PHONETICS LABORATORY
LUND UNIVERSITY

1666.

WORKING PAPERS

5-1971

MICHAEL STUDDERT-KENNEDY KERSTIN HADDING	Auditory and linguistic processes in the perception of intonation contours	1
ANDERS LÖFQVIST	Some observations on supraglottal air pressure	27

PHONETICS LABORATORY
LUND UNIVERSITY

WORKING PAPERS
5·1971

MICHAEL STUDDERT-KENNEDY, KERSTIN HADDING — Auditory and linguistic processes in the perception of intonation contours 1

ANDERS LÖFQVIST — Some observations on supraglottal air pressure 27

Phonetics Laboratory
1971

Staff:

Kerstin Hadding	Professor of Phonetics
Eva Gårding	Docent
Kurt Johansson	Lecturer
Gösta Bruce	Lecturer
Robert Bannert	Assistant
Helmut Jankóvic	Assistant (on leave)
Eva Wigforss	Assistant
Ulf Herrström	Engineer
Gunvor Hwargård	Secretary
Majastina Lindgren	Secretary, librarian
Eva Wällstedt	Secretary

Research and Part Time Teachers:

Boris Larnert
Per Lindblad
Anders Löfqvist
Kerstin Nauclér
Lennart Norman
Sidney Wood

Correspondence concerning this series should be addressed to:

Robert Bannert
Phonetics Laboratory
Lund University
Kävlingevägen 20
S-222 40 LUND
Sweden

Previous numbers and contents:

WP 1 • 1969	K. HADDING M. HIRANO T. SMITH	Electromyographic study of lip activity in Swedish CV:C and CVC: syllables
	KURT JOHANSSON	Försök avseende vokaltransitions riktning och dess betydelse för "plats"-distinktionen bland tonande klusiler
WP 2 • 1970	MONA LINDAU	Prosodic problems in a generative phonology of Swedish
WP 3 • 1970	GÖSTA BRUCE	Diphthongization in the Malmö dialect
	EVA GÅRDING	Word tones and larynx muscles
	KURT JOHANSSON	Perceptual experiments with Swedish disyllabic accent-1 and accent-2 words
WP 4 • 1971	ROBERT BANNERT	Hat das Deutsche zugrundeliegende Spiranten?
	EVA GÅRDING	Laryngeal boundary signals
	KARIN KITZING	Contrastive acoustic analysis of vowel phonemes, pronounced by some North German and South Swedish high school pupils (A summary)
	SIDNEY WOOD	A spectrographic study of allophonic variation and vowel reduction in West Greenlandic Eskimo

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_o) 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_o 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_o 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_o 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_o 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_o 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_o 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_o at one point in the utterance for a high f_o elsewhere. For example, an utterance with a relatively high f_o at peak or turning-point required a smaller terminal rise to be heard as a question than an utterance with relatively low f_o 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_o 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 spectrographic 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_0.

The utterance "November" [no'vɛmbɚ] 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 cps 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 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_0 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

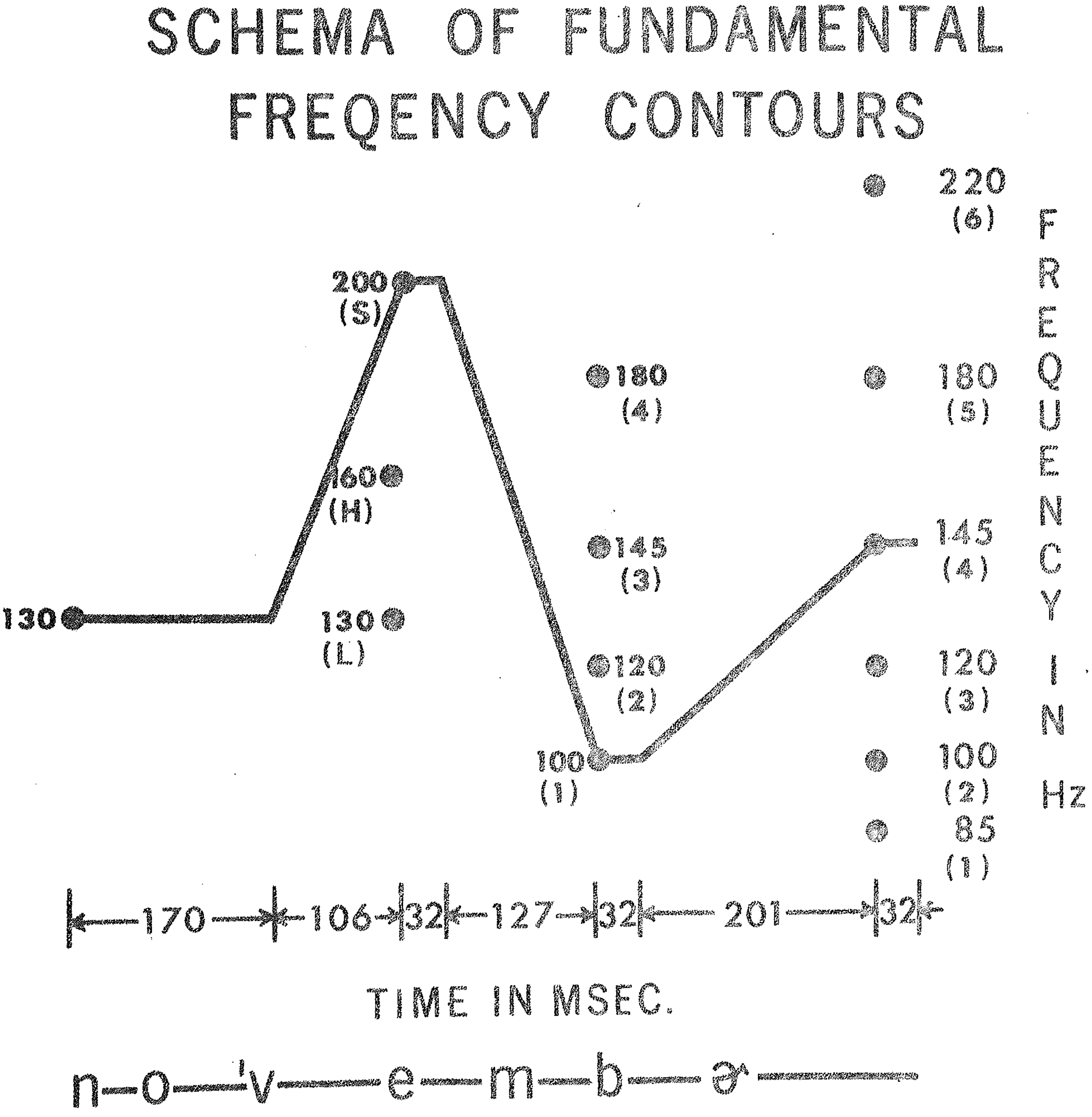


Figure 1. Schema of fundamental frequency contours imposed on the utterance "November" [noˈvɛmbɚ].

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

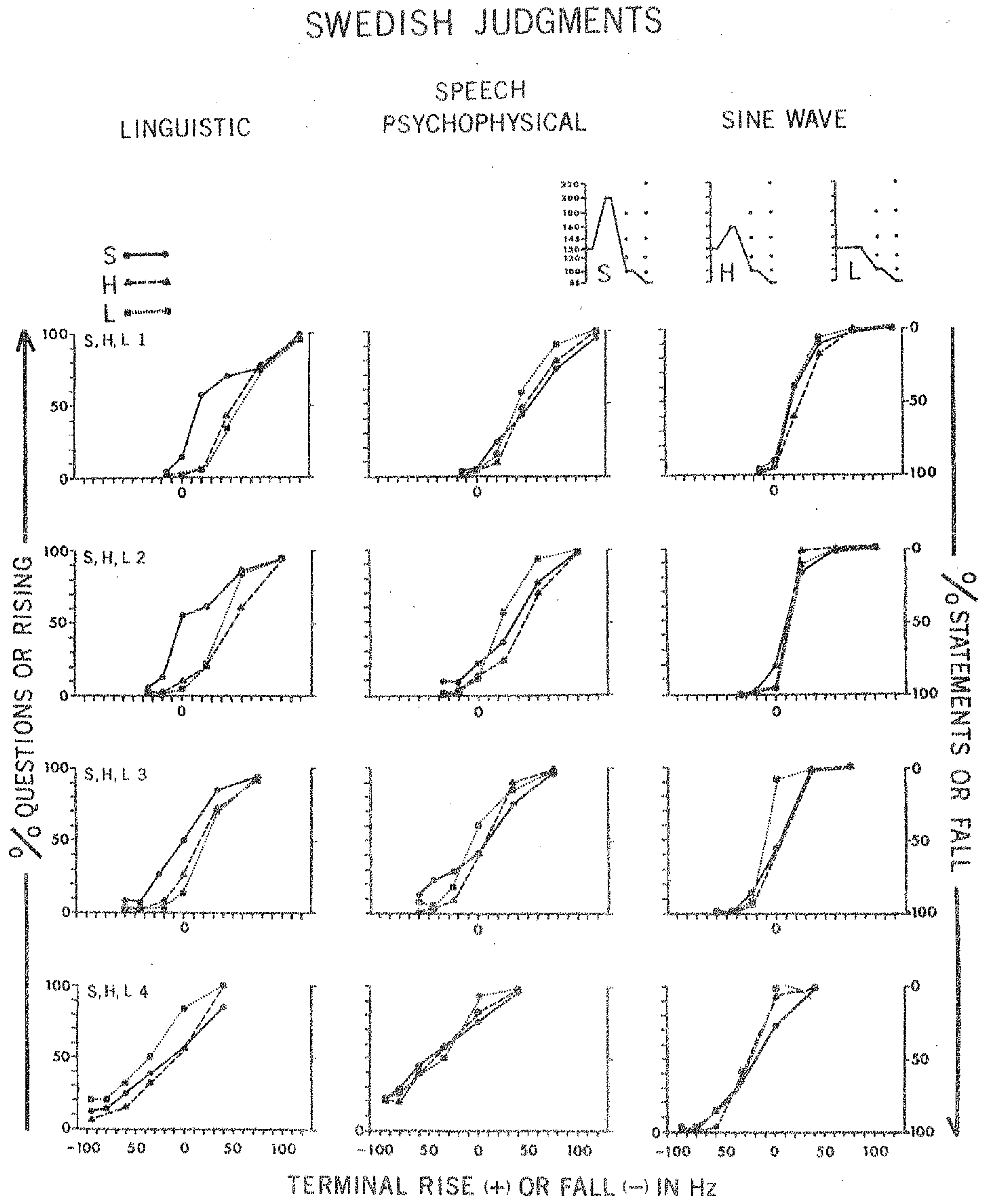


Figure 2. Percentages of question or rise responses (left-axis) and statement or fall responses (right-axis) plotted as functions of terminal glide in Hz. Peak values are constant across rows and turning-points are parameters of the curves. For Swedish subjects.

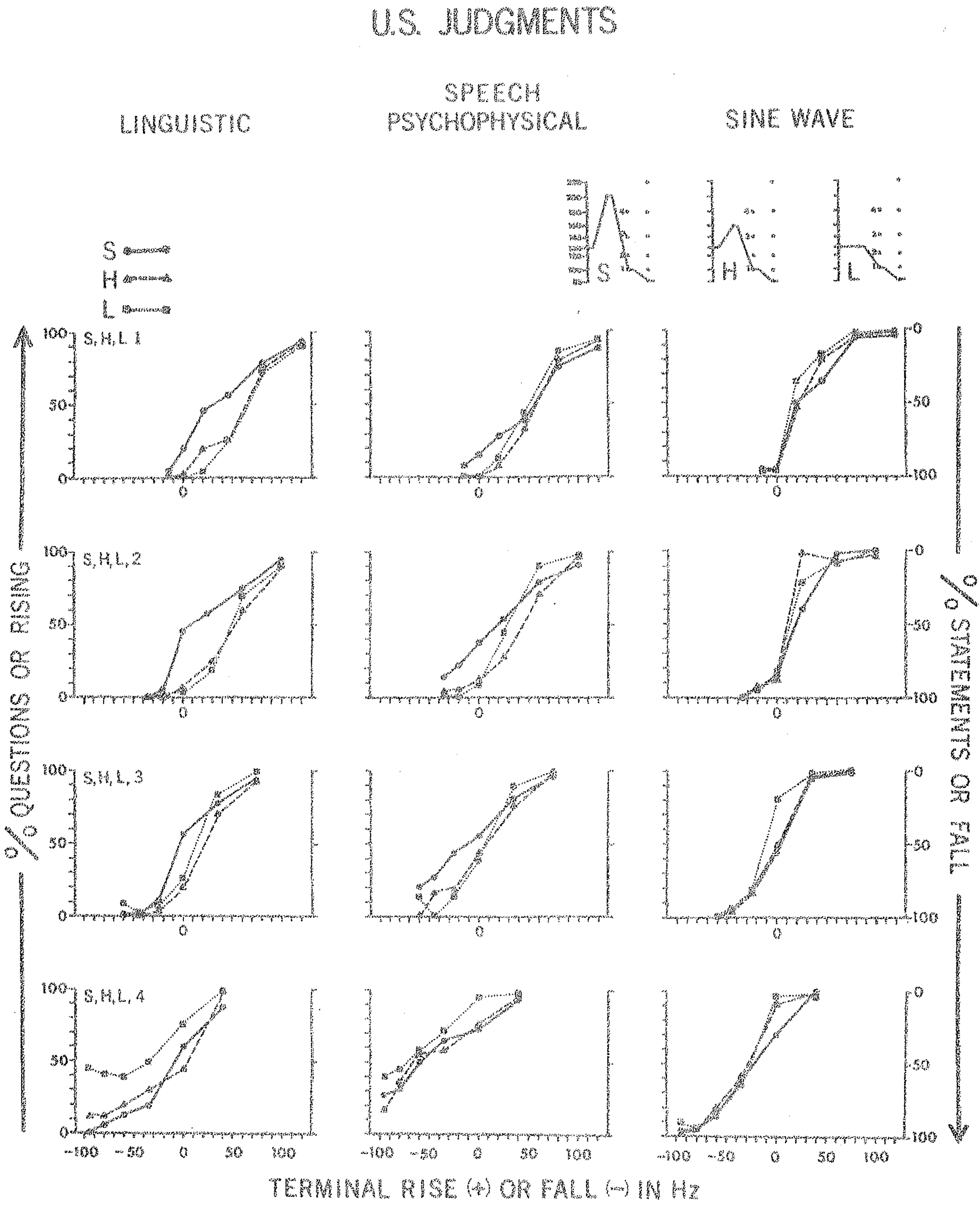


Figure 3. As for Figure 2, for the American subjects.

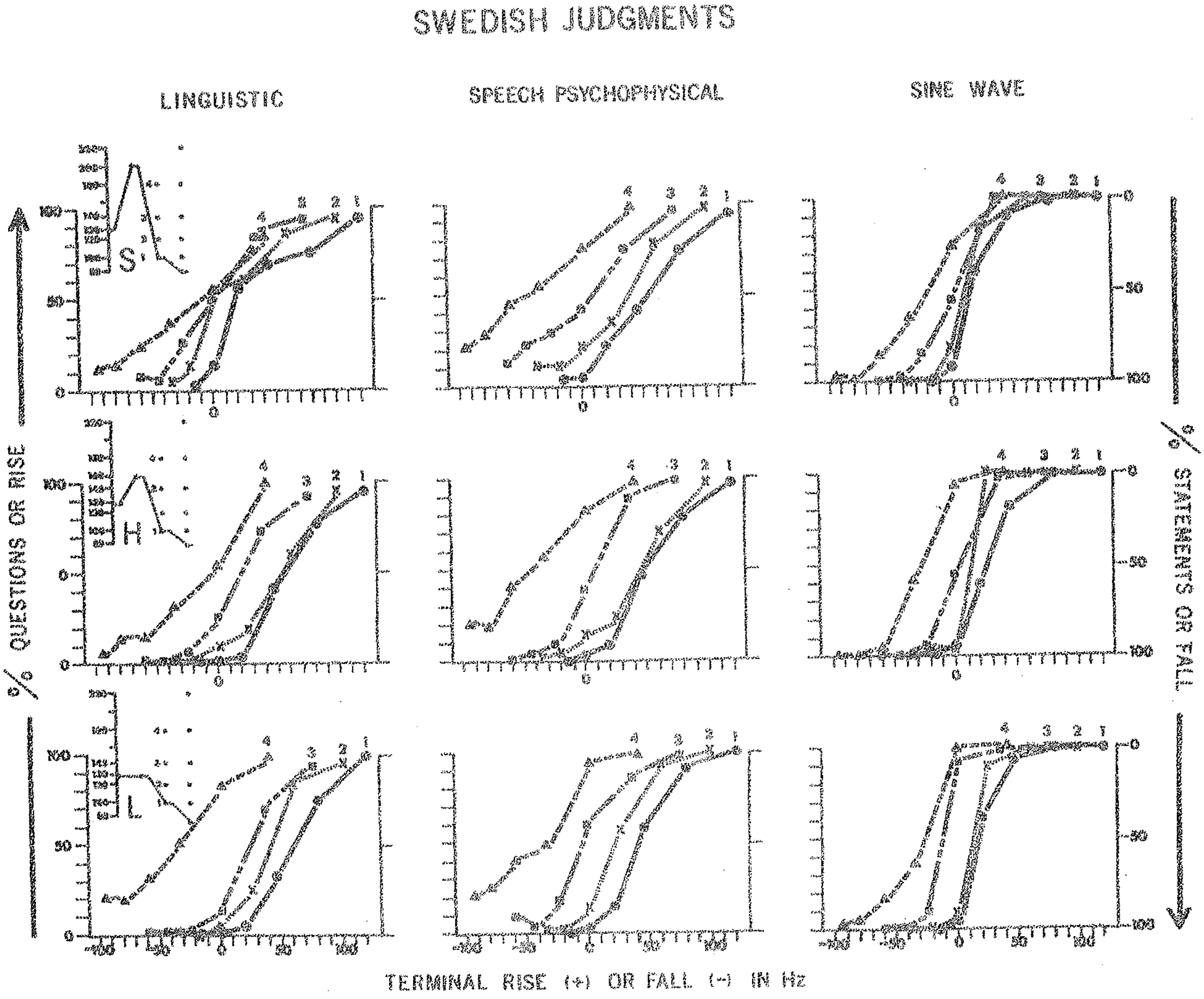


Figure 4. Percentages of question or rise responses (left-axis) and statement or fall responses (right-axis) plotted as functions of terminal glide in Hz. Turning-point values are constant across rows and peak values are parameters of the curves. For Swedish subjects.

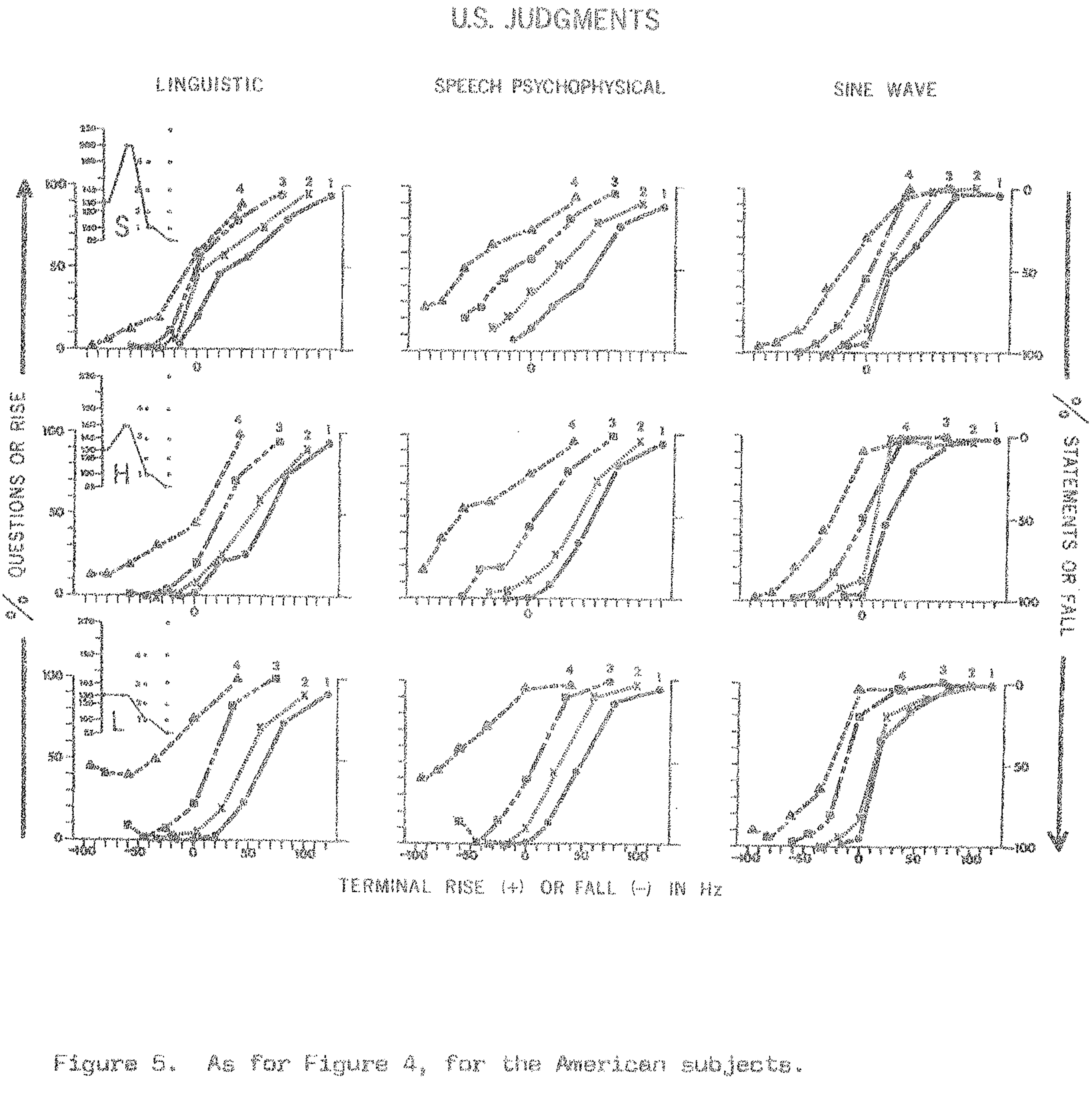


Figure 5. As for Figure 4, for the American subjects.

Schemata of Preferred Statement and Question Contours

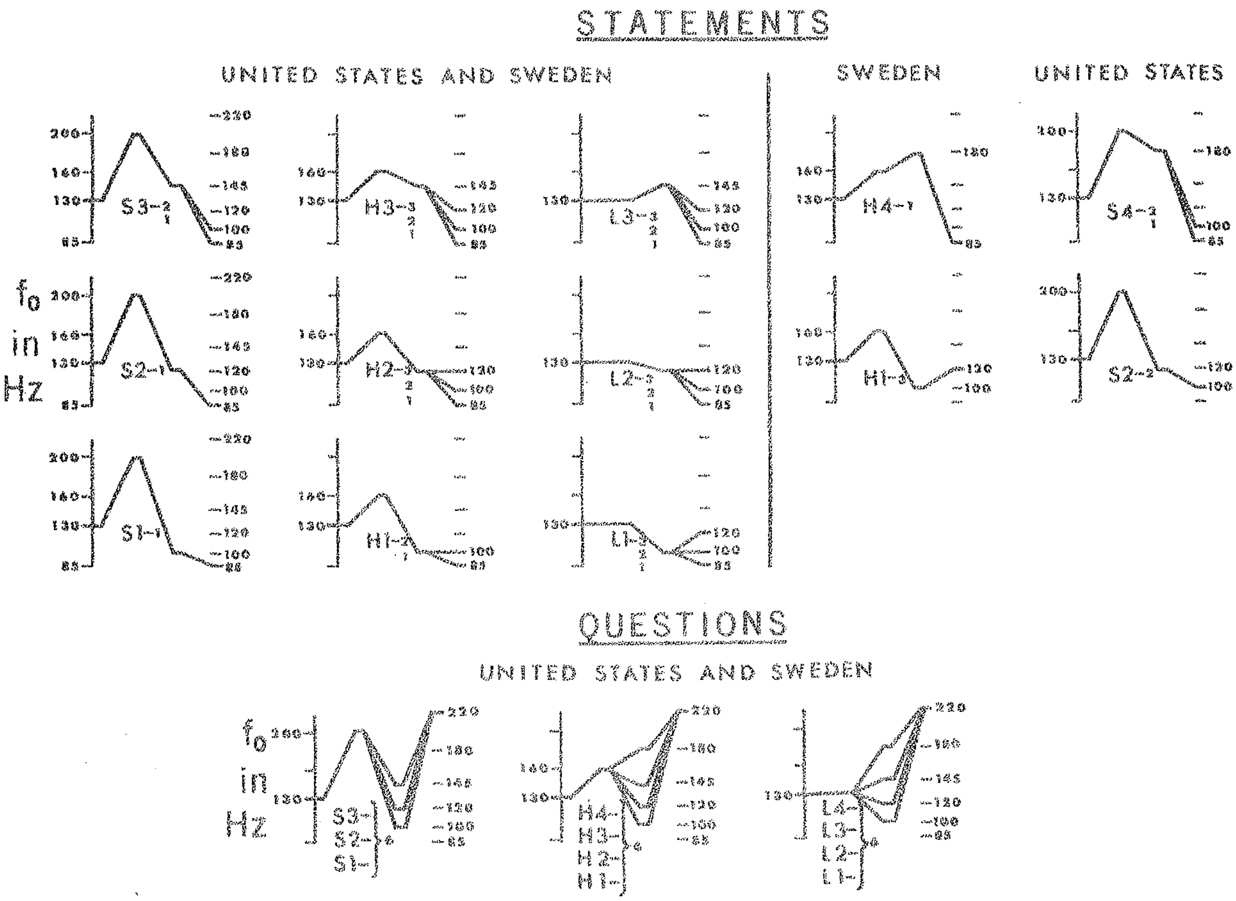


Figure 6. Included are all contours for which at least 90% of the judgments of a given language group were in a single category.

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 relatively 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_o, with turning-point f_o as parameter; in the second and fourth, they are graphed as functions of turning-point f_o, with peak f_o 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_o 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

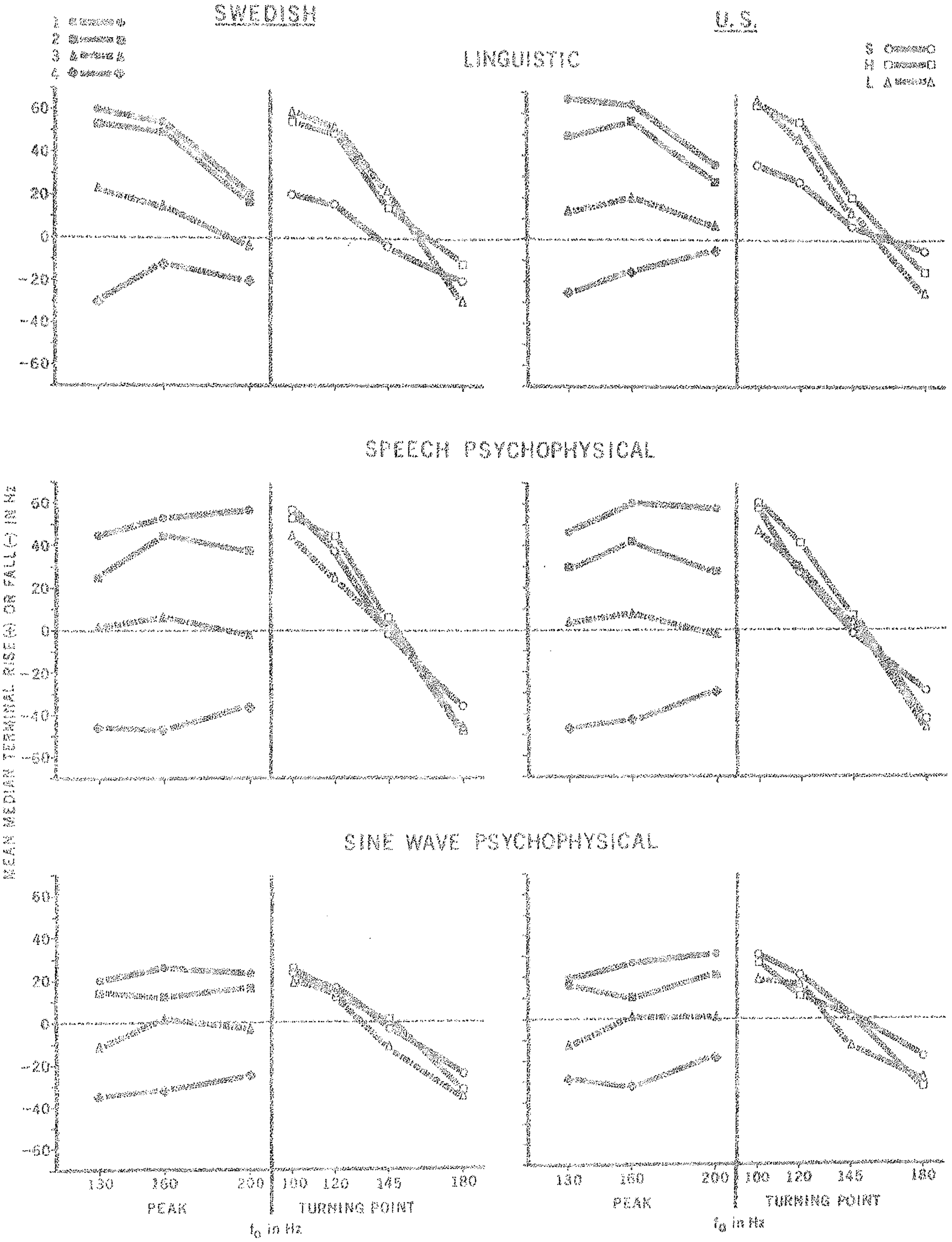


Figure 7. In the first and third columns mean medians are plotted as functions of peak f_0, with turning-point f_0 as parameter; in the second and fourth columns, they are plotted as functions of turning-point f_0, with peak f_0 as parameter.

(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_0 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_0 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

SOME OBSERVATIONS ON SUPRAGLOTTAL AIR PRESSURE

Anders Löfqvist

This is a preliminary report of some work in progress on supraglottal air pressure during speech production. It presents data on peak pressure, rise time, pressure decay and duration of oral closure for Swedish stops pronounced under different stress conditions in different positions. Only one speaker has been investigated. The study should be considered explorative and is intended to serve as a base for further research.

METHOD

The air pressure was measured with a differential pressure transducer EMT 33. A plastic tube with an inner diameter of 1,9 mm and 20 cm long was fitted to the transducer and introduced into the oropharynx through the nose and held in the same position during the recordings. This arrangement was chosen because it was felt to minimize interference with articulation and also because it made possible the study of velar sounds. It is possible, however, that the position of the tube with the plane of the opening perpendicular to the air flow might give spurious recordings due to stagnation pressures as pointed out by Hardy (1965). Although one might argue that the air flow for the sounds studied, stops, is minimal except at the release, the possibility should be kept in mind in interpreting the data.

No critical resonances were found in the recording system and the frequency response was judged sufficient for the purpose of the study. For a detailed technical discussion of pressure recordings, see Fry (1960). During the recording sessions it was sometimes necessary to blow air through the tube to expel mucous. By using the same carrier sentence in all the recordings it was possible to check the recording system to ensure that its damping remained the same. The output from the transducer was amplified by an electromanometer EMT 31 and recorded on a Mingograph together with the speech signal; a tape recording was simultaneously made of the speech for acoustic analysis.

The pressure recording system was calibrated with a water manometer and all pressure measurements are given in mm of H_2O. In measuring the pressure for the voiced plosives the pressure trace with the oscillations due to vocal fold activity was bisected and the pressure at the midline was measured.

The following measurements were made:

Peak pressure; this was measured from the peak of the pressure curve to the base line.

Rise time; in the litterature one finds several ways of measuring rise time. In the present study three methods have been used. The first one follows Lisker (1970) and rise time is defined as "the duration of the interval during which pressure builds up to its peak value". In the second case rise time was defined according to Subtelny et al. (1966): "Rise time was defined as the interval between initiation of pressure elevation and the earliest point at which a relatively stable elevated oral pressure was attained." The third measure of rise time was taken from Fischer-Jørgensen (1969) and is the interval during which pressure builds up to 85 % of its peak value.

Pressure decay; this was taken as the interval between the point where the pressure started to fall rapidly and the first point at which supraglottal pressure returned to base pressure.

Duration of oral closure; the duration of the oral closure was taken as the interval between initial pressure elevation and the point at which supraglottal pressure started to fall rapidly. This measure only gives an approximate indication of the duration of the oral closure because the pressure presumably begins to rise before complete closure is attained. More reliable data on the duration of oral closure can be obtained from air flow traces or by special techniques as described by Karlsson and Nord (1970) and Slis (1971).

MATERIAL

The speech material consisted of nonsense words containing different plosives in different positions under different stress conditions. The words were:

1. 'CɑːC, where $C_1 = C_2$ = /p, t, k, b, d, g/;
2. Ca'Cɑː, with the same consonants as above and stress on the last syllable;
3. 'Cɑ́ːCən, with consonants as before and stress on the first syllable; the word had the acute tonal accent or accent 1, in the following denoted by ´;
4. 'Cɑ̀ːCən, with consonants as before and stress on the first syllable; the word had the grave tonal accent or accent 2, in the following denoted by ` .

The words were placed in the carrier phrase "Ja sa ... igen" (I said ... again) except for the words of type 1 which were pronounced in the phrase "Ja sa ... " and thus occured in sentence final position.

A Swedish male speaker - with a dialect of Southwestern Swedish (Västgötska) - pronounced all the words 12 times. The speaker was told to use the same intensity and speech rate during all the utterances. Differences in intensity and speech rate might contribute to variations in the data as they have been shown to affect supraglottal air pressure, cf. Subtelny et al. (1966), Arkebauer et al. (1967).

Earlier investigations of the same parameters for the same class of sounds from other languages do not show a complete agreement in their results. The variations might be due to differences in the composition of the material, differences in intensity and speech rate and also reflect differences between different languages. Even if the same language has been studied there are some variations between different investigations. This is evident from the table below where mean values have been calculated of peak pressure for Am. English stops taken from different investigations. The studies which have been used for the calculations are Subtelny et al. (1966), Arkebauer et al. (1967), Malécot (1968), and Lisker (1970). Values are given in mm of H_2O.

		initial			medial			final	
		ptk	bdg		ptk	bdg		ptk	bdg
Subtelny et al.		75	62		70	53		65	30
Arkebauer et al.		67	30		83	40		63	26
Malécot	C'V	83	69	VC'V	91	65	'VC	57	31
	CV	72	60	'VCV	83	34	VC	57	31
Lisker		57	53	VC'V	68	61		47	15
				'VCV	71	36			

We see that there are not only absolute differences but also differences between the relative values for the two categories of stops in different positions. The tendency seems to be that the voiceless plosives have higher peak pressure than their voiced counterparts. With regard to position the voiceless plosives have the highest pressure in initial position, lower pressure in medial position and the lowest pressure in final position. The voiced plosives tend to have the highest pressure in initial position, lower pressure medially and lowest in final position.

For the three other parameters I will not make any survey of earlier investigations at this point, as they are somewhat difficult to compare. I will make some comparisons and further comments below when I present my own results.

RESULTS AND DISCUSSION

As the material processed as yet is rather limited and explorative in nature data will only be given as frequency distributions and mean values. No further statistical discussion will be presented. Measurements are made in H_2O for pressure and in msec. for duration.

All the examples of the voiced plosives /bdg/ in the study were pronounced with vibrations of the vocal folds as can be seen from the oscillograms. In the following discussion the terms voiced and unvoiced thus refer to two classes of speech sounds differentiated by their respective mode of glottal activity.

Peak pressure

Diagrams of peak pressure variations are presented in Fig. 1-2 and mean values are summarized in Table I.

Table I. Mean peak pressure in mm of H_2O.

	Initial			Medial		Final	
	ptk	bdg		ptk	bdg	ptk	bdg
'CV:	72	47	V'CV:	73	43	59	32
CV	69	48	'V́:CV	73	35		
			'V̀:CV	64	28		

From Table I the following conclusions can be drawn.

Voiceless plosives have higher supraglottal pressure than their voiced counterparts. This is clear in all positions and from Fig. 1 and 2 we see that there is some overlapping between the values for peak pressure for the two sets of plosives only in initial position.

The stress affects supraglottal pressure. The voiceless plosives have higher peak pressure in stressed than in unstressed syllables. The only exception to this is the medial unstressed plosive in words with tonal accent 1 where the pressure is as high as in a stressed syllable.

The voiced plosives show almost the same influence of stress upon peak pressure but in initial position the peak pressure is almost the same irrespective of stress.

If we look at the effects of position upon supraglottal pressure we see in Table I that the voiceless plosives have their highest pressure in medial position and the lowest pressure in final position. The difference between the peak pressure in initial and medial stressed syllable is very small, however, and we can also see that the tonal accent of the word affects the peak pressure of the medial unstressed plosives.

The voiced plosives have highest pressure initially and lowest pressure in medial position. It is, however, difficult to draw any firm conclusions of the effect of position upon peak pressure as it is obviously interconnected with the effects of stress and tonal word accent.

An interesting fact is the difference we find in Table I between the medial plosives in words with tonal accent 1 and the same plosives in words with tonal accent 2: the medial plosive has higher pressure if the word has tonal accent 1 than if it has tonal accent 2. For the initial stops the tonal accent of the word has no influence on peak pressure.

The most obvious difference between the two sets of plosive that could explain the higher supraglottal pressure for the voiceless ones is the glottal activity; vibrations for the voiced set and no vibrations for the unvoiced set. The glottal resistance is higher when the glottis is in a voiced position. The increased glottal resistance for the voiced plosives as compared to the unvoiced reduces the air flow through the glottis and prevents the supraglottal pressure to reach the same level for the two sets of sounds, cf. Fischer-Jørgensen (1963).

Besides the difference in glottal resistance there might also be different articulatory adjustments during the production of voiced and voiceless plosives. If the vocal cords are to vibrate there must exist a pressure drop across the glottis. When the vocal tract is obstructed in the oral cavity as is the case during the production of stops, the pressure drop would tend to diminish and the vocal folds would cease to vibrate. In order to maintain the pressure drop during the closure of a voiced plosive various mechanisms have been proposed, cf. Rothenberg (1968) for a discussion and also Hudgins and Stetson (1935), Perkell (1968), Kent and Moll (1969), Berti (1971).

The mechanisms suggested are incomplete velopharyngeal closure, passive expansion of the supraglottal cavities and active expansion of the same cavities. Whether any of these mechanisms is actually used is not completely known at present but they would all affect supraglottal air pressure and tend to reduce it. That supraglottal pressure is higher in stressed than in unstressed syllables can be a result of the increase in subglottal pressure that has been shown to take place in a stressed syllable (Ladefoged 1967).

The relationship between tonal accent and supraglottal air pressure is suitable for further investigations which also should include recordings of subglottal pressure. The problem of the Swedish word accent has not yet been covered in all its aspects. There is a general agreement that variations in the fundamental frequency are important for the accent distinction, cf. Malmberg (1963), Hadding-Koch (1962), Öhman (1968), Gårding (1970), Johansson (1970).

Besides the F_0 variations, which according to Malmberg are the most important perceptual cues, there are also differences of duration and intensity. The pattern of the variations in fundamental frequency differs between various dialects. For the speaker investigated F_0 and intensity curves for the two tonal accents are given in Fig. 3. In the stressed syllable the F_0 shows a fall in words with tonal accent 1 and has a rising-falling pattern in tonal accent 2. The second, unstressed, syllable has a rise in the fundamental frequency for the two tonal accents.

The observed differences in supraglottal pressure to some extent reflect differences in subglottal pressure. During the closure of a voiceless plosive the subglottal and supraglottal pressures are almost equal, cf. Netsell (1969), Scully (1969), Ladefoged (1967, 1968).

We do not find any striking differences between the F_0 and intensity curves for the second, unstressed, vowel in words with tonal accent 1 as compared with the same parameters for the same vowel in words with tonal accent 2. The fundamental frequency shows the same rising pattern. There is a small difference in F_0 at the beginning of the second vowel: here F_0 is approximately 5 Hz higher in words with tonal accent 1 than in words with tonal accent 2. This difference might be caused by the higher pressure but variations in the fundamental frequency depend both upon subglottal pressure and laryngeal adjustments, cf. Ohala (1970).

Further work on this problem should include material from several dialects in order to find out whether the mentioned differences are systematic or accidental and confined to the speaker investigated.

Rise time

Fig. 4 and 5 give some representative examples of supraglottal air pressure curves for unvoiced and voiced plosives. During the production of an unvoiced plosive the air pressure rises quickly until it attains a plateau where it stays more or less stable until the release when it drops rapidly; it is, however, not always the case that the pressure stays at the same level after the plateau has been attained: it might as well continue to rise until the release.

The pressure curve for a voiced plosive most often show a continuous rise during the whole closure period. Sometimes some fluctuations up and down can be seen during the rise, cf. also Fig. 10.

Table II gives mean values of rise time defined as the interval during which pressure builds up to its peak value; frequency distributions for the same measure are given in Fig. 6 and 7.

Table II. Mean duration of the interval during which pressure builds up to its peak value. Msec.

	initial			medial		final	
	ptk	bdg		ptk	bdg	ptk	bdg
'CV:	154	130	V'CV:	74	101	203	155
CV	135	134	'V́:CV	166	101		
			'V̀:CV	196	117		

We see from Table II that rise time is shorter for the voiced than for the unvoiced plosives. The only exception is in medial stressed position.

The effects of stress and position are not quite clear. Rise time is, however, longest in final position, and shortest in medial position.

Rise time for the medial plosives also tend to be shorter if the word has tonal accent 1 than if it has tonal accent 2.

The shorter rise time for the voiced plosives obviously reflects the difference in the shape of the pressure curve and also the shorter duration of the closure for a voiced plosive as compared to a voiceless plosive.

Table III presents mean duration of the interval between the initiation of pressure elevation and the earliest point at which a relatively stable elevated pressure is attained. Only data for the voiceless plosives are given as the voiced plosives do not consistently show any period of stable elevated pressure.

Table III. Mean duration of the interval between the initiation of pressure elevation and the earliest point at which a relatively stable elevated pressure is attained. Msec.

	initial		medial	final
	ptk		ptk	ptk
'CV:	50	V'CV:	48	41
CV	50	'V́:CV	46	
		'V̀:CV	46	

Evidently the point of stable elevated pressure is attained in about the same time irrespective of stress and position. This reflects the very quick initial pressure rise for an unvoiced plosive.

Fig. 8 and 9 and Table IV present data on rise time taken as the duration of the interval during which pressure builds up to 85 % of its peak value.

Table IV. Mean duration of the interval during which pressure builds up to 85 % of its peak value. Msec.

	initial			medial		final	
	ptk	bdg		ptk	bdg	ptk	bdg
'CV:	57	112	V'CV:	37	81	65	136
CV	46	119	'V́:CV	39	83		
			'V̀:CV	86	91		

With this measure of rise time we see that the unvoiced plosives have shorter rise time than the voiced ones.

The effects of stress, position and tonal accent interact. The tendency seems to be that rise time is shorter in stressed than in unstressed position. This holds for the voiced plosives but not completely for the unvoiced as in initial position rise time is shorter for the unstressed than for the stressed unvoiced plosive.

In final position we find the longest rise time, in medial position the shortest.

The tonal accent of the word also affects the rise time: it is shorter for the medial plosive if the word has tonal accent 1 than if it has tonal accent 2.

The different measures of rise time give different results and which one that should be preferred depends on the aim of the study. It seems that the last one given gives the intuitively correct result as it takes into account the initial quick pressure rise that occurs for the unvoiced plosives. This is ignored by the first measure of rise time used above.

Two more comments will be made on the shape of the pressure curve. In the material investigated there are some cases of unvoiced plosives where peak pressure is attained immediately after the initial rise and where the supraglottal pressure stays at the peak level during the whole closure. This only occurs for /p/ and /t/, never for /k/. It happens most often in medial stressed syllable where it is the case 11 out of 12 times for /p/ and 10 out of 12 times for /t/.

If we want to explain the differences in the shape of the pressure curve, - cf. Fig. 10 for some more examples of traces for voiced plosives -, we will have to take into account the difference in glottal activity, possible variations in the volume of the supraglottal cavities and subglottal activity.

As to subglottal activity the studies of Ladefoged (1967, 1968) do not permit us to draw any conclusions whether the respiratory system actually increases the subglottal pressure in the production of a voiceless plosive. The rise in supraglottal pressure towards the end of the occlusion for unvoiced plosives could perhaps be explained in this way; see also Rothenberg (1968) for a discussion of this point and Sears and Newsom Davis (1968), Campbell, Agostini and Newsom Davis (1970) for relevant material on the functioning of the respiratory system.

In connection with Fig. 10 it should also be pointed out that the drop in supraglottal pressure that occurs at the beginning of the oral closure for /g/ is very systematic for the speaker investigated and has also been noted for other speakers.

Pressure decay

Mean values of pressure decay are given in Table V and frequency distributions of the data in Fig. 11.

Table V. Mean pressure decay. Msec.

	initial			medial		final	
	ptk	bdg		ptk	bdg	ptk	bdg
'CV:	57	19	V'CV:	63	20	202	29
CV	62	22	'V̆:CV	22	21		
			'V́:CV	23	22		

From Table V we see that the decay is rapid for the voiced plosives irrespective of stress and position.

The voiceless plosives have a rather slow decay in initial position, in medial stressed position, and in final position. In the other positions the decay time for the unvoiced plosives is as rapid as that for a voiced plosive.

The tonal accent of the word has no noticeable influence upon pressure decay.

The results in Table V agree well with those found for English by Lisker (1970).

The measure of pressure decay that has been used is not the same as "aspiration" or the "voicing lag" of Lisker and Abramson (1964) because the vocal folds start to vibrate before the pressure has reached the base pressure. It is, however, of the same dimension as the status of the **glottis** seems to be critical for it. If the glottis is open air can flow up through it and keep the supraglottal air pressure at a high level.

If on the other hand glottis is in a voiced position and the glottal resistance is high the air flow through the glottis will be minimal and the supraglottal pressure will decay at a rapid rate, cf. Fant (1960) p. 272 ff.

Pressure decay might also be influenced by the supraglottal resistance at the place of articulation. During the closure this resistance is infinite and at the release it drops abruptly. The speed with which the supraglottal resistance falls will reflect the speed of the articulators and this might be reflected in the decay of the supraglottal pressure. Keeping other factors constant one might assume that the decay will be proportional to the rate with which the articulatory resistance falls.

In further studies of the interaction of glottal resistance and supraglottal resistance and their effects on pressure decay the findings of Vencov (1968) should be taken into account.

Duration of oral closure

Data on the duration of oral closure are presented in Fig. 12 and 13 and also in Table VI.

Table VI. Mean duration of oral closure. Msec.

	initial			medial		final	
	ptk	bdg		ptk	bdg	ptk	bdg
'CV:	182	137	V'CV:	114	106	283	164
CV	190	141	'V́:CV	191	108		
			'V̀:CV	223	123		

We see that the closure period is longer for the unvoiced than for the voiced plosives.

Closure duration also seems to be shorter in a stressed than in an unstressed syllable.

For the voiced plosives we find that the closure is shortest in medial position; it is a little longer in initial position and longest in final position.

The voiceless plosives have the longest closure duration in final position but it is hard to draw any conclusions for medial and initial position because other factors also play a role.

The closure duration of the medial unstressed plosives is markedly influenced by the tonal accent of the word in which they occur. They have longer closure in words with tonal accent 2 than in words with tonal accent 1.

The difference in closure duration between voiced and unvoiced plosives corresponds to that found by Karlsson and Nord (1970) for Swedish. For Dutch the same kind of variations has been noted by Slis (1970, 1971) whereas Kent and Moll (1969) did not find any difference in closure duration due to difference in voicing for American English stops.

The fact that closure duration is shorter in a stressed than in an unstressed syllable is not in agreement with the findings of Slis (1971) for Dutch. Karlsson and Nord (1970), who studied Swedish, state that an unvoiced plosive after a stressed vowel is lengthened as compared to the same plosive before a stressed vowel. They found the shortest closure durations for stops between two unstressed vowels.

As matters seem to be rather complicated it seems unwise to draw any firm conclusions on the basis of the present material.

One further point appears from the present data: there seems to be an inverse relationship between closure duration and pressure decay: when closure duration is short, as in a stressed syllable, pressure decay is slow and the other way round.

REFERENCES

Arkebauer H.J., Hixon T.J., and Hardy J.C. 1967. Peak intraoral air pressures during speech. *J. Speech and Hear. Res.* 10:196-208

Berti F.B. 1971. The velopharyngeal mechanism: An electromyographic study. A preliminary report. *Haskins Laboratory Status Report* 25/26:117-129

Campbell E.J.M., Agostini E., and Newsom Davis J. 1970. The respiratory muscles. London: Lloyd-Luke

Fant G. 1960. Acoustic theory of speech production. S'Gravenhage: Mouton

Fischer-Jørgensen E. 1963. Beobachtungen über den Zusammenhang Zwischen Stimmhaftigkeit und intraoralem Luftdruck. *Z. Phonetik, Sprachwiss. und Kommunikationsforschung* 16:19-36

Fischer-Jørgensen E. 1968. Voicing, tenseness and aspiration in stop consonants with special reference to French and Danish. *ARIPUC* 3:63-114

Fry D.L. 1960. Physiologic recording by modern instruments with particular reference to pressure recording. *Physiol. Rev.* 40:753-788

Gårding E. 1970. Word tones and larynx muscles. *Working Papers* 3:20-46. Phonetics Lab. Lund University

Hadding-Koch K. 1962. Notes on the Swedish word tones. *Proc. IVth Intl. Congr. Phonet. Sci.* p. 630-638

Hardy J.C. 1965. Air flow and air pressure studies. *ASHA Reports* 1:141

Hudgins C.V. and Stetson R.H. 1935. Voicing of consonants by depression of the larynx. *Arch. néerl. phon. exp.* 11:1-28

Johansson K. 1970. Perceptual experiments with Swedish disyllabic accent-1 and accent-2 words. *Working Papers* 3:47-74. Phonetics Lab. Lund University

Karlsson I. and Nord L. 1970. A new method of recording occlusion applied to the study of Swedish stops. *STL-QPSR* 2/3:8-18

Kent R.D. and Moll K.L. 1969. Vocal-tract characteristics of the stop cognates. *J. Acoust. Soc. Am.* 46:1549-1555

Ladefoged P. 1967. Three areas of experimental phonetics. Oxford University Press

Ladefoged P. 1968. Linguistic aspects of respiratory phenomena. Ann. N.Y. Acad. Sci. 155:141-151

Lisker L. 1970. Supraglottal air pressure in the production of English stops. Language and Speech 13:215-230

Lisker L. and Abramson A. 1964. A cross-language study of voicing in initial stops: Acoustical measurements. Word 20:384-422

Malécot A. 1966. The effectiveness of intra-oral air-pressure-pulse parameters in distinguishing between stop cognates. Phonetica 14:65-81

Malécot A. 1968. The force of articulation of American stops and fricatives as a function of position. Phonetica 18:95-102

Malmberg B. 1963. Structural linguistics and human communication. Berlin: Springer

Netsell R. 1969. Subglottal and intraoral air pressures during the intervocalic contrast of /t/ and /d/. Phonetica 20:68-73

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

Perkell J.S. 1969. Physiology of speech production: Results and implications of a quantitative cineradiographic study. Cambridge, Mass.: The M.I.T. Press

Rothenberg M. 1968. The breath-stream dynamics of simple-released-plosive production. Bibl. Phonetica. Vol. 6. Basel: Karger

Sears T.A. and Newsom Davis J. 1968. The control of respiratory muscles during voluntary breathing. Ann. N.Y. Acad. Sci. 155:183:190

Scully C. 1969. Problems in the interpretation of pressure and air flow data in speech. Phonetics Department Report No. 2:53-92. University of Leeds

Slis I.H. 1970. Articulatory measurements on voiced, voiceless, and nasal consonants. Phonetica 21:193-210

Slis I.H. 1971. Articulatory effort and its durational and electromyographic correlates. Phonetica 23:171-188

Subtelny J.D., Worth J.H., and Sakuda M. 1966. Intraoral pressure and rate of flow during speech. J. Speech and Hear. Res. 9:498-518

Vencov A.V. 1968. A mechanism for production of voiced and voiceless intervocalic consonants. Z. Phonetik, Sprachwiss. und Kommunikationsforschung 21:140-144

Öhman S.E.G. 1967. Word and sentence intonation: A quantitative model. STL-QPSR 2/3:20-54

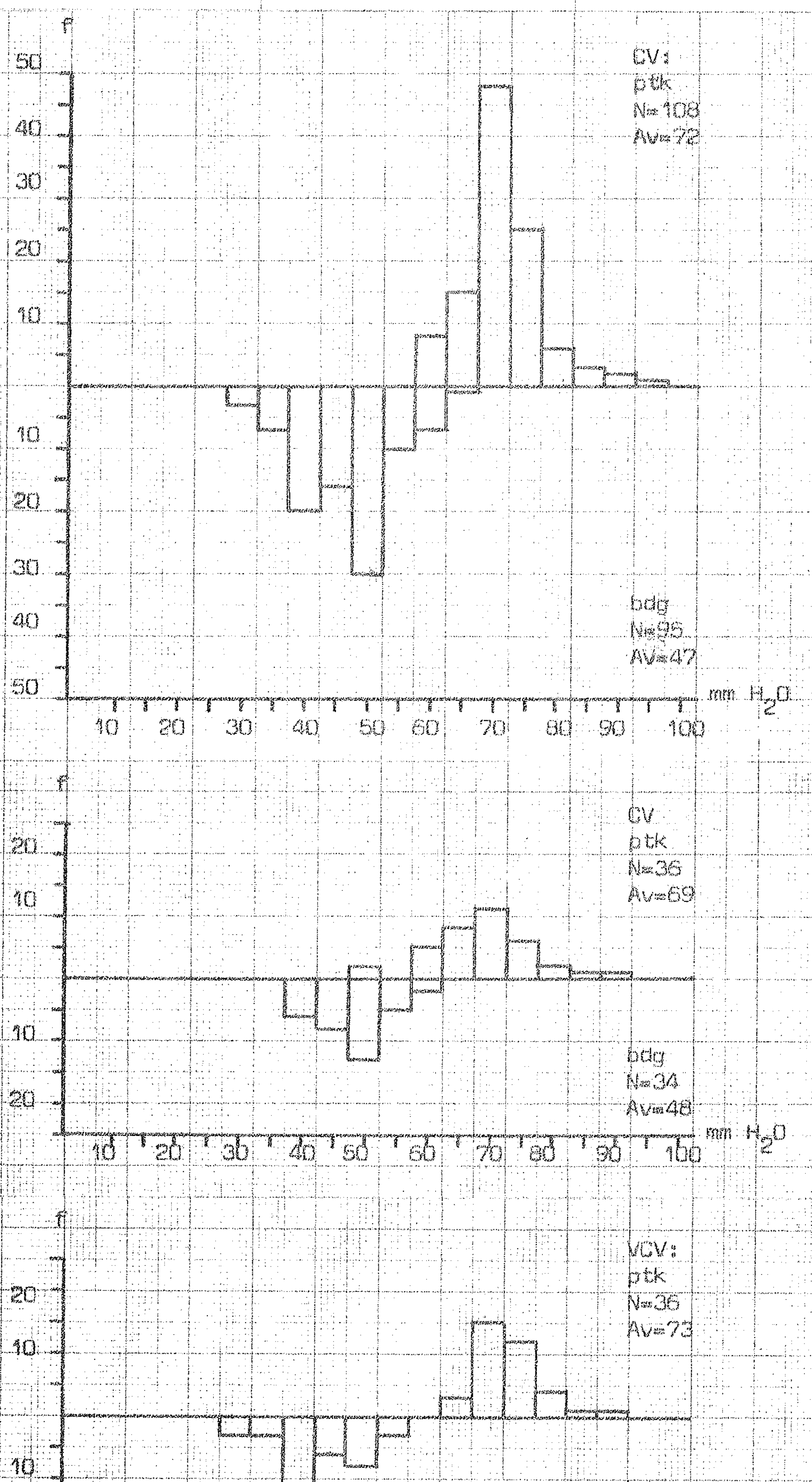


Fig. 1. Frequency distributions for peak pressure values.

7325 01 - 514 A4 - 1 x 1 mm
ESSELTE
4441

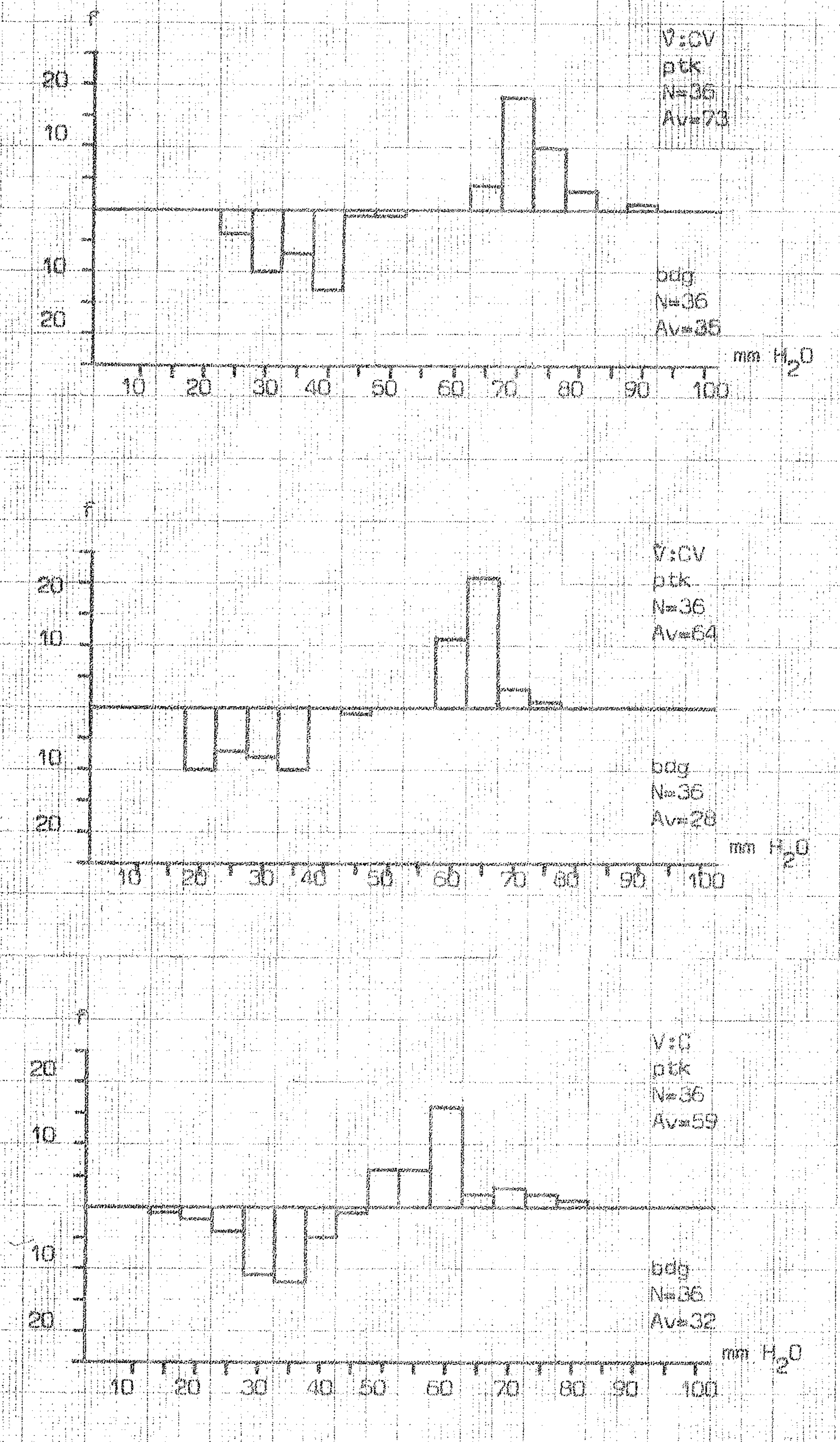


Fig. 2. Frequency distributions for peak pressure values.

ESSELTE
4441

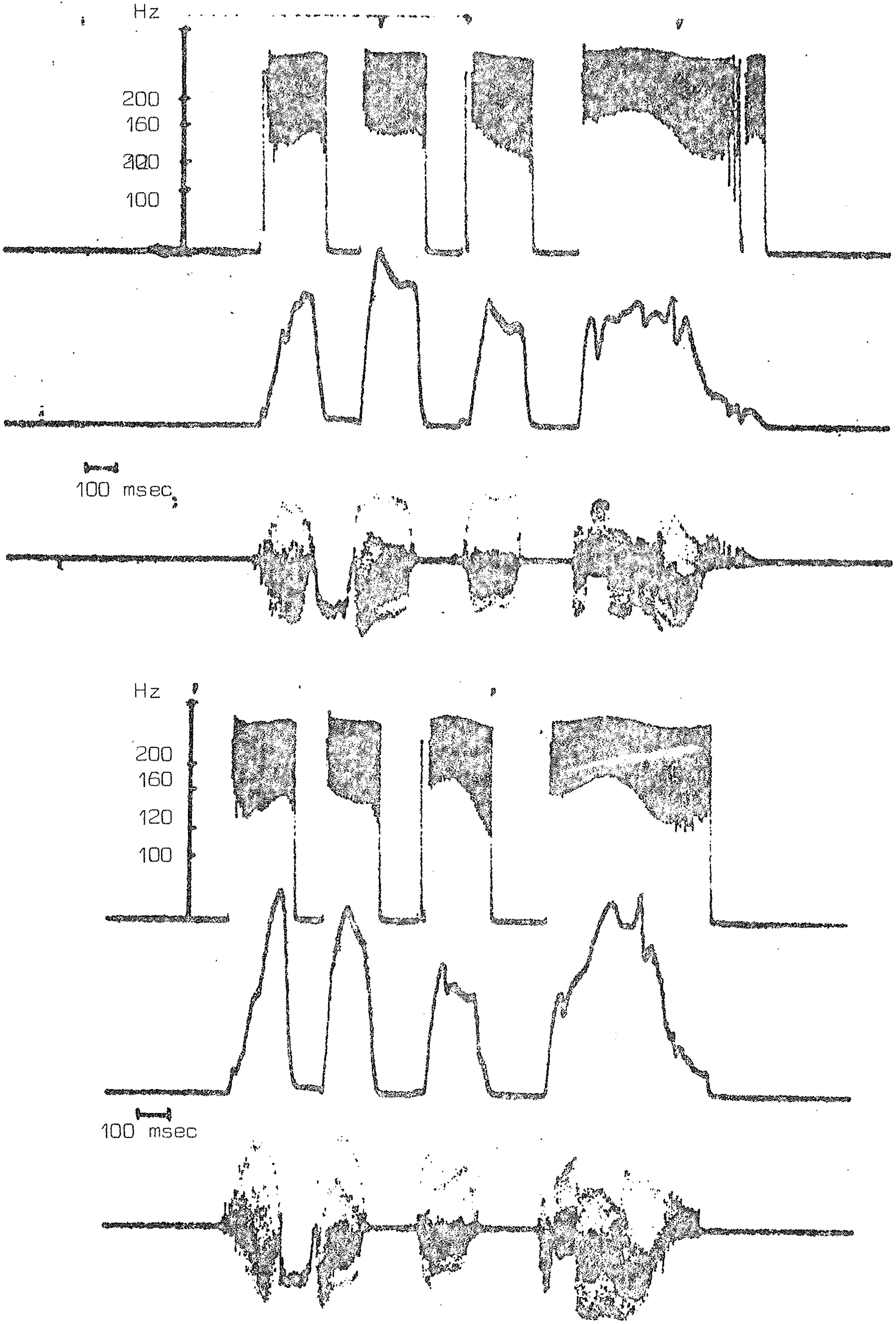


Fig. 3. F_0 curve, intensity curve and duplex oscillogram for the utterances "Ja sa pápen igen" (above) and "Ja sa pàpen igen" (below).

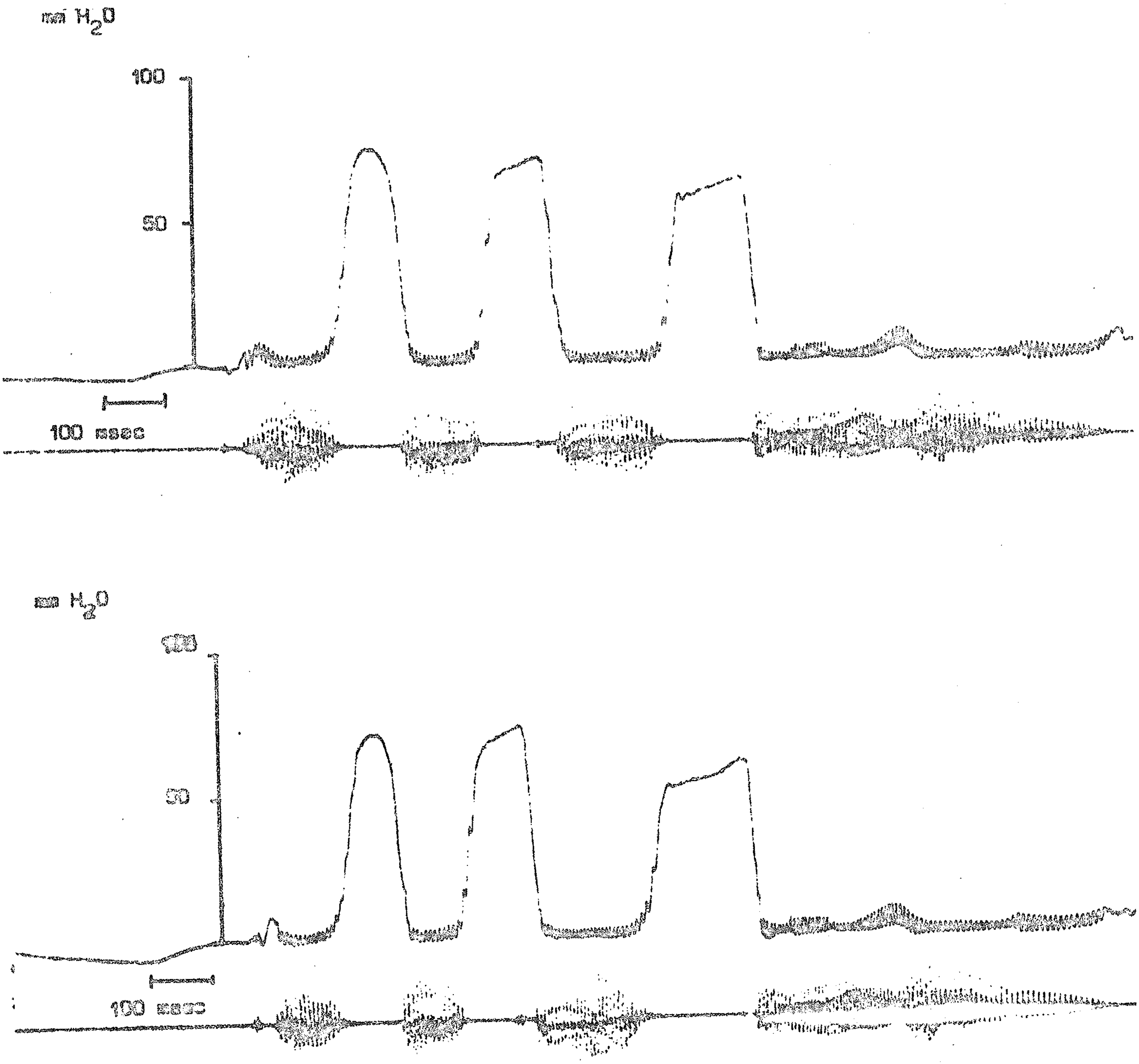


Fig. 4. Supraglottal air pressure during the utterances "Ja sa kǻken igen" (above) and "Ja sa kå̀ken igen" (below).

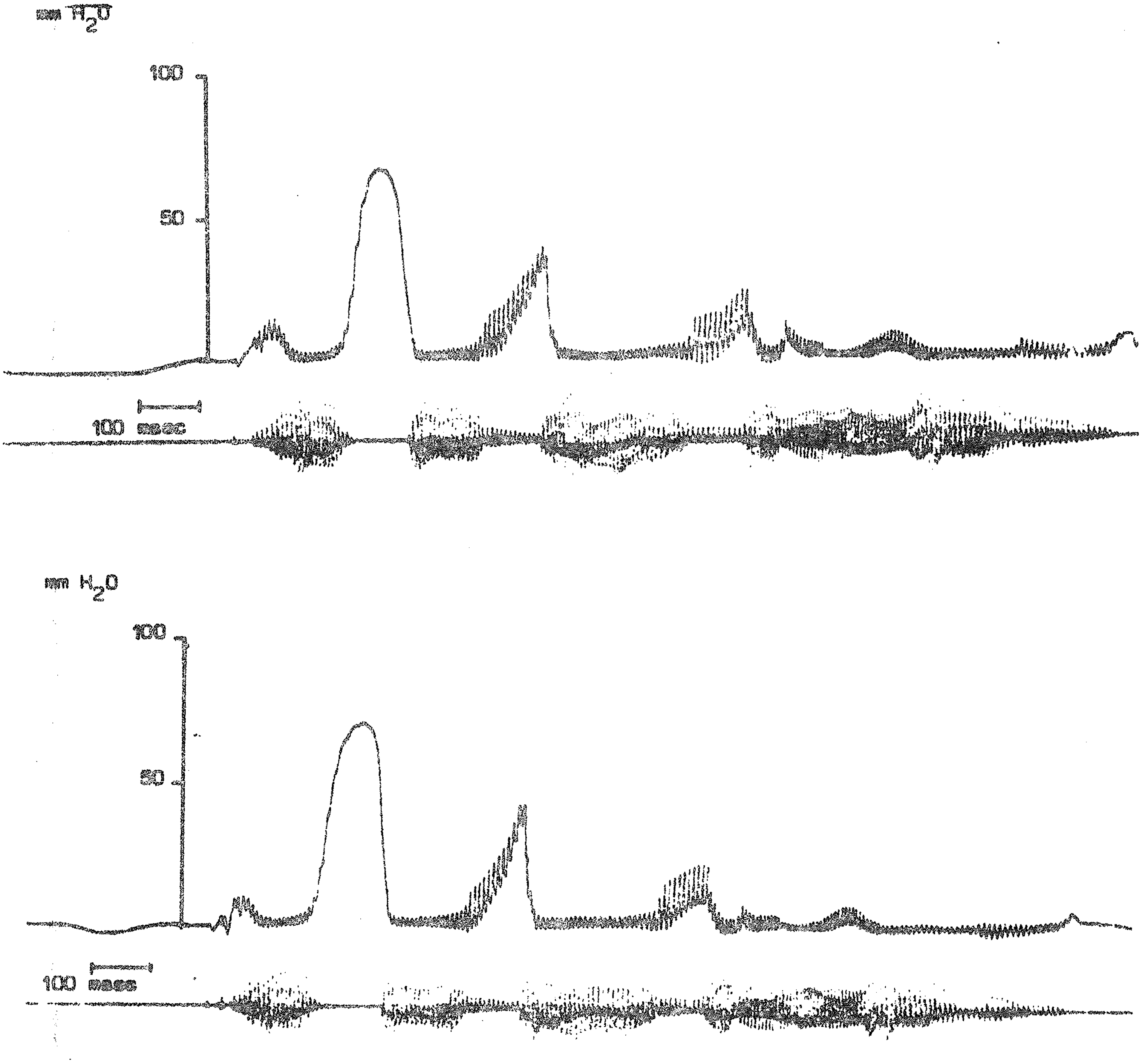


Fig. 5. Supraglottal air pressure during the utterances "Ja sa gágen igen" (above) and "Ja sa gâgen igen" (below).

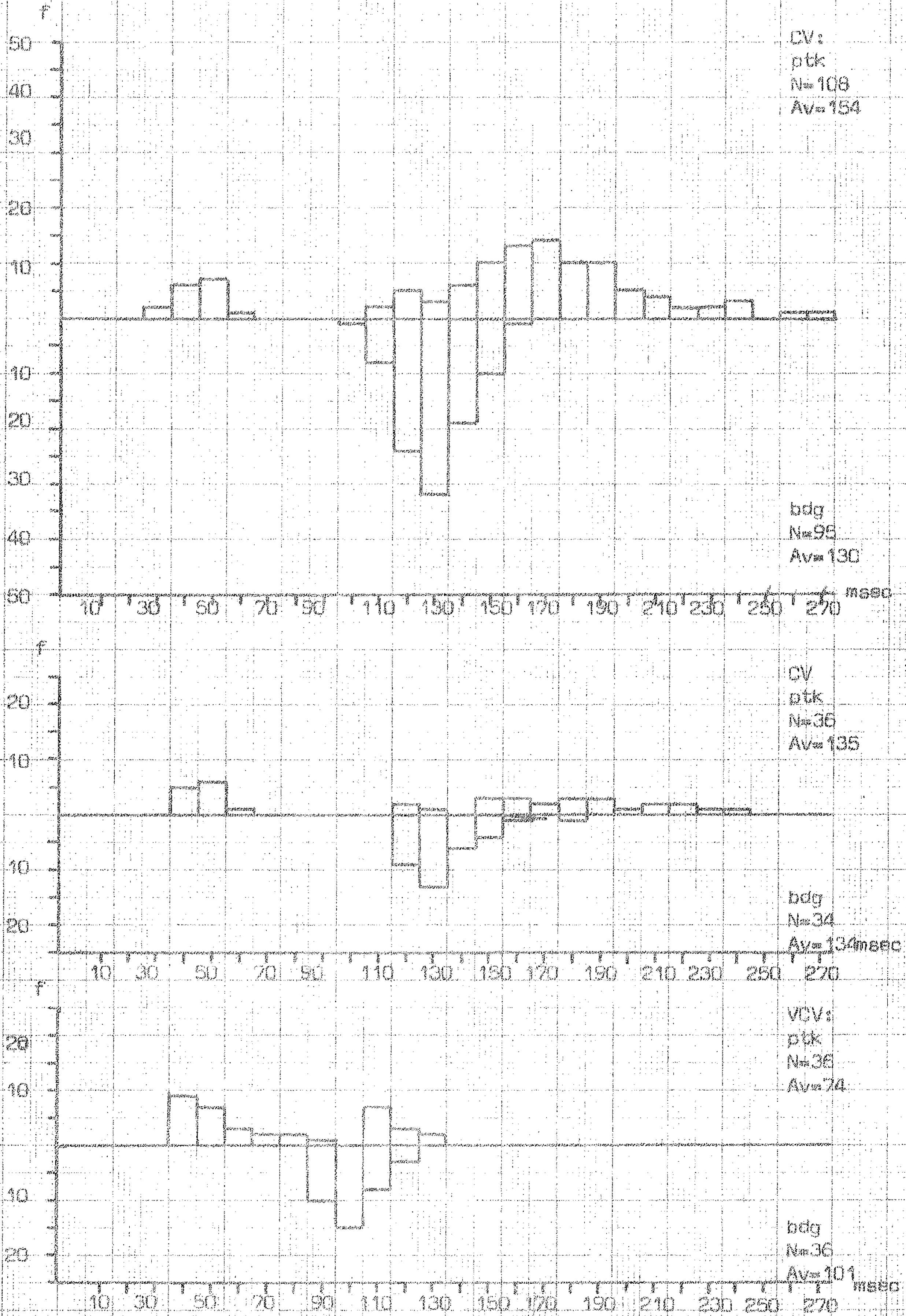


Fig. 6. Frequency distributions for the duration of the interval during which pressure builds up to its peak value.

ESSELTE
4441

II

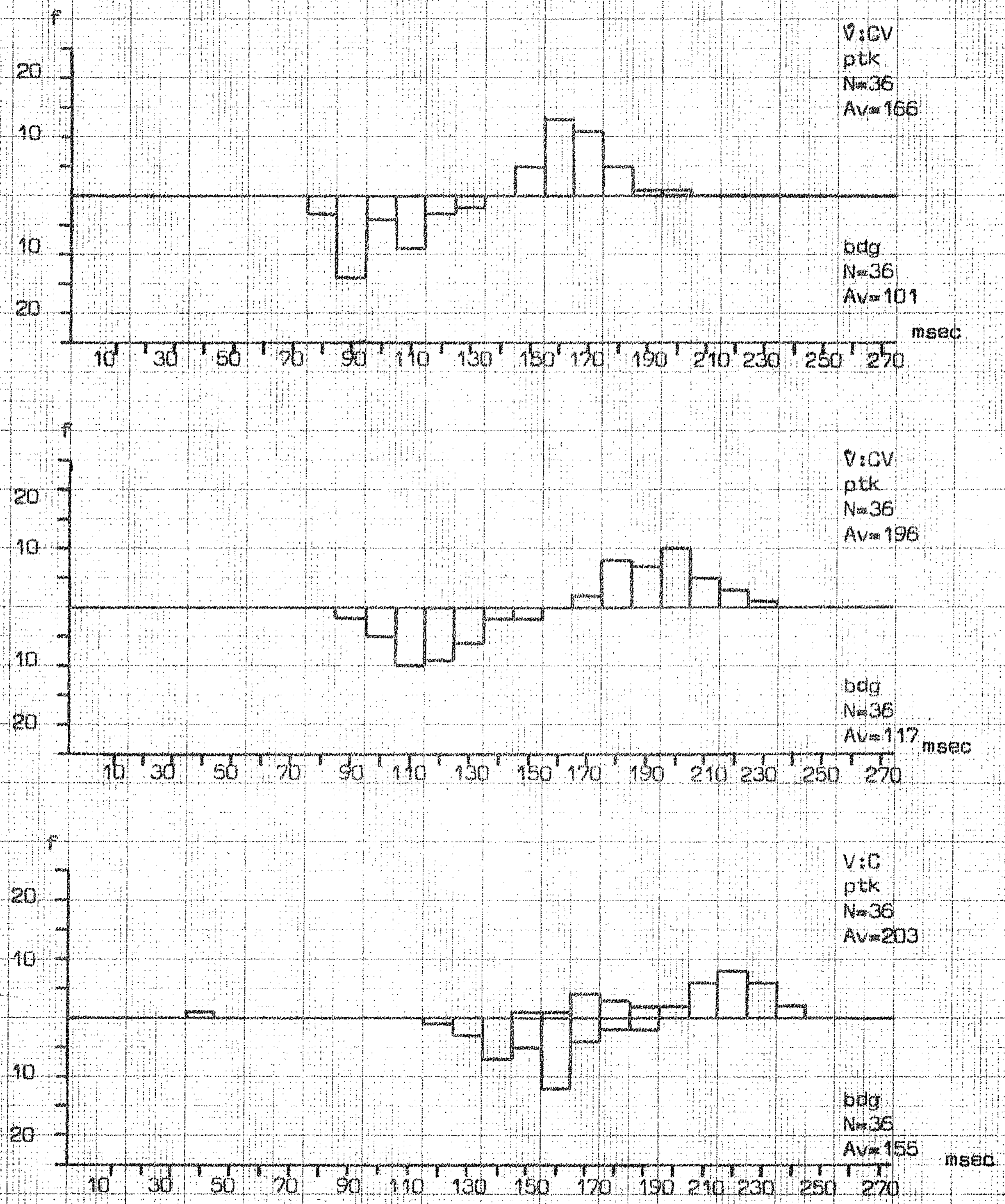


Fig. 7. Frequency distributions for the duration of the interval during which pressure builds up to its peak value.

SIS 73 25 01 - 514 A4 - 1 x 1 mm
ESSELTE
4441

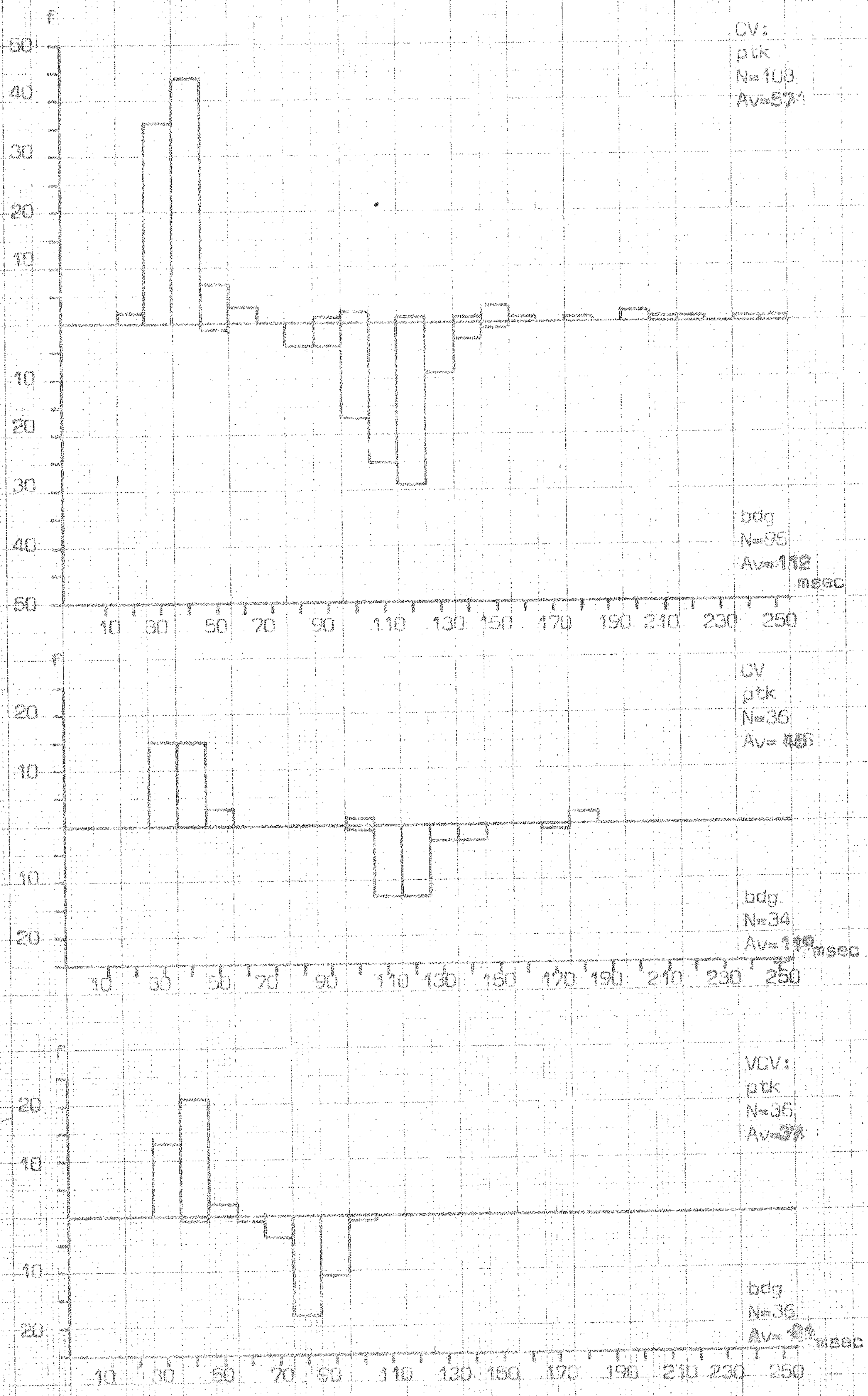


Fig. 8. Frequency distributions for the duration of the interval during which pressure builds up to 85% of its peak value.

79250 - S4 A4 - 1x1mm
ESSELTE
4441

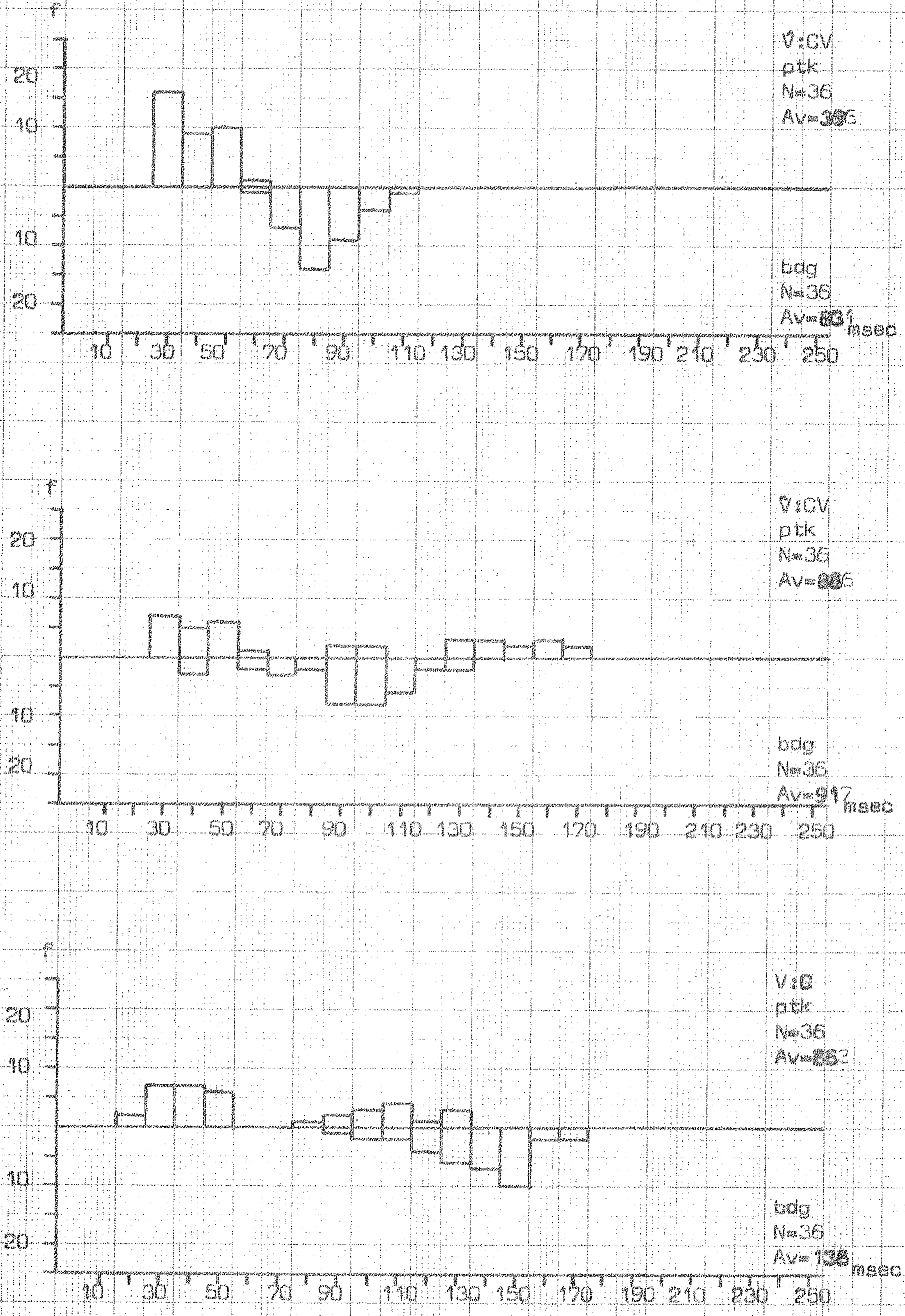
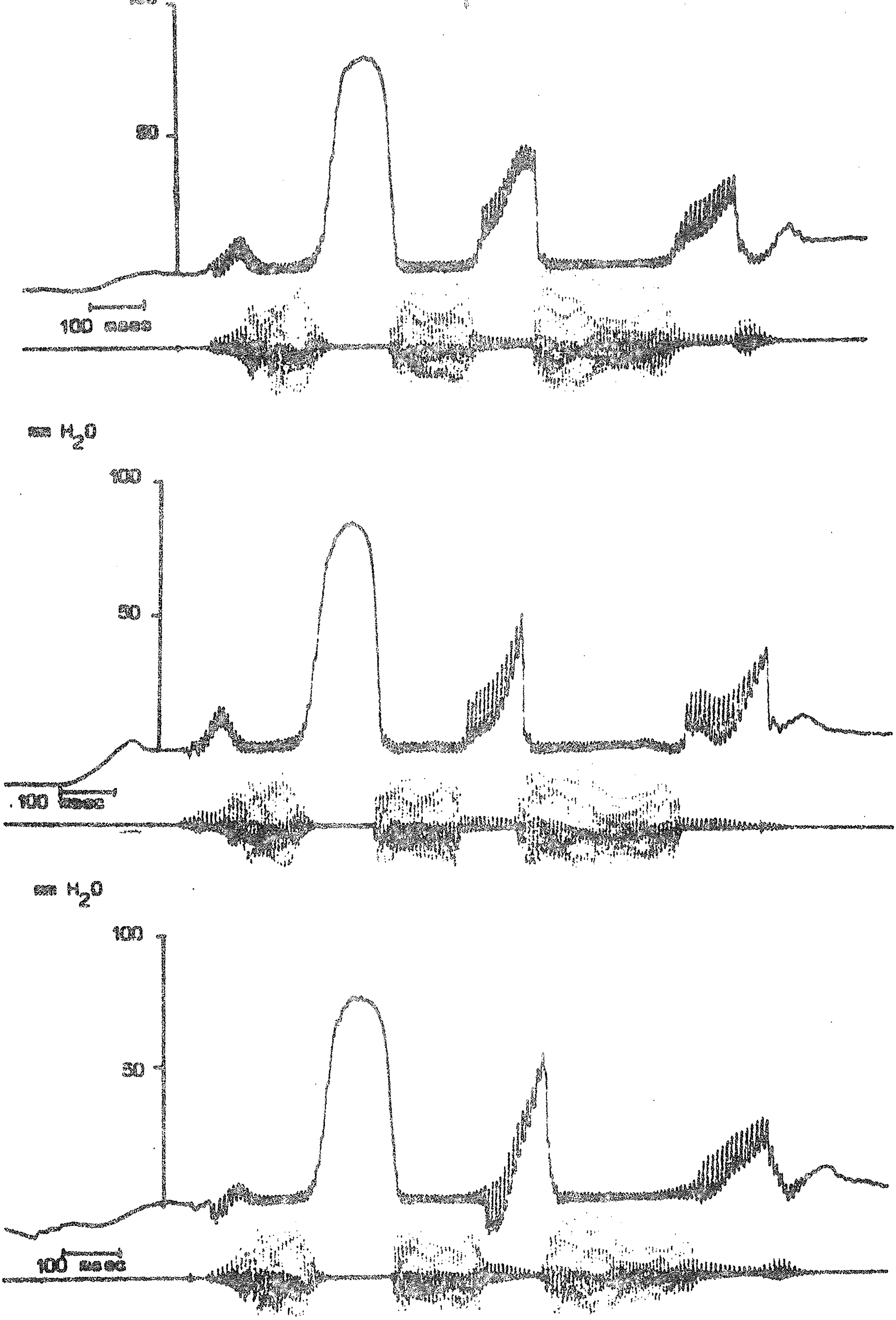


Fig. 9. Frequency distributions for the duration of the interval during which pressure builds up to 85% of its peak value.

732501 - 514 A4 - 1 x 1 mm
ESSELTE
4441



- Fig. 10. Supraglottal air pressure during the utterances "Ja sa bab" (top), "Ja sa ded" (middle) and "Ja sa gag" (bottom).

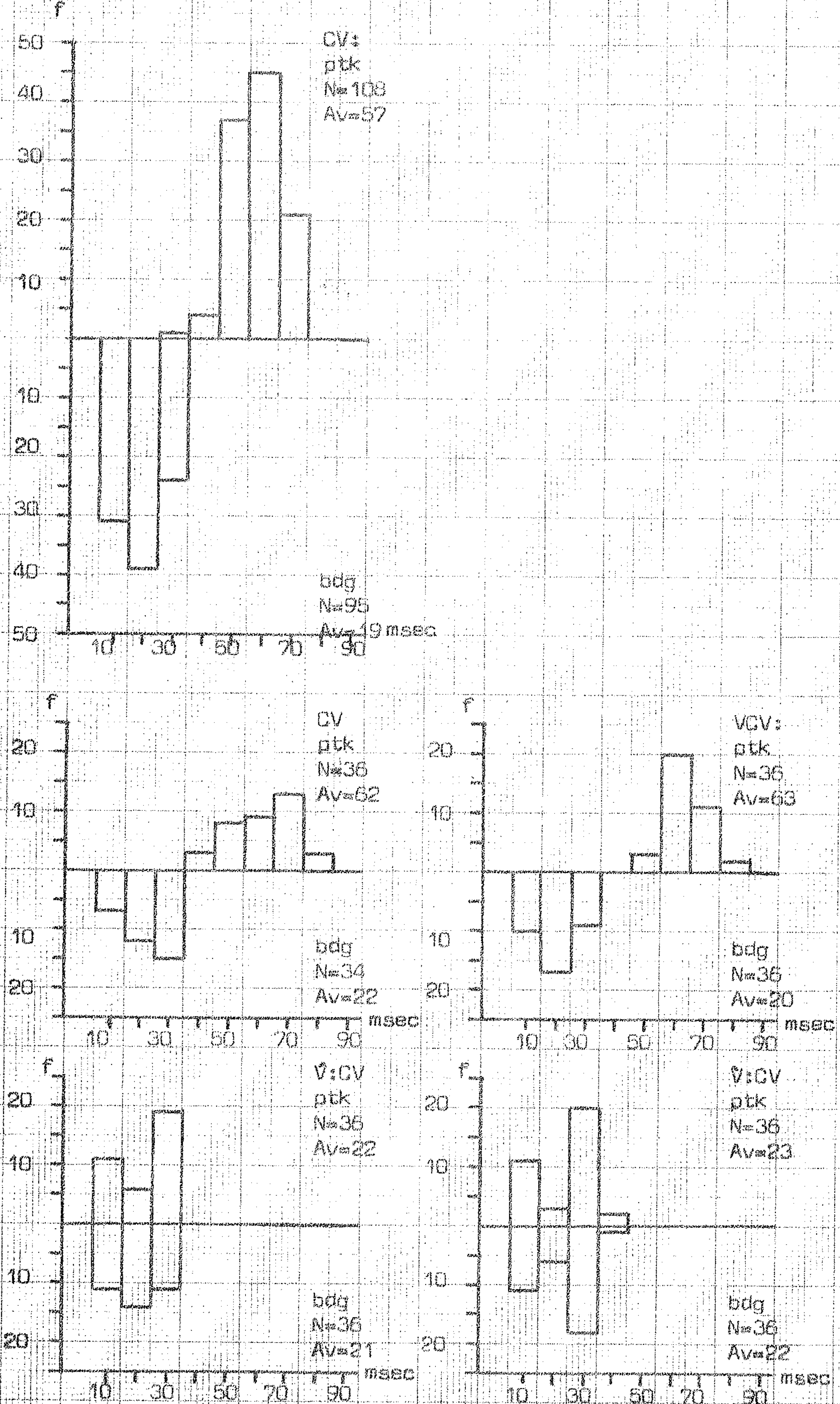


Fig. 11. Frequency distributions for pressure decay.

SIS 73 25 01 - 514 A4 - 1 x 1 mm

ESSELTE
4441

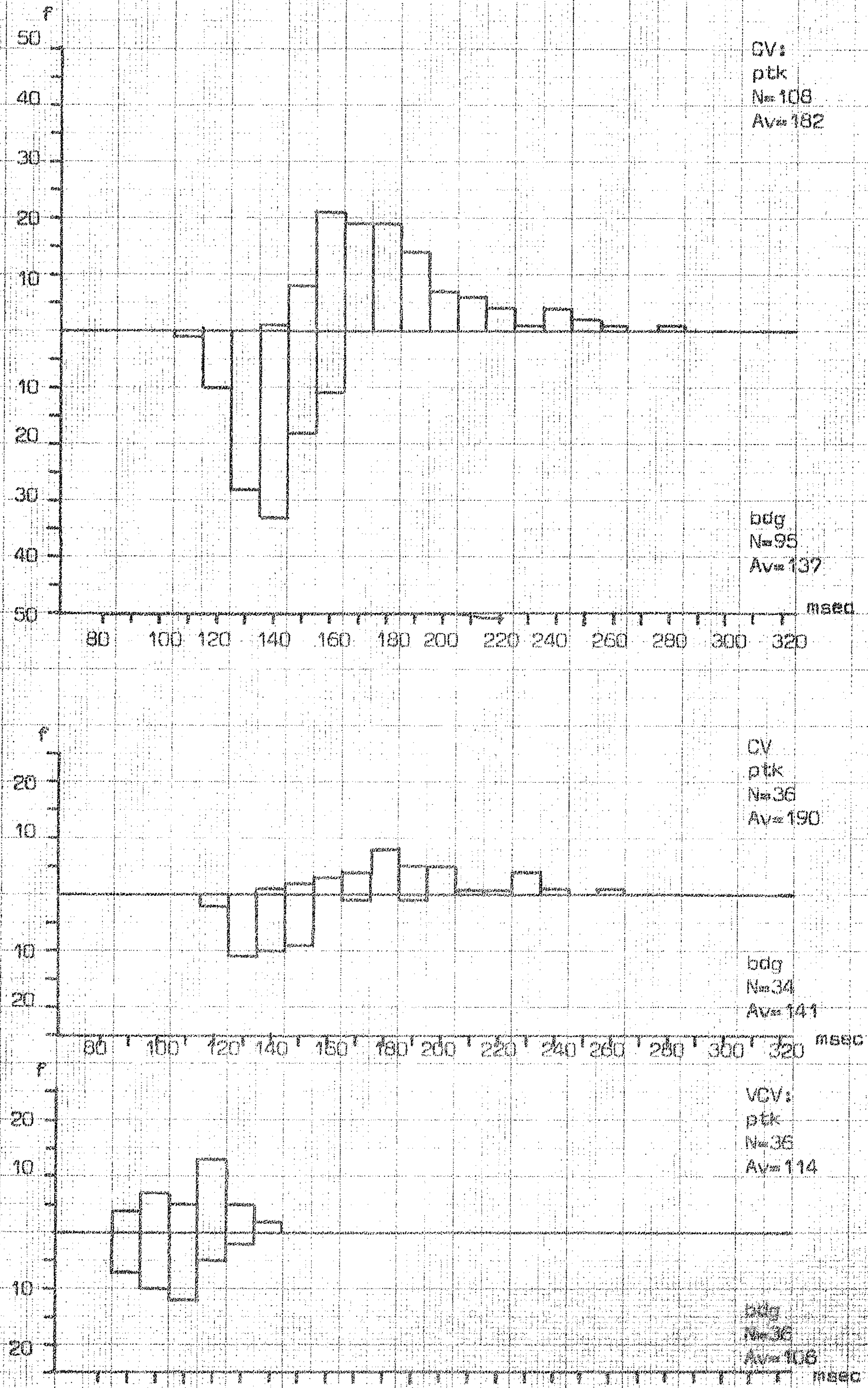


Fig. 12. Frequency distributions for duration.

732501 - 514 A4 - 1x1 mm
ESSELTE

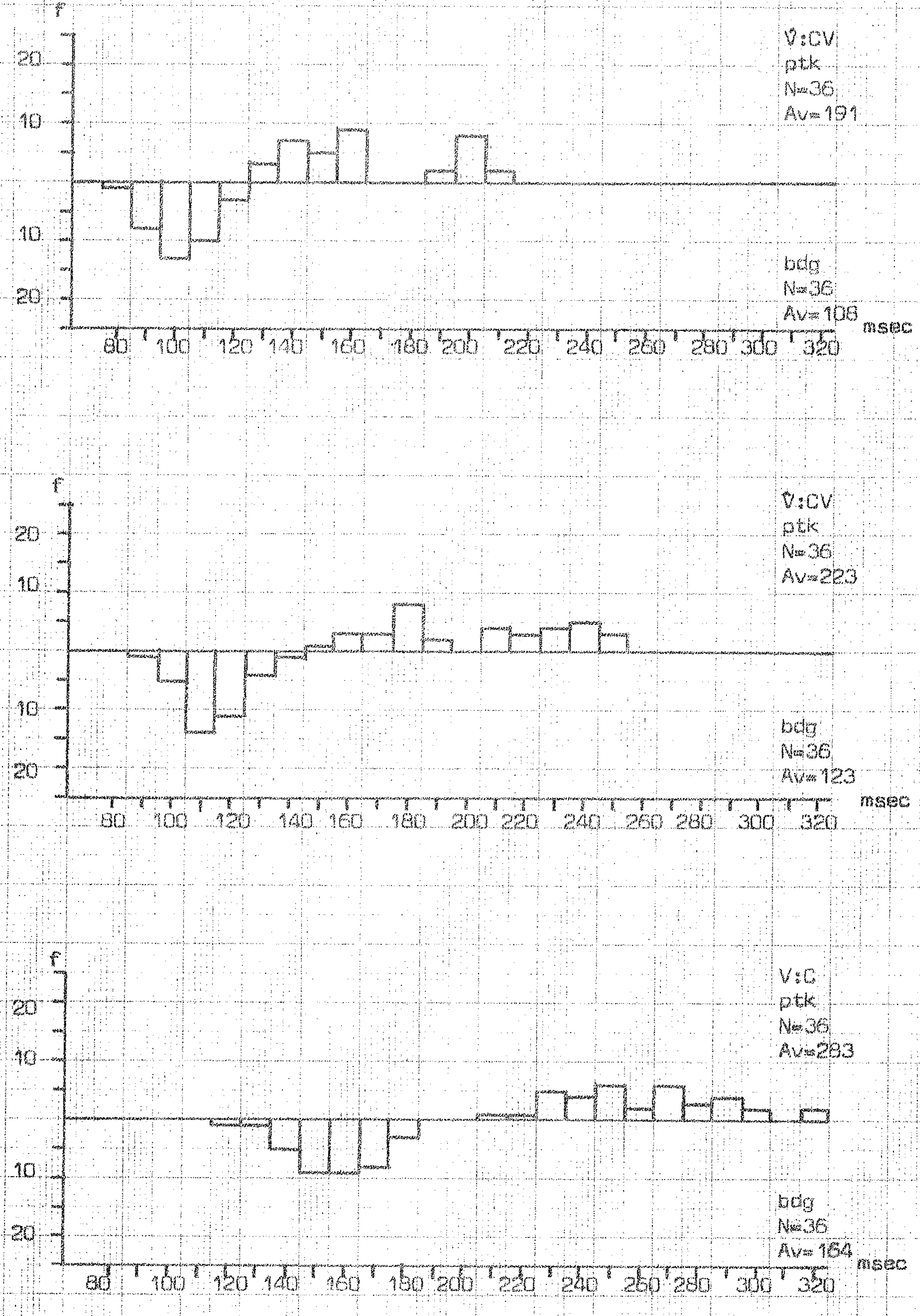


Fig. 13. Frequency distributions for duration.

ESSELTE 4441
732501 · 514 A4 · 1×1 mm