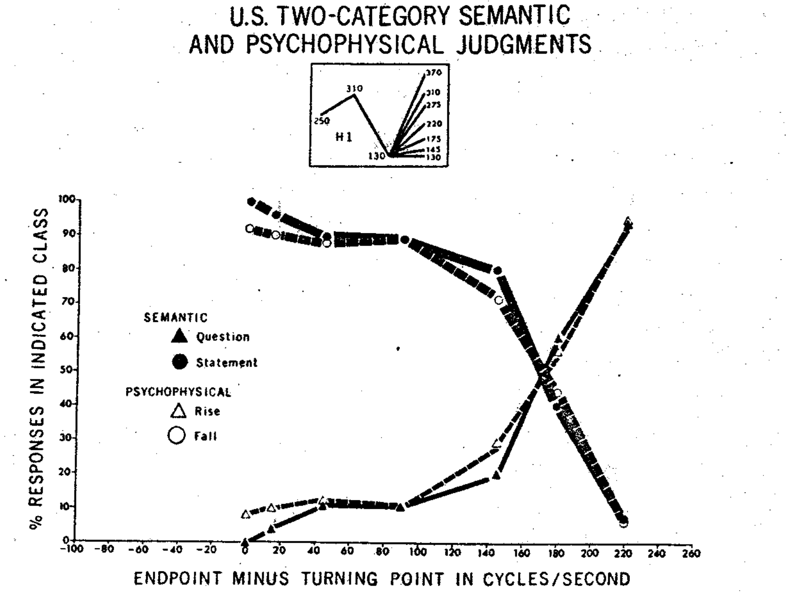

Figure 5. Two-category semantic (solid line) and psychophysical (hatched line) data for American subjects on the *H1* stimulus series. Percentage of responses in indicated class is plotted against terminal rise or fall in cycles/second of fundamental frequency.

themselves, it is not clear from the present study whether this may not have been an artefact of the use of typical Swedish contours in constructing the stimuli.

Further experiments, designed to clarify this and other points, are now in progress. In these experiments we have constructed the stimuli with a device that permits even finer control of the fundamental frequency than in the present experiment (the Digital Spectrum Manipulator at Haskins Laboratories, see Cooper, 1964). Our new stimuli include typical American English, as well as typical Swedish, contours. We are, for instance, adding one lower peak and one higher turning point. This will enable us, among other things, to create stimuli with the more or less continuously rising contour characteristic of the American English question. We will also be able to determine whether a high level fundamental frequency has the same effect of "talking to self" (our third category) as had the comparatively low level contour of the *H3*-series. That is to say, we want to see whether it is the *low* pitch or the more or less *level* pitch, that is essential to the third category judgment.

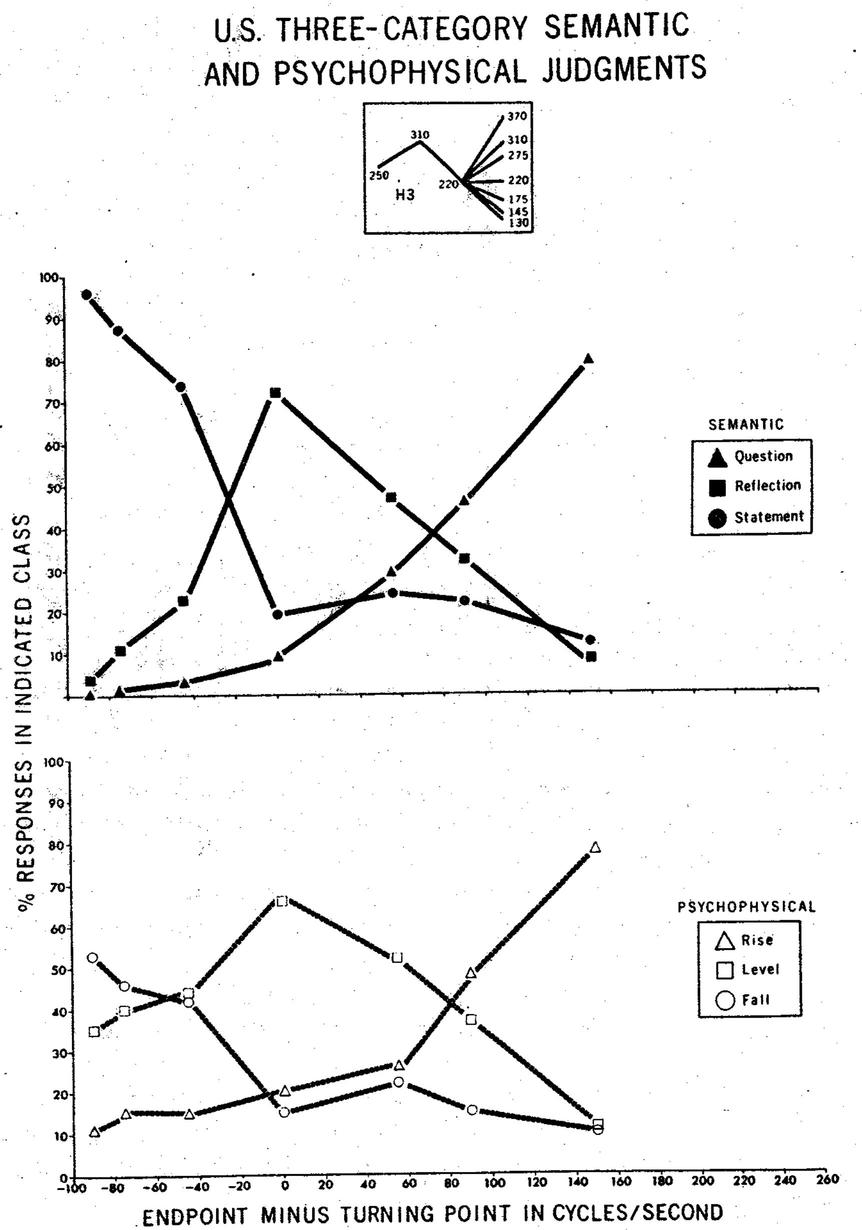

Figure 6. Three-category semantic data (above) and psychophysical data (below) for American subjects. Percentage of responses in indicated class is plotted against terminal rise or fall in cycles/second of fundamental frequency.

Terminal rises and falls will be more evenly represented in our new tests. In the present study we investigated the range between a slight fall and a high rise. This was due to the fact that polite statements in Swedish may have a final rise and as moderate final rises may occur with statements also in American English (Uldall, 1962), we had expected the dividing line between statement and question responses to lie somewhere between a slight rise and a high rise.

Our new series of stimuli also include a parallel set with a slowly falling intensity on the final syllable of the utterance. In the present series, intensity is kept fairly level, a fact which may have reduced the number of statement (and fall) responses, since statements in the two languages are usually spoken with a final dying-away of intensity.

Finally, we are collecting psychophysical data on pure tone contours identical with the fundamental frequency contours of the speech in order to compare the responses given to these non-speech stimuli with those given to corresponding speech stimuli.

Authors' addresses: Kerstin Hadding-Koch, Fonetiska Institutionen, Kävlingevägen 20, Lund, Sweden and Michael Studdert-Kennedy, Haskins Laboratories, 305 E. 43rd St., New York 17, N.Y., U.S.A.

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. Bolinger, D. L.: A theory of pitch accent in English, Word *14*: 109—149 (1955).
2. Borst, J. M. and Cooper, F. S.: Speech research devices based on a channel vocoder, J. Acoust. Soc. Amer. (1957), p. 777 (abstract).
3. Cooper, F. S.: Instrumental methods for research in phonetics. Paper given at the 5th Int. Congr. Phonetic Sciences, August 16—23 (Münster 1964).
4. Denes, P.: A preliminary investigation of certain aspects of intonation, Language and Speech *2*: 106—122 (1959).
5. Denes, P. and Milton-Williams, J.: Further studies in intonation, Language and Speech *5*: 1—14 (1962).
6. Gårding, E.: A study of the perception of some American English intonation contours. Paper read before 75th Meeting Mod. Lang. Ass. Amer., 28 Dec. (Philadelphia 1960).
7. Hadding-Koch, K.: Acoustico-phonetic studies in the intonation of Southern Swedish (Lund 1961).

8. Hadding-Koch, K. and Studdert-Kennedy, M.: Intonation contours evaluated by American and Swedish test subjects. Paper read at the 5th Int. Congr. Phonetic Sciences, August 16—23 (Münster 1964).
9. Hadding-Koch, K. and Studdert-Kennedy, M.: An experimental study of some intonation contours, Phonetica *11*:3/4 (1964).
10. Hermann, E.: Probleme der Frage. Nachrichten v. der Akademie der Wissenschaften in Göttingen *3—4* (1942).
11. Pike, K. L.: The intonation of American English (Ann Arbor 1945).
12. Stevens, S. S. and Davis, H.: Hearing, its psychology and physiology (New York 1938).
13. Uldall, E. T.: Ambiguity: question or statement? or "Are you asking me or telling me?", Proc. 4th Int. Congr. Phonetic Sciences, Helsinki 1961, pp. 779—783 (The Hague 1962).