Reprinted from WORD
Vol. 13, No. 1, April, 1957
Printed in France

ACOUSTIC CUES FOR THE PERCEPTION OF INITIAL /w, j, r, l/ IN ENGLISH*

J. D. O'CONNOR,** L. J. GERSTMAN, A. M. LIBERMAN,***
P. C. DELATTRE,**** and F. S. COOPER

In investigating the acoustic cues for the perception of speech it is usually convenient and often necessary to study only a relatively small group of phonemes in any single experiment. Thus, our own earlier research and that of our colleagues at Haskins Laboratories has dealt separately with the stop, nasal, and fricative consonants.[1] In the experiments to be reported here we have chosen to study /w, j, r, l/ as a group and to try to find the physical stimuli essential to the recognition of these phonemes in (absolute) initial position before vowels.

The selection of /w, j, r, l/ is necessarily arbitrary, at least in some degree; it is not random, however, since these phonemes have certain articulatory and distributional properties which tend to set them off from the other English consonants. From an articulatory standpoint, they differ from the constrictive consonants (stops, fricatives) in the degree of oral stricture present,[2] and

* 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.

** Also, University College, London.
*** Also, University of Connecticut.
**** Also, University of Colorado.

[1] A. M. Liberman, P. C. Delattre, F. S. Cooper, The role of selected stimulus-variables in the perception of the unvoiced stop consonants, *American Journal of Psychology*, 65.497-516 (1952); A. M. Liberman, P. C. Delattre, F. S. Cooper, L. J. Gerstman, The role of consonant-vowel transitions in the perception of the stop and nasal consonants, *Psychological Monographs*, 68. No. 8, Whole No. 379 (1954); Katherine S. Harris, Cues for the identification of the fricatives of American English, *Journal of the Acoustical Society of America*, 26.952 (1954).

[2] The majority of the allophones of /w,j,r,l/ are voiced oral resonants, and the occurrence of voiceless fricative allophones of all four phonemes in the same phonetic environment, i.e., after a voiceless stop or fricative in the same syllable, whilst no doubt connected with their phonemic distribution, is an added mark of their coherence as a class.

from the nasals by their oral articulation. On the basis of distributional characteristics, too, /w, j, r, l/ can be distinguished as a group from other sub-classes of English consonants. Their most obvious distinguishing mark is that they, and they alone, can constitute the third member of an initial three-term consonant cluster—for example in words like *screw*, *splint*, *skew*, and *square*. In other initial clusters these consonants must occupy the position immediately before the vowel and cannot have any other consonant intervening between them and the vowel.[3] Similarly, where they occur in final consonant clusters they must, again, occupy the place nearest the vowel, as in words like *melts* and *birds*.[4] The one exception to this is in words such as *snarl* or *world*, where /r/ comes between the vowel and /l/. We can, therefore, generalize the distribution of these sounds by saying that they must occupy a position in the syllable in immediate contact with the vowel (or with the vowel plus /r/), and that they are the only sounds permitted as the third term of a three-term initial consonant cluster.[5]

The fact that /w, j, r, l/ can be grouped together on articulatory and distributional grounds permits us to hope that they will have certain significant acoustic features in common. We should expect, then, that some of the acoustic cues which enable the listener to distinguish within the class will be found to lie on the same acoustic dimension, and, further, that these common acoustic features will serve, at least in part, to distinguish this class of phonemes from other classes.

METHOD

The inspection of acoustic spectrograms is a rewarding study which has already made a notable contribution to linguistic work, and it was upon such study that we relied for our preliminary observations. If a spectrogram has a fault, however, it is in the

[3] In many, perhaps most, dialects of English this would be an exclusive class-marker but, in the case of those dialects where "new" = /nu/, the phoneme /n/ shares this distributional characteristic.

[4] This sets them off from all English phonemes except /ŋ/ which must also occupy the immediately post-vocalic position, although it is quite unlike /w, j,r,l/ in its general distribution.

[5] For a more detailed statement of the combinatory latitudes of /w,j,r,l/ see J. D. O'Connor and J. L. M. Trim, Vowel, consonant, and syllable — a phonological definition, *Word*, 9.103-122 (1953).

abundance of its revelations, which lead an investigator to wonder what, in a given stretch of speech, is linguistically basic and what dispensable. Often, too, the acoustic features are only vaguely shown on a spectrogram, making it difficult to know the precise characteristics of those features which are revealed. These difficulties can to some extent be overcome by use of the pattern playback, a machine with which hand-painted patterns, resembling spectrograms more or less closely, may be converted to sound.[6] This provides a convenient method for making experimental modifications in various presumably important parts of the pattern and then evaluating the effects of these changes on the sound as heard. The present investigation was based primarily on this technique.

We were concerned in this experiment with the recognition of phoneme sequences as such. This is, of course, only one stage in the process by which speech is understood: the sounds perceived must be set against the pattern of both linguistic and extra-linguistic probabilities and a balance struck. In the experiments being reported here we have minimized the effect on intelligibility of any but the immediate sound contexts by working exclusively with nonsense syllables of the pattern CV. The vowels used were seven in number, corresponding approximately to the cardinal vowels [i, e, ε, a, ɔ, o, u].[7]

The experiments fall into two parts. In Part I we explored a wide variety of acoustic variations and attempted to determine, always by our own listening, which of these variations were important for the perception of /w, j, r, l/ and which were not. It is difficult to know precisely how many patterns and pattern changes were made and listened to, but the number is in the thousands. This is an important consideration only because it means that we were unable to control perfectly the contexts in which, and occasions on which, we listened to and judged the

[6] For a description of the pattern playback see F. S. Cooper, Spectrum analysis, *Journal of the Acoustical Society of America*, 22.761-762 (1950) ; F. S. Cooper, A. M. Liberman, J. M. Borst, The interconversion of audible and visible patterns as a basis for research in the perception of speech, *Proceedings of the National Academy of Sciences*, 37.318-325 (1951); F. S. Cooper, Some instrumental aids to research on speech, pp. 46-53 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).

[7] See P. C. Delattre, A. M. Liberman, F. S. Cooper, Voyelles synthétiques à deux formantes et voyelles cardinales, *Le Maître phonétique*, 96.30-37 (1951).

various patterns. The possibility exists, therefore, that our judgments were made against a background or standard that varied somewhat from week to week and from one context to another. As a partial control against this possibility, and also in order to have judgments from a group of naive listeners, we carried out the experiment reported in Part II. For the purposes of the second part of the study we sampled from the various acoustic dimensions we had found to be important in Part I, arranged these stimuli in a random order, and presented them to naive college students for judgment as /w/, /j/, /r/, or /l/.

PART I

Figure 1 shows patterns that contain the cues we believe to be most important for the perception of /w, j, r, l/. These patterns are, in general, like those we have found in previous studies to be appropriate for other CV syllables in that they consist primarily of formant transitions, or frequency shifts, followed by a steady state.[8] It had been found in the earlier studies that the transitions contain important, and in many cases sufficient, cues for the perception of the consonants,[9] while, as is well known, vowel color depends largely on the frequency position of the steady-state formants.[10] As can be seen from these patterns, the distinctions among /w, j, r, l/ depend primarily on the transitions of the second and third formants. The second-formant transition is sufficient to distinguish /w/ from /r-l/ from /j/: for /w/ this transition originates

[8] We have drawn all our transitions as straight lines even though the transitions of real speech are necessarily curvilinear. Using straight lines enables us to control the patterns more precisely, and it does not significantly alter the auditory impression.

[9] See Liberman, Delattre, Cooper, Gerstman, *op. cit.*

[10] This has been demonstrated by several methods : from a statistical analysis of the vowel productions of many speakers by G. E. Peterson and H. L. Barney, Control methods used in a study of the vowels, *Journal of the Acoustical Society of America*, 24. 175-184 (1952); from tests constructed at the pattern playback by P. C. Delattre, A. M. Liberman, F. S. Cooper, L. J. Gerstman, An experimental study of the acoustic determinants of vowel color: observations on one- and two-formant vowels synthesized from spectrographic patterns, *Word*, 8.195-210 (1952); and from vowel synthesis employing electronic analogs of the vocal tract by several investigators. See: H. K. Dunn, The calculation of vowel resonances, and an electrical vocal tract, *Journal of the Acoustical Society of America*, 22.740-753 (1950); K. N. Stevens, S. Kasowski, C. G. M. Fant, An electrical analog of the vocal tract, *Journal of the Acoustical Society of America*, 25.734-742 (1953); K. N. Stevens and A. S. House, Development of a quantitative description of vowel articulation, *Journal of the Acoustical Society of America*, 27.484-493 (1955).

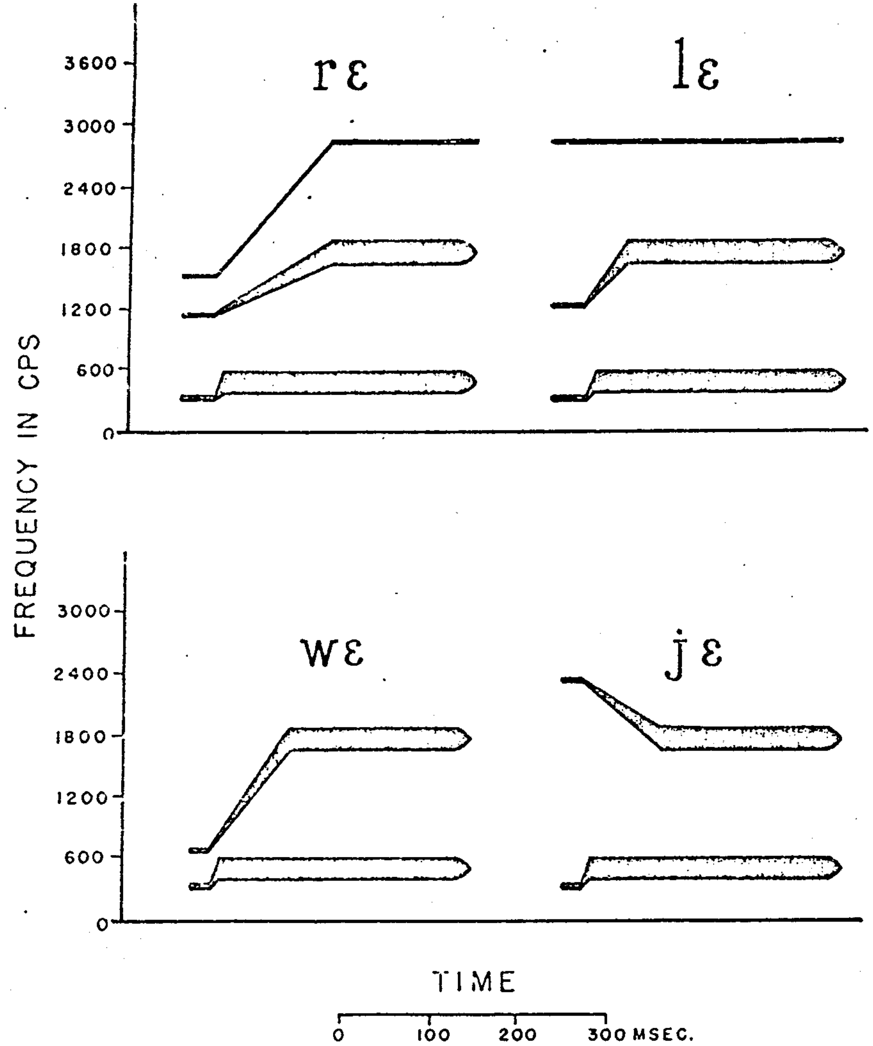

Fig. 1. **Hand-painted spectrograms that produce reasonable approximations to /rɛ/, /lɛ/, /wɛ/, and /jɛ/ when converted into sound by the pattern playback.**

at a low frequency, for /r/ and /l/ it begins nearer the middle of our frequency range, while /j/ requires that the transition start high. The distinction between /l/ and /r/ seems to depend primarily on the third-formant transition: for /r/ the third formant begins at a point just slightly above the second formant and then

rises to its steady-state level, while for /l/ the third formant starts at a level at least as high as the steady-state.

There is also at the beginning of each formant a relatively short steady state which, for convenience, we will refer to as a "steady-state onset." These steady-state onsets represent what we have previously referred to as the consonant loci,[11] and they are, quite obviously, the acoustic counterparts of the starting points of the consonant articulation. It is of considerable interest that the loci are explicit for /w, j, r, l/ because in this respect these four phonemes are different as a class from the stop and nasal consonants. In the case of the stops (and presumably for the nasal consonants, too) we had found for the second formant that the transition may not begin at the locus but must rather start at a point somewhat delayed in time. The steady-state onsets we see in Fig. 1 indicate that for /w, j, r, l/ the formants not only begin at the loci but spend from 30 to 50 msec. there before proceeding to the steady-state positions of the vowel.

Steady-state onsets similar to those of Fig. 1 were found in an earlier study to be necessary for the synthesis of the nasal consonants in initial position, but in this respect /w, j, r, l/ are very different from the nasal consonants in that the steady-state onsets are always continuous with the transitions for the former, but typically differ in frequency from the starting points of the transitions for the latter. This is so, presumably, because in the case of the nasal consonants the steady-state onset is produced as a consequence of the addition of the nasal resonator and is little affected by the (buccal) articulatory position, which nevertheless determines the starting frequency of the formant transitions. With /w, j, r, l/ the steady-state onsets reflect the start of the buccal articulation.

Having made these rather general comments about the acoustic cues that distinguish among /w, j, r, l/, and about certain characteristics of these four phonemes as a class, we should like to proceed to a more detailed discussion of our results.

Number of formants needed.

As shown in Fig. 1, a distinction may be made at the outset between /w, j/ on the one hand, and /r, l/ on the other, in that

[11] For a discussion of the "locus" concept, see P. C. Delattre, A. M. Liberman, F. S. Cooper, Acoustic loci and transitional cues for consonants, *Journal of the Acoustical Society of America*, 27.769-773 (1955).

the former can be synthesized satisfactorily with only two formants, while the latter require three. In the case of /w, j/ this is no more than might be expected, bearing in mind the well-known possibility of two-formant vowel synthesis, and the articulatory relationship between [i] and /j/ and between [u] and /w/. We find for /r/ and /l/, however, that a third formant is clearly necessary in order to synthesize these two consonants before the full range of vowels. As can be seen in Fig. 1, and as we will have occasion to discuss in greater detail later, the second-formant starting points are not very different for /r/ and /l/—they both begin in the middle of our frequency scale—so it is not surprising that we cannot distinguish these phonemes by means of two formants alone. This does not mean that the more-or-less common /r-l/ second formant will produce, with all vowels, a phoneme which is indifferently /r/ or /l/. In fact, we find with two-formant patterns that when the second-formant transition rises from the /r-l/ region to the frequency levels appropriate for front vowels we hear /r/, but when the transition falls to a back vowel we hear /l/.

Duration of steady-state onsets.

Steady-state onsets were found to be useful, not for any distinctive function they might have within our group of consonants, but to avoid a potential confusion with clusters of Stop+/w, j, r, l/. When the onset phase was omitted and the transition begun immediately, the resulting stimulus appeared to have an explosive beginning, a stop of some kind being clearly apprehended though not always identified with certainty. The duration of steady-state necessary to avoid this clustering effect differs as between /r, l/ and /w, j/. For /r/ and /l/ a steady-state onset period of about 50 or 60 msec. was needed—below this the explosive beginning started to make itself heard—whilst at greater durations, from 70 msec. upwards, the effect was of syllabicity, often leading to the identification of a schwa vowel before the consonant. For /j/ and /w/ the onset duration could apparently be smaller; 30 msec. was sufficient to eliminate the explosive beginning, and 40 msec. upwards gave the effect of a full vowel, /i/ or /u/, despite a reduction in intensity of the steady-state onsets *vis-à-vis* the following vowel formants.

Recent work suggests that the steady-state onsets can perhaps be dispensed with for /w/ and /j/ without causing an explosive beginning to the syllable, provided that the first-formant transition begins at a sufficiently high frequency (discussed in the next section). We have also found it possible to hear /r/ without initial steady states, provided that the intensity onset is gradual, but we have never been able to synthesize an acceptable /l/ using transitions alone. We note that none of the four phones is harmed by the presence of appropriate steady-state onsets, and indeed, that spectrograms typically show, if not a steady state preceding the transitions, at least a very much slower rate of frequency change than pertains to the transition phase.

Starting frequencies of the transitions (steady-state onset frequencies).

In the preceding section we observed that appropriate steady-state onsets serve mainly to distinguish the class /w, j, r, l/ from other groups of phonemes. Since these onsets have durations up to 60 msec., it is not inconceivable that they might also themselves be identifiable as /w/, /j/, /r/, or /l/, but our observations suggest that for the most part they are not; rather, it is the transitions, particularly those of the second and third formants, which play the main part in distinguishing among the four phonemes. Steady-state /w/ and /j/ are quite out of the question since the absence of transitions can result only in /u/ and /i/. The onset of /l/ contains some identifying information, but naive listeners have considerable difficulty in perceiving the steady state as /l/. For /r/ there is more information in the steady-state onset, to the extent that the onset approximates an /r/-colored vowel.

Bearing in mind that the transition directions and extents are the most important cues for distinguishing among /w, j, r, l/, we will nevertheless specify these variables by reference to the frequency positions of the steady-state onsets, since they are always continuous in frequency with the start of the transition. These starting points provide a convenient way of describing the transitions because the onsets remain relatively fixed while transition directions and extents shift radically with changes in the formant positions of the following vowel.[12]

First-formant onset. It will be obvious from the patterns of Fig. 1 that the onset of the first formant does not distinguish among /w, j, r, l/ since, for the same following vowel, its form is unchanged throughout. It does, however, serve the very impor-

[12] Delattre, Liberman, Cooper, (1955), *op. cit.*

tant function of distinguishing /l/ from the nasal consonants. If the onset of the first-formant transition is located at 240 cps, the resulting sound will be identified more often as a nasal than as /l/ even when the second and third formants are given their best /l/ values, whereas with the first-formant onset at 360 cps, or above, the identification will be /l/. The best /l/ results were obtained with the steady-state onset located at or slightly below the frequency of the first formant of the following vowel, but no lower than 360 cps. In subsequent work the first-formant onset was placed at 480 cps for [a], and 360 cps for the remaining vowels.

For both /w/ and /j/ one would expect the frequency of the first-formant onset to be in the neighborhood of 240 cps (the first-formant frequency of both /i/ and /u/), and this proved to be well founded. In fact, there was not much freedom to deviate from that value lest either of two complications occur: when the steady-state onset was located at 120 cps, the semivowels acquired stop-like beginnings, /bw/ or /gj/, an effect that we have discussed in other papers;[13] when the onset was placed much above 360 cps, the impression of /w/ was destroyed, while a non-speechlike whistle was introduced into /j/.

The influence of the first-formant starting frequency on /r/ was small. If a pattern contained second- and third-formant transitions satisfactory for /r/, the only difference made by a first-formant onset between 120 and 600 cps was a difference of "color", corresponding at the bottom of the range to extreme lip-rounding and at the top to lip-spreading.

We see, then, that /l/ imposes more stringent requirements upon the first-formant onset frequency than the other three phonemes. However, when we use values appropriate for /l/ we also satisfy the restrictions imposed by /w, j, r/.

Second-formant onset. The onset frequency of the second-formant transition has a very considerable effect on the perception of /w, j, r, l/. Other things being equal, high frequencies (about

[13] Delattre, Liberman, Cooper, (1955), *op. cit.;* A. M. Liberman, P. C. Delattre, L. J. Gerstman, F. S. Cooper, Tempo of frequency change as a cue for distinguishing classes of speech sounds, *Journal of Experimental Psychology*, 52.127-137 (1956). The first-formant locus for the stop consonants was found to be at the lowest frequency attainable with the pattern playback (120 cps). Any higher frequency caused the stops to sound less like stops and more like semivowels. This effect was slight, howe- ver, as compared to the strong influence of transition duration: patterns were heard as semivowels or stops depending on whether the transitions were more or less than 50 msec.

2760 cps) give /j/, low frequencies (about 600 cps) give /w/, and intermediate frequencies give either /l/ or /r/. The frequency ranges of /l/ and /r/ overlap.

For /w/ and /j/ the starting frequencies are, again as one would expect, in the neighborhood of the second formants of [u] and [i], respectively, but the opener and closer semi-vocalic allophones are clearly reflected in the differing ranges before different vowels. Before the vowel [i], the lower limit of the /j/ range has the same frequency as the second formant of the vowel (2760 cps), the glide effect being obtained by the transition of the first formant and the gradual increase in the intensity of the transitions; more satisfactory results are obtained, however, with a higher second-formant onset frequency and consequent falling transition. Before [e] (second formant at 2160 cps), the lower limit of the range is 2280 cps. One cannot start /j/ below 2280 cps for any following vowel without producing the effect of a glide from [ø] or [y]. There is virtually no upper limit for /j/, 3600 cps still being entirely satisfactory.

The frequency range of the second-formant onset for /w/ before [u] (second formant at 720 cps), is the very low and restricted one of 360-480 cps; before [o] on the other hand the upper limit is 600 cps, and for the remaining vowels, 840 cps.

For /r/ the frequency range of the second-formant onset before [i, e, ɛ] is 840-1560 cps, before [ɔ] 840-1200 cps, and before [o, u] 600-1200 cps. In the lower part of the range the acoustic counterpart of labio-velarization is heard, leading to confusion between /r/ and /w/ below the lower limit, whilst in the higher part of the range an effect of palatalization is heard.

For /l/ before [i, e, ɛ] the frequency range is 960-1800 cps, before [a] 840-1800 cps, and before [ɔ, o, u] it is 840-1680 cps. In the lower part of these ranges the /l/ was of a dark or velarized variety, the lower the frequency the darker the /l/; at the higher frequencies the /l/ was clearer or more palatalized, until, above the upper limits here given, the effect of laterality was lost, and was replaced by a vowel glide from [e] or from [i].

We can therefore distinguish /w/, /l, r/, and /j/ by reference to the second formant. Although the starting point of the second-formant transition should be somewhat higher for /l/ than for /r/, this does not provide a reliable differentia between the two phones.

Third-formant onset. The third formant is the crucial factor, as we suggested earlier, in distinguishing /l/ from /r/, but contributes little to the perception of /w/ and /j/. The third-formant starting frequency for /l/ is close to that of the vowel third formant, whereas the third-formant onset of /r/ needs to be lower in frequency, fairly close to the second-formant onset. This being the case, it is possible to pass from /l/ to /r/ by no other change in the pattern than a gradual lowering of the starting frequency of the third formant. If /w/ and /j/ are to have a third formant, the transition can be similar to that required for /l/, although as we have pointed out, such a third formant contributes very little to the perception of the semivowels.

We have already mentioned that the need for third-formant transitions in distinguishing /l/ from /r/ varies as between front and back vowels, so it should not be surprising to find these differences reflected in the permissible ranges of third-formant starting frequencies. For the vowels [i, e, ε], where the second-formant transition is rising and only two formants are required to hear /r/, the third-formant onset for /r/ can vary from 840 cps (the lower limit for the second-formant onset) to as high as 1920 cps; this means in effect that the frequency ranges of the second- and third-formant onsets are virtually identical for /r/, whereas the third-formant onset for /l/ must be no lower than the third formant of the vowel. For the vowels [ɔ, o, u], where the second-formant transition is falling and a two-formant pattern is heard as /l/, the restrictions are now reversed: the third-formant onset for /l/ may be as low as 1920 cps, but for /r/ it may be no higher than 1680 cps. It is possible, of course, to find starting frequencies for the second and third formants which cancel each other, so that neither /r/ nor /l/ is perceived with clarity; we may make /r/ and /l/ maximally discriminable, however, by placing the third-formant onset either as low as the second-formant onset permits, or as high as the third formant of the vowel.

Precise specification of formant starting frequencies. The reader will have noted that we summarized our observations about the starting frequencies of the second- and third-formant transitions in terms of ranges rather than as specific values. Thus, for example, we placed the second-formant onset frequency of /re/ between 840 and 1560 cps, but did not say that any particular value between these extremes gave a "better" /re/ than any other.

We believe that this is the most realistic way of stating our findings, since, in different idiolects and dialects, there are many phonetically discriminable sounds which are nevertheless readily identifiable as /r/. Similarly, two different frequencies within

our range will produce sounds which are distinguishable, but which are, in our estimation, both identifiable as /r/. We do not claim that every sound produced from different frequencies within a stated range makes an equally natural impression on us, or would do so on a wider selection of listeners—a very clear [l] initially may sound as odd to some listeners as a very dark [l] does to others—but we do believe that every such sound is recognizable phonemically.

The limits of the ranges are usually difficult to draw, since a given stimulus, evaluated phonetically, may contain auditory features reminiscent of both, say, /w/ and /r/, or /r/ and /l/; further, a stimulus identified by the phonetician as [ð] might well be categorized by the naive listener as /j/, for lack of any other oppositional category in which to place it, even though [ð] is not normally found as an allophone of /j/. We have always tried to draw the line at a point where the identification seems to us to be still positive; the same identification might be made beyond the limits we have laid down, but we would expect more confusion in these areas and more uniform responses within our ranges.

Our observations indicate that the range of starting frequencies is shifted somewhat (on the frequency scale) according to the following vowel, and this is no doubt a function of conditioned allophonic variation : we should suppose that to produce a given allophone we must change the starting frequency as we change the vowel. However, our frequency ranges reflect a good deal more than the allophones of a single idiolect or even of a single dialect, within which we would not necessarily expect so wide a range of frequencies.

Since our own judgments, whilst basically phonemic, are necessarily also partly phonetic, it will be particularly interesting to see, in Part II, whether the shifting of starting points in relation to the following vowel is confirmed by naive listeners judging on what we assume to be a purely phonemic basis.

Transition duration.

Transition duration is another factor in the perception of /w, j, r, l/. Its role seems to be similar to that of the steady-state onsets in that it does not serve to distinguish among the four phonemes (with the minor exception to be noted below), but does aid in differentiating /w, j, r, l/ as a class from other groups of consonants. If the transitions are of too brief duration, there is confusion with nasals and stops, while with transitions that occupy too much time in moving from the steady-state onsets to the steady-state vowel formants, there is the danger of losing the consonant impression entirely in favor of a vowel of changing color. Between these extremes there is a middle range of durations which serves reasonably well for all four consonants, although the particular values that give the most realistic sounds are somewhat different for each phoneme.

A duration of 100 msec. is suitable for the second- and third-formant transitions of all four phonemes. Briefer values give slightly better /l/ identifications, while longer ones help /r/. When /l/ is made as brief as 30 msec., there is the possibility of confusion with the nasal consonants, but at 60-70 msec. this ambiguity has disappeared, leaving /l/ at its most generally satisfactory value. When the transitions of an otherwise satisfactory /r/ pattern are reduced to 50 msec. or less, one hears a retroflex flapped sound, but, on the other hand, durations of up to 300 msec. do not destroy the /r/ effect.

Experiments with two-formant /w/ and /j/ showed that durations of 50 and 100 msec. were satisfactory, though the latter gave a somewhat more realistic sound; below 50 msec. the effect was of a vowel plus stop or flap.[14] At 150 msec. /j/ was still good, but an otherwise satisfactory /w/ pattern, before the vowels [i, e, ɛ], was heard as /wr/; this effect is undoubtedly due to the large rising transition of the second formant coupled with a comparatively slow rate of frequency change. The rising second-formant transition is often sufficient to produce an /r/ without the parallel rising third-formant transition that we have found to be so potent a cue for /r/. We find that a relatively broad rising second formant produces nearly the same effect as second- and third-formant rising transitions running parallel and in near contiguity. As one would expect, the /wr/ effect disappears when the transition durations are made briefer (thus speeding up the rate of frequency change), or when a straight third formant is added.

The first-formant transition for /w, j, r/ may have a duration identical to those of the second and third formants. For /l/, however, there is something of a special problem, since even at the

[14] These effects correspond exactly to those obtained with naive subjects who were able to hear the series /bɛ, wɛ, uɛ/ or /gɛ, jɛ, iɛ/ solely as a function of changes in transition duration. See Liberman, Delattre, Gerstman, Cooper, *op. cit.*, and Footnote 13.

briefest durations used with second and third formants (50 msec.), the first-formant transition seemed to be too gradual, giving the impression of a vocalic glide. It is apparently not without significance that spectrograms of spoken /l/ show a very abrupt frequency change between the first-formant onset and the following vowel. We sought to obtain this effect synthetically by reducing the transition duration to 10 msec., with the result that /l/ identifications were definitely improved. We found, further, that the patterns for /w, j, r/ were not adversely affected by this special first-formant transition, and have incorporated it into all the patterns used in Part II.

PART II

In this section we shall describe the results we obtained when we presented some of the stimuli of Part I to phonetically naive listeners for judgment as /w, j, r, l/. We are concerned here only with the cues that distinguish among these four phonemes. Accordingly, we have varied only those aspects of the pattern that had been found in our earlier investigation (see Part I) to be important for those distinctions, and we have asked our listeners to restrict their choices to /w/, /j/, /r/, or /l/. In regard to those acoustic variables which distinguish these phonemes from other classes, we have selected values which favored /w, j, r, l/ and held them constant in all the stimulus patterns.

The patterns seen in Fig. 2 illustrate, for the vowel [e], the stimulus variables we used in this experiment. As shown in the top row, there were six second-formant transitions, so chosen as to cover the range that had been found earlier to be important for /w, j, r, l/. We should note, however, that the range is sampled in rather large steps. By converting these two-formant patterns into sound we produced six of our test stimuli. Another twelve stimuli were made by combining each of these various second-formant transitions with a straight third formant and with a rising third formant as shown in the middle and bottom rows of Fig. 2. The straight third formant and the rising third formant had seemed on the basis of our earlier research to be reasonably appropriate for /l/ and /r/, respectively. There was, then, a total of 18 stimulus patterns for the vowel [e]. In the bottom row of the figure we see that the rising third-formant transitions always started at a point just slightly above the starting point of the second-formant transitions.

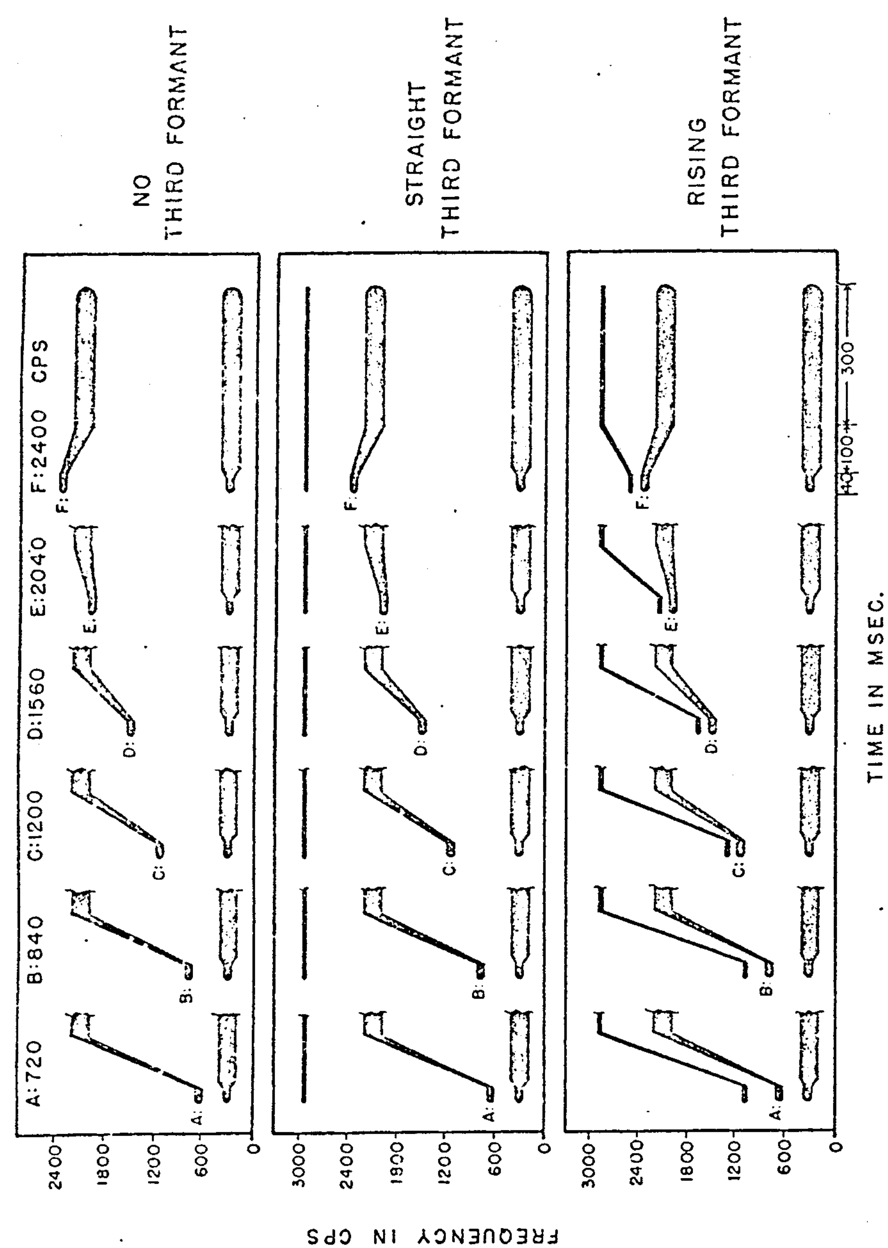

Fig. 2. Schematic representation of the stimulus variables used in Part II. The patterns shown are appropriate only for the vowel [e]. At the extreme right of each row are complete, hand-painted spectrograms illustrating some actual patterns from which the test stimuli were produced.

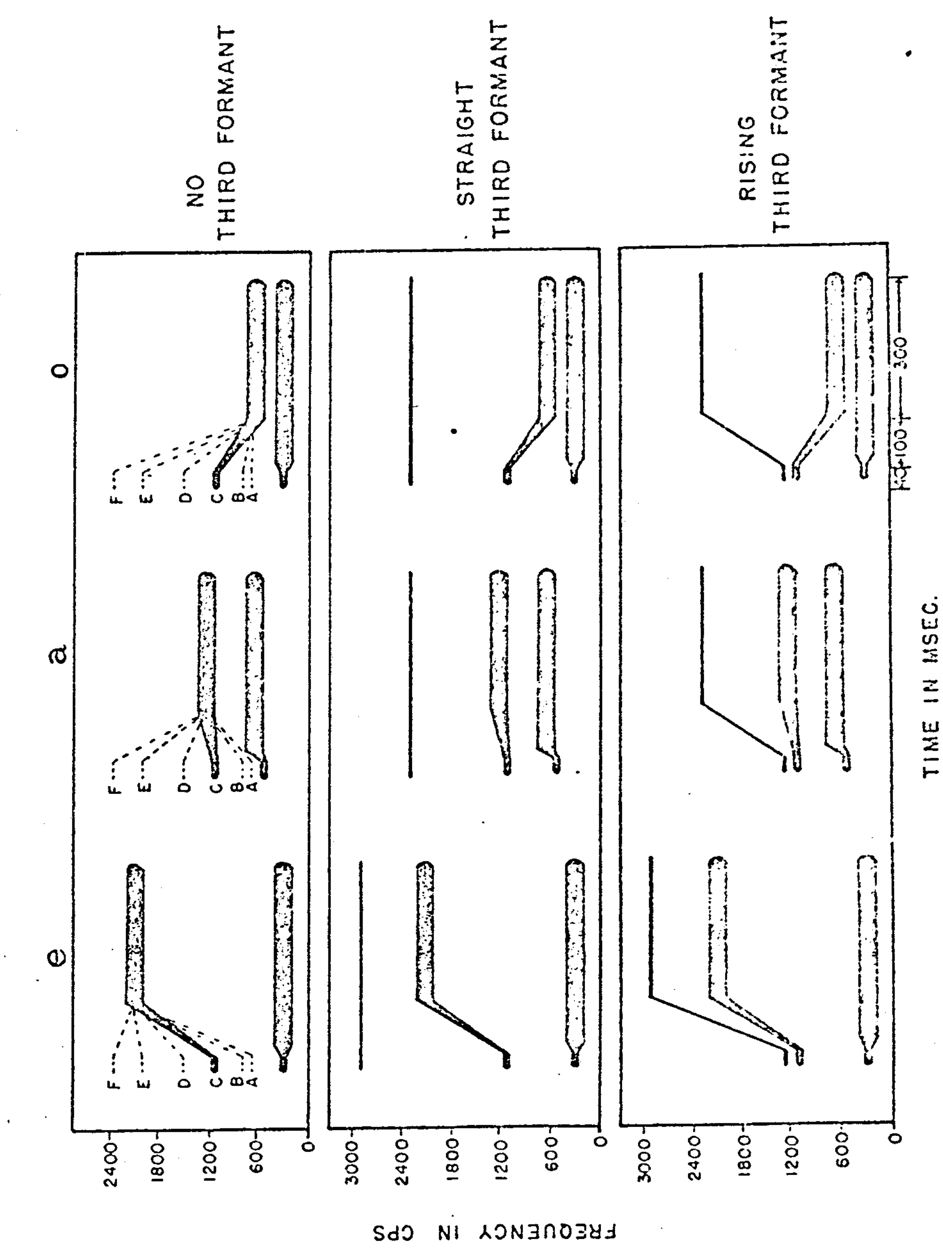

Fig. 3. Hand-painted spectrograms illustrating the stimulus variables of Part II with all three vowels [e, a, o].

There were, in addition, 18 stimulus patterns each for the vowels [a] and [o], constructed in essentially the same fashion as the stimuli for [e] in Fig. 2. Figure 3 summarizes the stimuli for [e, a, o] together. In the top row we see that the starting points of the second-formant transitions were identical for all three vowels. The middle and bottom rows illustrate, respectively, straight and rising third formants for a single second-formant starting point.

These patterns were converted into sound and presented in random order to a group of 44 phonetically naive listeners, all of whom were undergraduate students at the University of Connecticut. Each listener was asked to identify every sound as /w/, /j/, /r/, or /l/, and to guess if necessary.

The results of this experiment are to be found in the judgments of our listeners, all of which are shown in Fig. 4. There we have plotted the /w, j, r, l/ responses, expressed as per cents, against the second-formant onset frequencies, which are labelled on each abscissa by letter and identified by frequency in the key at the bottom of the figure. The three third-formant conditions (no third formant, straight third formant, and rising third formant) are the parameters, coded as shown in the upper right-hand corner. To avoid the difficulties in reading that would have been caused by many more or less overlapping curves, we have put the curves into separate plots according to responses (columns) and vowels (rows).

In general, the results obtained here are similar to those we described in Part I. The starting point of the second-formant transition distinguishes /w/, /r-l/, and /j/. This formant starts at a low frequency for /w/, at a middle frequency for /r/ and /l/, and at a high frequency for /j/. The third formant tends to distinguish /r/ from /l/, /r/ being helped by the rising third formant and /l/ by the straight third formant. Straight and rising third formants have little effect on /w/ and /j/, except that the rising third formant, in favoring /r/, tends to reduce the number of /w/ responses.

As we pointed out earlier, the stimuli that were presented for judgment in this experiment represent a rather large-step sampling of the various ranges, and it is, therefore, idle to draw conclusions which depend upon exact comparisons of the various curves, particularly in regard to the positions and heights of their peaks. However, some of these more specific effects are sufficiently large, and sufficiently consistent within the experiment, that we can

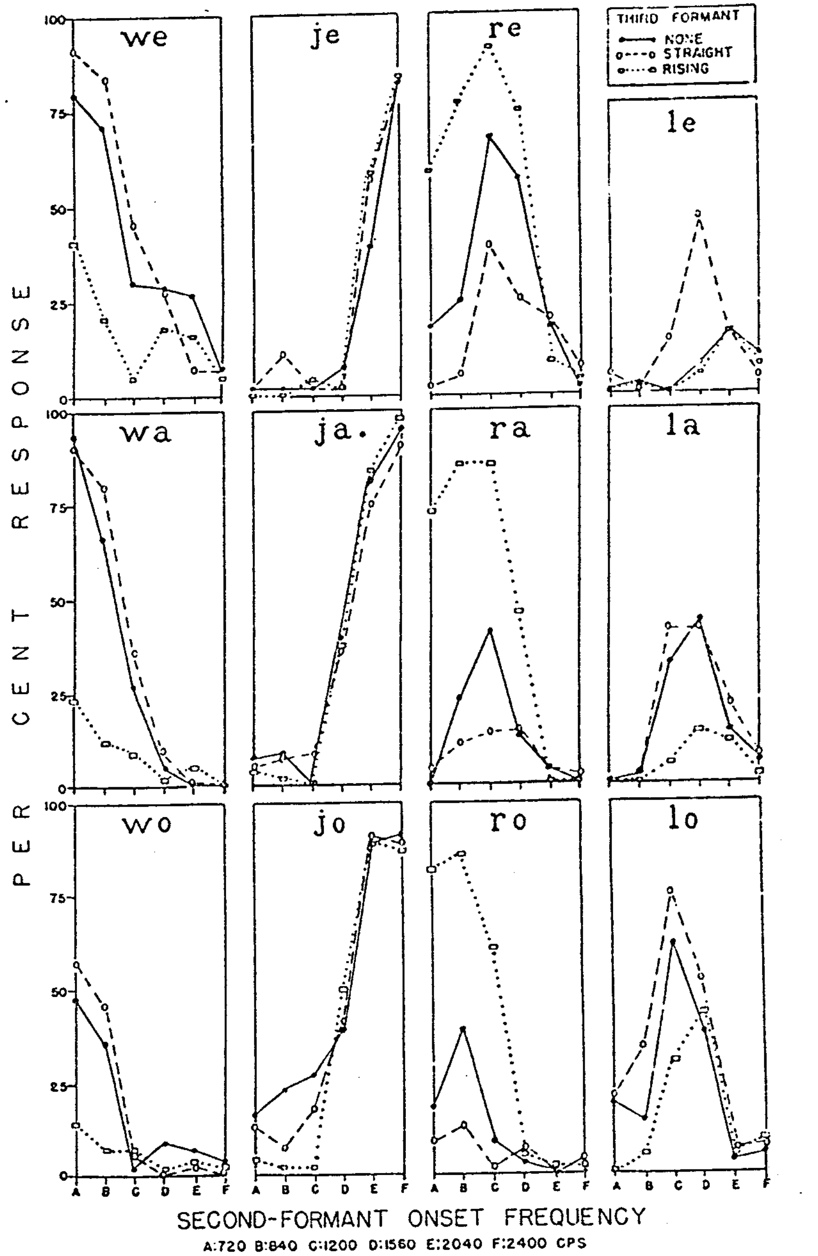

Fig. 4. Responses of 44 phonetically untrained listeners, plotted as a function of second-formant onset frequency for each of the three third-formant conditions.

be reasonably confident that they would be reproduced, at least in their grosser aspects, if the experiment were repeated with a much finer sampling of the stimulus range.

Perhaps the most important of these more specific results is the tendency of the peaks of both the /l/ and /r/ curves to occur at different positions on the abscissa as we go from the vowel [e] through [a] to [o]. These shifts in the response peaks mean that to produce the best /r/'s and /l/'s we must adjust the starting frequency of the second-formant transition according to the frequency level of the second formant of the following vowel. The starting point of the second-formant transition should be somewhat higher when the second formant of the vowel is high (as in [e]), and it should be somewhat lower when the vowel formant is low (as in [o]). We see, however, that the variations in the starting points of the transitions are quite small in relation to the frequency range through which the vowel formant moves. All of this confirms what we had previously observed and reported in Part I, and considerably increases our confidence in the result. The result itself is of some general importance, we believe, for when we take into account that /r/ and /l/ transitions begin at the actual loci (as many other consonant transitions do not), we see that we have here a clear case in which the frequency position of a consonant locus varies with changes in the frequency level of the vowel formant. This variation in the position of the locus is relatively small, which is fortunate for the utility of the locus concept as a simplifying assumption, but it clearly occurs, and we have remarked on it at some length because it is the first direct evidence we have had that a consonant locus will behave in this manner.[15]

The judgments of our naive listeners tend to agree with our own impressions (as reported in Part I) in several other particulars.

[15] K. N. Stevens and A. S. House have demonstrated that precisely this kind of locus movement can be expected, on the basis of their calculations, to occur for the second-formant loci of /b/ and of /g/: Studies of formant transitions using a vocal tract analog, *Journal of the Acoustical Society of America*, 28.578-585 (1956). In our own earlier research on the second-formant loci of /b/, /d/, and /g/, we had to contend with the difficulties arising from the fact that the second-formant transitions do not (and, indeed, cannot) begin at the locus, but must only point to it. As a result, we had to rely on a series of straight second formants (at various frequency levels) in order to find the /b/, /d/, and /g/ loci. By the very nature of that procedure, it was impossible to detect the kind of locus movement that we have here found with /r/ and /l/ and that Stevens and House would expect to find with /b/ and /g/.

We see in Fig. 4, for example, that the best starting point for the second-formant transition is at a somewhat higher frequency for /l/ than for /r/, though this difference would not appear from these data to provide a very good basis, in itself, for the perceived distinction between these two phonemes. It is also reasonably clear from the results with naive listeners that with two formants alone, /l/ is relatively good with the back vowels, while /r/ is relatively good with the front vowels.

It had been our feeling in working with these sounds that our /l/ was, at best, inferior to the other three phonemes, but we were somewhat surprised to discover how very poor our naive listeners found it. We believe that certain rather detailed changes in the first-formant steady-state onset and transition will considerably improve the /l/ for our naive listeners.

We see in Fig. 4 that /w/ before the vowel [o] was rather poor. This is not surprising when we consider that there were no patterns for [o] in which the starting point of the second-formant transition was below the steady state of the vowel (720 cps). Starting the transition at a somewhat lower frequency will probably improve the /wo/ considerably.

Haskins Laboratories.
New York, N. Y.