CROSSLANGUAGE STUDY OF VOWEL PERCEPTION

K. N. STEVENS
Massachusetts Institute of Technology
A. M. LIBERMANN and M. STUDDERT-KENNEDY
Haskins Laboratories, New York
S. E. G. ÖHMAN
Royal Institute of Technology, Stockholm

Reprinted from
LANGUAGE AND SPEECH
Vol. 12, Part 1, January-March 1969, pp. 1 - 23

CROSSLANGUAGE STUDY OF VOWEL PERCEPTION*

K. N. STEVENS
Massachusetts Institute of Technology
A. M. LIBERMANN** and M. STUDDERT-KENNEDY***
Haskins Laboratories, New York
and
S. E. G. ÖHMAN
Royal Institute of Technology, Stockholm

This study examines the discrimination and identification of synthetic rounded and unrounded vowels by speakers of two languages (English and Swedish). The unrounded vowels are phonemic in both languages, whereas the rounded vowels are phonemic only in Swedish. A subsidiary aim of the study is to compare the perception of the synthetic vowels with that of synthetic consonant-vowel syllables in which the consonants are stops arranged along a continuum from /b/ to /d/ to /g/. The data indicate that the ability of subjects to discriminate between the vowels is relatively independent of their linguistic experience: Swedish and American English subjects exhibit similar performance in the discrimination tests, though they have somewhat different identification functions. The discrimination functions are characterized by peaks and valleys, suggesting that listeners can discriminate given shifts in the vowel formant frequencies more readily in some vowel regions than in others. Comparison of the data on stop-consonant and vowel perception is consistent with earlier findings: the number of discriminable tokens along the stop-consonant continuum is roughly equal to the number of absolutely identifiable items (three in this case); on the other hand, the number of discriminable vowels is much greater than the number that can be absolutely identified. The data are in accord with the view that a human listener uses different modes in the perception of steady-state vowels and stop consonants.

This study examined the identification and discrimination of synthetic rounded and unrounded vowels by speakers of English and Swedish. Both rounded and unrounded vowels of the experiment occur as phonemes in Swedish; only the unrounded vowels are phonemic in English. Comparison of the results obtained with the two

* *This work was supported in part by grants to the Research Laboratory of Electronics from the U.S. Air Force Cambridge Research Laboratories, Office of Aerospace Research (Contract No. AF19(628)-5661) and from the National Institutes of Health (Grant NB-04332-05), and in part by grants to Haskins Laboratories from the National Science Foundation and the National Institute of Child Health and Human Development. The early phases of this work were carried out at the Royal Institute of Technology in Stockholm, Sweden. The authors acknowledge with gratitude the co-operation of Dr. Gunnar Fant in providing helpful advice and in making the facilities of his laboratory available for the generation of the stimuli and for the experiments with Swedish subjects.*

** *Also, University of Connecticut, Storrs, Connecticut and Yale University, New Haven, Connecticut.*

*** *Also, University of Pennsylvania, Philadelphia, Pennsylvania.*

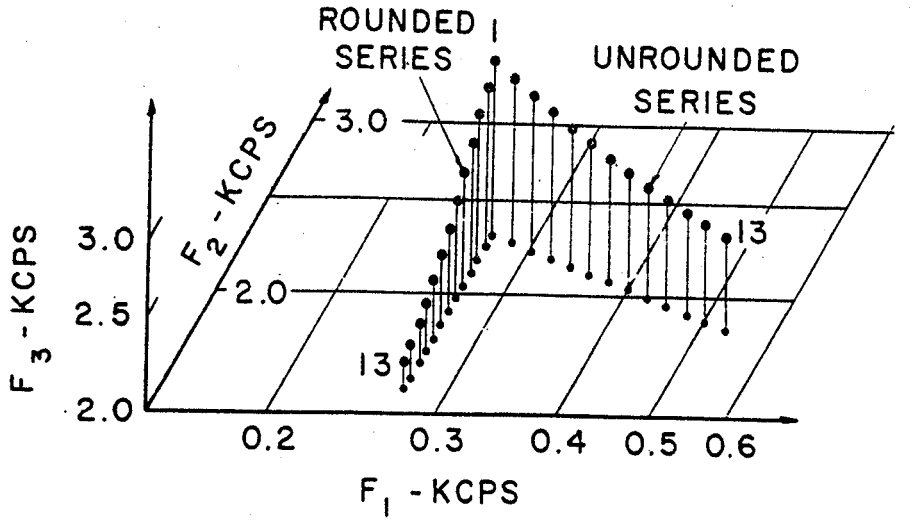

Fig. 1. Arrangement of the stimuli in the rounded and unrounded vowel series. The values of the first three formant frequencies are plotted on the three axes.

groups of listeners gave evidence concerning the effects of linguistic experience on vowel perception.

A subsidiary aim was to repeat, with certain changes and improvements in the stimuli, earlier experiments comparing the perception of (synthetic) steady-state vowels and stop consonants (Liberman, Harris, Hoffman and Griffith, 1957 ; Griffith, 1958 ; Liberman, Harris, Eimas, Lisker and Bastian, 1961 ; Liberman, Harris, Kinney and Lane, 1961 ; Bastian, Eimas and Liberman, 1961 ; Fry, Abramson, Eimas and Liberman, 1962 ; Eimas, 1963).

Description of stimuli

The stimuli were 25 vowels synthesized by the OVE II speech synthesizer at the Speech Transmission Laboratory of the Royal Institute of Technology in Stockholm (Fant, Mártony, Rengman and Risberg, 1962). In Fig. 1 are shown the frequencies of the first three formants of these vowels. The axes represent the formant frequencies F_1, F_2 and F_3 along logarithmic scales. Thirteen of the vowels, designated the *unrounded series,* were arranged so that the first three formants varied in approximately equal logarithmic steps through three unrounded vowels—roughly /i/, /e/, /ä/ in Swedish and /i/, /ɪ/, /ɛ/ in English. Stimulus number 1 of this series also

TABLE 1

Formant frequencies for vowel stimuli, and differences in formant frequencies between adjacent stimuli. Stimuli 13 through 1 constitute the unrounded series ; stimuli 1, R2, R3 . . . R13 constitute the rounded series.

	FORMANT FREQUENCIES (cps.)			CHANGES IN FORMANT FREQUENCIES (cps.)		
STIMULUS NUMBER	F_1	F_2	F_3	ΔF_1	ΔF_2	ΔF_3
13	530.5	1858	2492			
12	501.5	1898	2518	+ 29	− 40	− 26
11	472.5	1926	2544	+ 29	− 28	− 26
10	444.5	1966	2556	+ 28	− 40	− 12
9	419.5	1999	2581	+ 25	− 33	− 25
8	396.5	2032	2628	+ 23	− 33	− 47
7	374	2070	2666	+ 22.5	− 38	− 38
6	353.5	2103	2719	+ 20.5	− 33	− 53
5	336	2144	2776	+ 17.5	− 41	− 57
4	315	2180	2836	+ 21	− 36	− 60
3	298.5	2226	2902	+ 16.5	− 46	− 66
2	285.5	2262	2960	+ 13	− 36	− 58
1	270.5	2300	3019	+ 15	− 38	− 59
R2	271	2228	2918	+ 0.5	− 72	− 101
R3	269.5	2160	2822	− 1.5	− 68	− 96
R4	269	2099	2706	− 0.5	− 61	− 116
R5	270	2042	2610	+ 1	− 57	− 96
R6	270	1981	2496	0	− 61	− 114
R7	269	1925	2412	− 1	− 56	− 84
R8	270	1860	2338	+ 1	− 65	− 74
R9	269.5	1803	2282	− 0.5	− 57	− 56
R10	271.5	1748	2227	+ 2	− 55	− 55
R11	272.5	1699	2185	+ 1	− 49	− 42
R12	269.5	1650	2148	− 3	− 49	− 37
R13	270	1603	2124	+ 0.5	− 47	− 24

represented the end point of a *rounded series* in which the first formant remained fixed at 270 cps., while the second and third formants varied in approximately equal logarithmic steps from the unrounded /i/ through the rounded Swedish /y/ to the rounded /u/.

The values of the first three formant frequencies for each of the 25 stimuli, together with the differences between corresponding formant frequencies for adjacent stimuli along the continuum, are listed in Table 1. The tabulated values of ΔF_1, ΔF_2 and ΔF_3 indicate that the spacing between stimuli for the unrounded and rounded series varied monotonically, although there were small deviations from uniform spacing. These deviations arose because the formant frequencies for each stimulus could only be set to within a few cps.

The bandwidths of the first three formants for all stimuli were fixed at 60, 80 and 100 cps., respectively. Fourth and fifth formant frequencies were fixed, and a standardized correction for poles above the fifth was used. All stimuli were generated with the same falling pattern of inflection of the fundamental frequency from 125 to 80 cps. The duration of each stimulus was 300 msec., and the level of the excitation signal representing the glottal output was the same for all stimuli. The rise and decay times of the amplitude of the excitation were both 50 msec. Time functions representing the amplitude and frequency of the glottal excitation are shown in Fig. 2. While the amplitude of the glottal excitation was the same from one stimulus to the next in the series, there were changes in overall intensity of the stimuli from one end of a series to the other as the formant frequencies changed. Such changes in overall intensity are an automatic consequence of the shifts in formant frequencies, and occur in natural speech. For the unrounded series this change was greatest, and amounted to about 3 db. from stimulus 1 to stimulus 13 ; for a change of three steps along the continuum (the maximum number of steps used in obtaining a discrimination response) the difference in level was never more than 1 db.

Experimental procedure

In the preparation of identification and discrimination tests, each of the 25 vowels was recorded on magnetic tape. A tape loop of each item was then prepared, and from the tape loops a number of recorded copies of each vowel was made. These were cut into magnetic-tape segments for each vowel, and the segments were then spliced together to form the identification and discrimination tests. All recording, copying and presentation of the material in actual test sessions were carried out on uniformly calibrated Ampex tape recorders.

Two identification tests were prepared—one for the unrounded series and one for the rounded series. The order of stimulus presentation in each test was quasi-random, adjusted so that each of the 13 stimuli followed every other one (including itself) once, giving a total of 13 presentations of each stimulus. This arrangement was used in order to distribute evenly any context effects on the identification of vowels

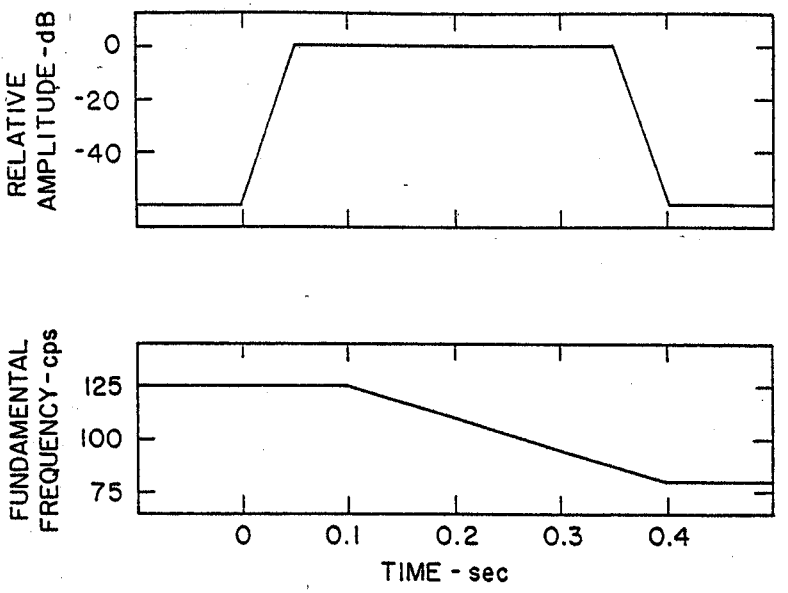

Fig. 2. Time functions used to control the amplitude and frequency of the pulses simulating the glottal excitation for the synthesized vowels.

presented in sequence. Each 169-item test was divided into two roughly equal parts, and the test items in each part were preceded and followed by several stimulus items that were not scored, in order to minimize end effects in the tests.

Two sets of ABX discrimination tests were assembled—one set for the unrounded series and one for the rounded series. Members of each AB stimulus pair were separated by either one, two or three steps along the vowel continuum. For a 13-point continuum there are 66 such pairs, and 132 appropriate ABX triads. Each test consisted of a randomly ordered sequence of these 132 items (preceded and followed by a few buffer items); three such tests with different random orders were prepared for each of the two vowel series.

All tests were administered through binaural headphones to groups of American English and Swedish subjects. For the unrounded vowel identification tests, the Swedish listeners were required to identify each vowel as one of the Swedish vowels /i/, /e/ or /ä/; for the rounded series the response ensemble was /i/, /y/ and /ʉ/. Sample words were written in the instructions as illustrations of the use of these vowels in Stockholm Swedish. Since the two rounded vowels /y/ and /ʉ/ do not occur in American English, and since it was necessary to equalize the response tasks in the rounded and unrounded identification tests, the American subjects were instructed to identify the stimuli in both tests by numbers rather than by phonetic symbols. For the unrounded vowels, the possible responses were /i/, /ɪ/ and /ɛ/, numbered 1, 2

and 3, respectively ; they were identified for the listeners with sample words containing the vowels. For the rounded vowel tests a recording was played (before the test began) of a Swedish speaker producing samples of each of the vowels and identifying them by number.

Subjects

Eleven female Swedish listeners, aged 18-25 years, took the identification and discrimination tests. All but one of these subjects had spent most of their lives in Stockholm, and we considered them to be a uniform group as far as their dialect of Swedish was concerned. The American English subjects were female undergraduate college students, aged 17-20 years, at a university in New York. These listeners had lived in various regions of the United States, and their linguistic backgrounds were not as homogeneous as those of the Swedish group. None of the American listeners could speak Swedish, but some of them had college-level familiarity with other Western European languages.

Initially, 11 American English subjects participated in both identification and discrimination tests. Analysis of the identification data from the unrounded vowels (i.e., the vowels with which the subjects were presumably familiar) for these listeners showed considerable individual differences. Some subjects did not give consistent responses to certain vowels ; others provided well-defined identification functions and for certain stimuli gave the same response on every presentation. Analysis of the discrimination data for these two sub-groups of American English listeners (i.e., the highly consistent identifiers and the less consistent identifiers) indicated that there was no significant difference in the discrimination functions for the two groups: the level of performance on the discrimination tests for a given subject did not seem to be related to level of performance on the identification tests. (This point will be discussed in more detail later.) In order to obtain more reliable identification functions for American English, an additional group of 28 undergraduate subjects was given only the identification tests. The results to be reported here represent discrimination data for the initial 11 subjects and identification data for the larger group of 28 subjects.

Results

Identification Functions : Unrounded Vowels

Average identification functions for the two groups of listeners for the *unrounded* series are shown in Fig. 3. The American English listeners are in less agreement than the Swedish listeners on the identification of these vowels, particularly of the vowel /ɪ/. A similar degree of response variability was displayed by the initial American English group of 11 subjects who participated in both identification and discrimination tests.

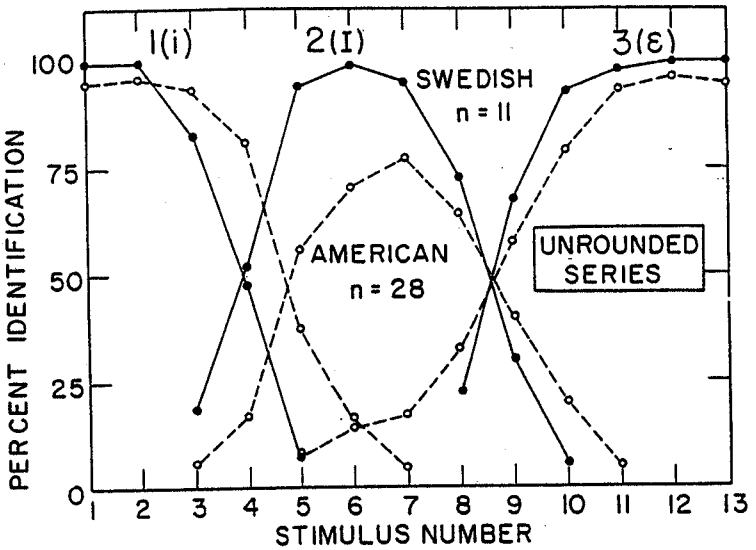

Fig. 3. Identification functions for the unrounded vowel series. The categories were /i e ä/ for the Swedish listeners and /i ɪ ɛ/ for the American listeners.

Some of the American variability in this and in the rounded vowel identification test may stem from difficulties with the number names for the vowels.

To obtain a group of Americans more nearly comparable in response consistency to the Swedish listeners, and to see the relation, if any, between response consistency and phoneme boundaries, we divided the 28 American subjects into two sub-groups: those who had, in each of the three vowel categories, assigned the same phoneme label to one or more stimuli at least 90 per cent of the time, and those who had met an analogous criterion of consistency at the lower level of 60 per cent. There were 11 subjects in the former sub-group (Group 1) and 13 in the latter (Group 2). The identification functions for these two sub-groups are shown in Fig. 4. The cross-over points between phonemes for the more (Group 1) and less (Group 2) consistent subjects differ by no more than half a step. The response consistency of the better American sub-group is similar to that of the Swedish listeners.

Differences in phoneme location for the Swedish and American groups occur primarily for the vowel /ɪ/. Whether we look at the more consistent sub-group of Americans or at the whole group, we find the cross-over point (phoneme boundary) between /i/ and /ɪ/ to be approximately one step higher on the continuum for the Americans than for the Swedes. The phoneme centre (point of greatest agreement) for /ɪ/ is also one step higher for the Americans.

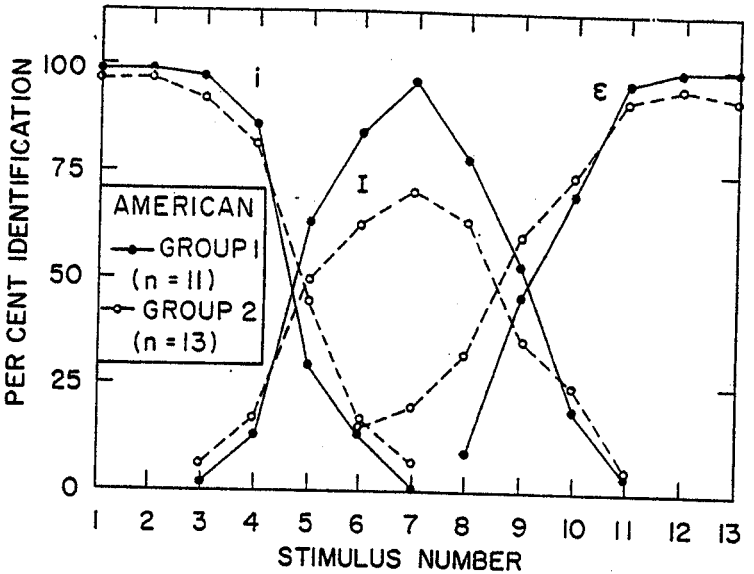

Fig. 4. Identification functions for two groups of American listeners for the unrounded vowels. Group 1 listeners were the more consistent identifiers and Group 2 listeners were the less consistent ones.

Identification Functions : Rounded Vowels

The identification curves for the rounded vowels, displayed in Fig. 5, again show a less consistent performance for the American English listeners than for the Swedish, though the Swedish listeners themselves have some difficulty with the centre vowel, /y/. If the data are examined for the two sub-groups of American listeners considered in connection with Fig. 4, the sub-group that gives less consistent responses for the unrounded vowels also tends to give less consistent responses for the rounded, but the difference between the two groups is not nearly as marked: both American sub-groups are less consistent than the Swedish group, particularly for /y/ and /ʉ/. This is to be expected, of course, since these rounded vowels are not in the phoneme system of American English.

Differences in phoneme location for the Swedish and American listeners are again evident. The cross-overs from /i/ to /y/ and from /y/ to /ʉ/ are approximately one step higher on the continuum for the American than for the Swedish listeners. The point of greatest agreement on /y/ is two steps higher for the Americans.

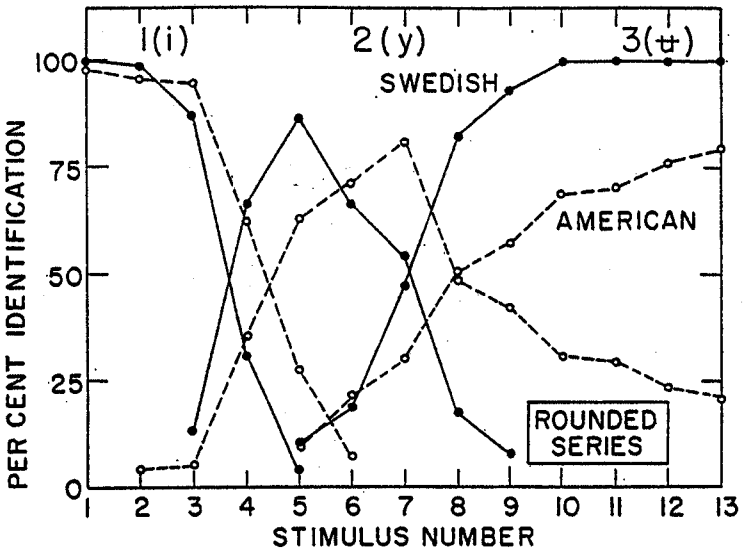

Fig. 5. Identification functions for the rounded vowel series. The categories were /i y ʉ/ for the Swedish listeners, whereas the American listeners categorized the vowels by number.

Context Effects in the Identification Functions

More detailed analysis of the identification data for both sets of vowels indicates that the identification of a given vowel depends to some extent on the context in which it appears in the test. Such effects of context have been noted by Fry, Abramson, Eimas and Liberman (1962) and by Eimas (1963). For example, a stimulus located near a boundary (such as stimulus 4 in the unrounded series) would be more likely to be identified as being in category 1 if it were preceded by a stimulus numbered higher than 4 and in category 2 if it were preceded by stimulus 1, 2 or 3. One general effect of such contextual influences would be to shift phoneme boundaries that are located close to the end of a series (such as the boundary between /i/ and /y/ in Fig. 5) in a direction toward that end of the series (to the left in Fig. 5). The data indicate, however, that this shift is less than one step.

Discrimination Functions: Unrounded Vowels

Results of the discrimination tests on unrounded vowels for the Swedish and American English listeners are compared in Fig. 6. The three pairs of curves represent

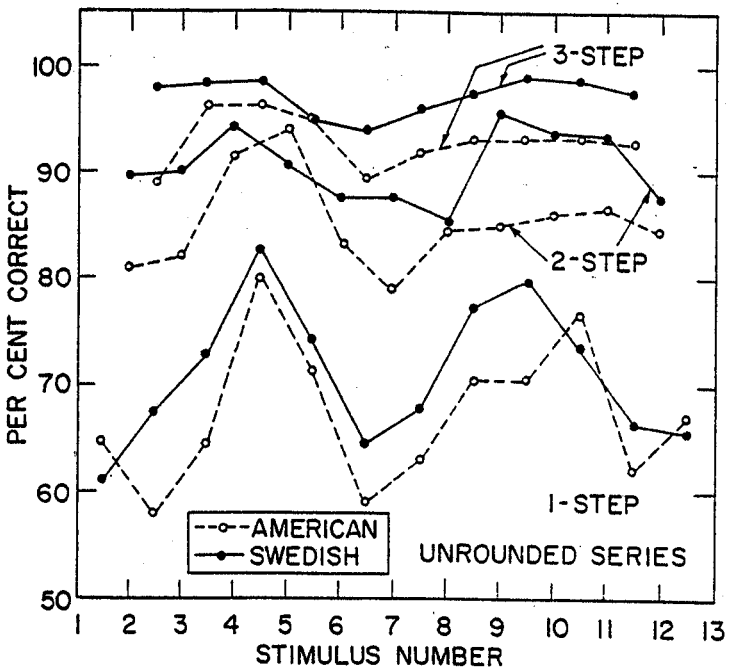

Fig. 6. Discrimination functions for the unrounded vowels for Swedish and American listeners.

judgments for AB pairs separated by one, two and three steps on the 13-point continuum. The general effect of increasing the size of the stimulus differences to be judged is to raise the overall level of the curves and to reduce their variability. For both groups of listeners the curves show peaks and valleys: the valleys correspond to the centres of the vowel regions established from the identification curves, the peaks to the phoneme boundaries. But the correspondence is rough, and differences in identification functions for the two groups, such as in the location of the boundary between /i/ and its neighbour, are not reflected by consistent differences in discrimination functions.

The American listeners discriminate less well than the Swedish, perhaps because their tape-recorded tests were dubbed from those heard by the Swedish listeners. Slight random tape speed variations, even within the specified limits of a good tape recorder, could cause deterioration in test results for the discrimination tests, since the formant frequency differences are quite small.

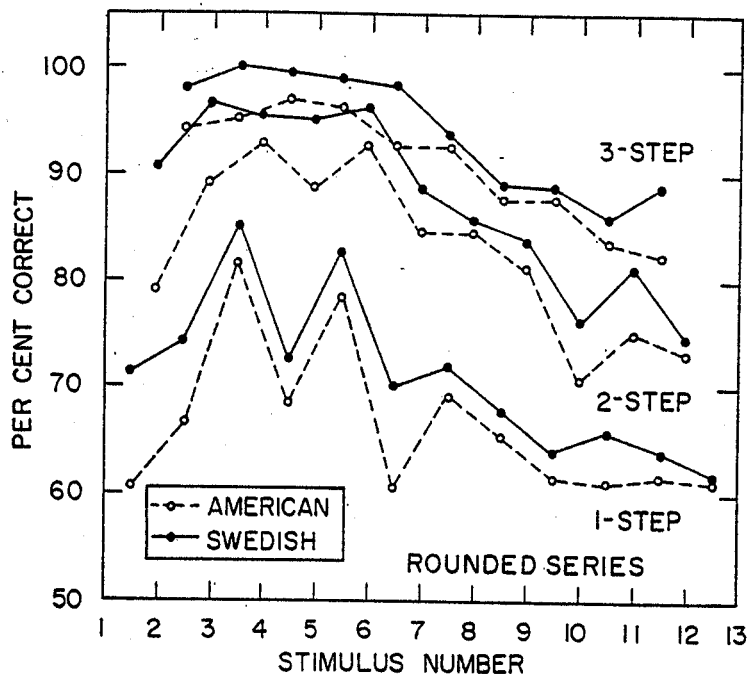

Fig. 7. Discrimination functions for the rounded vowels for Swedish and American listeners.

Discrimination Functions : Rounded Vowels

The discrimination curves for the rounded vowels, shown in Fig. 7, are very similar for the two groups of listeners. The similarity in shape is particularly striking in view of the different linguistic backgrounds of the two groups, as clearly reflected in their identification functions. The implication that linguistic experience has little effect on vowel discrimination will be taken up in the discussion section.

Peaks and valleys occur, at least in the 1-step and 2-step functions, although they are not as sharply defined as for the unrounded series. The minima in the discrimination functions appear in regions where the identification functions are greatest for Swedish listeners: in the vicinity of stimuli 1 and 2 for /i/, stimulus 5 for /y/, and stimuli 9 through 13 for /ʉ/. All curves show a marked decline in discriminative performance over the upper half of the continuum, the region roughly occupied by the vowel /ʉ/ in the Swedish identification functions.

As with the unrounded vowels, the American listeners discriminate less well than the Swedish, perhaps again due to deterioration of the stimuli with additional dubbing.

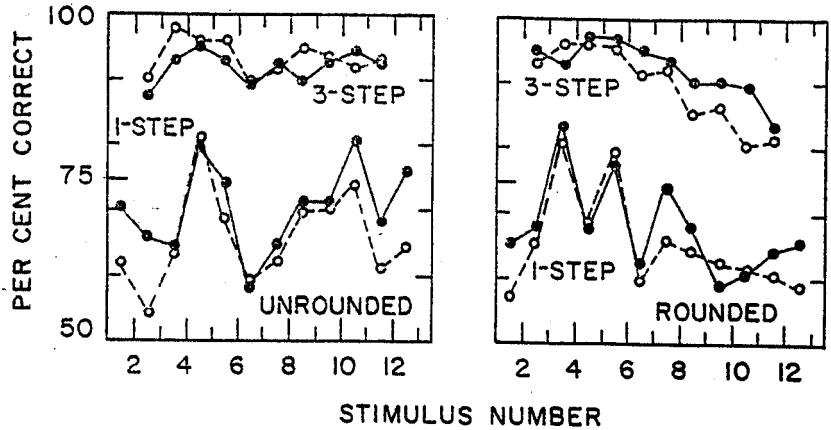

Fig. 8. Discrimination functions for unrounded and rounded vowels for two sub-groups of American listeners: (1) those who were most consistent in their identification of the unrounded vowels (solid curves), and (2) those who were least consistent in their identification of the unrounded vowels (dashed curves).

Relation Between Identification and Discrimination Functions for More and Less Consistent American Subjects

As earlier remarked, phoneme identifications of both rounded and unrounded vowels were made far less consistently by the American than by the Swedish listeners, and some American subjects were more consistent identifiers than others. It is appropriate, then, once again to divide the American subjects into two sub-groups, putting the more consistent identifiers in the one sub-group and the less consistent identifiers in the other, and to compare the discrimination functions for these two sub-groups. The results of that division, for the 11 subjects on whom discrimination functions were obtained, are shown in Fig. 8. We see that the discrimination functions for the two sub-groups are nearly identical, both for the unrounded and for the rounded vowels, indicating no relation between a subject's ability to identify these stimuli as phonemes and his ability to discriminate them as being different on any basis whatsoever.

Discussion

If we extrapolate our results to other vowels and to other languages, the conclusion to be drawn is that, for vowels presented in isolation, the listeners' linguistic experience has essentially no effect upon their ability to discriminate small differences in the vowel formant frequencies. Relevant to this conclusion, and more or less in keeping with it,

is the fact that, in the data of the American listeners, there is no relation between the consistency with which they identify the stimuli and the accuracy with which they discriminate them.

Studies of consonantal identification and discrimination (Liberman, Harris, Hoffman and Griffith, 1957) have suggested that fluctuations in discrimination functions, similar to those of this study, might be attributed to listeners' learned phoneme identifications.[1] Here, however, since linguistic experience seems to have no effect upon the shape of the discrimination functions, it is tempting to speculate that functions of the type shown in Figs. 6 and 7 provide evidence for the manner in which the human auditory mechanism discriminates vowels, whatever system of vowels listeners have learned to produce and identify. In what follows we elaborate this hypothesis for the two vowel series of this experiment.

Unrounded vowels

Certain peaks and valleys occur in the unrounded vowel discrimination functions of both Swedish and American listeners. In Fig. 6 the low level of the one-step discrimination function in the vicinity of stimuli 1 through 3 and the relatively high level in the range of stimuli 4 and 5 indicate that all listeners, regardless of their language, tend to have difficulty discriminating between stimuli 1 and 2 or between 2 and 3, but can more easily discriminate between stimuli 4 and 5. In some sense, stimuli 1, 2 and 3 lie close together along a perceptual continuum and tend to sound alike, whereas stimuli 4 and 5 lie farther apart. We believe that, as with the rounded vowels, these differences in discriminability are not a result of experience with linguistic categories, but are, rather, a reflection of some basic property of the auditory mechanism. We would suggest, further, that such a property would favour the formation of a linguistic category in the vicinity of stimuli 1 through 3. That would be to say that this part of the discrimination function is not a consequence of the linguistic category but a cause of it.

Let us suppose that the articulatory mechanism is capable of generating any sound along the acoustic continuum represented by the abscissa in Fig. 6, and that certain regions along this continuum are to be assigned to particular linguistic categories. Since stimuli lying in a valley of the discrimination function tend to sound alike, there

[1] *In some studies of consonant perception (Liberman, Harris, Eimas, Lisker and Bastian, 1961; Liberman, Harris, Kinney and Lane, 1961) listeners were asked to discriminate small differences in an acoustic variable under two conditions: (1) when it cued a phonemic distinction and (2) when, in some non-speech context, it did not. Differences in the results were attributed to the effects of experience with the speech signals. There was, however, no part of the experiment—for example, a cross-language comparison—that might have provided direct evidence for that conclusion. An alternative interpretation is that, independently of experience or learning, the perceptual processing was carried out by a special speech mechanism in the former case and by a non-speech auditory mechanism in the latter.*

is no need for precise control of the articulatory configuration that generates sounds in this valley. This reduction of the articulatory demands could predispose a language to evolve linguistic categories centred in regions where discrimination is poor, and bounded by regions where discrimination is relatively good. Coupled with this favourable set of conditions for perception in the region of stimuli 1 through 3 in Fig. 6 there also exists, of course, a well-defined articulatory configuration corresponding to the vowel: an extreme high front tongue position.

The combination of perceptual and articulatory conditions appropriate for the formation of a linguistic category suggests that such a category may exist in many different languages. For this particular range of stimuli (centred on stimulus 1 in Figs. 6 and 7) this is indeed true: the vowel /i/ exists in most, if not all, languages of the world. There are slight differences in the region of vowel space in which /i/ is centred in different languages (depending, for example, upon whether there is a rounded version of /i/ in the language, as in Swedish), but presumably there is no language in which the region penetrates beyond the "barrier" represented by the discrimination peak in Fig. 6. Evidence for this tendency for a series of vowel-like stimuli to be perceived in a non-continuous fashion has also been reported by Chistovich, Fant, de Serpa-Leitão and Tjernlund (1966).

In addition to the minima in the discrimination functions in the vicinity of stimulus 1 (corresponding to the vowel /i/), there also exist minima in the regions of stimuli 6 and 7 and of stimuli 12 and 13. These minima in the discrimination functions also provide "natural" regions within which vowels should have a high probability of occurring in language—in this case the vowels /e/ and /æ/, which are known to occur in many languages.[2]

Rounded Vowels

Peaks and valleys in the rounded vowel discrimination functions are less distinct. There is a peak between stimuli 3 and 4 (Fig. 7) as one passes from the unrounded /i/ to the rounded /y/, suggesting some sort of natural perceptual boundary between these two vowels. This perceptual boundary occurs in the region where there is onset of rounding. Regions of poorer discrimination corresponding to the rounded vowels /y/ (stimuli 4 and 5) and /ʉ/ (stimuli 7 through 13) are also in evidence, but the minima are not well-defined. The minima in the discrimination functions corresponding

[2] *The fact that American English listeners associate these regions with the lax vowels /ɪ/ and /ɛ/ is coincidental. The situation in American English is confounded by the fact that the acoustic manifestation of the phoneme /e/ is not a steady vowel but is diphthongal, while /ɪ/ and /ɛ/ are lax vowels whose identification normally depends upon temporal as well as spectral cues. Presumably stimulus 13 (or a stimulus slightly beyond it, i.e., with higher F_1 and lower F_2) would have been identified by American English listeners as /æ/, if such a category had been available to them in the identification test.*

to the rounded vowel /y/ are quite narrow (and, indeed, there is no minimum in the 3-step function of Fig. 7). Between the centres of the vowels /y/ and /ʉ/ the identification functions show that there is a reasonably broad region in which neither response is unanimous. Thus it might be expected that there is not a well-defined peak in discrimination in this region. The low level of discrimination in the vicinity of stimuli 9 and 13 presumably creates conditions appropriate for the location of the vowel phoneme /ʉ/. It is of interest to note, however, that the vowels /y/ and /ʉ/ occur comparatively rarely in language. It might be suggested that the properties of the auditory mechanism do not favour these vowel categories as strongly as categories such as /i/, /e/ and /æ/.

We return to this discussion below after reporting the results of the subsidiary experiment on stop-consonant perception.

A SUBSIDIARY EXPERIMENT: STOP-CONSONANT PERCEPTION COMPARED WITH UNROUNDED VOWEL PERCEPTION FOR AMERICAN SUBJECTS

Earlier studies (referred to in the introduction) have compared stop consonant and unrounded vowel perception, but with synthetic speech sounds characterized by only two formants. The present experiment added appropriate variations in the third formant and, in other ways, made the sounds more realistic.

Description of Stimuli and Procedure

A set of stop-consonant stimuli was synthesized on the same OVE II synthesizer that was used for generating the vowels. The set consisted of 13 consonant-vowel syllables of 300 msec. duration. The final 260 msec. of each stimulus was a steady vowel having the quality of American English /ɛ/, with the first three formants fixed at 700, 1550 and 2600 cps. During the initial 40 msec., the first three formants underwent transitions along parabolic contours, from specified starting frequencies to the vowel formant frequencies. The starting frequency for the first formant was 220 cps. in all cases. For the second and third formants the starting frequencies are shown in Fig. 9. Stimulus 1 had second- and third-formant transitions beginning at almost the same frequency, between the second and third formant frequencies of the following vowel; for succeeding stimuli in the series, the second-formant starting frequency decreased in equal steps of 85 cps. The third-formant transition began at successively higher frequencies for stimuli 1 through 7, while for stimuli 7 through 13 its starting point gradually decreased. This change in formant transitions from stimuli 1 to 13 is the pattern that would be expected if the location of consonantal constriction were moved in steps from a velar position through an alveolar position to a labial position. All stimuli in the series were characterized by a rapid onset of glottal excitation and by a falling pattern of inflection of the fundamental frequency from 125 to 80 cps. Spectrograms of three of the stimuli (2, 6 and 12) are shown in Fig. 10.

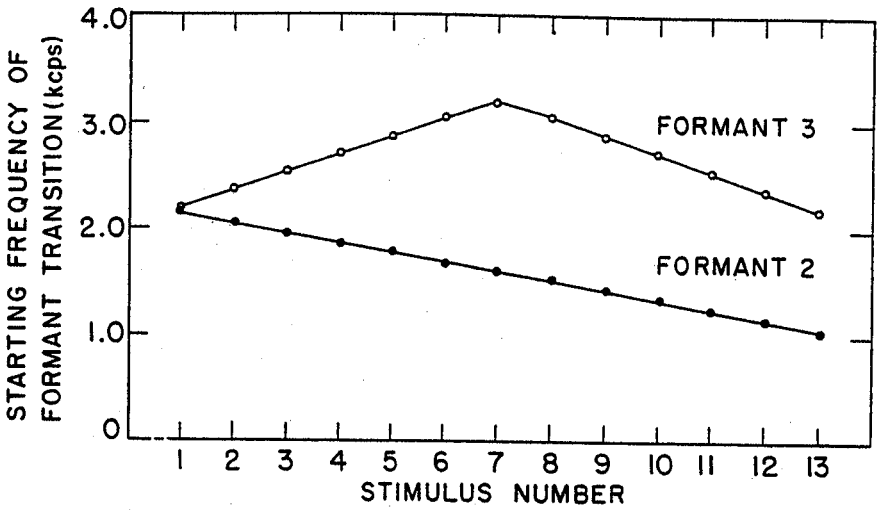

Fig. 9. Starting frequencies of the transitions of the second and third formants for the 13 stimuli in the stop consonant series.

From these 13 stimuli a 169-item identification test was prepared in the manner described for each series of vowel stimuli. Three 132-item ABX discrimination tests analogous to those for the vowel stimuli were also assembled. These two types of test were presented over headphones to a group of eight American English undergraduate listeners. For the identification test, the listeners were instructed to identify each syllabic stimulus as beginning with one of the stop consonants /b/, /d/ or /g/.

Results and Discussion

The solid lines in Fig. 11 give the discrimination functions obtained with the synthetic stop consonants. It will be remembered that both the vowels and the stops were synthesized so as to lie at 12 regularly spaced intervals along a physical continuum encompassing three adjacent phonemes. Since the inter-phoneme ranges are divided into the same number of equally spaced steps, we have a basis for comparing the discrimination functions of these two kinds of speech sounds. To make that comparison we have reproduced the discrimination data for the unrounded vowels (for the American subjects) in Fig. 11 ; these data are indicated by dashed lines.

We have already called attention to the peaks and valleys in the vowel functions and have noted that the valleys occur in the centres of phoneme regions. The considerably more prominent valleys in the stop-consonant functions also occur in the centres of the phoneme regions, as one can see by examining the phoneme identification

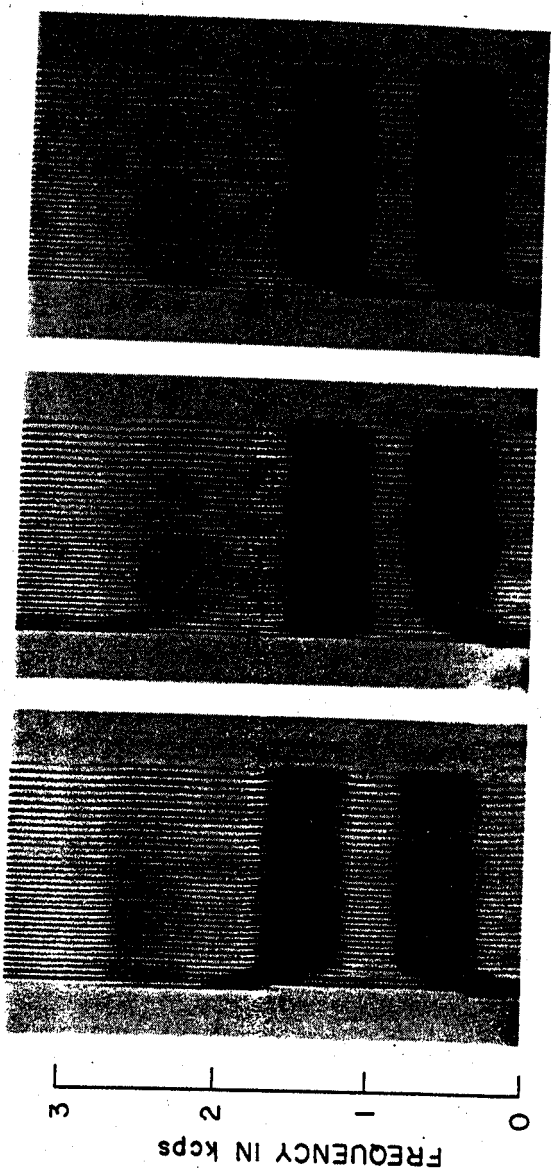

Fig. 10. Spectrograms of stimuli 2 (left), 6 (middle) and 12 (right) of the stop-consonant series.

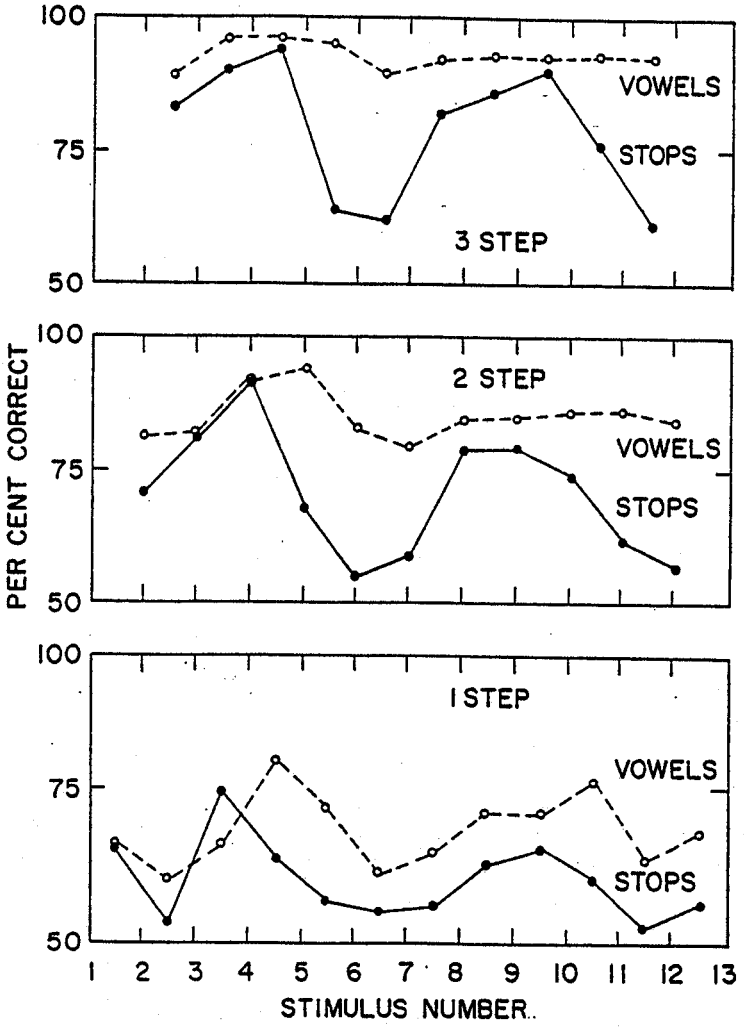

Fig. 11. Discrimination functions for the stop-consonant stimuli are compared with the corresponding functions for stimuli in the unrounded vowel series, for American listeners.

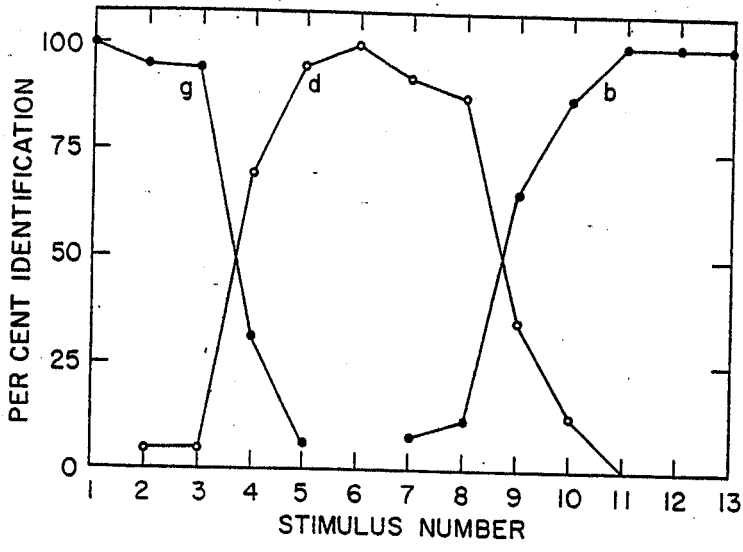

Fig. 12. Identification functions for the stop-consonant stimuli. Average data for eight listeners.

data of Fig. 12. The first discrimination peak, in the vicinity of stimulus 4, is seen from the identification functions of Fig. 12 to correspond to the region where phoneme identifications are changing from /g/ to /d/; the second peak, between stimulus 8 and stimulus 10, is where identifications change from /d/ to /b/.

Discrimination of the vowels is better, in general, than that of the stops. Between phoneme centres a greater number of different vowel tokens than of stop tokens is heard. The differences, particularly for the two- and three-step curves, are greatest in the regions corresponding to the centres of the phoneme classes. At phoneme boundaries, discrimination of stops is almost as good as that of vowels, but between boundaries it falls, as vowel discrimination does not, to almost chance levels.

The implications of these differences between vowel and stop discrimination may be drawn out by comparing the obtained data with those that would be expected on the extreme assumption that listeners can discriminate the stimuli no better than they can identify them as phonemes—that is, that they can only hear phonemically and cannot detect intra-phonemic differences. In yet other words, the assumption is that when subjects are asked to discriminate these stimuli—to say whether the X of an ABX triad is identical with A or with B—they absolutely identify each signal as one of the three phonemes and make their "discrimination" judgments accordingly. In that case the probability that two stimuli will be correctly discriminated is equal to the probability that they are identified as different phonemes, and the "expected" discrimination functions may be computed from the identification functions. The procedure

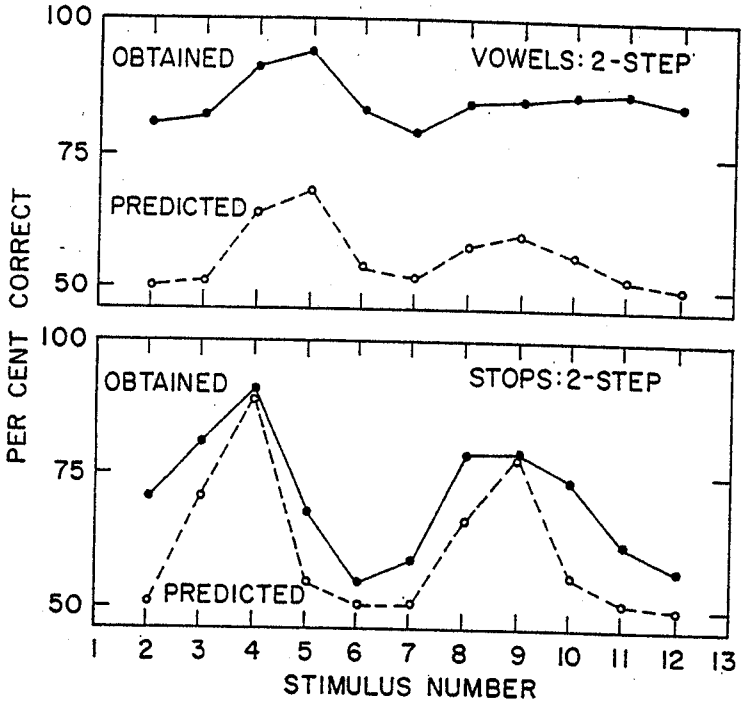

Fig. 13. Comparison of obtained and predicted discrimination functions for the stop-consonant and unrounded vowel stimuli (American listeners). Data for two-step discrimination only are shown.

for making these computations has been described by Liberman, Harris, Hoffman and Griffith (1957). Discrimination functions computed on that basis for the two-step stimulus differences are shown by the dashed lines in Fig. 13. The obtained discrimination functions are redrawn for comparison with the computed ones.

The predicted discrimination functions for the vowels lie well below the obtained ones, but the corresponding curves for the stops match each other more closely. This is to say, it was possible to discriminate many more of the vowel stimuli than could be reliably identified, a result that reflects a kind of perception we have earlier called continuous. For the consonants, on the other hand, perception tended to be categorial in the sense that discrimination was only a little better than absolute identification. Of these two outcomes, the one with the vowels is, of course, the more similar to what one ordinarily finds in the perception of non-speech sounds that lie on a physical continuum (Pollack, 1952 ; Pollack, 1953 ; Miller, 1956 ; Eimas, 1963).

The differences between vowel and stop-consonant perception found here with three-formant patterns are qualitatively similar to findings that have been reported

previously in research with synthetic speech generated with only two formants, although our data differ in some details. The vowel stimuli in the experiments of Fry, Abramson, Eimas and Liberman (1962) encompassed the vowels /ɪ ɛ æ/ rather than /i ɪ ɛ/, but the range was covered by 12 steps, as in our experiment. The stimuli of Fry *et al.*, due to the way in which they were generated, displayed greater variability in intensity and in spacing along the continuum than those of the present study. Perhaps as a result, they were more easily discriminated: the overall level of discrimination was higher and dips in the discrimination functions around phoneme centres, such as we found in our study, did not appear so clearly in theirs. Nevertheless, the two sets of data are quite similar, especially by contrast with the results obtained here and in other experiments on the perception of the stops (Liberman, Harris, Hoffman and Griffith, 1957; Griffith, 1958): the level of discrimination of the vowels tends, in general, to be high, and listeners discriminate many more stimuli than they identify as phonemes.

Further discussion

Having found that discrimination of steady-state vowels seems unaffected by linguistic experience, we are, of course, interested to know whether the same result will be obtained with other classes of speech sounds—particularly in view of evidence that the perception of some speech sounds, e.g., stops, differs from that of steady-state vowels. We have seen in this paper that listeners tend to discriminate the stops little better than they identify them absolutely as phonemes, while the vowels, as in the usual psycho-physical situation, are discriminated much better than they are identified. Other evidence that stops and steady-state vowels may be perceived in different ways, and, indeed, in different parts of the brain, stems from experiments on the perception of competing stimuli presented simultaneously to the two ears. With competing stimuli contrasting in only one phoneme, Shankweiler and Studdert-Kennedy (1967a, 1967b) found that stops presented to the right ear (hence primarily to the left hemisphere) were more often perceived than those presented to the left, but that there was no advantage to either ear for vowels. These differences are paralleled by differences in the nature of the acoustic cues for the stops and the vowels (Liberman, Cooper, Shankweiler, and Studdert-Kennedy, 1967). For the stops there appears to be a relatively complex relation between the phoneme and its representation as sound, and the acoustic cues reside in a signal characterized by rapid change. One may argue, as has been argued elsewhere (Liberman, Cooper, Shankweiler and Studdert-Kennedy, 1967), that a special decoding mechanism would be required to perceive these segments. For the steady-state vowels, on the other hand, there would be no need for a special decoder, since their acoustic attributes are relatively simple. All the foregoing makes us wonder whether the results of our cross-language comparison of steady-state vowel discrimination will be obtained with speech sounds that may be, like the stops, perceived in a different mode.

We may ask, too, about the perception of vowels rapidly articulated in normal phonetic contexts. There is in such cases an "undershooting" of the formant targets (Lindblom, 1963 ; Stevens, House and Paul, 1966) and a merging of cues for successive phonemes. The perception of such vowels may involve a more complex kind of decoding, of the type used in the perception of stops. That they are so perceived is suggested by the results in a recent study by Stevens (1968). He used synthetic vowels similar to the unrounded set of the study reported in this paper, but placed them between an initial /b/ and a final /l/. The result was a tendency away from the continuous perception of the steady-state vowels and toward the categorial perception of the stops. We should like to know whether our rounded vowels—the ones that are familiar to the Swedish listeners but not the Americans—would have been discriminated differently by Swedish and American listeners if our vowels had been, like those Stevens used, in some dynamic context.[3]

It may, of course, turn out that discrimination functions for speech stimuli other than steady vowels are similar for listeners with various linguistic backgrounds. If these "universal" discrimination functions were characterized by well-defined peaks and valleys such as those illustrated in Fig. 11, then experimental justification would be available for the anchoring of linguistic features in the perceptual as well as in the articulatory domain. Further discrimination experiments of this type would lead, then, to a clearer understanding of the nature of the various distinctive features that play a role in language.

References

BASTIAN, J., EIMAS, P. D. and LIBERMAN, A. M. (1961). Identification and discrimination of a phonemic contrast induced by silent interval. *J. acoust. Soc. Amer.*, 33, 842 (Abstract).

BENNETT, D. C. (1965). Vowel duration and spectral form as cues in the recognition of English and German vowels. Unpublished M.A. thesis, University of London.

CHISTOVICH, L., FANT, G., DE SERPA-LEITAO, A. and TJERNLUND, P. (1966). Mimicking of synthetic vowels. Quarterly Progress and Status Report No. 2, Speech Transmission Laboratory, Royal Institute of Technology, Stockholm, 1.

CHISTOVICH, L., FANT, G. and DE SERPA-LEITAO, A. (1966). Mimicking and perception of synthetic vowels, Part II. Quarterly Progress and Status Report No. 3, Speech Transmission Laboratory, Royal Institute of Technology, Stockholm, 1.

[3] *There is some experimental evidence to suggest that the perception of vowels in a consonant-vowel-consonant context depends upon the linguistic experience of the listeners. In a study of the relative importance of duration and spectral form on vowel perception by English and German listeners, Bennett (1965) showed that German listeners assigned more weight to duration in the categorization of unfamiliar vowels, whereas English listeners assigned more weight to spectral form.*

EIMAS, P. D. (1963). The relation between identification and discrimination along speech and non-speech continua. *Language and Speech*, 6, 206.

FANT, G., MARTONY, J., RENGMAN, U. and RISBERG, A. (1962). OVE II synthesis strategy. *Proc. of Speech Communication Seminar*, Royal Institute of Technology, Stockholm, paper F5.

FRY, D. B., ABRAMSON, A. S., EIMAS, P. D. and LIBERMAN, A. M. (1962). The identification and discrimination of synthetic vowels. *Language and Speech*, 5, 171.

GRIFFITH, B. C. (1958). A study of the relation between phoneme labelling and discriminability in the perception of synthetic stop consonants. Unpublished Ph.D. Dissertation, Univ. of Connecticut.

LIBERMAN, A. M., HARRIS, K. S., EIMAS, P., LISKER, L. and BASTIAN, J. (1961). An effect of learning on speech perception: the discrimination of durations of silence with and without phonemic significance. *Language and Speech*, 4, 175.

LIBERMAN, A. M., HARRIS, K. S., HOFFMAN, H. S. and GRIFFITH, B. C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *J. exp. Psychol.*, 54, 358.

LIBERMAN, A. M., HARRIS, K. S., KINNEY, J. A. and LANE, H. (1961). The discrimination of relative onset-time of the components of certain speech and nonspeech patterns. *J. exp. Psychol.*, 61, 379.

LIBERMAN, A. M., COOPER, F. S., SHANKWEILER, D. P. and STUDDERT-KENNEDY, M. (1967). Perception of the speech code. *Psychol. Rev.*, 74, 431.

LINDBLOM, B. (1963). Spectrographic study of vowel reduction. *J. acoust. Soc. Amer.*, 35, 1773.

MILLER, G. A. (1956). The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychol. Rev.*, 63, 81.

POLLACK, I. (1952). The information of elementary auditory displays. *J. acoust. Soc. Amer.*, 24, 745.

POLLACK, I. (1953). The information of elementary auditory displays, II. *J. acoust. Soc. Amer.*, 25, 765.

SHANKWEILER, D. and STUDDERT-KENNEDY, M. (1967a). An analysis of perceptual confusions in identification of dichotically presented CVC syllables. *J. acoust. Soc. Amer.*, 41, 1581 (Abstract).

SHANKWEILER, D. and STUDDERT-KENNEDY, M. (1967b). Identification of consonants and vowels presented to left and right ears. *Quart. J. exper. Psych.*, 19, 59.

STEVENS, K. N. (1968). On the relations between speech movements and speech perception. *Zeitschr. f. Phon.*, 21, 102.

STEVENS, K. N., HOUSE, A. S. and PAUL, A. P. (1966). Acoustical description of syllabic nuclei: An interpretation in terms of a dynamic model of articulation. *J. acoust. Soc. Amer.*, 40, 123.