Reprinted from The Journal of the Acoustical Society of America, Vol. 27, No. 4, 769–773, July, 1955

Acoustic Loci and Transitional Cues for Consonants*

Pierre C. Delattre,† Alvin M. Liberman,‡ and Franklin S. Cooper
Haskins Laboratories, New York, New York

(Received April 8, 1955)

Previous studies with synthetic speech have shown that second-formant transitions are cues for the perception of the stop and nasal consonants. The results of those experiments can be simplified if it is assumed that each consonant has a characteristic and fixed frequency position, or locus, for the second formant, corresponding to the relatively fixed place of production of the consonant. On that basis, the transitions may be regarded as "movements" from the locus to the steady state of the vowel.

The experiments reported in this paper provide additional evidence concerning the existence and positions of these second-formant loci for the voiced stops, *b*, *d*, and *g*. There appears to be a locus for *d* at 1800 cps and for *b* at 720 cps. A locus for *g* can be demonstrated only when the adjoining vowel has its second formant above about 1200 cps; below that level no *g* locus was found.

The results of these experiments indicate that, for the voiced stops, the transition cannot begin at the locus and go from there to the steady-state level of the vowel. Rather, if we are to hear the appropriate consonant, the first part of the transition must be silent. The voiced stops are best synthesized by making the duration of the silent interval equal to the duration of the transition itself.

An experiment on the first formant revealed that its locus is the same for *b*, *d*, and *g*.

In an earlier experiment[1,2] we undertook to find out whether the transitions (frequency shifts) of the second formant—often seen in spectrograms in the region where consonant and vowel join—can be cues for the identification of the voiced stop consonants. For that purpose we prepared a series of simplified, hand-painted spectrograms of transition-plus-vowel, then converted these patterns into sound and played the recordings to naive listeners for judgment as *b*, *d*, or *g*. The agreement among the listeners was, in general, sufficient to show that transitions of the second formant can serve as cues for the identification of the stops and, also, to enable us to select, for each vowel, the particular transitions that best produced each of the stop consonant phones. These transitions are shown in Fig. 1.

We found in further experiments[2] that these same second-formant transitions can serve as cues for the unvoiced stops (*p–t–k*) and the nasal consonants (*m–n–ŋ*), provided, of course, that the synthetic patterns are otherwise changed to contain appropriate acoustic cues for the voiceless and nasal manners of production. Moreover, and more important for the purposes of this paper, the results of these experiments plainly indicated a relationship between second-formant transition and articulatory place of production. Thus, the same second-formant transitions that had been found to produce *b* proved to be appropriate also for the synthesis of *p* and *m*, which, like *b*, are articulated at the lips; the second-formant transitions that produced *d* produced the consonants *t* and *n*, which have in common with *d* an articulatory place of production at the alveols; and, similarly, the second-formant transitions were found to be essentially the same for *g*, *k*, and ŋ, which are all produced at the velum.

It is an obvious assumption that the transitions seen in spectrograms reflect the changes in cavity size and shape caused by the movements of the articulators, and if we further assume that the relation between articulation and sound is not too complex, we should suppose, on the basis of the evidence of the preceding paragraph, that the second-formant transitions rather directly represent the articulatory movements *from* the place of production of the consonant *to* the position for the following vowel. Since the articulatory place of production of each consonant is, for the most part, fixed, we might expect to find that there is correspondingly a fixed frequency position—or "locus"—for its second formant; we could then rather simply describe the various second-formant transitions as movements from this acoustic locus to the steady-state level of the vowel, wherever that might be.[3] As may be seen in Fig. 1, the various transitions that produce the best *d* with each of the seven vowels do, in fact, appear to be coming from the same general region, and on the assumption that the first part of the acoustic transition is somehow missing, one may suppose that the transitions originate from precisely the same point. Clearly, *d* is the best case. For *b* the transitions all appear to be coming from some point low on the frequency scale, but an exact position for the *b* locus is not evident. In the case of *g*,

* This work was supported, in part, by the Carnegie Corporation of New York and, in part, by Department of Defense Contract. Some of the results of this research were reported in a paper read before the Acoustical Society of America on October 15, 1953.

† Also at the University of Colorado, Boulder, Colorado.
‡ Also at the University of Connecticut, Storrs, Connecticut.

1 Cooper, Delattre, Liberman, Borst, and Gerstman, J. Acoust. Soc. Am. 24, 597–606 (1952).

2 Liberman, Delattre, Cooper, and Gerstman, Psychol. Monogr. 68, No. 8, 1–13 (1954).

3 We do not wish to restrict the concept of locus to the second formant, nor do we mean to relate it exclusively, on the articulatory side, to place of production. By "locus" we mean simply a place on the frequency scale at which a transition begins or to which it may be assumed to "point." We have found this to be a useful concept, since, for first and second formants, there appear to be many fewer loci than there are transitions.

The locus is in certain respects similar to the concept of the "hub" as developed by Potter, Kopp, and Green. See Potter, Kopp, and Green, *Visible Speech* (D. Van Nostrand Company, Inc., New York, 1947), pp. 39–51.

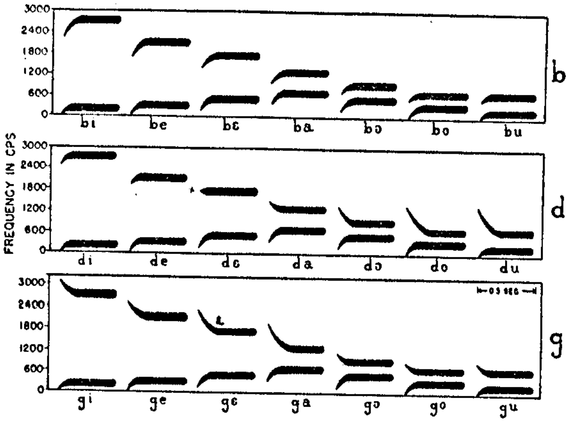

FIG. 1. Synthetic spectrograms showing second-formant transitions that produce the voiced stops before various vowels.

there would appear to be a single high-frequency locus for the front vowels *i*, *e*, ɛ, and the mid-vowel *a*; but for the back vowels ɔ, *o*, and *u* the acoustic pattern breaks sharply, and it is obvious that the same *g* locus cannot serve for all vowels. In this connection it is known that the articulatory place of production of *g* is displaced somewhat according to the vowel that follows it, but there is no evidence that there is in this displacement the kind of discontinuity that occurs at the acoustic level in the sudden and large shift of the *g* transition. It would appear, then, that in this particular instance the relationship between place of production and sound has become rather complex, and a simple correspondence between this articulatory variable and a second-formant acoustic locus is not found.

In the series of experiments to be reported here we have, first, undertaken to collect additional evidence concerning the existence and position of the second-formant loci for *b*, *d*, and *g*, and, in particular, to determine whether these loci are independent of vowel color as, indeed, they must be if the concept is to have any utility; second, we have tried to determine whether, in the case of the stops, the locus can be the actual starting point for the transition, or whether, alternatively, the locus is a place to which the transition may only point; and, third, we have collected evidence concerning a first-formant locus.

APPARATUS AND GENERAL PROCEDURE

All the acoustic stimuli used in this study were produced by converting hand-painted spectrograms into sound. The special-purpose playback that accomplishes this conversion has been described in earlier papers.[4,5] It produces 50 beams of light, separately modulated at each of the first 50 harmonics of a 120-cycle fundamental, and spreads them across the hand-painted spectrogram in such manner that the frequency of the modulated light at any point corresponds approximately to the frequency level of the place at which it strikes the spectrogram. The painted portions of the spectrogram reflect the appropriately modulated beams of light to a phototube whose current is amplified and fed to a loudspeaker.

As shown in Fig. 2, the hand-painted patterns consisted of two formants, each of which included three contiguous harmonics of the 120-cycle fundamental. The intensity of the central harmonic of the formant was 6 db more than the two outlying ones; the frequency of that harmonic is used in specifying the frequency position of the formant. All transitions of either formant were always painted and heard in initial position in the syllable. A transition is called "rising" or "falling" according to whether it originates at a frequency lower than (rising) or higher than (falling) the steady state of the corresponding formant of the vowel.

SECOND-FORMANT LOCI

The purposes of this part of the investigation were to find the positions of the second-formant loci of the

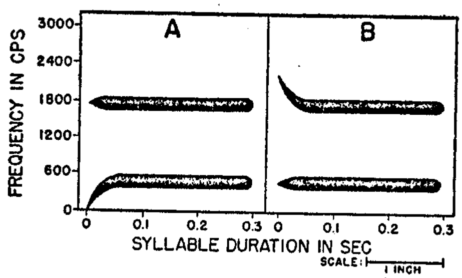

FIG. 2. Scale drawings of sample two-formant patterns used in this study. (A) A rising transition in the first formant and a straight transition in the second. (B) A straight transition in the first formant and a falling transition in the second.

stop consonants, and to test whether their existence and position are independent of vowel color and, also, of the extent of first-formant transition. Accordingly, we prepared the series of patterns shown schematically in Fig. 3, and converted them to sound for evaluation by ear.

As shown in the figure, each stimulus pattern had a straight transition of the second formant and some degree of rising transition of the first formant. This arrangement was dictated by two considerations: first, the possibility that the initial part of the transition is not sounded, in which case we should suppose that only a straight second formant can "point" precisely to the frequency position of the locus[6]; and, second, the fact

[4] F. S. Cooper, J. Acoust. Soc. Am. 22, 761–762 (1950).
[5] Cooper, Liberman, and Borst, Proc. Natl. Acad. Sci. 37, 318–325 (1951).

[6] Potter, Kopp, and Green (pp. 81–103 of reference given in footnote 3) located the "hub" of each of the stop consonants by a technique which obviously takes account of the same consideration. They made spectrograms of each of the stops paired with a variety of vowels and then looked for those patterns in which the second formant was straight. The syllables *dæ* (as in "dad") and *bʊ* (as in "book") yielded straight second formants, and they concluded that the hub of *d* is in the same position as the hub of æ and that, in similar fashion, *b* goes with ʊ. The hub of *g* was found to be variable.

that with zero transition of the second formant, a consonant will be heard, if at all, only when the first formant is curved.

When the first and second formants of Fig. 3 are paired in all combinations, 65 vowels are produced, comprising a wide variety of colors and including many that do not correspond to known speech sounds. The extent of first-formant transitions was varied as shown in the figure. Each of the two-formant patterns was converted into sound and listened to carefully by the authors of this paper, who identified and evaluated each sound as *b*, *d*, or *g*. When the judgments thus obtained are appropriately tabulated, the following conclusions emerge:

(1) Rather clear stop consonants are heard at particular positions of the second formant. The best *g* is produced by a second formant at 3000 cps, the best *d* at 1800 cps, and the best *b* at 720 cps.[7,8] We shall suppose that these three frequencies represent the acoustic loci of *g*, *d*, and *b*, respectively. At other frequency levels of the straight second formant a stop-like sound is heard, the identity of which is more or less clear depending on its nearness to one of the three frequencies given above. At about 1320 cps the sound is indifferently *b*, *d*, or *g*.

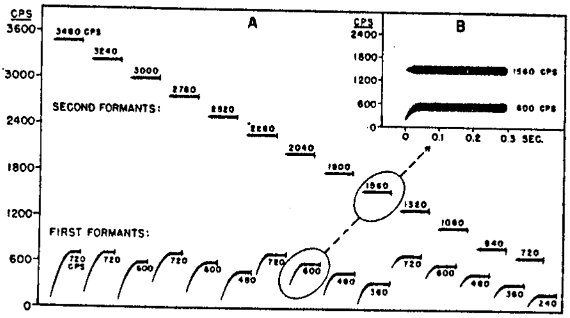

FIG. 3. Schematic display of the stimuli used in finding the second-formant loci of *b*, *d*, and *g*. (A) Frequency positions of the straight second formants and the various first formants with which each was paired. When first and second formants were less than 240 cps apart they were permitted to overlap. (B) A typical test pattern, made up of the first and second formants circled in (A).

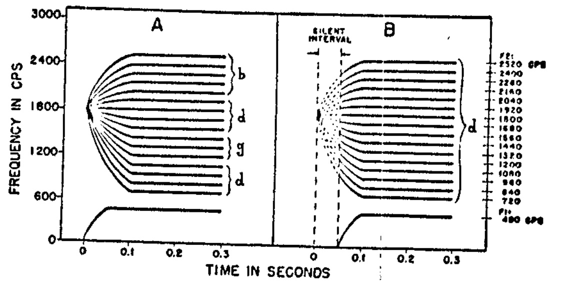

FIG. 4. Stimulus patterns (shown schematically) and identifications with and without a silent interval between the second-formant locus and the onset of the transition. (A) Second-formant transitions that originate at the *d* locus and go to various steady-state levels, together with the first formant with which each was paired. (B) The same patterns, except that a silent interval of 50 msec has been introduced between the locus and the start of the transition. Note that there is no silent interval in the first formant, but that it has been displaced along the time scale so that its onset is, as in (A), simultaneous with that of the second formant. Similar adjustments in time of onset were made for all the silent intervals tested in this experiment. The introduction of a silent interval into the first formant always weakened the consonant, but did not affect its identity.

(2) The steady-state level of the first formant has essentially no effect on either the strength or identity of the consonant impression, with the exception that when the straight second formant is about midway between the *g* locus (at 3000 cps) and the *d* locus (at 1800 cps), raising or lowering the level of the first formant tends to push the sound toward *d* or *g*. Otherwise, it appears that the second-formant loci are independent of the changes in vowel color that are produced by varying the position of the first formant.

(3) The extent of the first-formant transition has little or no effect on the identity of the stop consonant. Such variations do, however, affect the strength of the consonant impression. As was pointed out earlier, the first formant must have some degree of rising transition if a consonant is to be heard at all when the second formant is straight. Our observations in this experiment point additionally to the conclusion that in the case of the voiced stops the consonant impression is stronger as the first-formant transition is larger. The strongest stop is obtained when the first formant starts at the lowest frequency (120 cps) and rises from that point to the steady-state level appropriate for the following vowel.

THE LOCUS AND THE START OF THE TRANSITION

This part of the investigation was designed to determine whether the transitions can start from the locus and move to the steady-state level of the vowel, or whether the transitions must only point to this locus, as they appeared to do in Fig. 1. For that purpose we used the locus values that had been found in the first part of this investigation, and, making the assumption that the transition can actually originate at the locus, we prepared a series of patterns like those shown schematically in A of Fig. 4. There we have, with a

[7] One would infer from the graphs of Fig. 1 that the locus of *b* must be somewhere below the second formant of *u* (720 cps), since the best *bu*, as indicated by the amount of agreement among our naive listeners in the earlier transition study, was formed when the second formant had a rising transition. We believe that the discrepancy between that result and the present one is to be attributed to differences of detail in the patterns used in the two studies.

[8] As is indicated in Fig. 3, the straight second formants were spaced at intervals of 240 cps. After it had begun to appear that the stop consonant loci were in the vicinity of 3000, 1800, and 720 cps, we experimented, on an exploratory basis, with straight second formants 120 cps on either side of each of these three values, and found that none of the stops was significantly improved by these adjustments.

Of the three stops that are produced when the straight second formants are at the loci, the *d* (at 1800 cps) is the most compelling, the *b* (at 720 cps) is slightly less so, and *g* (at 3000 cps) is, perhaps, the least satisfying.

fixed lower formant, a choice of second formants which all originate at the *d* locus, i.e., at 1800 cps, and move from that point to their respective steady-state positions. When these patterns are sounded—the fixed first formant with each of the second formants in turn—we do not hear *d* in every case. Rather, we hear *b* when the steady state of the second formant is in the range 2520 cps through 2040 cps, then *d* from 1920 cps through 1560 cps. With second-formant levels from 1440 cps through 1200 cps, *g* is heard, and then *d* again when the second-formant level goes below about 1200 cps.

We find, however, that if we erase the first 50 msec of the transition, creating a silent interval between the locus and the start of the transition, as shown in *B* of the figure, then reasonably good *d*'s are heard in all cases. A silent interval less than 50 msec does not produce *d* at all steady-state levels of the second formant, and intervals greater than 50 msec also fail, at least at some second-formant levels, to give *d*. At we pointed out earlier in this paper, the second-formant locus of a consonant presumably reflects the articulatory place of production, and the transition can be assumed to show the movement from that place to the articulatory position appropriate for the following vowel. The fact that the transition serves best if it does not begin at the locus might be taken as an indication that no appreciable sound is produced until at least part of the articulatory movement has been completed.

In all the patterns of Fig. 4, the time interval between the locus and the steady state of the second formant was 100 msec. We should suppose that this corresponds to some particular rate of articulation. To find out what might happen at other articulatory rates we have repeated the procedures described above with sets of patterns in which the total interval between locus and steady state of the vowel was 40, 60, 80, and 120 msec; that is, for these additional total intervals we prepared and listened to patterns in which the transition went all the way to the locus, and also to patterns in which various amounts of the initial part of the transition had been erased. For total intervals of 80 and 120 msec the results are almost identical with those that were obtained with a total interval of 100 msec, except that the "best" silent intervals (that is, the ones at which *d* is most clearly heard at all steady-state levels of the second formant) are about 40 and 60 msec, respectively. From these values (together with that of 50 msec which was best for a total interval of 100 msec) it would appear that the best silent interval is approximately half the total interval. When the total interval is 60 msec or less, we do not get good *d*'s at any silent intervals.

We have repeated these procedures with the *b* locus (at 720 cps) and with the *g* locus (at 3000 cps). It is reasonably clear in these cases, as it was with *d*, that the transition cannot start at the locus and go all the way to the steady state. The length of the silent interval that gives the best results seems to depend on the total duration of the interval (from locus to steady state), the best silent interval, as with *d*, is approximately half the total interval. With *b* and *g* the best results are obtained (with the appropriate silent intervals) when the total interval is 80 or 100 msec. Both these sounds are relatively poor (at all silent intervals) with a total duration of 120 msec; at total durations of 60 and 40 msec, *g* suffers rather more loss in clarity than *b*.

The results obtained with *b* and *g* are in certain other respects different from those that were found with *d*, and they are also, perhaps, somewhat less definite. When the transition started at the *d* locus, consonants having other than the *d* (alveolar) place of production were clearly heard at some levels of the second-formant steady state. (Thus, as shown in A of Fig. 4, *b* was heard at relatively high levels of the second formant, and *g* was heard when the steady-state level of the second formant was in the range 1440 cps to 1200 cps.) When the transition starts at the *b* locus, however, we hear *bw* (which has the same place of production as *b*) for steady-state levels from 2520 cps through 1440 cps, and then from 1220 cps through 960 cps we hear something that sounds very vaguely like *gw*. The alveolar consonant (*d*) is not heard at all in the *b* series, and the *g* in the *gw* cluster is very weak. Thus, the impossibility of starting the transition at the locus is less strikingly demonstrated for *b* than it was for *d*, and the improvement in the *b* which results from the introduction of a silent interval is, accordingly, less dramatic than the effects that are produced when a silent interval precedes the *d* transitions.

The results obtained with the *g* locus were different from those with *b* and *d* in that there is, apparently, no silent interval that will produce *g* at all steady-state levels of the second formant. In the best case—that is, with the best silent interval—one hears *g* from a steady state of 2520 to one of approximately 1200 cps. Below the latter value we hear *d*. This result is not surprising, since it was quite clear from our earlier data on transitions as cues that the same *g* locus could hardly apply to all vowels.

FIRST-FORMANT LOCUS

In the first part of this investigation, which was concerned primarily with finding the second-formant loci, it appeared that the best stop consonants were produced when the curved first formant started at the lowest possible frequency. The purpose of the third part of the study was to explore further the problem of the first formant, using straight first formants and curved second formants. The straight first formant will presumably serve here, as the straight second formant did in the first part of the investigation, to avoid the problems introduced by the possibility that the transition does not reach all the way to the locus; the transition in the second formant will be necessary, as was the first-formant transition of the earlier experiment, to produce a consonant effect.

The experimental patterns are shown schematically in Fig. 5. Four first formants were used having frequencies appropriate for the first formants of the cardinal vowels *i*, *e*, ɛ, and *a*. These were paired with various second formants as shown in the figure and explained in the legend. The second formants always had transitions that rose or fell through four harmonics (of the 120-cycle fundamental). These patterns were converted into sound by the playback and judged by the authors of this paper.

As we should have expected from the results we had previously obtained in our work on second-formant transitions, the listeners heard *g* or *d* when the transition of the second formant was falling, and *b* when it was rising. (With falling transitions of the second formant, *g* was heard for steady-state levels from 3000 to 2280 cps; between 2280 and 1320 cps the sound could be identified either as *g* or *d*; and below about 1320 cps it was clearly *d*.) It will be remembered, however, that our primary interest was not in the second formant, but rather in the first, and, more particularly, in the effects of its frequency level. In this connection we found that the stop consonant—whether *b*, *d*, or *g*—was best when the first formant was at the lowest position (240 cps). When the first formant was raised from 240 cps, there was, apart from the change in vowel color, a weakening of the stop consonant; however, the identification of the stop as *b*, *d*, or *g* was not affected by the frequency level of the first formant. It would appear, then, that the locus of the first formant is at 240 cps for all the voiced stops, but inasmuch as we did not, and, indeed, with our playback could not, center the first formant much lower than 240 cps, we should rather conclude that the first-formant locus is somewhere between that value and zero.

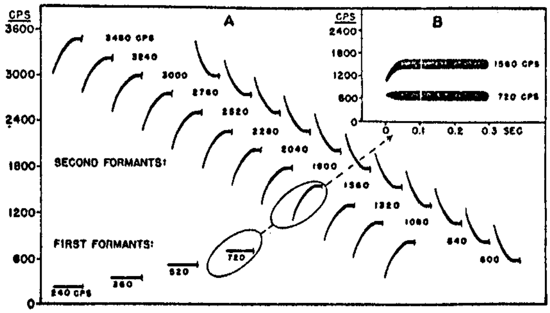

FIG. 5. Schematic display of the stimuli used in finding the first-formant locus of *b*, *d*, and *g*. (A) Frequency positions of the straight first formants and the various second formants with which they were paired. All combinations of first and second formants were used, except for eight cases in which the two formants were so close together as to overlap. The formant shown at 520 cps is composed, in slightly unequal parts, of the fourth harmonic at 480 cps and the fifth harmonic at 600 cps; 520 cps is an estimate of its equivalent frequency.

DISCUSSION

The experiments reported here were concerned only with the voiced stops, *b*, *d*, and *g*. We know, however, that the same second-formant transition that produces *b*, for example, will also produce other consonants, such as *p* and *m*, which have the same articulatory place of production. From some experiments now in progress it appears, further, that with an appropriate lengthening of its duration, this same transition will produce the semivowel *w*. We should suppose, then, that *b*, *p*, *m*, and *w* might have the same second-formant locus, which would correspond, as it were, to their common place of production, and that we might generalize the results of this study by assuming that the second-formant loci we found here are appropriate not only for *b*, *d*, and *g*, but, more broadly, for the three places of production (bilabial, alveolar, and velar) that these stop consonants represent.

Although we expect that consonants with the same place of production will be found to have the same second-formant locus, we do not think that they will necessarily have the same best silent interval. In the case of the stops it is clear that approximately the first half of the total interval from locus to steady state must be silent. With a semivowel like *w*, on the other hand, it appears on the basis of exploratory work that the second-formant transition can be made to go all the way from the locus to the steady state of the vowel without adversely affecting the identifiability of the sound—indeed, it may well be that the best semivowel is made in this way.

The results of these experiments indicate that the locus of the first formant is not different for *b*, *d*, and *g*. We might guess, then, that the first-formant locus has little or nothing to do with *place* of production. Evidence from experiments now in progress suggests rather that it is closely related to the articulatory dimensions of *manner*.

We know from earlier experiments that third-formant transitions are cues for the identification of the stop consonants according to place of production, and we might expect, therefore, that there would be a third-formant locus for each of the stops. We have been trying to find these loci by procedures analogous to those used in the present study, that is, by varying the frequency position of a straight third formant. These procedures have yielded some evidence that the third-formant loci do exist. The results are less clear than for the second-formant loci, however, and it appears that additional and more sensitive techniques will be required.