Reprinted from JOURNAL OF EXPERIMENTAL PSYCHOLOGY,
Vol. 52, No. 2, August, 1956
Printed in U. S. A.

TEMPO OF FREQUENCY CHANGE AS A CUE FOR DISTINGUISHING CLASSES OF SPEECH SOUNDS[1]

ALVIN M. LIBERMAN,[2] PIERRE C. DELATTRE,[3] LOUIS J. GERSTMAN, AND FRANKLIN S. COOPER

Haskins Laboratories, New York City

In recent experiments with synthetic speech (**3, 5, 6, 7**) we have isolated many of the acoustic cues by which a listener identifies various consonants of American English. Some of the cues that we have found to be important for the voiced stops (*b*, *d*, *g*), the voiceless stops (*p*, *t*, *k*), and the nasal consonants (*m*, *n*, *ŋ*) are illustrated in the hand-painted spectrograms of Fig. 1. These spectrographic patterns are very highly simplified and schematized by comparison with spectrograms of real speech, yet capable of producing fair approximations to the intended consonant-vowel syllables when converted into sound by an appropriate playback instrument. The vowel, which is *a* in all cases, is given by the two concentrations of acoustic energy, or "formants" as they are called, the one centering at about 720 cps and the other at 1320 cps. The consonant cues, with which we will be more concerned, begin at the left-hand side of each pattern and extend to the point at which the formants assume the steady state that characterizes the vowel.

As can be seen in the hand-painted spectrograms, the acoustic cues that distinguish the columns of the figure (i.e., *b–p–m* from *d–t–n* from *g–k–ŋ*) are the direction and extent of the relatively rapid shifts in the frequency position of the second (higher) formant. These frequency shifts, or "transitions," are typically found in spectrograms of real speech at the junction of consonant and vowel.[4] By converting patterns like those of Fig. 1 into sound, we have found that the second-formant transitions can, in fact, be cues for the perceived distinctions among the three classes, *b–p–m*, *d–t–n*, and *g–k–ŋ*. There are other acoustic cues for these same classes—we have discovered that appropriate transitions of the third formant, for example, contribute to the identification of these sounds—but our results indicate that the second-formant transitions are very nearly sufficient.

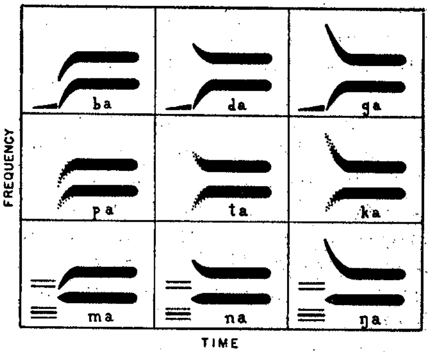

FIG. 1. Hand-painted spectrograms of nine consonant-vowel syllables, showing some of the principal acoustic cues for the perception of the stop and nasal consonants.

---

[1] This research was supported in part by the Carnegie Corporation of New York and in part by the Department of Defense in connection with Contract DA49-170-sc-1642.

[2] Also at the University of Connecticut.

[3] Also at the University of Colorado.

[4] The transitions presumably reflect the changes in the oral cavities that necessarily occur as the articulators move from the position of the consonant to that of the vowel.

We know rather less about the cues that distinguish the patterns that form the rows (i.e., *b–d–g*, *p–t–k*, *m–n–ŋ*). Though we have isolated several cues for the distinction between the voiced (*b–d–g*) and voiceless (*p–t–k*) stops, we have not yet investigated the problem carefully enough to know which of these are most important That omission need not concern us here, however, since for this paper the most relevant consideration is the evidence that each of these cues is essentially constant within a row (of Fig. 1) and more or less capable of distinguishing that row from others. As shown in the figure, one cue that marks the voiced stops (*b–d–g*) as a class is the low-frequency voice bar that immediately precedes the transitions. The voiceless stops probably depend in part on the presence of aspiration (i.e., noise) in place of harmonics in the first part of the transition. For the distinction between the nasal consonants (*m–n–ŋ*) and stops (*b–d–g*, *p–t–k*) the short, steady-state resonance seen at the beginning of each of the nasal consonants in the bottom row is quite adequate.

The patterns of Fig. 1 provide the basis for a table of acoustic elements out of which many of the sounds of speech can be made.[5] The table is interesting, we think, because it indicates that a small number of cues on one acoustic dimension (second-formant transition) combine in all possible pairs with various acoustic markers (the constant resonance of the nasal consonants, for example) to produce some of the highly distinctive sounds of speech. Thus, highly identifiable stimuli are created out of the wholly unpatterned combination of simple and discrete elements. We expect that the table of Fig. 1 will ultimately be expanded to include all the consonant sounds. It is one of the purposes of this study to take a step in that direction.[6]

Exploratory work suggested that we might convert the patterns of Fig. 1 into different speech sounds by varying only the tempo of the transitions, thus pointing to time as another of the stimulus dimensions which may be important in the perception of the individual phones of speech. As the transitions are progressively slowed, the pattern for the stop consonant *b* plus the vowel *a* (as in *bottle*) begins to be heard as the semivowel *w* plus *a* (as in *wobble*), and then as a vowel

---

[5] The 3 × 3 table of Fig. 1, in which the patterns are arranged in terms of the acoustic cues, parallels a commonly accepted phonetic classification according to the articulatory dimensions of *place* and *manner* of articulation. The three columns correspond to three places of production (i.e., points along the oral tract at which the consonant closure is made), *p–b–m* being produced at the lips, *t–d–n* at the alveolar ridge, and *k–g–ŋ* at the velum. Manner of production refers to the particular articulatory features (for example, the presence or absence of voicing or nasality) that are common to the sounds in any given row.

[6] Students of language have long found it useful to describe each consonant as an articulatory event in which one of a small number of places of production is combined with one manner. The specification of place and manner is, then, sufficient to describe a consonant uniquely. To the extent that the relation between articulation and sound is not too complex, we should expect that the categories of place and manner might be equally appropriate for a classification of the acoustic characteristics. That such a classification has utility for a study of speech perception has recently been shown in an important experiment by Miller and Nicely (8). They took account of the confusions that occur when English consonants are heard against progressively increasing amounts of noise and found, for example, that a listener will continue to identify a sound correctly as being a member of a certain manner class even though he does not hear it accurately in regard to its place of production. In general, their results indicated that the place and manner cues, whatever their nature, may be perceived quite independently of each other.

which changes color from *u* to *a* (as in *too odd*). Similarly *ga* (as in *goggle*) goes to *ja* (as in *yacht*), and then to *ia* (as in *theology*). The remaining stop, *d*, goes through comparable transformations, producing (with the vowel *a*) the semivowel *ɥ* (as in French *nuage*) and then the vowel-of-changing-color *ya* (as in French *cru a*).[7]

The experiments to be reported here were designed to extend our exploratory observations concerning transition tempo as a cue for the perceived distinctions among stop consonants, semivowels, and vowels of changing color.

EXPERIMENT I

In this part of the study we have tried to determine whether changing the tempo of first- and second-formant transitions is sufficient to convert the syllables *bε* and *gε* (stop consonant plus vowel) into *wε* and *jε* (semivowel plus vowel) and, in the extreme, into *uε* and *iε* (vowels of changing color). The semivowel (plus vowel) *ɥε* and vowel-of-changing-color *yε*, which presumably result from a slowing of the *d* transition, are not familiar to our American listeners and were for that reason excluded from the experiment.

The vowels of changing color are very different linguistically (and perceptually) from stops and semivowels in that they are not single linguistic units and occur only in situations in which one of the vowels belongs to one syllable and the other to the next. We thought it appropriate, therefore, to deal separately with the two-category distinction between stop and semivowel and the three-category choice among stop, semivowel, and vowel of changing color.

Method

General aspects of procedure and apparatus.—In these experiments hand-painted spectrograms have been used as a basis for creating and controlling speech-like sounds. To convert the spectrograms into sound—an obviously necessary step in this method—we take advantage of a special-purpose instrument called a pattern playback. Descriptions of this instrument, together with discussions of our method, are to be found in earlier papers **(1, 2, 4)**.

The playback employs a variable-density tone wheel to modulate the light from a mercury arc, producing a fundamental of 120 cps and all its harmonics through the fiftieth at 6000 cps. The modulated light beams are imaged on the spectrogram, and are so spread across it as to match its frequency scale. As the hand-painted spectrogram is moved through the light, the white paint reflects beams whose modulation frequencies correspond to the position of the paint on the frequency scale of the spectrogram. The reflected beams are led by plastic light guides to a phototube, the current of which is amplified and converted to sound.

Stimuli.—To vary the tempo of the transitions, we prepared (for conversion into sound) spectrographic patterns like those shown in Fig. 2. The series of patterns illustrated in the top and bottom rows were designed to produce *bε*, *wε*, *uε*, and *gε*, *jε*, *iε*, respectively. In each series, the duration of first- and second-formant transitions was varied from 10 msec. to 300 msec. in steps of 10 msec. When we vary the duration of the transition, we are also, of course, varying the rate of transition (that is to say, the frequency shift per unit time) so long as the frequency extent of the transition remains constant. For convenience, we will present specific values in terms of duration; however, we will continue to speak of this variable generally as tempo.

As can be seen in Fig. 2 the first- and second-formant transitions were varied together, the durations of the two being always equal in any one sound. This approximates what would seem, on the basis of an inspection of spectrograms, to happen in real speech.

The steady-state portion of the pattern has a duration of 300 msec. in all cases, and its formants are so placed as to produce an approximation to the vowel *ε*. By using this vowel, it

---

[7] In articulatory terms *b* and *w* have a common place of production at the lips; *d* and *ɥ* are both produced at the alveols. The semivowel *j* is normally articulated at a point slightly forward of the velar *g*; however, there is in English no semivowel whose place of articulation is closer to *g*.

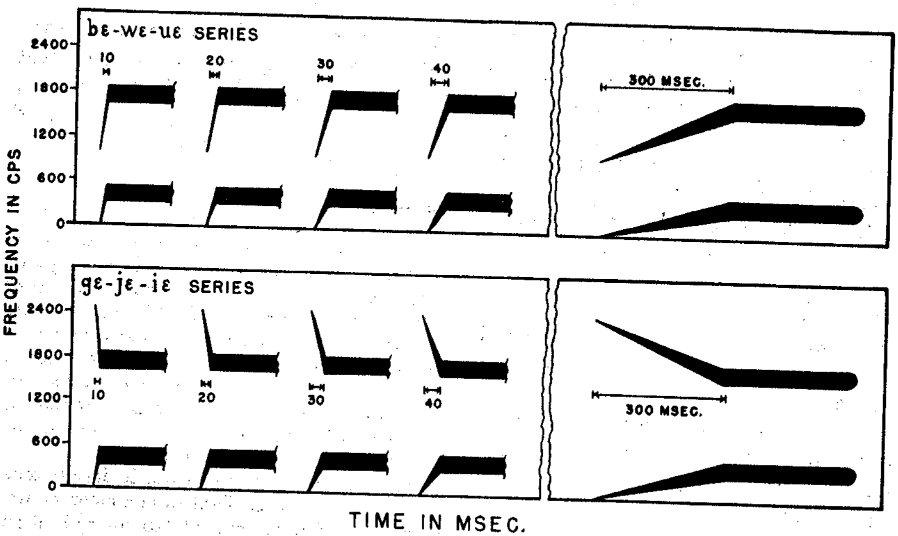

FIG. 2. Illustrations of the spectrographic patterns used to produce the stimuli of Exp. I. The first four patterns in each row show how the tempo of the transitions was varied. At the extreme right of each row is a complete stimulus pattern, i.e., transition plus steady-state vowel, for the longest duration of transition tested.

has been possible to make the frequency shifts for the *bɛ–wɛ–uɛ* series exactly equal in extent to those that produce *gɛ–jɛ–iɛ*. The *b* transition begins 720 cps below, and the *g* transition 720 cps above, the steady state of the second formant of the vowel (1800 cps). The first-formant transition begins at 120 cps and rises to the steady state of the first formant at 480 cps.

The transitions of actual speech are not, of course, so angular as those shown in Fig. 2. We have found, however, that drawing the transitions as we have in that figure does not adversely affect the synthetic speech, and it does, for obvious reasons, enable us to control transition durations more precisely. Each formant consisted of a central strong harmonic and two flanking harmonics of lower intensity.

Presentation of stimuli.—As was pointed out in the introduction to Exp. I, we had thought it desirable to experiment separately with the two-category distinction between stop and semivowel and the three-category choice among stop, semivowel, and vowel of changing color. To investigate the distinction between stop and semivowel we assembled two sets of test stimuli. One, which we will call Test B–W, included 12 patterns from the series illustrated in the top row of Fig. 2 (rising second-formant transitions). The transition durations of these 12 patterns ranged from 10 to 120 msec. in steps of 10 msec., and had seemed on the basis of exploratory work to cover quite adequately the range from *bɛ* through *wɛ*. The other set of stimuli, which we will call Test G–J, included the 12 corresponding patterns from the bottom row of Fig. 2 (falling second-formant transitions). These had appeared in exploratory work to produce sounds that ranged from *gɛ* through *jɛ*.

For the three-category choice among *bɛ*, *wɛ*, and *uɛ*, we used all the patterns of the series illustrated in the top row to make up a set of stimuli that we will refer to as Test B–W–U. The series in the bottom row was used in connection with the distinctions among *gɛ*, *jɛ*, and *iɛ*, and will be referred to as Test G–J–I. For each of these two tests there was a total of 30 stimuli with transition durations that ranged from 10 to 300 msec. in steps of 10 msec.

The spectrographic patterns in each of the four tests were converted into sound and recorded on magnetic tape. By cutting and splicing the magnetic tape we assembled two random orders of stimulus presentation for each test. The signals were arranged on the magnetic tape in such a way that each stimulus would be presented and then repeated after an interval of 1 sec. There was an interval of 6 sec. between successive pairs of stimuli (i.e., a sound and its repetition). This interval provided time for *S* to make and record his judgment. The *S*s heard and identified each stimulus only once in Tests B–W and G–J. Half the *S*s heard the stimuli in one random order, and half in the other. The *S*s who identified the stimuli of Tests B–W–U and G–J–I made two judgments of each stimulus, each of these tests having been given to the *Ss* once in each random order.

The *Ss* were told that each stimulus would be a synthetically produced syllable consisting of an initial speech sound followed in all cases by the vowel *ɛ*. They were asked to identify only the initial sound, and to limit their responses to the choices offered by *E*. For tests B–W, G–J, B–W–U, and G–J–I these choices were *b* or *w*, *g* or *j*, *b*, *w*, or *u*, and *g*, *j*, or *i*, respectively. Examples of these sounds, in syllabic contexts with the vowel *ɛ*, were given. The *Ss* were urged to make an identification of every stimulus, even though their judgments might in some cases be guesses. Before *Ss* made judgments in any test they were asked to listen to the first four stimuli of the series in order that they might become familiar with the nature of the sounds and the general method of presentation.

The four tests were variously interspersed among other tests involving synthetic speech (for example, groups of stimuli set up to study the fricative and nasal consonants), and sometimes two tests of the present experiment were given to the same group of *Ss* at the same session. The only restriction was that Test B–W was never paired with Test B–W–U, and, similarly, that Test G–J was never paired with Test G–J–I in the same session. An analysis of the results showed essentially no effect of the context in which a particular test was given (or of the random order used), so in reporting the results we have combined all the judgments for each test.

Subjects.—A total of 168 paid volunteers, all of them undergraduate students at the University of Connecticut, served as *Ss*. Of this group, 59 took Test B–W and 60 took Test G–J. The stimuli of Tests B–W–U and G–J–I were presented to groups of 41 and 49 listeners, respectively. Of the 41 listeners for Test B–W–U, 19 had previously heard and identified the stimuli of Test G–J; 22 of the 49 listeners for Test G–J–I had previously served in Test B–W. Prior to serving in this experiment, none of the *Ss* had had experience in identifying the synthetic speech sounds produced by the pattern playback.

Results

Figure 3 shows that our listeners were able to use the tempo of first-

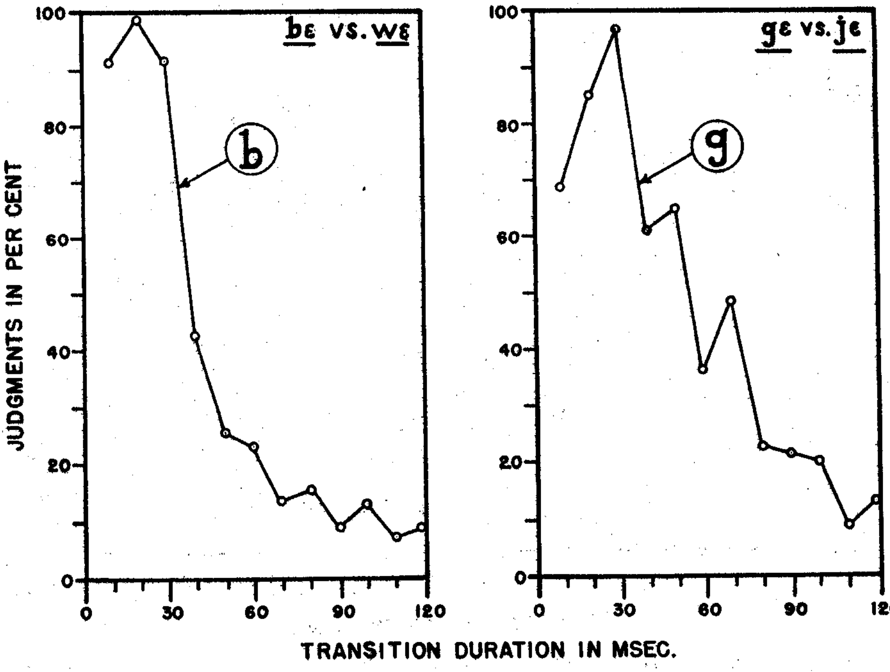

FIG. 3. The distinction between stop consonant and semivowel as a function of the tempo of transition. These curves show the percentage of stop consonant responses, and are based on the judgments of separate groups of 59 and 60 listeners for the *b–w* and *g–j* distinctions, respectively. Tempo is here expressed in terms of duration.

and second-formant transitions as a cue for distinguishing between stop consonant and semivowel. The change from *b* to *w* occurred when the duration of the transitions reached 40 msec., while *g* changed to *j* in the neighborhood of 50–60 msec. It is apparent from both curves that the very shortest duration of transition did not produce the best stop consonant.

Figure 4 shows the judgments we obtained when the range of transition tempos was increased and a third judgment category (*uε* or *iε*) was added. Apparently, our *S*s can reliably discriminate the three categories, stop, semivowel, and vowel of changing color on the basis of tempo of transition alone. It is also apparent, however, that the amount of agreement among our subjects was not so great in the three-category as in the two-category situations (cf. Fig. 3). The expansion of the stimulus range and the inclusion of the third judgment category obviously had the least effect on the responses to the stimuli at the short-duration end of the scale.

EXPERIMENT II

All the results of Exp. I, which indicated that transition tempo can distinguish stop consonant from semivowel from vowel of changing color, were obtained with the vowel *ε* following the initial transition. In Exp. II we have tried to find out how listeners respond to variations in transition tempo when vowels other than *ε* constitute the second part of the syllable. This experiment was

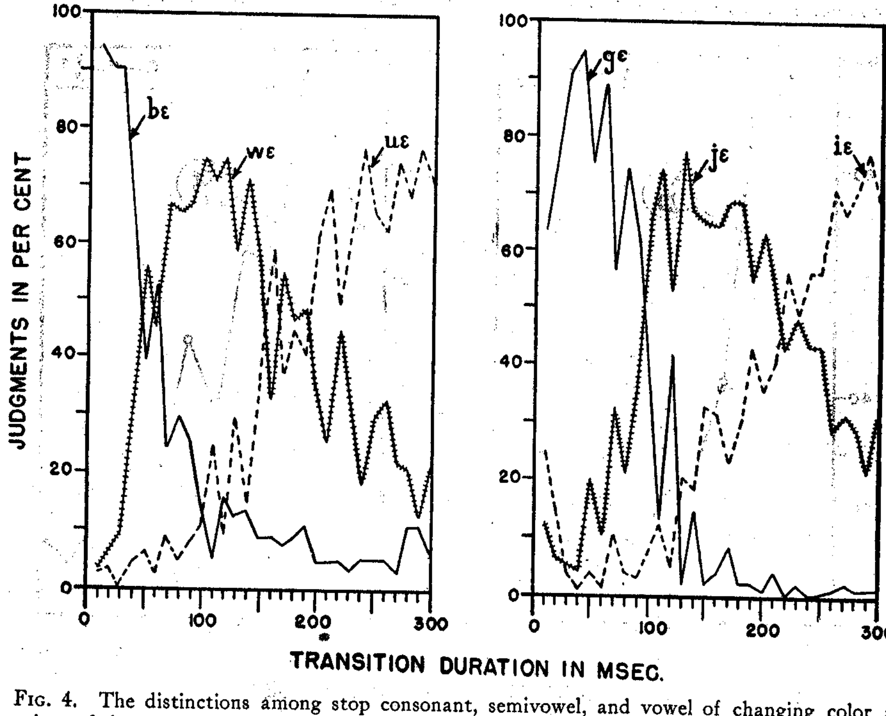

FIG. 4. The distinctions among stop consonant, semivowel, and vowel of changing color as functions of the tempo of transition. The curves representing the *bε–wε–uε* and *gε–jε–iε* responses are based on the judgments of separate groups of 41 and 49 listeners, respectively. Two judgments of all stimuli were obtained from each listener. As in Fig. 3, tempo is expressed in terms of duration.

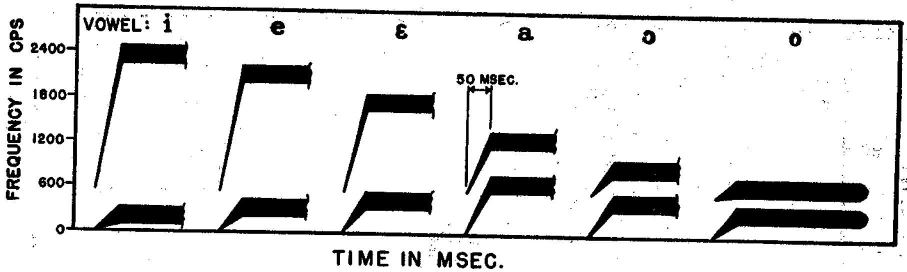

Fig. 5. Illustrations of the spectrographic patterns used to produce the stimuli of Exp. II. One only of the 15 transition durations, viz., 50 msec., is shown with each of the vowels. A complete stimulus pattern is shown at the extreme right.

concerned only with the distinction between the stop consonant *b* and the semivowel *w*. We have omitted the *g–j* distinction because to produce *g* with a variety of vowels requires a radical shift in the frequency level at which the second-formant transition begins, and introduces complications which are largely irrelevant to our present purposes.

In Exp. I rate and duration of transition were varied together. This was unavoidable, of course, because the extent of the transition—that is to say, the frequency range through which the formant moved—remained constant. To produce the various vowels that have been used in Exp. II we have had, necessarily, to put the steady-state formants at several different frequency levels. By starting the transitions from the same point for all these vowels, as can be done with *b*, we have been able to vary the extent of the transitions (from vowel to vowel) and thus to separate the rate and duration aspects of transition tempo. The results of this experiment should, therefore, help to determine which of these variables—rate or duration—is the controlling cue.

Method

Stimuli.—In Fig. 5 are samples of the spectrographic patterns that were used to produce the stimuli of this experiment. These patterns are the same as those of the B–W test of Exp. I except in three respects. The principal difference concerns the steady-state vowel. In Exp. I the first and second formants of the steady-state portion of the syllable were always set at frequency levels that would produce synthetic approximations to the vowel *ɛ*. (See Fig. 2). In Exp. II there are six different steady-state levels of the formants, these levels being appropriate to the vowels *i*, *e*, *ɛ*, *a*, *ɔ*, and *o*. The second difference is in the frequency at which the second-formant transitions begin—1080 cps in Exp. I and 600 cps in Exp. II. It was necessary in Exp. II to lower the starting point of the second-formant transition in order to produce *b* and *w* with vowels whose second-formant frequencies are as low as those of *ɔ* and *o*. For syllables that contain the other vowels, *i*, *e*, *ɛ*, and *a*, it is not necessary to start the second formant at so low a frequency, but, as we pointed out above, it is possible to do so.[8] We thought in this case that it would be desirable to start all the second-formant transi-

---

[8] To produce the very best *b*'s we must start the second-formant transition at levels higher than 600 cps, especially when the *b* is followed by a front vowel such as *i*, *e*, or *ɛ*. When we start this transition from a point as low as 600 cps, we not only produce a somewhat inferior *b*, but we tend in some cases to add at least a suggestion of *bw*. This is to say that starting the second-formant transition at or near the *b–w* locus is itself a cue for the semivowel *w*. For a discussion of "locus" see (4).

The addition of a small amount of *w* is counterbalanced, perhaps, by the fact that starting the first formant at 120 cps, as we do in all the patterns of this experiment, provides a cue for the stop consonant *b* as opposed to the semivowel *w*. To synthesize the best semivowel we must start the first formant at a somewhat higher frequency.

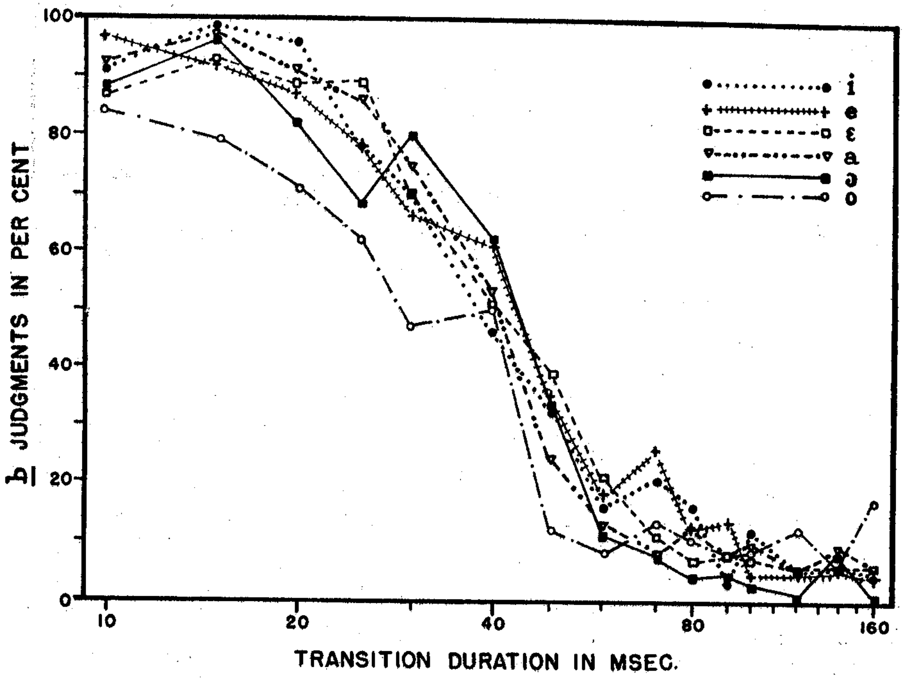

Fig. 6. The distinction between *b* and *w* with various vowels. These curves show the percentage of *b* judgments as a function of transition tempo when tempo is expressed as duration of transition. Each of 38 listeners judged all stimuli twice. Duration is scaled logarithmically to make these curves more directly comparable with those of Fig. 7 and 8.

tions at the same low frequency because in so doing we produced the greatest variation in extent of transition from vowel to vowel and thus obtained the greatest separation of rate and duration. The third difference between the patterns of Exp. I and those of Exp. II concerns the particular transition durations we chose to test. The results of Exp. I (see Fig. 3) had suggested that it would be most appropriate to sample the range 10 to 160 msec., using smaller steps at the short end and larger steps at the long end. Accordingly, we selected for use in Exp. II the following 15 values of transition duration for each vowel: 10, 15, 20, 25, 30, 40, 50, 60, 70, 80, 90, 100, 120, 140, and 160 msec.

Presentation of stimuli.—The 90 stimuli of Exp. II (15 transition durations times six vowels) were converted into sound on the pattern playback and recorded on magnetic tape. By cutting and splicing the magnetic tape we prepared two random orders of these sounds. The presentation of the stimuli was exactly as it had been in the B–W test of Exp. I, except that each *S* heard and judged the stimuli once in each random order. Thus, two judgments of each stimulus were obtained from each *S*.

Subjects.—There were 38 *S*s, all of whom were paid volunteer undergraduates at the University of Connecticut. None of the *S*s had previously served in Exp. I or had any other opportunity to hear the synthetic speech sounds produced by the pattern playback.

Results

We see in Fig. 6 that transition tempo is sufficient to distinguish the stop consonant *b* from the semivowel *w* with each of a wide variety of vowels. A comparison of the curves of Fig. 6 with those obtained for the vowel *ɛ* alone (see Fig. 3 in Exp. I) shows no very great difference.

When our listeners' judgments are plotted against transition duration, as in Fig. 6, the curves for the various vowels are very nearly superimposed. We may reasonably wonder what these curves will look like when the judgments are plotted against rate of transition, since, as we pointed out earlier, rate and duration can be separated in this experiment. Such plots are shown, separately for second and first formants, in Fig. 7 and 8.

It will be remembered that our stimulus patterns were so drawn as to make the duration of first- and second-formant transitions always equal in any one pattern. Transition rates, on the other hand, are different for first and second formants inasmuch as the transition extents differ. In considering the responses as a function of rate, we must, therefore, deal with first and second formants separately.

We should note in regard to Fig. 8 that duration and rate of first-formant transition covary for the vowels *e–o* and, also, for *ε–ɔ*, since, as can be seen in the stimulus patterns of Fig. 5, the extent of first-formant transition is the same within each of these two pairs. It follows, then, that if the data for *e* and *o*, or for *ε* and *ɔ*, should yield similar curves when plotted against duration, they would necessarily produce similar curves against rate of first-formant transition.

Clearly the curves for the various vowels are not so nearly coincident for either of the rate plots as they are when the abscissa is laid out in terms of duration. This result would appear to support an assumption that duration of transition is the essential component of what we have been calling "tempo"—that is to say, that duration rather than rate is the controlling cue.

In assessing the relative importance of duration and rate we should take account of the fact that the rates of first- and second-formant transitions move in opposite directions in the vowel series from

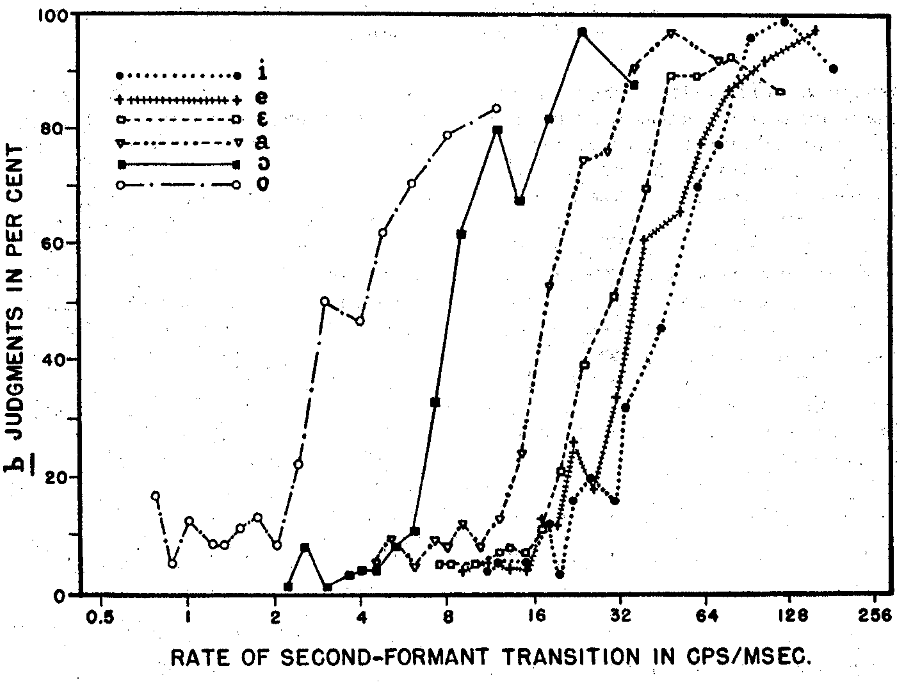

FIG. 7. The same data as shown in Fig 6, here plotted against rate of second-formant transition. Scaling the rates logarithmically serves to equate the distance on the abscissa occupied by the range of rates for each of the vowels.

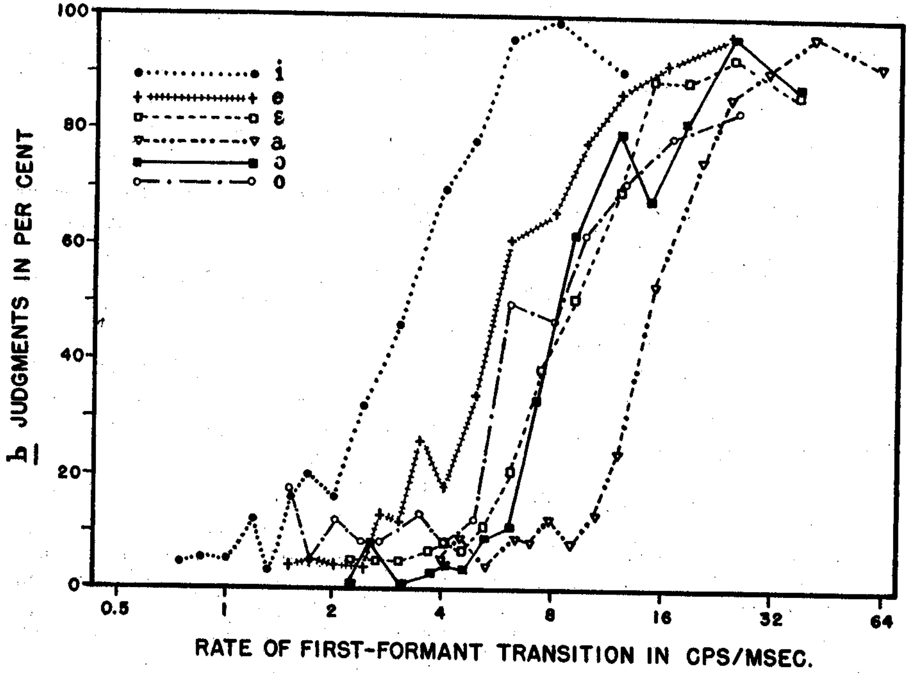

FIG. 8. The same data as shown in Fig. 6 and 7, here plotted against rate of first-formant transition.

i through *a*. (The extent of the second-formant transition decreases from 1800 cps for *i* to 720 cps for *a*, while, for the same change in vowel, the extent of the first-formant transition increases from 120 cps to 600 cps. Hence, for a given duration of transition, the rate of the second-formant transition decreases and the rate of the first-formant transition increases as the vowel is changed from *i* through *e* and *ε* to *a*.) It is possible, if unlikely, that the rate cues produce effects which cancel each other in such a way as to generate curves that appear to be invariant with duration.

For the vowels *a*, *ɔ*, and *o*, the extents (and hence the rates) of first- and second-formant transitions vary in the same direction, so we ought with these vowels to be able to make a less ambiguous comparison of the roles of duration and rate. If we look at the results for these three vowels, we see that the curves appear to be more closely bunched when plotted against duration than when plotted against rate, though in these cases the relatively small change in frequency extent of transition provides something less than an ideal basis for comparison.

Obviously, the clearest separation of rate and duration could be made if we were to hold the transition of one formant constant in all respects while varying the rate, duration, and extent of the other. A series of exploratory studies made it clear that for distinguishing stop and semivowel the tempo of the second-formant transition is considerably more important than that of the first. Nevertheless, we failed to produce a highly realistic series from stop to semivowel with one transition fixed, and we have therefore been unable by this means to obtain direct evidence in regard to rate vs. duration.

We have found in exploratory work that the conversion of the stop consonant *b* to the semivowel *w* is most effectively accomplished if, in addition to slowing the transitions, we also make other changes in the spectrographic pattern. The accurate specification of these other cues and an evaluation of their relative contributions have yet to be made. At the present time we can only say that we have so far found no cue that promises to be more important than the time cue we have isolated in the present experiment. In any event, it is clear from the data reported here that the tempo or, more specifically, the duration of transition is a sufficient cue for distinguishing stop from semivowel from vowel of changing color. Thus, duration is to be added to the list of dimensions that are important for the discrimination and identification of individual speech sounds.

SUMMARY

It had been shown in earlier research that the direction and extent of second-formant transitions enable listeners to distinguish speech sounds *within* each of the three classes, voiceless stops (*p–t–k*), voiced stops (*b–d–g*), and nasal consonants (*m–n–ŋ*). In distinguishing *among* these classes, listeners apparently depend on certain other cues, such as the presence of a nasal resonance or voicing, each of which serves as an acoustic marker for its class. In the first part of the present experiment it was found that, with all other things equal, the tempo of the transitions was sufficient to distinguish members of the class of voiced stop consonants from corresponding members of the classes semivowels and vowels of changing color. In the syllable consisting of the stop consonant *b* plus the vowel *ɛ*, the stop *b* was transformed into the semivowel *w* when the duration of first- and second-formant transitions exceeded 40 msec.; the corresponding change from *gɛ* to *jɛ* occurred at 50 or 60 msec. Further increases in the duration of the transitions caused *wɛ* and *jɛ* to become the vowels-of-changing-color *uɛ* and *iɛ*, respectively. The shift from semivowel to vowel of changing color was much less sharp than the change from stop to semivowel.

In the second part of the study it was found that transition tempo served equally well for distinguishing *b* from *w* with many vowels other than the one (*ɛ*) used in obtaining the results just described. By investigating the tempo cue with a variety of vowels, it was possible (as it had not been in the first part of the study) to assess the relative contributions of duration and rate of transition. The results suggest that duration is the significant aspect of the tempo cue.

REFERENCES

1. COOPER, F. S. Spectrum analysis. *J. acoust. Soc. Amer.*, 1950, **22**, 761–762.
2. COOPER, F. S. Some instrumental aids to research on speech. In *Report of the fourth annual round table meeting on linguistics and language teaching.* Washington, D. C.: Institute of Languages and Linguistics, Georgetown University, 1953. Pp. 46–53.
3. COOPER, F. S., DELATTRE, P. C., LIBERMAN, A. M., BORST, J. M., & GERSTMAN, L. J. Some experiments on the perception of synthetic speech sounds. *J. acoust. Soc. Amer.*, 1952, **24**, 597–606.
4. COOPER, F. S., LIBERMAN, A. M., & BORST, J. M. The interconversion of audible and visible patterns as a basis for research in the perception of speech. *Proc. Nat. Acad. Sci.*, 1951, **37**, 318–325.
5. DELATTRE, P. C., LIBERMAN, A. M., & COOPER, F. S. Acoustic loci and transitional cues for consonants. *J. acoust. Soc. Amer.*, 1955, **27**, 769–773.
6. LIBERMAN, A. M., DELATTRE, P. C., & COOPER, F. S. The role of selected stimulus-variables in the perception of the unvoiced stop consonants. *Amer. J. Psychol.*, 1952, **65**, 497–516.
7. LIBERMAN, A. M., DELATTRE, P. C., COOPER, F. S., & GERSTMAN, L. J. The role of consonant-vowel transitions in the perception of the stop and nasal consonants. *Psychol. Monogr.*, 1954, **68**, No. 8 (Whole No. 379).
8. MILLER, G. A., & NICELY, P. E. An analysis of perceptual confusions among some English consonants. *J. acoust. Soc. Amer.*, 1955, **27**, 338–352.

(Received February 20, 1956)