ARE YOU ASKING ME, TELLING ME OR TALKING TO YOURSELF?

Kerstin Hadding* and Michael Studdert-Kennedy**

Haskins Laboratories, New Haven

In a study of Swedish intonation, Hadding-Koch (1961) distinguished among three functional categories of utterance and their correlated fundamental frequency (f_0) contours. The first category ("question") occurred when a speaker wanted an answer from a listener; it was characterized by a relatively high f_0 at the stress peak and a rising terminal glide. The second category ("statement") occurred when a speaker wanted a listener to believe or agree with him; it was characterized by a lower f_0 at the stress peak and a falling glide. Later perceptual studies of synthetic speech, in which the f_0 contour of an utterance was systematically varied, have largely supported these descriptive analyses (Hadding-Koch and Studdert-Kennedy 1964, 1965a and b; Studdert-Kennedy and Hadding 1972; in press. Listeners tended to classify contours with an apparent terminal rise and/or high f_0 at the stress as questions, contours with an apparent terminal fall and/or low f_0 at the stress as statements (cf. Uldall, 1962).

The third category of utterance, described by Hadding-Koch, had a level terminal glide ("terminal sustain"). With a relatively even and moderately high overall f_0, this type of contour occurred when the speaker was musing or talking to himself. With various other f_0 patterns in earlier sections of the contour, level terminal glides also occurred in exclamations and in some other type of utterances expressing a somewhat emotional reaction. These are not treated here. Common to all these contexts is the fact that the speaker was not primarily interested in eliciting a listener's response - in fact, no listener need be present at all. Moravcsik (1971) quotes Householder as differentiating "statements which disclaim knowledge, but

* Also Lund University, Sweden

** Also Graduate Center and Queen's College, City University of New York

exhibit indifference towards obtaining it from real questions by a feature (±Hearer) indicating hearer's involvement" (p. 81, fn 1). We would like to propose a similar feature though with a somewhat different definition.

As a first step, the present study was intended to assess the perceptual validity of the third category. The hypotheses were that (1) listeners can reliably identify fundamental frequency contours which display a level terminal glide rather than a terminal rise or fall, (2) listeners can reliably form a category of utterances defined by the speaker's talking to himself rather than addressing a listener, (3) "talking-to-self judgments, if they occur, are made of contours characterized by a moderate, even f_0, ending with a level glide.

Method

The stimuli were those used in a previous study (Studdert-Kennedy and Hadding, 1972; in press). They were prepared by means of the Haskins Laboratories Digital Spectrum Manipulator (DSM) (Cooper, 1965). This device provides a spectrographic display of a 19-channel vocoder analysis, digitized to 6 bits at 10 msec 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 displays the time course of the fundamental frequency of the utterance, since it was by manipulating the contents of this channel that we varied f_0.

The utterance "November" [no'vɛmbɚr] was spoken by an American male voice into the vocoder and stored in the DSM. F_0 was then manipulated over a range from 85 Hz to 220 Hz. The f_0 values at the most important points of the contours (starting point, peak, turning point, and end point) were chosen to represent four different f_0 levels of a speaker with a range from 65 Hz to 250 Hz. The four levels were based on a

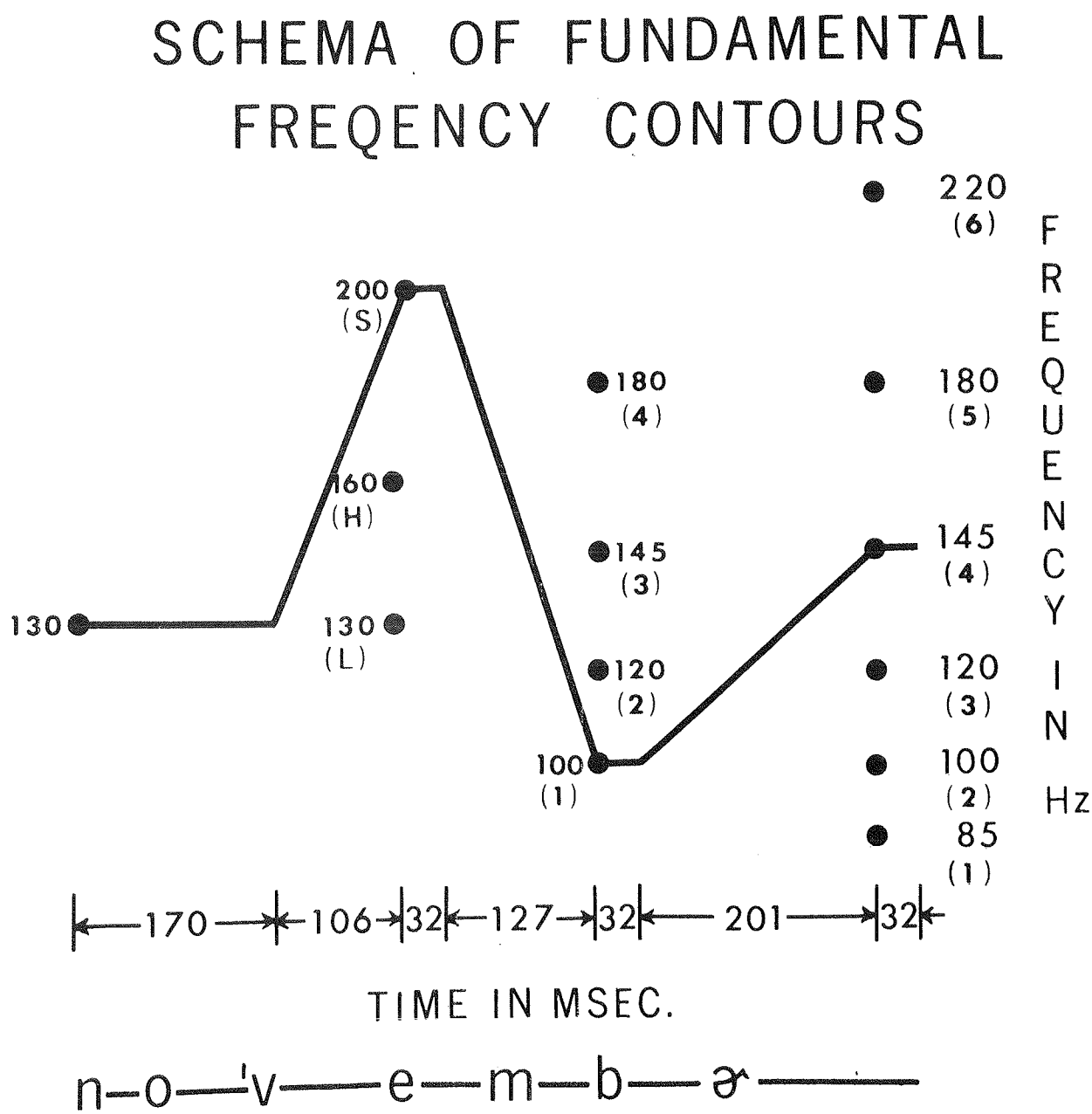

Figure 1. Schemata of fundamental frequency contours imposed on the utterance "November" [no'vɛmbɚr]

previous analysis of a long sample of speech by a speaker with this particular range (Hadding-Koch, 1961, pp. 110 ff.).

The contours are schematized in Figure 1. All contours start on a f_0 of 130 Hz, sustained for 170 msec, over the first syllable (the "pre-contour"). They then move, during 106 msec, to one of three "peaks": 130 Hz (L, or low), 160 Hz (H, or high), 200 Hz (S, or superhigh). They proceed, during 117 msec, to one of four turning points: 100 Hz (1), 120 Hz (2), 145 Hz (3), 180 Hz (4). Finally, they proceed, during 201 msec, to one of six end-points: 85 Hz (1), 100 Hz (2), 120 Hz (3), 145 Hz (4), 180 Hz (5), and 200 Hz (6). Peak, turning point, and end point are each sustained for 32 msec. The combination of three peaks, four turning points, and six end points yields 72 contours, each specified by a letter and two digits (e.g., S14 for the contour of Figure 1) and each lasting 700 msec.

The 72 contours were recorded on magnetic tape from the output of the vocoder in two forms: (1) carried on a speech wave [no'vɛmbər], (2) as a frequency-modulated sine wave. Each set of 72 was spliced into five different random orders with a five-second interval between stimuli and a ten-second pause after every tenth stimulus.

A group of 22 Swedish graduate and undergraduate volunteers (10 of whom had served as subjects in our earlier experiments) was tested in a series of three sessions, each lasting about 45 minutes. They listened to the tests over a loudspeaker in a quiet room. In a given session they heard the five test orders for one type of stimulus only. All subjects heard the sine wave stimuli in their first session (so as to reduce the possible influence of speech mechanisms on judgments of nonspeech stimuli). Half the group then made linguistic judgments of the speech stimuli in their second session, psychophysical judgments of the same stimuli in their third session; half the group took the tests in reverse order.

In the sine wave session and in the speech psychophysical session, subjects were asked to listen to the terminal glide of each contour and to judge whether it was rising, falling or level in pitch. In the linguistic session, they were asked to picture three situations: a speaker addressing a question to a listener, a speaker making a statement to a listener, and a speaker not addressing a listener, but talking to himself. The subjects' task was then to listen to each utterance and assign it to its appropriate category: question, statement or "talking-to-self". The third category is not, of course, logically exclusive of the first two, and proved difficult to explain. Nonetheless, subjects agreed to try to use it and, in the event, were able to do so with fair consistency.[1]

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. Figure 2 presents the sine wave, speech psychophysical and linguistic results for the three series of contours (H3, L3, L2) in which at least one contour was judged as expressing "talking-to-self" on more than 50 % of the group's judgments. Percentages of fall, level and rise judgments (sine wave and speech psychophysical) or of statement, "talking-to-self" and question judgments (linguistic) are plotted against terminal glide, measured as rise (positive) or fall (negative) in Hz from turning point to end point. Each data point represents a percentage of 110 judgments (22 subjects judged each contour five times).

Consider, first, the sine wave results (Figure 2, left column). For each series of contours the only contour judged more than 50 % of the time to be terminally level in pitch is the contour for which the ter-

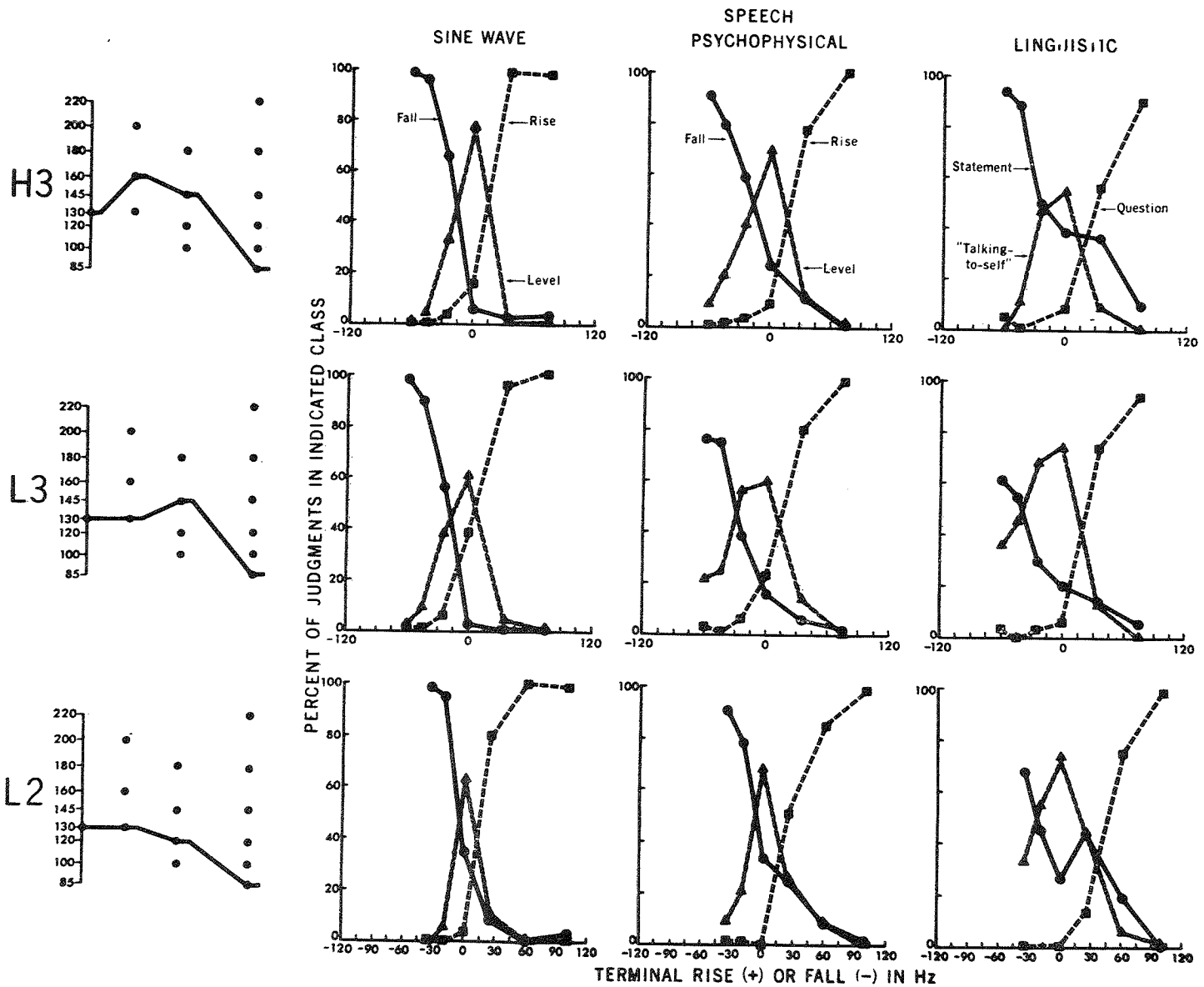

Figure 2. Percentages of fall, level and rise responses (sine wave and speech psychophysical) or of statement, "talking-to-self" and question responses (linguistic) as a function of terminal glide in Hz. Data for 22 subjects on the three series of contours for which at least one contour was judged "talking-to-self" on more than 50 % of the group's judgments.

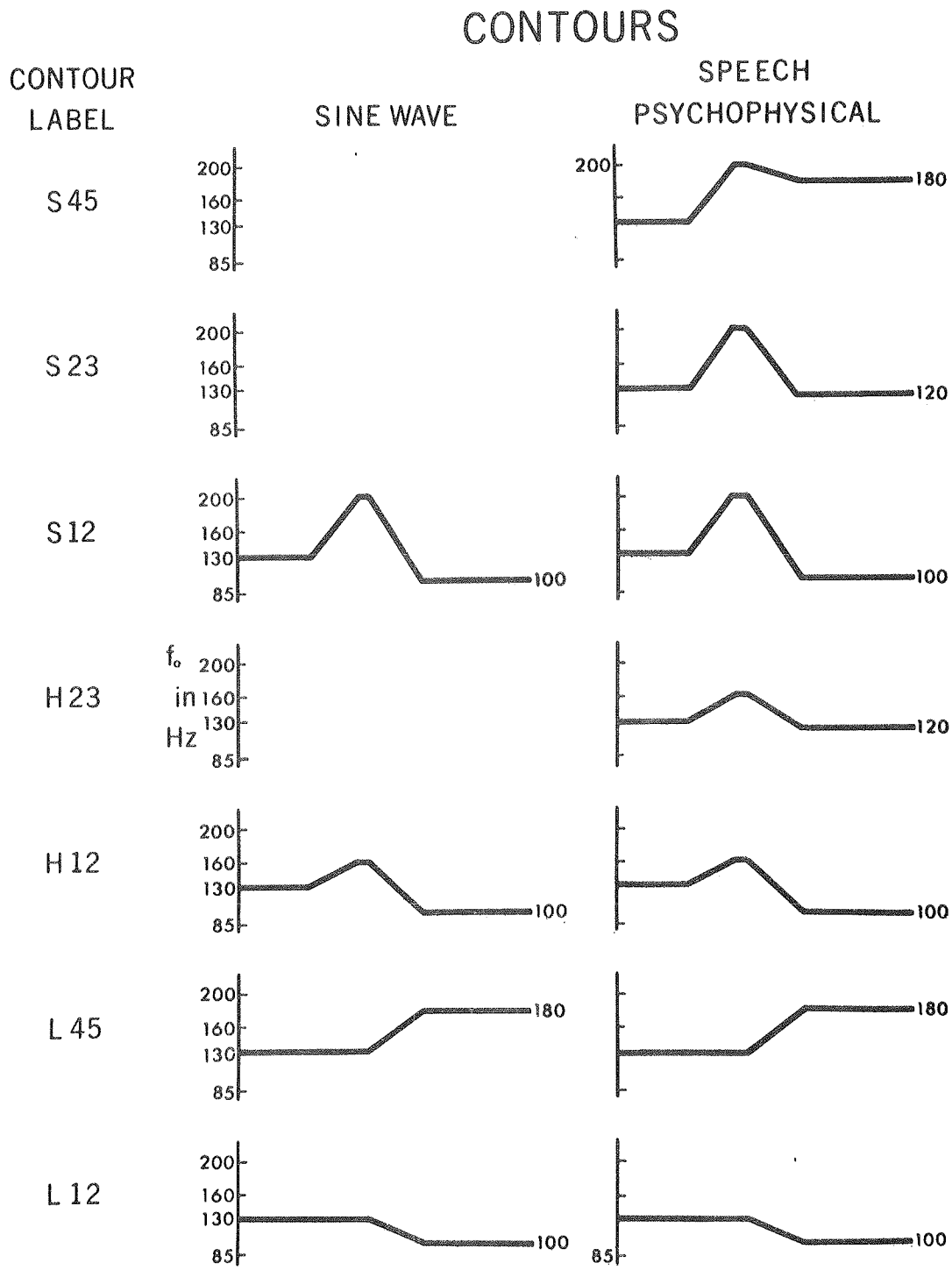

Figure 4: Schemata of all terminally level contours judged "level" on less than 50 % of the group's judgments.

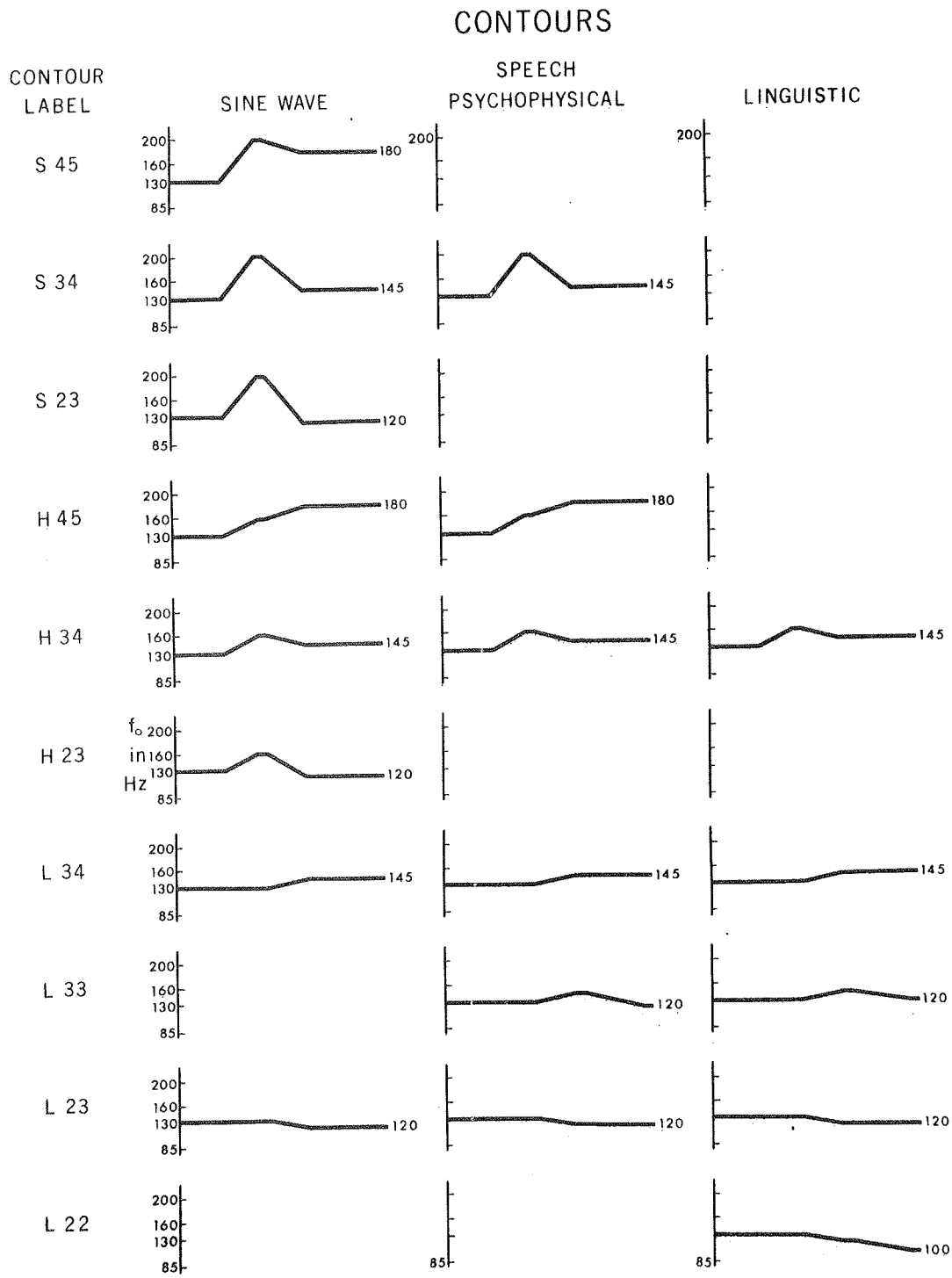

Figure 3. Schemata of all contours judged "level" (sine wave and speech psychophysical) or "talking-to-self" (linguistic) on more than 50 % of the group's judgments.

minal f_0 glide was, in fact, level. "Level" judgments increase and decrease systematically on either side of this zero value, with a stronger tendency to hear a slight fall as level than a slight rise. Since "level" judgments never reached 100 %, either for the terminally level contours of Figure 2 or for the nine other terminally level contours presented, it is evident that listeners did not find the judgment easy. However, their errors were primarily "misses" rather than "false alarms". That is to say, while four of the twelve terminally level contours failed to draw more than 50 % "level" judgments, none of them drew as many as 50 % "fall" or "rise" judgments, and none of the sixty terminally rising or falling contours drew as many as 50 % "level" judgments.

These results are summarized in Figures 3 and 4. Figure 3 (left column) sketches the eight sine wave contours judged "level" more than 50 % of the time. Figure 4 (left column) sketches the four terminally level sine wave contours for which "level" judgments did not reach 50 %. Note that three of the latter (S12, H12, L12) display a fall from the peak to a turning point 30 Hz below the onset level of the contour; one (L45) displays a rise from the peak to a turning point 50 Hz above the onset level of the contour.

Figure 2 (center column) presents speech psychophysical results. In each graph it is again the terminally level contour that collects the highest percentage of "level" judgments. But the spread of "level" judgments over terminally falling contours is clearly broader than for the corresponding sine wave contours. This is particularly noticeable for the L3 series, where one terminally falling contour (L33, middle row) actually draws 56 % "level" judgments. Nonetheless, this is the only "false alarm", so that, with five of the twelve terminally level contours being judged "level" more than 50 % of the time, the errors were again primarily "misses". Figures 3 (center column) and 4 (right column)

summarize these results.

Figure 2 (right column) presents the linguistic judgments. In each series it is the terminally level contour that draws the highest percentage of "talking-to-self" judgments. But there is a clear tendency for these judgments to invade the statement category. In one series (L3, middle row) the invasion matches quite strikingly that made by "level" judgments into the "fall" category of the speech psychophysical data. However, the tendency appears in all three series so that each has a terminally falling contour that draws close to 50 % "talking-to-self" judgments: H33 (46 %), L33 (58 %), L22 (55%). Figure 3 (right column) summarizes these results. Note the weight of "talking-to-self" judgments in the moderate to low stress peak series. No contour in the S-series meets the 50 % criterion, only one in the H-series, four in the L-series. The latter include two contours with level terminal glides, two with terminal glides that fall by 20-25 Hz.

Finally, we note that, while the preferred question contours of our previous study (Studdert-Kennedy and Hadding, 1972; in press) were totally unaffected by the introduction of a third category, the preferred statement contours did not fare so well. Nine of the twenty-three statement contours on which subjects displayed at least 90 % agreement in the previous study dropped below that level in the present study. Three of these (L33, L22, L23) were among the five contours collecting more than 50 % "talking-to-self" judgments.

Discussion

Listeners to brief (700 msec) frequency modulated sine wave contours can, with some reliability, identify those that sustain a level frequency over the last 265 msec. But their performance is not perfect. While they seldom hear a rising or falling terminal glide as level, they do with

fair frequency hear a level glide as rising or falling. They tend to be misled not by the initial rise to the peak, but by the rise of fall from peak to turning-point, that is, by the movement of the contour during the 127 msec immediately preceeding the terminal sustain: Figure 4 (left column) shows that two of the four terminally level contours that were "missed" display no onset-to-peak movement, but all four display a movement of at least 30 Hz from peak to turning-point, and end on a frequency at least 30 Hz above or below the precontour level of 130 Hz. We may therefore say that, exactly as in our previous study (Studdert-Kennedy and Hadding, 1972; in press), listeners seem to use the precontour as an anchor and then have difficulty in separating the terminal glide from the immediately preceding section of the contour if that section displays a marked movement to a point well above or well below the anchor.

The speech psychophysical results display a similar pattern. All four of the contours missed in the sine wave judgments were also missed in the speech psychophysical, and three more were added. Two of those added (S23, H23) display a strong fall from peak to turning point, but end on a frequency only 10 Hz below the precontour level; the other (S45) displays a fall of only 20 Hz from peak to turning point, but ends on a level 50 Hz above the precontour. In other words, there is clear overlap between sine wave and speech psychophysical data.

Even where the two sets of data do not agree, as in the tendency for listeners to judge certain terminally falling speech wave contours as "level", the errors would seem to arise from the same source as the sine wave errors, namely, from a simple inability to separate terminal glide from earlier sections of the contour. Thus, if the first 467 msec of the contour are relatively level (as in the H3 and L3 series, where all frequency variations between onset and turning point are within 30 Hz of

the precontour) listeners may fail to detect the slight terminal fall and then judge it to be "level" (see Fig. 2, center column). In other words, they do not, as might be predicted from the analysis-by-synthesis model of Lieberman (1967), accept glide as level when the stress peak is exceptionally high, but rather when the entire section of the contour preceding the terminal glide is relatively low and level.

Turning to the linguistic data, we may say that listeners are indeed able to identify an utterance as that of a speaker talking to himself and that they may even do so with more consistency than they make the corresponding psychophysical judgment (see Fig. 2: L34, L23). Nonetheless, they are not perfectly consistent. One reason for this is that the categories statement and "talking-to-self" are not mutually exclusive, and so compete for certain contours. This is evidenced by the tendency of "talking-to-self" judgments to take over the statement category at level terminal glide (see Fig. 2: L23, Linguistic) and by the fact that three of the four contours collecting more than 50 % "talking-to-self" judgments in the present study drew more than 90 % statement judgments in our previous study. Combined with this, a second factor may have contributed to listener uncertainty: intensity. Talking to ourselves we speak softly. But all utterances in the present study were of equal intensity so that a listener, choosing between the two competing categories, may have been pushed toward statement by a relative intensity more apt for addressing others than self.

In considering the linguistic results, we should bear in mind that, while psychophysical judgments were made on the terminal glide, linguistic judgments were made on the entire contour. If, therefore, sine wave, speech psychophysical and linguistic judgments coincide, we may reasonably conclude that terminal glide controlled linguistic decision. From Figure 2 and 3 it is evident that, as far as the third category ("level"

or "talking-to-self") is concerned, the three groups of judgments do coincide on certain contours that exhibit a level (H34, L34, L23) or slightly falling (H33, L33) terminal glide. This agreement confirms the importance of the terminal glide in linguistic judgments of intonation contours. While our previous study gave clear evidence of the connection between terminal rise/fall and judgments of question/statement, the present study demonstrates a clear connection between terminal sustain and judgments of "talking-to-self".

However, terminal glide is not the only determinant of linguistic decision. Figure 3 shows that one terminally contour (L22) was judged as "talking-to-self" more than 50 % of the time, but did not reach criterion on speech wave "level" judgments, while two speech wave contours (S34, H45), correctly heard as "level" more than 50 % of the time did not reach criterion on "talking-to-self" judgments. In fact, of the five acceptable "talking-to-self" contours, four display no stress peak (L34, L33, L23, L22), one displays a moderate peak but then drops to within 15 Hz of the precontour level (H34). Evidently, we expect people talking to themselves not only to end their utterances with a level (or slightly falling) glide, but also to maintain an even, low to moderate pitch over earlier sections of the contour. The initial hypothesis is therefore largely confirmed.

To sum up, this study has provided experimental support for the validity of the third category described by Hadding-Koch (1961), and for the adoption of a new prosodic feature, [±Listener], implemented by variations in fundamental frequency and, perhaps, intensity. The communicative function of the feature [+Listener] is presumably to draw and hold a listener's attention. Further evidence of its operation and of its functional development might be gained from systematic study of "egocentric" and "other-directed" speech in young children.

Footnote

We might have avoided some of the difficulties in the linguistic session, by asking subjects to use only two categories: talking to a listener and "talking-to-self". However, we wished to compare the results with those for the psychophysical sessions, and two-category psychophysical data would have concealed potentially interesting information on the subjects' capacities for discriminating terminally level from terminally rising or falling glides.

Acknowledgement. This work was supported in part by a grant to Haskins Laboratories from the National Institute of Child Health and Human Development.

References

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

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

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. 1965a. Intonation contours evaluated by American and Swedish test subjects. Proc. Vth Intl. Congr. Phonetic Sci., Münster 1964. Basel, 326-331

Hadding-Koch K. and Studdert-Kennedy M. 1965b. A study of semantic and psychophysical test responses to controlled variations in fundamental frequency. Studia linguistica XVII, 65-76

Lieberman P. 1967. Intonation, Perception, and Language. Cambridge, Mass.: The M.I.T. Press

Moravcsik E.A. 1971. Some crosslinguistic generalizations about yes-no questions and their answers. Working Papers on Language Universals 7, 45-181

Studdert-Kennedy M. and Hadding K. 1972. Further experimental studies of f_0 contours. Proc. VIIth Intl. Congr. Phonetic Sci., Montreal 1971. The Hague: Mouton, 1024-1031

Studdert-Kennedy M. and Hadding K. In press. Auditory and linguistic processes in the perception of intonation contours. Language and Speech (Also in Lund University Working Papers 5, 1971 and in Haskins Laboratories, SR-72, 1971)

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