Reprinted from THE JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA, Vol. 30, No. 2, 122–126, February, 1958

Effect of Third-Formant Transitions on the Perception of the Voiced Stop Consonants*

KATHERINE SAFFORD HARRIS, HOWARD S. HOFFMAN,† ALVIN M. LIBERMAN,‡ PIERRE C. DELATTRE,§ AND FRANKLIN S. COOPER
Haskins Laboratories, New York, New York
(Received November 19, 1957)

Experiments using synthetic speech show that third-formant transitions are cues for the perception of /b,d,g/. Detailed results are presented for a variety of third-formant transitions paired with each of a number of second-formant transitions in initial position before the vowels, /i/ and /æ/.

The results obtained with various third-formant transitions depend in part on the steady-state level of the third formant, implying the existence of third-formant loci analogous to those previously found for the first and second formants. The data of the present experiment are not sufficient to permit a specification of these loci.

The effects of third-formant cues are independent of the two-formant patterns to which they are added. When a third-formant cue enhances the perception of a particular phoneme, it typically does not do so equally at the expense of the other response alternatives.

WE have found in earlier studies with synthetic speech that formant transitions are cues for the identification of various consonants. Transitions of the third formant, with which this paper will be primarily concerned, have been investigated intensively in two studies[1,2] of the cues for /w,j,r,l/, where it was found that a listener must depend to a large extent on the third-formant transition in order to distinguish /r/ from /l/. Apart from the investigation of /w,j,r,l/, however, transitions of the third formant have received relatively little attention. In an earlier paper[3] that dealt with second-formant transitions as cues for the stop consonants, we described parenthetically the effects of adding three rather arbitrarily selected third-formant transitions. Our conclusion at that time was that transitions of the third formant did affect the perception of the consonant, though their effects were less important than those of the second-formant transitions. The purpose of the present paper, then, is to report the results of a more thorough investigation into the role of third-formant transitions in the perception of the voiced stops /b,d,g/.

SPECIAL PROBLEMS IN EXPERIMENTING WITH THIRD-FORMANT TRANSITIONS

When we undertake to find the effects of third-formant transitions, we encounter several special difficulties that affect the procedure of this experiment and limit the nature of the conclusions we can expect to draw. The first of these concerns the question: where—that is, at what frequency level—shall we put the third formant of the following vowel? This was never a problem in studying transitions of the second formant, since the frequency level at which one places this formant is rather precisely determined by the vowel being synthesized. Within rather broad limits the steady-state level of the third formant, on the other hand, has little or no effect on the phonemic identity or color of the vowel, with the result that an experimenter has considerable latitude in deciding where it shall go. Unfortunately for the generality of that experimenter's results, however, the steady-state level of the third formant *does* affect the way in which various third-formant transitions are heard. Thus with certain synthetic patterns that consist of formant transitions followed by steady-state formants, one can move the entire third formant, with its transition, up and down on the frequency scale and observe that the phonemic identity of the consonant changes, though that of the vowel does not. In the present experiment, we have used third-formant levels that had been found by Peterson and Barney[4] to be fairly typical of the particular American English vowels chosen for this experiment. This procedure gives us reasonable assurance that the results will have some generality, though we must of course be careful to describe the effects of the variable transitions with reference to the particular steady-state formants to which they are attached.

A better statement of the results will be possible when, and if, the consonant loci of the third formant are found. Consonant loci, which have so far been specified only for the first and second formants,[5] are the more-or-less fixed frequency positions, each characteristic of a particular consonant, at which the transitions begin or to which they may be assumed to

* 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-2159. The experiments reported were summarized in a paper read before a meeting of the Acoustical Society of America on May 24, 1957.
† Now at Pennsylvania State University, University Park, Pennsylvania.
‡ Also at the University of Connecticut, Storrs, Connecticut.
§ Also at the University of Colorado, Boulder, Colorado.
[1] O'Connor, Gerstman, Liberman, Delattre, and Cooper, Word 13, 24–43 (1957).
[2] L. Lisker, "Minimal cues for separating /w,r,l,y/ in intervocalic position," Word (to be published).
[3] Liberman, Delattre, Cooper, and Gerstman, Psychol. Monogr. 68, No. 8, 1–13 (1954).
[4] G. E. Peterson and H. L. Barney, J. Acoust. Soc. Am. 24, 175–184 (1952).
[5] Delattre, Liberman, and Cooper, J. Acoust. Soc. Am. 27, 769–773 (1955).

point. To the extent that these loci are independent of the steady-state frequency of the formant, they would permit one to describe the transition cues of the third formant in general terms, and without regard to the frequency level of the third formant of the vowel. It is not the primary purpose of this experiment to find the third-formant loci, and in this connection we have previously noted that they are very difficult to pin down,[5] but the results will possibly throw some light on their existence and positions.

A second problem, less serious and more obvious than the first, concerns the difficulty of isolating the third-formant transition so as to obtain a direct measure of the effect of this cue alone. The difficulty arises from the fact that the third formant cannot be presented without the first and second. The first formant presents no problem here, since its transitions do not affect the distinctions among /b,d,g/. However, transitions of the second formant are potent cues for distinguishing among these sounds, and their effects are not readily neutralized. Even when straight, the second formant will contribute appreciably to the perception of /b,d,g/ so long as the first formant has the rising transition that characrerizes the voiced stops as a class. We find it necessary, therefore, to combine each of the third-formant transitions with a variety of second-formant transitions. This not only increases greatly the number of stimuli to be tested, but also makes it difficult to know as directly and as exactly as we might wish just how much of a contribution the third formant is making relative to the second.

METHOD

A. Apparatus

All the stimuli of this experiment were produced with a pattern playback, a device which converts hand-painted spectrograms into sound. Descriptions of this instrument, together with accounts of certain general aspects of our procedure, are to be found in earlier papers.[6-8]

B. Stimuli

In (A) and (C) of Fig. 1 are examples of the hand-painted spectrograms from which the stimuli of this experiment were made. Each spectrogram was intended to synthesize a stop consonant-vowel syllable, the cues for the consonant being contained in the transitions of the various formants, while the differences in vowel color are given by the frequency positions of the formants in the steady-state parts of the pattern. Only two vowels were used: /æ/, as shown in the upper half of Fig. 1, and /i/, as shown at the bottom.

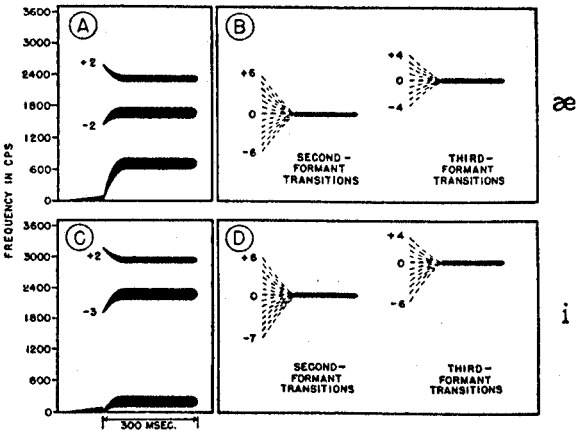

FIG. 1. The stimuli of the experiment. Parts (A) and (C) show scale drawings of sample patterns for the vowels /æ/ and /i/, respectively. Parts (B) and (D) show schematically the ranges of second- and third-formant transitions which were used for each of the vowels.

For convenience in description, the various formants will be designated by F (for formant) and a numeral (1, 2, or 3) to indicate whether it is the first, second, or third formant.

The F1 transition was constant in all the patterns. It always began at the lowest frequency the playback produces (120 cps) and proceeded from there to the steady state of F1. It has been found previously that such a transition serves primarily as a marker for the class of voiced stops; it appears to have essentially no effect on the perceived distinctions within the class. Another constant feature of the patterns is the low-frequency voice bar that immediately precedes the onset of the F1 transition. This too, is a marker for the class of voiced stops.

The F2 transitions, which are represented in (B) and (D) of Fig. 1, were selected on the basis of previous research and constitute, for each of the two vowels, a sampling of the range from /b/ through /d/ to /g/. The direction of the transition will be indicated by "plus" for all transitions which begin above the steady state and by "minus" for those that begin below the steady state. The extent of the transition will be specified by the number of 120-cy steps (i.e., harmonics of the fundamental frequency) which separate its point of origin from the steady state. Thus a transition of "−2" is one that begins two channels (240 cps) below the steady-state frequency of the formant.

As shown in Fig. 1, the transitions leading into /æ/ were varied through 13 steps (of 120 cps each) from −6 to +6. With the vowel, /i/, the range of F2 transitions was the same as with /æ/, except that there was, with /i/, an additional F2 transition of −7.

To provide a baseline for evaluating the effects of F3 transitions, we prepared for each of the two vowels a series of two-formant patterns that contained only the fixed F1 and the various F2 transitions.

[6] F. S. Cooper, J. Acoust. Soc. Am. 22, 761–762 (1950).
[7] Cooper, Liberman, and Borst, Proc. Natl. Acad. Sci. U. S. 37, 318–325 (1951).
[8] F. S. Cooper, "Some instrumental aids to research on speech," in *Report of the Fourth Annual Round Table Meeting on Linguistics and Language Teaching* (Institute of Languages and Linguistics, Georgetown University, Washington, D. C., 1953), pp. 46–53.

As we noted in the introduction, it is not possible to neutralize the effects of F2 transitions on the perceived distinctions among /b,d,g/; hence, to find the effects of F3 transitions, we have combined each of them with each of the F2 transitions described above. The range of F3 transitions was set on the basis of exploratory work. With the vowel /æ/ there were nine transitions of F3, as shown in (B) of Fig. 1. These transitions varied in steps of 120 cps from 480 cps below the steady state of F3 to 480 cps above this level, that is, from −4 to +4. With the vowel, /i/, we had found in exploratory work that the third-formant transition had to rise by as much as 720 cps to the steady state if we were to have a range of stimuli that would include all the possible effects of F3 transitions. Accordingly, as shown in (D) of Fig. 1, we have added transitions of −5 and −6.

There were, then, for the vowel, /æ/, 9 F3 transitions combined with 13 transitions of F2, or 117 three-formant patterns; there were, in addition, 13 two-formant patterns, making a total of 130 stimulus patterns. For /i/, 11 F3 transitions were combined with 14 F2 transitions. When added to the 14 two-formant patterns, this gave a total of 168 patterns.

We wished the synthetic vowels to approximate American English /i/, as in *beet*, and /æ/, as in *bat*. To produce these vowels we started with the values of formant frequency that Peterson and Barney[4] had obtained by measuring a large sample of utterances. These values were then adjusted slightly in order to make the vowels, as produced by the playback, sound right to our ears. The F1 and F2 frequencies which were used in the experiment were 780 cps and 1740 cps for /æ/, and 240 cps and 2400 cps for /i/.

The frequency levels of F3 were set as close to the Peterson and Barney values as the frequency resolution of the playback permitted. These were 2400 cps for /æ/ and 3000 cps for /i/.

We should have liked to include one of the back vowels, /ɔ/, /o/, or /u/, and so obtain a more representative sample of the vowel triangle. It is well known, however, that the third formant is typically very weak with the back vowels—the data of Peterson and Barney show that in the case of /ɔ/, for example, F3 is 34 db down relative to F1. We have done a rather large amount of exploratory work on F3 transitions with the back vowels, but at this point the only conclusion we are willing to venture is that the effects of F3 transitions are considerably smaller and more variable with the back vowels than with the front. We have, therefore, chosen to omit the back vowels from this particular experiment.

C. Presentation of Stimuli

All the spectrographic patterns described in the preceding section were converted into sound with the playback and recorded on magnetic tape. By cutting and splicing the magnetic tape we arranged the stimuli in a random order for presentation to the listeners.

The listeners were 101 undergraduate students at the University of Connecticut. One group of 50 judged the stimuli that included the F2 and F3 variations with the vowel, /æ/. A second group of 51 judged the stimuli with the vowel, /i/. All listeners were asked to identify each stimulus as /b/, /d/, or /g/, and to guess if necessary.

RESULTS

In the topmost row of Fig. 2 are the response distributions that show the effects on perception of various F2 transitions before the vowel, /æ/. These data, which were obtained from patterns without third formants, provide a baseline against which to compare the results for various added transitions of F3. We see that the F2 transitions from −6 to −2 were judged by almost all listeners as /b/. The /d/ curve has its peak at a transition of +1. From there to the higher values of plus transition, the /g/ responses increase and reach a very high maximum at the most extreme "falling" transition, +6. These results are in general very similar to those obtained and reported in an earlier investigation of F2 transitions as cues for the stop consonants.[3]

Of the three stops, only /d/ failed to achieve unanimous agreement among the listeners at some value of F2 transition. Except in the case of /d/, therefore, F3 transitions cannot produce an increase in identifiability as measured by the amount of agreement among the subjects. We can, however, reasonably hope to see the effects of F3 transitions in terms of changes in the shapes, sizes, and positions of the response distributions.

In the remaining response distributions of Fig. 2, F3 transitions are shown as the parameters. It is seen that plus transitions raise the /d/ peak to 100% agreement and, furthermore, increase the range of F2 transitions that are heard predominantly as /d/. A straight F3 also improves /d/ slightly, while minus transitions of F3 clearly harm this consonant, the effect being greater as the extent of the minus transition is larger. Minus transitions of F3 tend to increase the range of both /b/ and /g/ responses, but a close examination of the curves shows that a −2 transition of F3 increases the /b/ area most, and that −4 produces the largest area of /g/ responses. In general, however, the effect of the F3 transitions is greater on /g/ than on /b/.

Figure 3 shows the responses that were obtained with the patterns in which the various F2 and F3 transitions led into the vowel, /i/. As in the earlier figure, the topmost row shows the results that were obtained with patterns that included F1 and F2 only. There we see that F2 transitions of −7 to −4 yielded strong /b/ responses. The greatest number of /d/ responses occurred with an F2 transition of −1, but, at best, the /d/ response was weak. Transitions of +1 or more were heard by almost all of the listeners as /g/. These data, like those for /æ/ described above, are very similar to results obtained in earlier experiments on F2 transitions.[3] The effects of added F3 transitions seem to be limited to /b/ and /d/; there is little effect on /g/

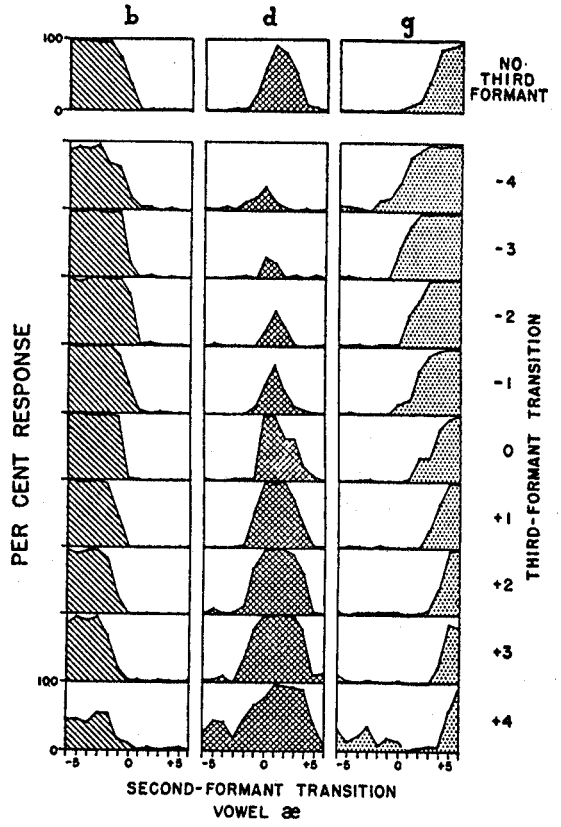

FIG. 2. Responses to patterns with variable second- and third-formant transitions leading into the vowel /æ/. Within each small rectangle, the percentage of responses /b/, /d/, or /g/ is plotted as a function of the value of the second-formant transition. The rows represent the various third-formant conditions. These data are based on the judgments of 50 subjects.

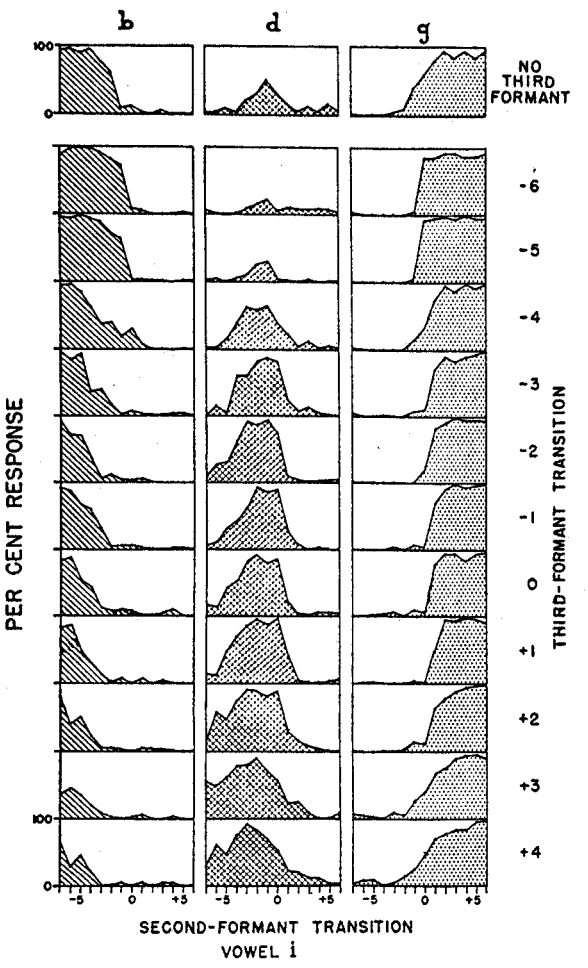

FIG. 3. Responses to patterns with various second- and third-formant transitions leading into the vowel /i/. Each point is based on the judgments of 51 subjects. The data are displayed as in Fig. 2.

responses. As F3 transition is varied from minus to plus, the /d/ judgments increase and the /b/ judgments decrease. Third formant transitions from −3 to +4 increase the number of /d/ responses.

Comparing the results obtained with the two vowels, we see that /d/ is favored in both cases by plus transitions of F3. There is a difference between the two sets of data, however, in that /d/ begins to improve at 0 transition for /æ/ but at −3 in the case of /i/. When we recall that the F3 steady state was 600 cy higher with /i/ than with /æ/, we see that the results obtained with the two vowels can be reconciled on the assumption that there is an F3 locus for /d/ somewhere between the F3 steady-state levels used for /i/ and /æ/—that is, between 3000 and 2400 cps. By this assumption, transitions from a /d/ locus would fall to the F3 of /æ/, but would rise to that of /i/. In any event the difference in steady-state levels and the possible existence of an F3 locus must be taken into account in any consideration of the difference between the results obtained with the two vowels. Indeed, we have preliminary evidence that when we raise the frequency of the /æ/ F3 to the level of the /i/ F3, we get results with /æ/ that begin to approximate those obtained with /i/.

We have performed a number of additional experiments in which we have moved F3 (and its transition) up and down on the frequency scale. These shifts usually affect the perception of the consonant—that is, the consonant produced by a particular F3 transition (other things equal) changes as the steady-state level of F3, and consequently the starting point of the transition, is moved up and down on the frequency scale. This reinforces and extends the observations already noted and provides more general evidence for the assumption that there are F3 loci analogous to those we have already found for F2. Although the results suggest that these loci exist, especially for /d/, the data are not yet sufficient to permit specification of their positions.

The results obtained in this study cannot throw very much light on the mechanism of cue addition, partly because, as we pointed out in the introduction, we cannot measure the effects of F3 independently, and partly because we obtained such a high measure of agreement among listeners with patterns containing

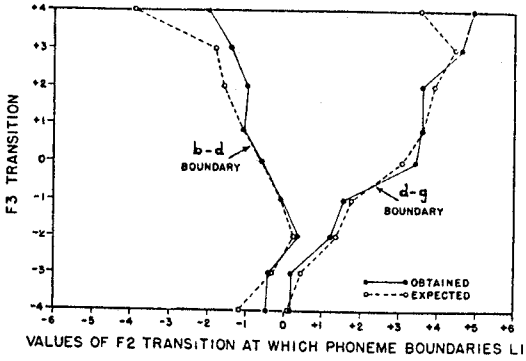

FIG. 4. Location of phoneme boundaries under various conditions of third-formant transition. The open circles show the location of the boundaries that are to be expected on the assumption that a shift in the boundary is the only effect of adding a third-formant transition. The filled circles show the boundary values that were actually obtained.

only F1 and F2 that it becomes somewhat difficult to detect the effects of adding the F3 transitions. It is nevertheless possible to see, especially in the case of /d/, that the effects of F2 and F3 transitions must be combining in some relatively simple way. We note in particular that the F2 transition that produces the best /d/ in a two-formant pattern continues to produce the best /d/ regardless of which F3 transition is added to it. Whether the effect of a particular F3 transition is to raise or to lower the peak of the /d/ response curve, the *position* of the peak on the F2 transition scale tends to remain relatively fixed.

To determine whether the cues are similarly independent for /b/ and /g/, we cannot profitably examine the peaks of the distributions, since, as one can see in Figs. 2 and 3, the /b/ and /g/ distributions are flat and extend to the ends of the stimulus scale. One can, however, suppose that if the cues are independent, then combining them in various ways will have no effect other than to change the positions of the boundaries between response distributions. This can be seen qualitatively by an examination of Figs. 2 and 3. A more precise test, for the data shown in Fig. 2, is described below.

If we assume that the only effect of adding a given F3 is to shift boundaries, then the amount of shift should be predictable from the change in the total number of /b/, /d/, and /g/ judgments when any given F3 transition is added to all the F2's. An example from the data may make this clear.

When an F3 transition of −3 was added to each of the two-formant patterns for the vowel /æ/, there was no change in the number of /b/ judgments, while the number of /g/ judgments was increased by 132, and the number of /d/ judgments was decreased by 132. Assuming that the only effect of adding F3 was at the boundary, we should expect that the /d-g/ boundary would have moved on the F2 continuum so as to decrease /d/ judgments by 132 and increase /g/ judgments by 132. Before F3 was added, the /d-g/ boundary, which we shall define as that point on the F2 transition continuum which is judged as /g/ 50% of the time, was at −2.1. To obtain the amount by which the boundary should have moved along the abscissa as a result of adding −3 F3, we divide 132 (the number of judgments added to /g/) by 50 (the number of subjects in the experiment), and obtain 2.6. Thus, on the boundary shift hypothesis, we should expect to find the /d-g/ boundary displaced 2.6 abscissa units in a direction such as to increase the /g/ area and decrease the /d/ area. This displacement would locate the expected /d-g/ boundary at +0.5 on the abscissa (the scale of F2 transitions); the actual position of this boundary, obtained from the data by linear interpolation, is +0.15 F2.

The positions of the expected and observed phoneme boundaries were calculated for each F3 condition by the procedure described in the example above. The results are shown in Fig. 4. It can be seen that the "expected" and "obtained" functions lie close together in most cases. The type of analysis described above for the vowel /æ/ has also been made for the /i/ data shown in Fig. 3. The "expected" and "obtained" functions, which are not shown in this paper, were again found to be quite similar. Apparently, then, the effect of combining cues is indeed simply to change boundary positions. In this sense the effects of F2 and F3 transitions are independent.

The display of the data shown in Fig. 4 enables us to examine further a point which we made in our initial discussion of the data for the two vowels. We noted then that adding a third formant did not have equal effects on all three phonemes. For the vowel /æ/, the effects of added F3 transitions were somewhat greater on /d/ and /g/ than on /b/. In terms of the present discussion, this means that there are large changes in the F2 position of the /d-g/ boundary as F3 is varied, while changes in the /b-d/ boundary are smaller. It can be seen in Fig. 4 that the /d-g/ boundary moves from about 0 to +5 (on the F2 axis) as F3 is varied, while the /b-d/ boundary moves from about +0.5 to −2.0; the displacement of the /b-d/ boundary is only about half as great as that of the /d-g/ boundary. The boundary effects of a given F3 transition can be considered as a measure of its cue value—that is, an F3 which moves a /d/ boundary in a direction such that the number of /d/ judgments is increased, is a /d/ cue. The fact that F3 transitions typically have unequal effects at the two boundaries means that a cue which enhances one response does not necessarily do so equally at the expense of the other response alternatives. Such a cue may be said not only to indicate to a listener what a speech sound *is*, but also to tell him what it *is not*. The latter information may be important in reducing equivocation when two cues are combined. If one conceives the various response alternatives as existing in some kind of multidimensional stimulus space, he may then assume that each cue element has components in more than one direction.