AUDITORY AND LINGUISTIC PROCESSES IN THE PERCEPTION OF INTONATION CONTOURS*

Michael Studdert-Kennedy and Kerstin Hadding**

The perception of spoken language may be conceived as a process conducted at several successive and simultaneous levels. Auditory, phonetic, phonological, syntactic and semantic processes form a hierarchy, but decisions from higher levels also feed back to correct or verify tentative decisions at lower levels and to construct the final percept. Suitable experiments (e.g. Warren, 1970) may demonstrate the control exercised by higher on lower level decisions, and the partial determination of phonetic shape by phonological and syntactic rules is readily assumed by some linguists (e.g. Chomsky and Halle, 1968, p. 24). However, the auditory level, itself a complex of interactive processes by which an acoustic signal is converted into a representation suitable for input to the phonetic component (Fourcin, 1971), is commonly taken to be relatively independent.

A few studies have questioned this assumption. Ladefoged and McKinney (1963), for example, showed that judgments of the loudness of words presented in a carrier sentence may be more closely related to the work done upon them in phonation, that is, to their degree of stress, than to their acoustic intensity. Allen (1971), replicating and extending the experiment, showed that both acoustic level and inferred vocal effort may serve as cues for the loudness of speech, and that individuals differ in the weight they

* This study will also appear in a forthcoming Report from Haskins Laboratories. Some of the results were reported at the VIIth International Congress of Phonetic Sciences in Montreal, August 1971. (Studdert-Kennedy and Hadding, 1971).

** Haskins Laboratories, New Haven, Connecticut. M. Studdert-Kennedy also at Graduate Center and Queens College, City University of New York and K. Hadding also at Lund University, Sweden.

assign to these cues. Evidently, loudness judgment may entail a relatively complex process of inference, drawing upon more than one level of analysis. The same may be true of pitch judgment: Hadding-Koch and Studdert-Kennedy (1963, 1964, 1965) found that auditory judgments of listeners asked to assess fundamental frequency (f_0) contours imposed synthetically on a carrier word, seemed to be influenced by linguistic decisions. The present experiment extends this earlier work and by examining the relations among sections of the f_0 contour used in judging an utterance as a question or statement, attempts a more detailed understanding of auditory-linguistic interaction in the perception of intonation contours.[1]

The starting-point for the study is the importance commonly attributed to the terminal glide as an acoustic cue for judgment of an utterance as a question or statement. Two related sets of questions present themselves. The first concerns the basis for auditory judgments of the glide. From our earlier study (Hadding-Koch and Studdert-Kennedy 1963, 1964, 1965) it was evident that listeners frequently judge a falling glide as rising and a rising glide as falling. Is the origin of this effect auditory (psychophysical) or linguistic? Our study left the question unanswered. There, we systematically manipulated the contour of an utterance by varying f_0 at the stress peak, at the "turning-point" before the terminal glide, and at the end-point. We then asked listeners to classify each contour as (1) question or statement (linguistic judgment), (2) having a terminal rise or fall (psychophysical judgment). The two tasks yielded remarkably similar results: whether judging the entire contour linguistically or its terminal glide psychophysically, listeners were influenced in similar ways by the overall pattern of the contour. The outcome suggested that auditory judgments may have been controlled, in part, by linguistic judgments. But the reverse interpretation--that linguistic judgments of the entire contour were controlled by auditory judgments of the terminal glide--is equally plausible as long as

we do not know the auditory capacity of listeners for judging the terminal glides of matched non-speech contours.

The present study attempts to resolve this ambiguity by including the necessary non-speech judgments. Effects observed only in the two types of speech judgment would then be compatible with the first interpretation, while effects observed in all three types of judgment would be compatible with the second.

At the same time, this study broaches a second, related set of questions. These concern the roles of the various sections of the contour in determining linguistic judgments. Previous studies, both naturalistic and experimental, have suggested that listeners make use of an entire contour, not simply of the terminal glide, in judging an utterance (see Gårding and Abramson, 1965; Hadding-Koch, 1961; Hadding-Koch and Studdert-Kennedy, 1963, 1964, 1965). For example, spectrographic analyses of Swedish speech have shown that, in this language, "yes-no" questions normally display not only a terminal rise, but also an overall higher f_0 than statements (Hadding-Koch, 1961). Other utterances in which the speaker wants to draw the listener's special attention also display an overall high f_0 and a terminal rise: in listening tests the labels "question", "surprise", "interest" have been found to be interchangeable (Hadding-Koch, 1961, pp. 126 ff.). If a speaker is not interested or is asking a question to which he thinks he knows the answer, his utterances tend to display a lower overall f_0 and a falling terminal glide, similar to those of statements.

The importance of the entire contour may be reflected in the phonetic description. If four f_0 levels are postulated, with arrows showing the direction of the terminal glide, the intonation contour of a typical Swedish "yes-no" question could be described with one number at the beginning of the utterance and two at the stress,[3] as 3 44 $2\uparrow^3$ (the superscript 3 indicates the endpoint of the terminal glide) or, if less "interested", as

2 33 2↑[3]. A neutural statement would be best described as 2 33 1↓, or even 2 22 1↓, though the latter might also indicate a certain indifference. Much the same statement contour is typical of American English. However, questions in this language are said to display a more or less continuously rising contour (Pike, 1945; Hockett, 1955) which might be described as 2 22 3↑[4] or 2 33 3↑[4]. Similar contours occur in Swedish echo-questions.[4]

These naturalistic observations of speech are, in general, consistent with results of our experimental study of perception (Hadding-Koch and Studdert-Kennedy, 1963, 1964, 1965). Swedish listeners selected a typical Swedish question (2 44 2↑) among their preferred question contours, and a lower contour with a level terminal glide (2 33 1→) among their preferred statements (they would probably have preferred 2 33 1↓ for a statement had this contour been included). The North American listeners also preferred 2 44 2↑ for a question and 2 33 1→ for a statement, but they were more uncertain (in less agreement with one another) than the Swedish listeners--perhaps because the contours were based on Swedish speech and did not include, for example, a typical American English question.

Granted, then, the importance of the entire contour, we may now ask how its various sections work together to control linguistic judgment. Here, let us recall a central finding of our previous study, namely that there was perceptual reciprocity among various sections of a contour: listeners would trade a high f_0 at one point in the utterance for a high f_0 elsewhere. For example, an utterance with a relatively high f_0 at peak or turning-point required a smaller terminal rise to be heard as a question than an utterance with relatively low f_0 at peak or turning-point. We may interpret this reciprocity in either of two ways. The first interpretation assigns only auditory status to peak and turning-point, and assumes their linguistic role to be indirect. Thus, an utterance is marked as question or statement by its apparent terminal glide. Earlier sections of the contour are important only

insofar as they alter (by some mechanism to be specified) listeners' perceptions of that glide, and thereby give rise to the observed reciprocity effects. Lieberman's (1967) account of our results rests squarely on these assumptions. He selects an "analysis-by-synthesis" mechanism to account for the reciprocity.

An alternative interpretation assigns a direct linguistic function to peak and turning-point. An utterance is marked as question or statement not only by its terminal glide, but also by the f_0 pattern over its earlier course. Listeners discover at least two acoustic cues within a contour, either or both of which may control their linguistic decision. The weighting of these cues (by some unknown mechanism) gives rise to the reciprocity observed in linguistic judgments.

A second purpose of this study was to distinguish between these accounts, again by extending our earlier work to include judgments of the terminal glides of matched non-speech contours. Effects present in all three types of judgment would then require the first interpretation, but would exclude an account, such as that of Lieberman (1967), that invoked specialized speech mechanisms. Effects present only in the two types of speech judgment would be compatible with both the first interpretation and Lieberman's hypothesized mechanism. Effects present only in the linguistic judgments would require the second interpretation.

Finally, an additional purpose of the study was to extend our cross-linguistic comparison of Swedish and American English listeners. We therefore enlarged the set of contours to include typical questions and statements from both American English and Swedish.

METHOD

The stimuli were prepared by means of the Haskins Laboratories Digital Spectrum Manipulator (DSM) (Cooper, 1964). This device provides a spectro-

graphic display of a 19-channel vocoder analysis, digitized to 6-bits at 10 millisecond intervals, and permits the experimenter to vary the contents of each cell in the frequency-time matrix, before resynthesis by the vocoder. For the present study we were interested in the channel that displayed the time course of the fundamental frequency of the utterance, since it was by manipulating the contents of this channel that we varied f_o.

The utterance "November" [no'vɛmbɚ] was spoken by an American male voice into the vocoder and stored in the DSM. F_o was then manipulated over a range from 85 cps to 220 Hz. The f_o values at the most important points of the contours (starting-point, peak, turning point and end point) were chosen to represent four different f_o levels of a speaker with a range from 65 Hz. to 250 Hz.. The four levels were based on a previous analysis of a long sample of speech by a speaker with this particular range (Hadding-Koch, 1961, p. 110 ff).[5]

The contours are schematized in Figure 1. They range between two poles that may be marked 2 44 3↑[4] and 2 11 1↓. All contours start on a f_o of 130 Hz (level 2), sustained for 170 msecs., over the first syllable.[6] They then move, during 106 msecs., to one of three peaks: 130 Hz. (L, or low, level 2), 160 Hz. (H, or high, level 3), 200 Hz. (S, or superhigh, level 4). They proceed, during 127 msecs., to one of four turning points: 100 Hz. (high level 1), 120 Hz. (level 2), 145 Hz. (low level 3), 180 Hz. (high level 3). Finally, they proceed, during 201 msecs., to one of six end-points: 85 Hz. (level 1), 100 Hz. (high level 1), 120 Hz., 145 Hz., 180 Hz., and 220 Hz. (level 4). Peak, turning-point and end-point are each sustained for 32 msecs. The combination of three peaks, four turning-points and six end-points yields 72 contours, each specified by a letter and two numbers (e.g. S24, L36) and each lasting 700 msecs.

The 72 contours were recorded on magnetic tape from the output of the vocoder in three forms: (1) carried on a speech-wave [no'vɛmbɚ], (2) as a

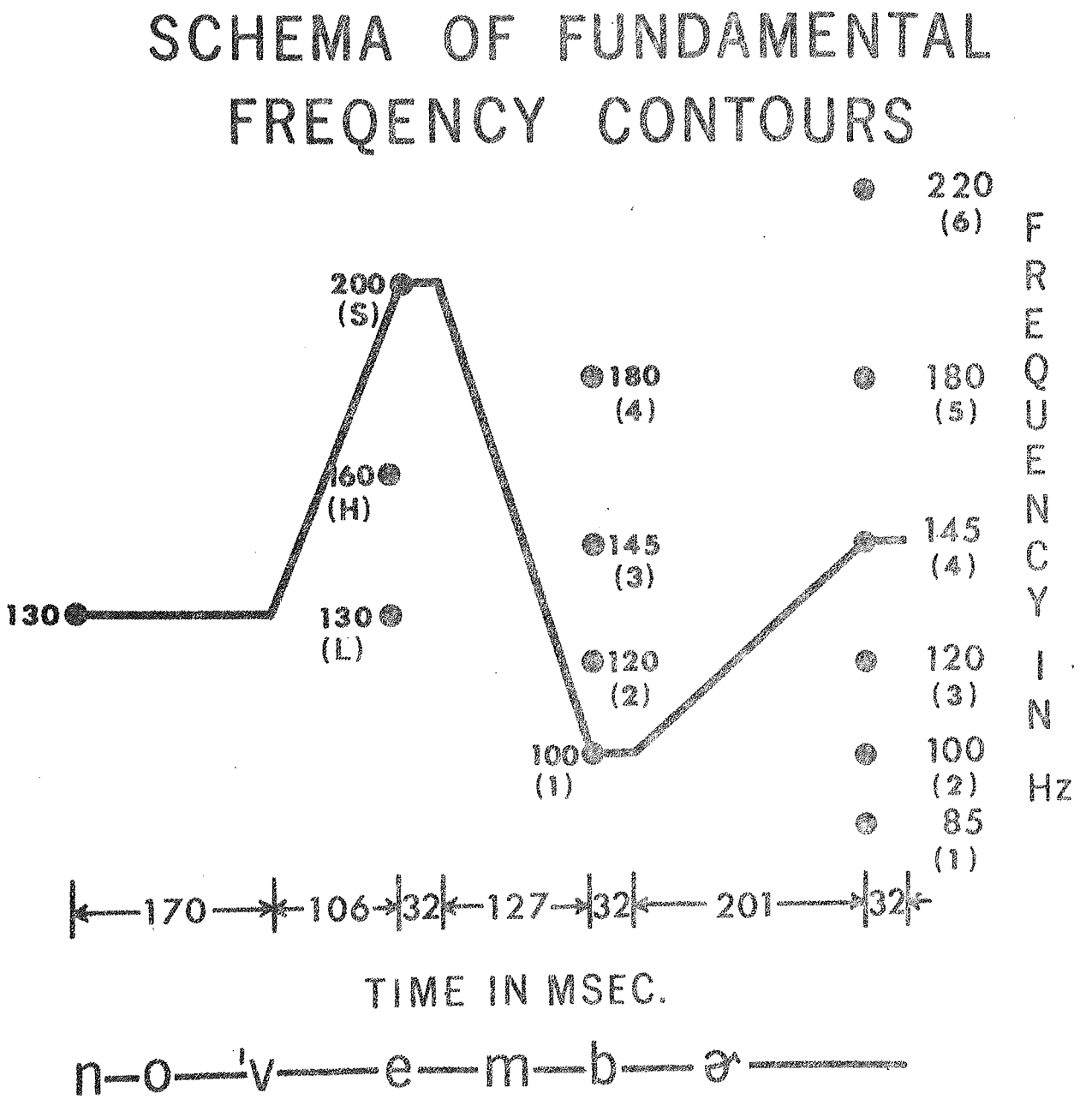

Figure 1. Schema of fundamental frequency contours imposed on the utterance "November" [nov&mbə].

frequency modulated sine wave, (3) as a frequency modulated train of pulses. Each set of 72 was spliced into 5 different random orders with a five-second interval between stimuli, a ten-second pause after every tenth stimulus, and presented to Swedish and U.S. subjects as described below.

Swedish Subjects. Twenty-two graduate and undergraduate volunteers were tested in three sessions, each lasting about forty-five minutes. They listened to the tests over a loud speaker at a comfortable listening level in a quiet room. In a given session they heard the five test orders for one type of stimulus only. They were divided into two groups of eleven. Both groups heard the sine-wave stimuli first; this was an important precaution intended to exclude any possible influence of speech mechanisms on judgments of the non-speech stimuli. In the second and third sessions both groups made psychophysical or linguistic judgments on the speech stimuli, group 1 in the order psychophysical-linguistic, group 2 in the reverse order. In the sine wave session and in the psychophysical speech session, subjects were asked to listen to the final glide of each contour and judge whether it was rising or falling. In the linguistic speech session subjects were asked to judge each contour as more like a question or more like a statement. For each contour, the procedure yielded 5 judgments by each subject under each condition, a total of 110 judgments in all.

U.S. Subjects. Sixteen female undergraduate paid volunteers were divided into two groups of eight. The procedure duplicated that followed with the Swedish subjects, except that the U.S. subjects listened to the tests over earphones in individual booths. The output of the phones was adjusted by means of a calibration tone to be approximately 75db SPL. These subjects also made psychophysical judgments on the pulse-train stimuli; these were counterbalanced with the sine waves in the first two sessions before the speech stimuli had been heard. The procedure yielded a total of 80 judgments on each contour under each condition.

RESULTS

No systematic differences between groups due to the order in which they made their judgments were observed. Data are therefore presented for the combined groups throughout. Figures 2 and 4 display the Swedish data, Figures 3 and 5 the U.S. data. In each figure the left column gives the linguistic, the middle column the speech psychophysical, and the right column the sine wave data.[6] Percentages of question add statement judgments (linguistic) or of rise and fall judgments (speech psychophysical and sine wave) are plotted against terminal glide, measured as rise (positive) or fall (negative) in Hz, from turning-point to end-point. In Figures 2 and 3 parameters of the curves are f_0 values at peaks (S, H, L), displayed for the four turning-point f_0 values from 1 (top) to 4 (bottom). In Figures 4 and 5 parameters of the curves are f_0 values at turning-points (1, 2, 3, 4) displayed for the three peak f_0 values of S (top), H (middle), and L (bottom).

Linguistic judgments

Cross-language comparisons

Before considering the acoustic variables controlling linguistic judgments, we will briefly compare Swedish and U.S. results. The main drift of the data is very similar for the two groups. A broad description of preferred statement and question contours for both groups can be given.

Statements. Figure 6 schematizes the most frequently preferred contours, those obtaining 90 % or better agreement. For all these contours, except two (L13; H13, Swedish only), the final f_0 of the terminal glide is the lowest f_0 of the utterance. In addition, the contours display at least one of the following: terminal fall, low or middle turning point (1, 2, 3), low or high peak (L, H). The range of preferred contours includes the 2 33 1↓ and 2 22 1↓ contours, suggested as typical by previous observations, but many others are equally acceptable. For example, the superhigh peak, even when

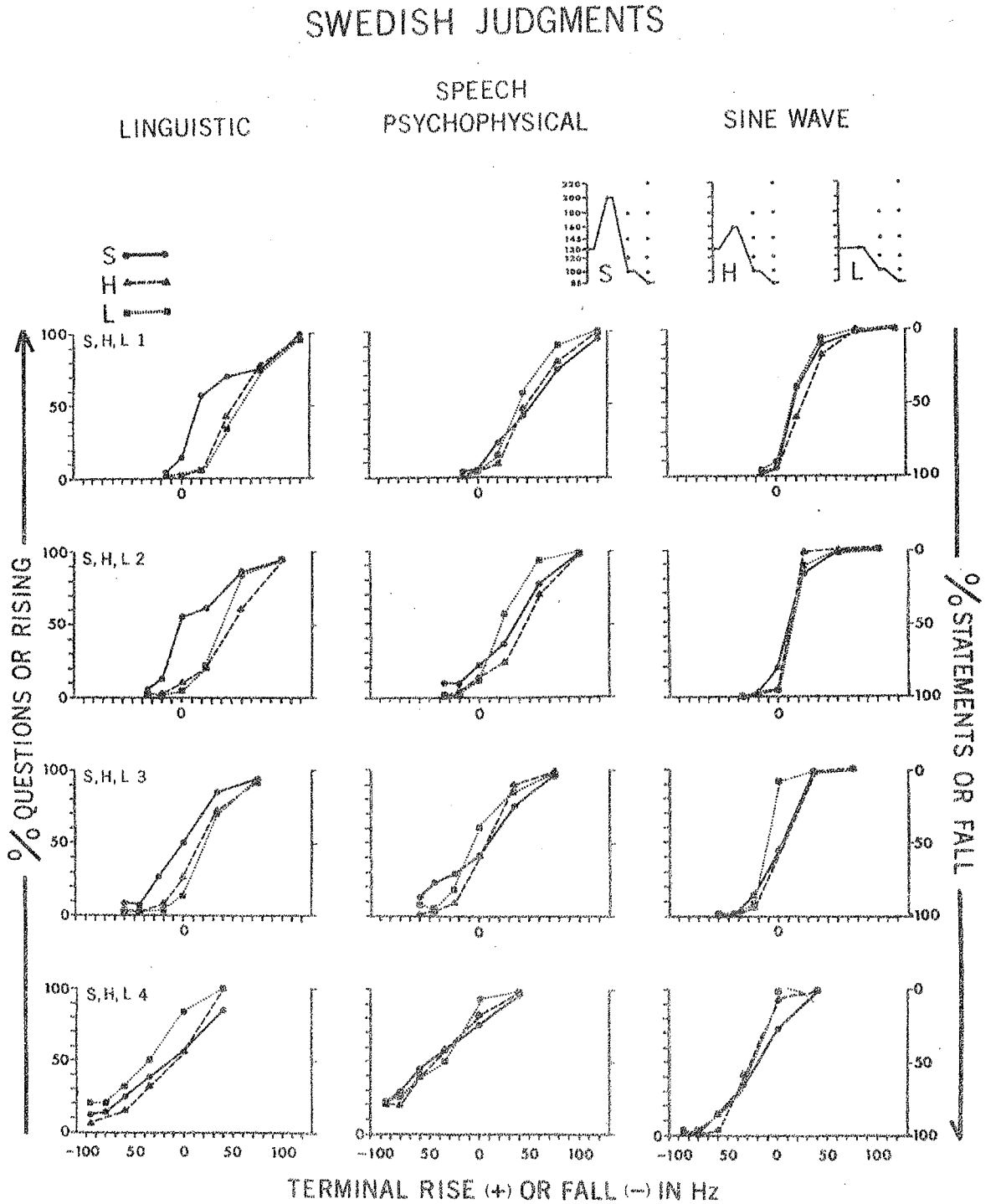

Figure 2. Percentages of question or rise responses (left-axis) and statement or fall responses (right-axis) plotted as functions of terminal glide in Hz. Peak values are constant across rows and turning-points are parameters of the curves. For Swedish subjects.

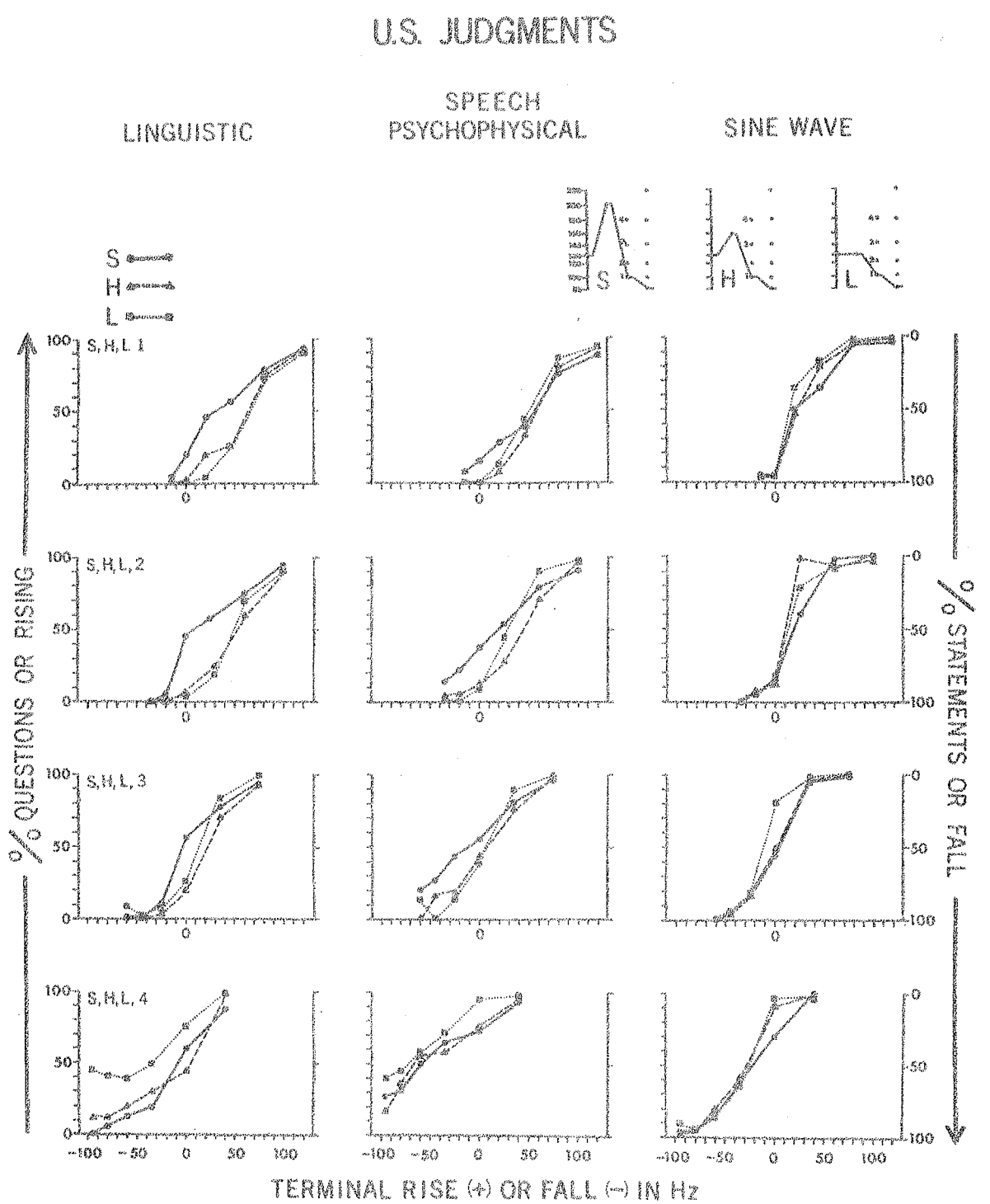

Figure 3. As for Figure 2, for the American subjects.

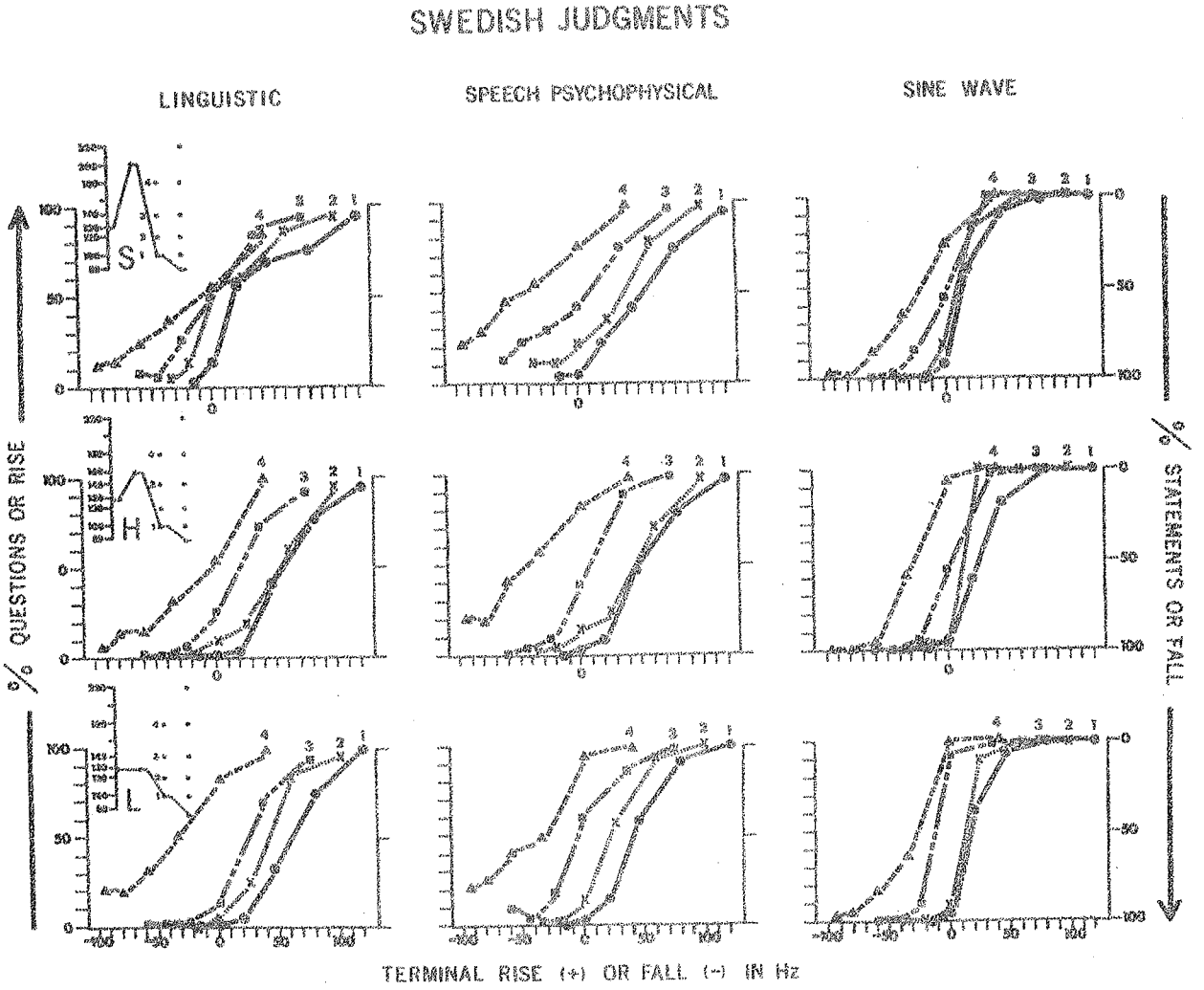

Figure 4. Percentages of question or rise responses (left-axis) and statement or fall responses (right-axis) plotted as functions of terminal glide in Hz. Turning-point values are constant across rows and peak values are parameters of the curves. For Swedish subjects.

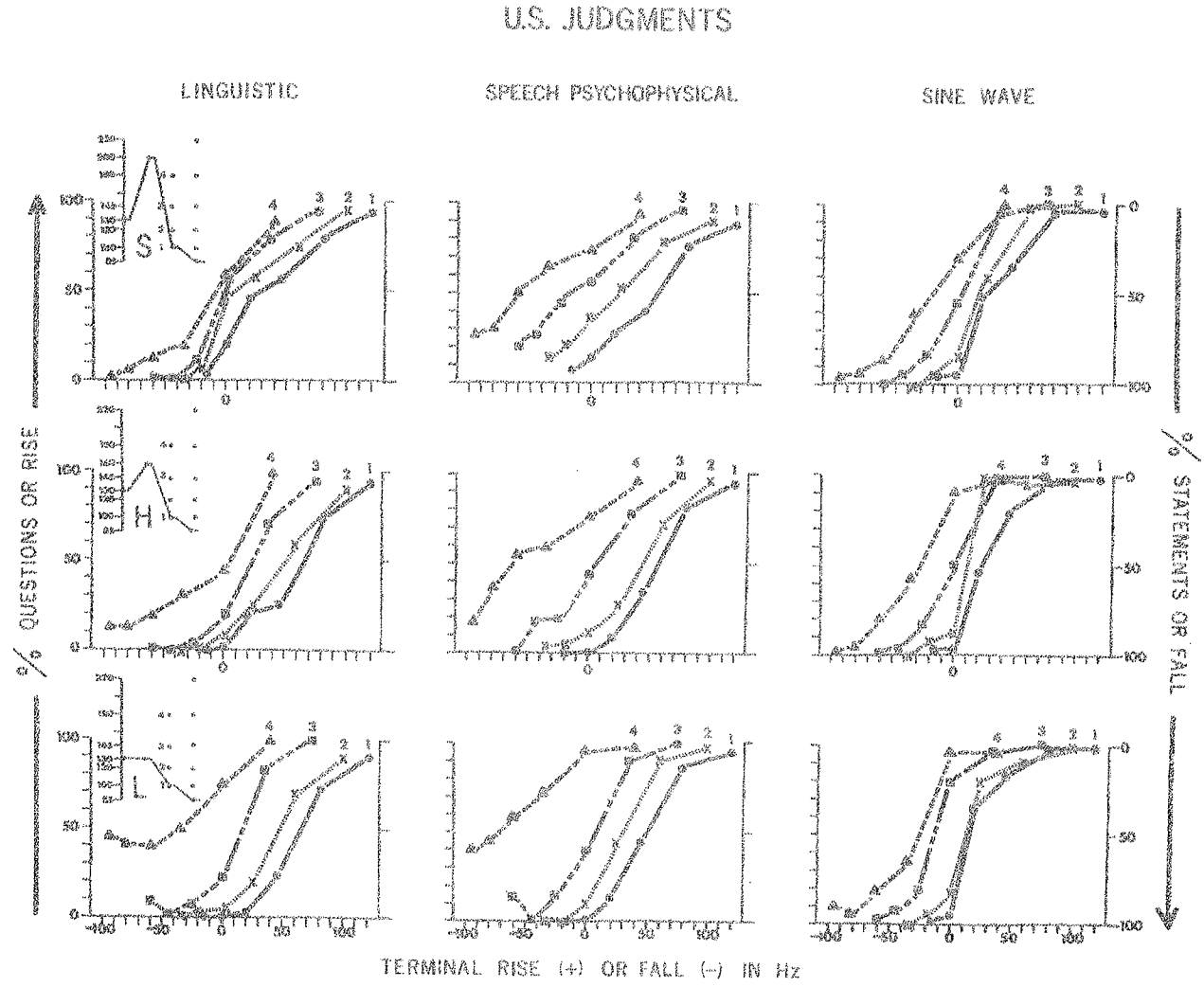

Figure 5. As for Figure 4, for the American subjects.

Schemata of Preferred Statement and Question Contours

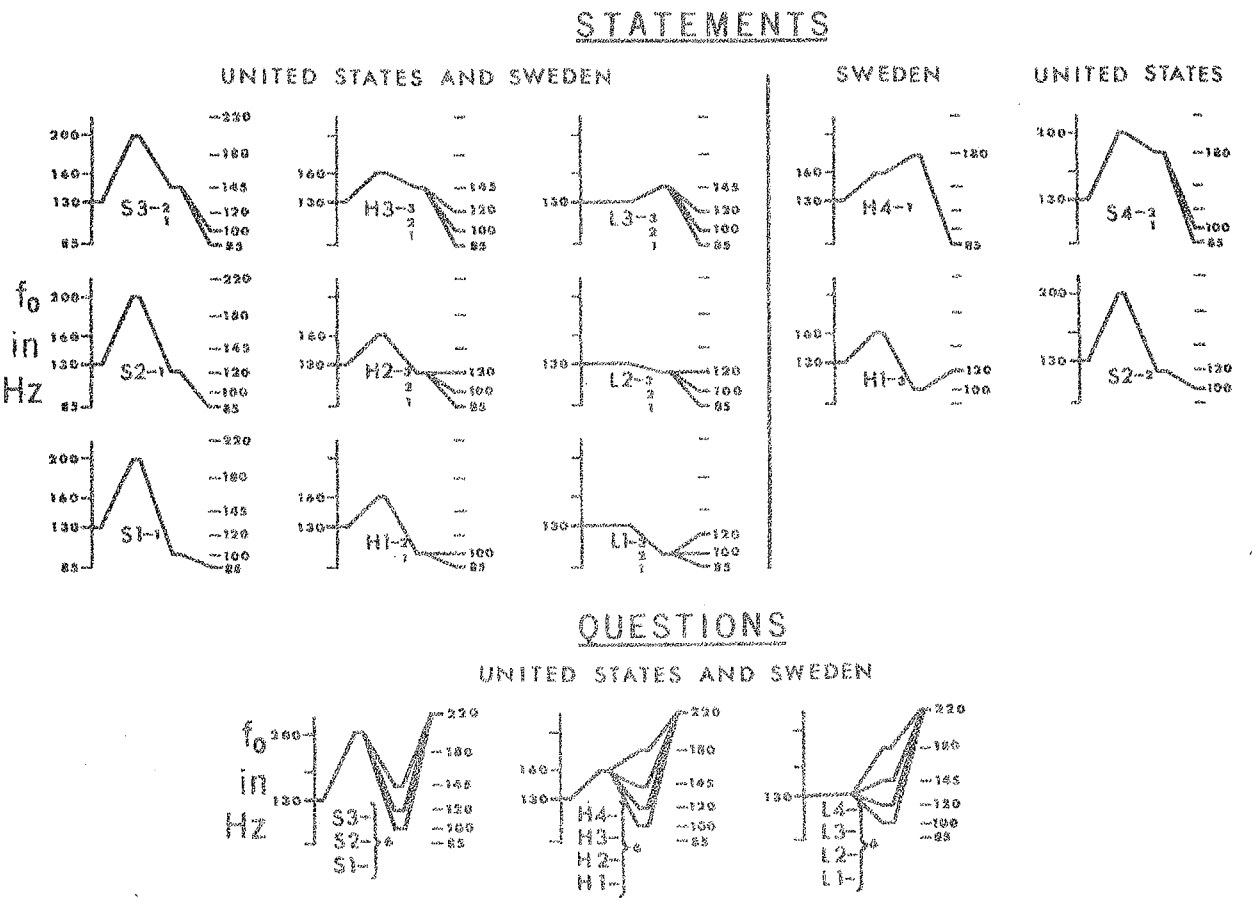

Figure 6. Included are all contours for which at least 90% of the judgments of a given language group were in a single category.

followed by a high (S4, US only) or moderately high (S3) turning-point, is accepted as a statement provided the terminal fall is large enough; the lower the turning-point (i.e. the larger the fall from the peak), the less the needed terminal fall (see S series, Figures 4 and 5). On the other hand, some terminally level contours (H23, H12, L23, L12) and even terminally rising contours (H13, Swedish only; L13) are also accepted as statements. Evidently the terminal fall is not essential, if preceding sections of the contour are low enough (L) or are falling from a moderate level (H).

Broadly, then, peak, turning-point and terminal glide engage in trading relations such that the contour of an acceptable statement has a low to high (rarely, and for US only, superhigh) peak and is, over some portion of its later course, low, falling or both. (Two anomalous series, H4 and L4, are discussed below under Swedish-U.S. differences.)

Questions. Figure 6 also schematizes contours obtaining 90 % or better agreement on a question judgment. For all these contours, the terminal glide is rising and the final pitch of the glide is the highest of the utterance (cf. Uldall, 1962, p. 780; Majewski and Blasdell, 1969). The range of preferred contours includes the expected continuously rising 2 22 $3\uparrow^4$ (L36, L46) and 2 33 $3\uparrow^4$ (H46) of American English and the superhigh peak contour 2 44 $2\uparrow^4$ (S26) of Swedish, but other contours are also accepted. For example, initially low and falling contours (L1, L2) are heard as questions, if the terminal rise is large enough. At the same time, even a terminally level contour (L45, Figures 2-5) gathers more than 80 % question judgments from both groups, when the preceding section of the contour has been steadily rising. In fact, this steady rise is a peculiarly powerful question cue that may quite override a large terminal fall that would otherwise cue a statement (cf. H4, L4, discussed below). Again there are trading relations among components of the contour, such that a generally accepted question displays either a rise from peak to turning-point (H4, L3, L4) and a rela-

tively small terminal rise, or a fall from peak to turning-point and a relatively large terminal rise.

Swedish-U.S. differences. As we have seen, the similarities between Swedish and U.S. judgments are more striking than the differences. The stimulus series included a number of contours presumably unfamiliar to one or other of both groups from their linguistic experience. Yet both groups were able to generalize such contours with more familiar patterns, classifying contours with a relatively high overall pitch as questions, contours with a relatively low overall pitch as statements. Nonetheless, small systematic differences are present.

(1) A comparison of Swedish and U.S. responses to the falling contours of the S2, S3, S4 series (Figures 4 and 5, top left) shows that U.S. subjects tended to give more statement responses than Swedish subjects. The effect is particularly marked for the S4 series on which Swedish statement judgments never reach 90 % agreement: a high peak with a high turning-point is difficult for Swedish subjects to hear as a statement. This may reflect the fact that Swedish statement intonation shows an earlier fall to a low level after stress than does English. At the same time, it may be taken as an indirect reflection of a Swedish preference for an overall high contour on questions, so that utterances displaying such a contour are difficult to hear as statements even when completed by a low terminal fall. It is true that the S4 series, which had been expected to collect a large number of question responses due to its overall high level, never obtained 90 % agreement on a question judgment from either group. But a control of these items revealed that they gave an impression of protest or indignation rather than of questioning, probably because the low precontour was heard in opposition to the rest of the utterance. A precontour on level 3 might have eliminated this impression and would also have been more similar to what actually occurs in Swedish questions. (cf. footnote 7).

(2) As was remarked above, the continuously rising contours (L4 and, to some extent, L3 and H4; see Figures 2 and 3, lower left) were readily accepted by both groups as questions, despite the fact that many of them are unlikely to occur in natural speech. L4, with its low peak rising 50 Hz. to the turning-point, and H4, with its high peak rising 20 Hz., were preferred to L3 with its low peak rising only 15 Hz. Furthermore, H4 and, especially, L4 elicited relatively few statement responses, even when their terminal glides were falling sharply. U.S. subjects identified these contours as statements even less frequently than the Swedish group. This may reflect the fact that the steadily rising question contour is more widely used in American English than in Swedish, and so might be peculiarly difficult for Americans to hear as a statement even when completed by a terminal fall.

In short, the differences between the two groups are small, but in directions predictable from linguistic analysis.

Variables controlling linguistic judgments.

Terminal glide is the single most powerful determinant of linguistic judgments. None of the highly preferred question contours and few of the highly preferred statement contours (Figure 6) lack the appropriate terminal rise or fall. Given a sufficiently extensive terminal glide, earlier sections of the contour have small importance. At the same time, Figures 2-5 show that f_o values at peak and turning-point may also play a role.

To provide a consistent criterion for the estimate of peak and turning-point effects, the median of the response distribution for each subject on each series was estimated. The median is the point of subjective equality, the value of the terminal glide at which subjects identify a given contour as a question or a statement 50 % of the time. In other words, it is the point of crossover from largely statement to largely question judgments.

The means of these medians, or crossover values, for the linguistic judgments are plotted in Figure 7 (row A) for Swedish subjects (left) and U.S. subjects (right). In the first and third plots mean medians are graphed as functions of peak f_0, with turning-point f_0 as parameter; in the second and fourth, they are graphed as functions of turning-point f_0, with peak f_0 as parameter.

Two cautions should be observed in studying these plots. First, it should be remembered that a median is a single value drawn from the center of its distribution. The relation between the medians of two distributions does not always accurately represent the relations between the upper and lower tails of those distributions. As long as two curves on any plot of Figures 2 to 5 are roughly parallel, the difference between their medians will give a reasonable estimate of their separation along the terminal glide axis. Where there are severe departures from the parallel, the appropriate plots of Figure 7 and of Figures 2 to 5 should be carefully read in conjunction. Second, it should be remembered that the mean of the medians of several distributions is not necessarily equal to the median of the combined distribution. Since the values of Figure 7 are the means of subject medians, they do not always agree exactly with the group median values read from Figures 2 to 5.

With these precautions in mind we return to row A of Figure 7. If the direction of the terminal glide were the sole determinant of linguistic judgments, we would expect all crossover values to fall at zero, the level of the dashed horizontal lines across Figure 7. In fact, crossover values deviate considerably from zero: both the direction and the extent of their deviation vary with peak and turning-point.

The peak effect (plots 1 and 3) is the smaller. For neither Swedish nor U.S. subjects does a change of peak f_0 from 130 Hz. to 160 Hz. (from L to H) have any consistent, significant effect. But a change from 160 Hz. to 200 Hz.

Mean Subject Medians Under the Three Experimental Conditions
for Swedish and American Subjects

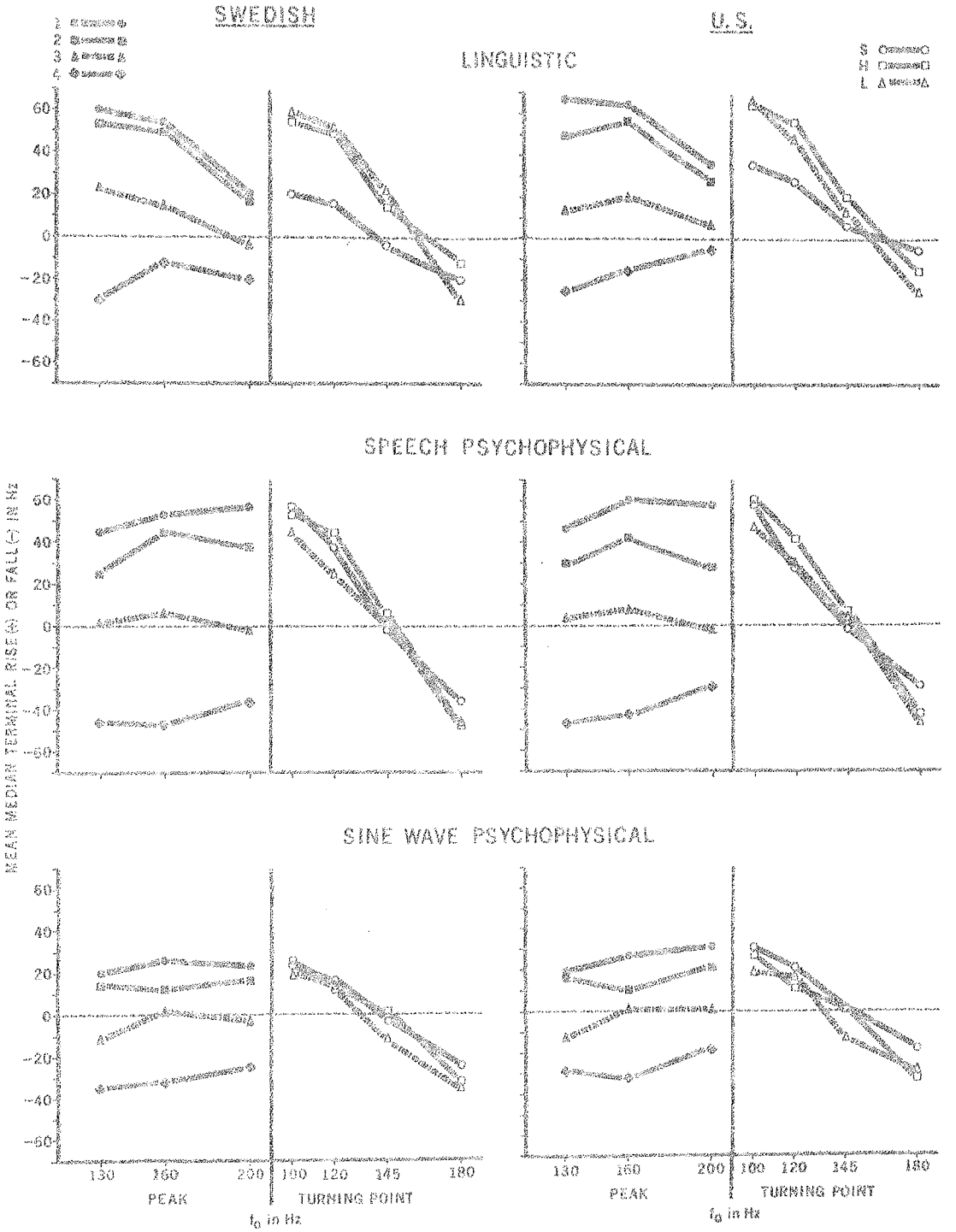

Figure 7. In the first and third columns mean medians are plotted as functions of peak f_0, with turning-point f_0 as parameter; in the second and fourth columns, they are plotted as functions of turning-point f_0, with peak f_0 as parameter.

(from H to S) does reliably reduce the crossover value for all contours, except that having a turning-point at 180 Hz. for the U.S. group. (This reversal is probably not reliable, as study of the bottom left plot of Figure 3 will suggest). These effects are statistically significant by matched pair t-tests between medians for turning-points 1, 2 and 3 in both groups ($p < .05$). They may be clearly seen in the left columns of Figures 2 and 3. Reading down the columns we note the leftward separation of the S curves. The separation is reduced for turning-point 3 and gives place to the L curve, with its steadily rising contour, for turning-point 4. We may also note that, as the terminal rise increases, the peak effect in the upper three plots disappears. In short, if the turning-point is at a low to middle f_o and the terminal rise is slight, a very high (level 4) peak at the stress leads to a significant increase in the number of questions heard and, by corollary, to a significant decrease in the number of statements.

The turning-point effect (plots 2 and 4 of Figure 7) is both larger and more consistent than the peak effect. For all values of peak f_o, an increase in turning-point f_o is associated with a decrease in crossover value. The decrease is significant by matched pair t-tests between medians ($p < .05$) for all turning-point shifts, except those from 100 to 120 Hz. for the Swedish S, H, and L curves and for the U.S. S and H curves. The effect is also considerably reduced, if the contour has a peak at 200 Hz. (S). (See top left plots of Figures 4 and 5). This again suggests that the high peak alone is a powerful question cue for both language groups.

Psychophysical judgments

Speech waves

Psychophysical judgments of the speech wave terminal glides differ from and resemble linguistic judgments of the entire utterance in important ways. The main difference may be seen in the center columns of Figures 2 and 3:

the effect of the high peak is absent from the Swedish data and much reduced in the U.S. data. The main similarity may be seen in the center columns of Figures 4 and 5: the turning-point effect is present and even more pronounced than in the linguistic judgments.

Figure 7 (row B) summarizes the data. The peak effects (plots 1 and 3) are inconsistent. An increase in peak f_o from 130 Hz. (L) to 160 Hz. (H) yields in every instance, except the high turning-point series for Swedish subjects, an increase rather than a decrease in the crossover value of the terminal rise. Two of these increases (for turning-points 1 and 2) are significant for both groups ($p < .05$ by a matched pair t-test between medians). On the other hand, an increase of peak f_o from 160 Hz to 200 Hz. yields, for the Swedish subjects, two increases, two decreases, neither of them significant. The absence of a consistent peak effect for the Swedish subjects is evident in the middle column of Figure 2. For the U.S. subjects, the picture is somewhat different: crossover values decrease from H to S for turning-points 1, 2 and 3 and increase for turning-point 4, exactly as in the linguistic data. The effects are reduced and statistically significant only for turning-point 2. But a trend is present and quite evident in the middle column of Figure 3.

The turning-point effect, on the other hand (center columns of Figures 4 and 5; plots 2 and 4, row B of Figure 7) is similar to and even more pronounced than the corresponding effect in the semantic data. All shifts are significant by matched pair t-tests ($p < .05$), except that from turning-point 1 to 2 in the Swedish L series. For both groups, the higher the turning-point, the smaller the terminal rise needed for a rise to be consistently heard. The similarity to the linguistic results is most marked for the H and the L series (second and third rows, Figures 4 and 5): H4 and L4 are again anomalous series, readily heard as rising even when the terminal glide is falling. In the S series the turning-point effect is even more pronounced than for the linguistic judgments.

Sine waves

From the steepened functions of Figures 2 to 5 (right-hand columns) it is evident that subjects were in better agreement on their sine wave than on their speech psychophysical or linguistic judgments. The two language groups are also in close agreement, which gives some confidence that the differences between their linguistic judgments are reliable.

Figures 2 and 3 (right-hand columns) show that the effect of the high peak is absent. As in the speech psychophysical data, low peak contours tend to be the most accurately judged, particularly by the Swedish. But the effects are neither fully consistent nor statistically significant (see plots 1 and 3, row C, Figure 7).

On the other hand, the turning-point effects (plots 2 and 4, row C, Figure 7) are clear, similar to those observed in the linguistic and speech psychophysical data, but considerably reduced. The effects are significant by matched pair t-tests ($p < .05$) for all turning-point shifts except those from 100 Hz. to 120 Hz. for the S and H curves in both groups, and may be seen in the right-hand columns of Figures 4 and 5. Note that H4 and L4 are no longer anomalous series.

DISCUSSION

Cross-language comparisons

There are striking similarities between Swedish and U.S. judgments of these intonation contours. Despite small, linguistically predictable differences, both groups tend to classify contours with a high peak or terminal rise as questions, contours with a low peak or terminal fall as statements. Hermann (1942) has pointed out the generality across languages, including Swedish, of a high pitch for questions (see also Hadding-Koch, 1961, especially pp. 119 ff.).Bolinger (1964), among others, has discussed

the apparently "universal tendency" to use a raised tone to indicate points of "interest" within utterances and also to indicate that more is to follow, as in questions (cf. Hadding-Koch, 1965). The data of this experiment are consistent with these "universal tendencies".

Perceptual relations within a contour

We are now in a position to resolve some of the uncertainties left by our previous study. Consider, first, the turning-point effect. Since this is present and significant under all three experimental conditions, we must assign it auditory status and assume that it takes linguistic effect indirectly by altering subjects' perceptions of the terminal glide. Furthermore, since it is present, even though reduced, in the sine wave data, our account of the process by which it affects perception of the terminal glide cannot invoke specialized mechanisms peculiar to speech.

We may gather some idea of the process from a study of plots 2 and 4 in row B, Figure 7 or of the center plots in Figures 4 and 5. The terminal glide of a contour, such as H1, with a strong fall from peak to turning-point (160 Hz. to 100 Hz.) requires a terminal rise of about 50 Hz. if it is to be judged 50 % of the time as rising; while the terminal glide of a contour, such as H4, with a steady rise for more than 200 msecs before the terminal glide, is heard as rising 50 % of the time, even when the glide is falling by about 50 Hz. Evidently listeners have difficulty in separating the terminal glide from earlier sections of the contour, if those earlier sections have a marked movement. The terminal glides of contours with a turning-point (145 Hz. in S3, H3, L3) close to the precontour level of 130 Hz. are more accurately perceived: the median values are close to zero in every plot of Figure 7, columns 2 and 4. Listeners are perhaps able to average across earlier sections of such contours, and establish an anchor against which terminal glide may be judged.

All this implies that later sections of the contours in this study (that is, roughly the last 400 msecs., from peak to turning-point to end point) were processed by listeners as a single unit, with attention focussed on the terminal glide. If a listener was able to separate the glide perceptually from the immediately preceding section (as in the S3, H3, L3 series), his linguistic judgments followed pretty well the traditional formulation of rise for questions, fall for statements. If he was not able to separate the glide, due to the difficulty--heightened perhaps for a complex signal-- of tracking a rapidly modulated frequency, relatively gross movements of the terminal glide were necessary for him to be sure whether he had heard a rise or a fall, a question or a statement.

Interpretation of the peak effect is more difficult. In our earlier study, the effect was clear in both linguistic and psychophysical judgments of both groups, though the Swedish were less consistent in their psychophysical judgments than the Americans. In this study, a peak effect is significantly present in linguistic judgments, totally absent from sine-wave judgments and, for speech psychophysical judgments, marginally present only for the Americans.

We will consider the speech psychophysical data below. Here, the important point is that the peak effect is reliably present in the linguistic, but absent from the sine-wave judgments. We may therefore, with reasonable certainty, reject an auditory (or psychophysical) account, and assign a direct linguistic function to the peak. Unlike turning-point variations, peak variations do not take linguistic effect by altering listeners' perceptions of the terminal glide. Rather, the peak is a distinct element to be weighed with the perceived terminal glide in determining the linguistic outcome.

We should note, in caution, that peak and terminal glide are not always simply additive in their effects. For example, a contour with a steady

rise from precontour to endpoint may require a relatively small terminal rise to be heard as a question, despite its low peak (e.g. L3 series). Here, it seems to be the overall sweep of the pattern that determines the judgment rather than the frequency levels of particular segments of the contour.

However, with few exceptions, two factors would seem to govern linguistic judgments of intonation contours, such as those of this study: fundamental frequency at the peak and perceived terminal glide. The entire contour is then interpreted as a unit with these factors in weighted combination, and with the heavier weight being assigned to the terminal glide. If a terminal fall is heard, the listener interprets the utterance as a statement, unless the fall was slight and he has also heard a very high peak; if a terminal rise is heard, the listener interprets the utterance as a question, unless the rise was slight and he has also heard an unusually low peak (cf. Greenberg, 1969, Ch. 2; Ohala, 1970, pp. 101 ff.).

Auditory-linguistic interactions

We turn, finally, to the speech psychophysical data. Our problem is to understand the instances in which speech psychophysical judgments follow the linguistic more closely than the sine-wave judgments. Obviously, these instances can only occur where linguistic judgments of the entire contour differ from auditory judgments of the terminal sine-wave glide, that is, where the contour carries some linguistically relevant cue other than terminal glide. For questions, such cues include a super-high peak or a monotonic rise from precontour to turning-point. Accordingly we find a tendency for speech psychophysical judgments to follow linguistic judgments in the superhigh (S) peak series (see Figure 3) and in the high turning-point series (see Figures 4 and 5). Consider, particularly, the results for speech contours of the H4 and L4 series. Listeners in both groups often judge these contours both as questions and as terminally rising, even though they are

able to hear that the corresponding sine-wave contours have terminal falls. Since listeners cannot have judged the contours to be questions because they heard a terminal rise, we are tempted to conclude that they heard the terminal rise because they judged the contours to be questions: linguistic decision determined auditory shape.

Before elaborating on this, it is important to remark that such effects do not always occur where they might be expected. For example, the peak effect was clearly present in the speech psychophysical judgments of both groups in our earlier study, but is reduced to a marginal effect in the American and has disappeared entirely from the Swedish speech psychophysical data of the present study. We can hardly therefore call on the effect to support a general account in terms of some specialized perceptual mechanism, such as that proposed by Lieberman (1967). At the same time, the results are evidently peculiar to speech and cannot be handled in purely auditory terms. What we need therefore is an account in terms of a process that may vary with experimental conditions and subjects.

An interesting hypothesis, suggested above, is that the results reflect the blend of serial and parallel processing that characterizes the perception of spoken language (and of other complex cognitive objects) (cf. Fry, 1956; Chistovich, et al., 1968; Studdert-Kennedy, in press). We may conceive the perceptual process as divided into stages (auditory, phonetic, phonological, etc.), but we must also suppose there to be feedback from higher to lower levels which may serve to correct or verify earlier decisions. Perceptual "correction" of an auditory or phonetic decision, in light of a higher linguistic decision, will presumably not occur if the lower decision is firm. Otherwise, we would not be able to deem the intonation of an actor "wrong", or understand a speaker, yet perceive his dialect to be unfamiliar. However, in difficult listening conditions and

under certain, as yet undefined, acoustic conditions, perceptual "correction", sufficient to produce a compelling phonetic illusion, may occur (Miller, 1956). Warren (1970; Warren and Obusek, 1971) has shown that listeners may clearly perceive a phonetic segment that has been excised from a recorded utterance and replaced by an extraneous sound (cough, buzz, tone) of the same duration. The important point is that listeners perceive the correct segment: the precise form of the phonetic illusion is determined not by the acou stic conditions alone, but also by higher order linguistic constraints.

Here, the illusion is auditory rather than phonetic, but a similar mechanism may be at work. Asked to interrupt his normal perceptual process at a pre-phonetic auditory stage, the listener falls back on his knowledge of the language. As we have seen, the single most powerful cue for question/statement judgments in this experiment was the terminal glide. Listeners evidently prefer, and presumably expect, a question to end with a rise, a statement with a fall (see Figure 6). However, earlier sections of the contour may also enter into the decision, and, if sufficiently marked, override an incompatible, but relatively weak terminal glide. Called upon to judge this glide, the listener then assigns it a value consonant with his linguistic decision. That is to say, if other factors dominate his linguistic decision, he may be led into non-veridical perception of the terminal glide.

The degree to which this happens might be expected to vary with the relative strength of the cues controlling linguistic decision. And in fact, just as the peak effect in the linguistic data was stronger for our first study than for our second, so too was the peak effect in the speech psychophysical data. Similarly, just as the question cue in the rising contours or the H4 and L4 series is stronger for the Americans than for the Swedish,

so too is the tendency toward non-veridical judgment of the terminal glide.

However, we should not expect to be able to develop a fully coherent account of our results in these terms, since we are ignorant of the limiting linguistic and acoustic conditions of the illusion. We are currently planning to broaden our understanding of the effect by taking advantage of what is known about the various acoustic cues to word stress (Fry, 1955, 1958). We might expect, for example, that, if linguistic decision can indeed determine auditory shape, syllables of equal duration, judged to be differently stressed on the basis of differences in either intensity or fundamental frequency, would also be judged of unequal length. The ultimate interest of the account is in its suggestion that the auditory level is not independent of higher levels, but is an integral part of the process by which we construct our perceptions of spoken language.

ACKNOWLEDGMENT

Work on this paper was supported in part by a grant to Haskins Laboratories from the National Institute for Child Health and Human Development, Washington, D.C.

FOOTNOTES

1. The acoustic correlates of intonation are said to be changes in one or more of three variables: fundamental frequency, intensity and duration, with variations in fundamental frequency over time being the strongest single cue (Bolinger, 1958; Denes, 1959; Fry, 1968; Lieberman, in press; Lehiste, 1970). The present study is concerned with only one of these variables, fundamental frequency, and the term "intonation contour" refers exclusively to contours of fundamental frequency.

2. Many workers who have reported, for various languages, that the same intonation is used in questions as in statements, seem to have been anxious to exclude all emotional "overtones" and therefore told their subjects to speak in a neutral voice. The result is that, in the absence of grammatical Q-markers, utterances sound like statements. A "neutral" intonation is not enough to convey, as sole cue, the impression of a question. If a question is asked merely for form's sake, with no particular interest in the answer, no difference in intonation is to be expected from that of a statement.

3. We write two numerals at the stress and one at the turning-point, even though they may be on the same "level" (intonation level, f_0 level), (cf. Hadding-Koch, 1961; Delattre, 1963; Hockett, 1955).

4. Compare the similar difference in intonation contours for French suggested by Léon, 1971.

5. One of the contentions of that study, based on a number of utterances in continuous speech by several Swedish subjects, was that every speaker has, in addition to a general speaking range, clusters of "favorite pitches" which he uses, for instance, on stressed segments of statements (represented by the H-peak in the present study), and a higher level

which he uses for questions and various expressions of "interest" (here represented by level 4; see Hadding-Koch, 1961; cf. also Bolinger, 1964).

Statements were found in that study to end on a low level, hesitant or exclamatory utterances higher up. Questions tended to have a terminal rise, usually from level 2, or a fall ending comparatively high. Questions were also generally spoken with an overall high f_0 compared to statements, a phenomenon that, according to the literature, occurs in many languages (Herman, 1942; Bolinger, 1964). The contour then often started high. Polite or friendly statements too might end with a final rise, but from a comparatively low level and with a moderate range (cf. Uldall, 1962).

6. Judgments of the modulated sine-waves and pulse trains by U.S. subjects were essentially identical. Accordingly, only sine-wave data are presented here.

7. We should probably have included a higher precontour, on level 3, to cover the question contours properly, since the large rise to the highest peak (from level 2 to level 4) gave some contours an unwanted and perhaps dominating effect of protest rather than question (cf. footnote 5). However, this would have meant a substantial increase in an already lengthy test.

REFERENCES

Bolinger D.L. 1958. A theory of pitch in English. Word 14,,109-149

Bolinger D.L. 1964. Intonation as a universal. Proc. IXth Intl. Cong. Linguistics. Cambridge, Mass. 1962. The Hague: Mouton 833-848

Chistovich L.A., Golusina A., Lublinskaja F., Malinnikova T., and Zukova M. 1968. Psychological methods in speech perception research. Z. Phon. Sprachwiss. u. Komm. Fschg. 21, 33-39

Chomsky N. and Halle M. 1968. The sound pattern of English. New York: Harper & Row

Cooper F.S. 1965. Instrumental methods for research in phonetics. Proc. Vth Intl. Cong. Phonet. Sci. Münster 1964. Basel. 142-171

Delattre P. 1963. Comparing the prosodic features of English, German, Spanish and French. IRAL 1, 193-210

Denes P. 1959. A preliminary investigation of certain aspects of intonation. Language and Speech 2, 106-122

Fourcin A.J. Perceptual mechanisms at the first level of processing. To appear in Proc. VIIth Intl. Cong. Phonet. Sci. Montréal 1971

Fry D.B. 1955. Duration and intensity as physical correlates of linguistic stress. J. Acoust. Soc. Am. 27, 765-768

Fry D.B. 1956. Perception and recognition in speech. In Halle M., Lunt H.G., and Van Schoonefeld C.H. (eds.),For Roman Jakobson. The Hague: Mouton. 169-173

Fry D.B. 1958. Experiments in the perception of stress. In Language and Speech 1, 126-152

Fry D.B. 1968. Prosodic phenomena. Manual of Phonetics, Malmberg B. (ed.), Amsterdam: North Holland Publ. Co. 365-410

Gårding E. and Abramson A.S. 1965. A study of the perception of some American English intonation contours. Studia Linguistica XIX, 61-79

Greenberg S.R. 1969. An experimental study of certain intonation contrasts in American English. UCLA Working Papers in Phonetics 13

Hadding-Koch K. 1961. Acoustico-phonetic studies in the intonation of Southern Swedish. Lund: Gleerups

Hadding-Koch K. 1965. On the physiological background of intonation. Studia Linguistica XIX, 55-60

Hadding-Koch K. and Studdert-Kennedy M. 1963. A study of semantic and psychophysical test responses to controlled variations in fundamental frequency. Studia Linguistica XVII, 65-76

Hadding-Koch K. and Studdert-Kennedy M. 1964. An experimental study of some intonation contours. Phonetica 11, 175-185

Hadding-Koch K. and Studdert-Kennedy M. 1965. Intonation contours evaluated by American and Swedish test subjects. Proc. Vth Intl. Cong. Phonet. Sci. Münster 1964. Basel. 326-331

Hermann E. 1942. Probleme der Frage. Nachrichten von der Akademie der Wissenschaften in Göttingen 3-4

Hockett C.F. 1955. A manual of phonology. Indiana Univ. Publ. in Anthropology and Linguistics 11

Ladefoged P. and McKinney N.P. 1963. Loudness, sound pressure and subglottal pressure in speech. J. Acoust. Soc. Am. 35, 454-460

Lehiste I. 1970. Suprasegmentals. Cambridge, Mass.: The M.I.T. Press

Léon P.R. Où en sont les études sur l'intonation. To appear in Proc. VIIth Intl. Congr. Phonet. Sci. Montréal 1971

Lieberman P. 1967. Intonation, perception, and language. Cambridge, Mass.: The M.I.T. Press

Lieberman P. (in press). A study of prosodic features. In Sebeok, T.A. (ed.), Current Trends in Linguistics, Vol. XII. The Hague: Mouton (Also in Haskins Laboratories SR 23, 1970)

Majewski W. and Blasdell R. 1969. Influence of fundamental frequency cues on the perception of some synthetic intonation contours. J. Acoust. Soc. Am. 45, 450-457

Ohala J. 1970. Aspects of the control and production of speech. UCLA Working Papers in Phonetics 15

Pike K.L. 1945. The intonation of American English. Ann Arbor

Studdert-Kennedy M. (in press). The perception of speech. In Sebeok, T.A. (ed.), Current Trends in Linguistics, Vol. XII. The Hague: Mouton (Also in Haskins Laboratories, SR 23, 1970)

Studdert-Kennedy M. and Hadding K. Further experimental studies of fundamental frequency contours. To appear in Proc. VIIth Intl. Cong. Phonet. Sci. Montréal 1971

Uldall E.T. 1962. Ambiguity: question or statement? or "Are you asking me or telling me?" Proc. IVth Intl. Cong. Phonet. Sci. Helsinki 1961. The Hague: Mouton 779-783

Warren R.M. 1970. Perceptual restoration of missing speech sounds. Science 167, 392-393

Warren R.M. and Obusek C.J. 1971. Speech perception and phonemic restorations. Perception and Psychophysics 9, 358-362