Acoustic Properties of Stop Consonants

M. Halle, G. W. Hughes, and J.-P. A. Radley

Citation: The Journal of the Acoustical Society of America **29**, 107 (1957); doi: 10.1121/1.1908634
View online: https://doi.org/10.1121/1.1908634
View Table of Contents: http://asa.scitation.org/toc/jas/29/1
Published by the Acoustical Society of America

Acoustic Properties of Stop Consonants*

M. HALLE,† G. W. HUGHES,‡ AND J.-P. A. RADLEY§
Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, Massachusetts

(Received September 6, 1956)

The two major cues for stop consonants, the burst of the stop release and the formant transitions in the adjacent vowel, were investigated. Detailed energy density spectra of the bursts were prepared. The transitions were studied by means of sonagrams. Possible criteria for identification were developed and tested. In order to assess the efficacy of the two types of cue, perceptual tests were conducted with isolated segments that contained either stop bursts or vowel transitions alone. Common acoustical properties of bursts and formant transitions are noted; differences as well as similarities are discussed in the light of different varieties of pitch judgments.

THE stop sounds, /p/ /t/ /k/ /b/ /g/, are produced by a complex of movements in the vocal tract. With the nasal cavity closed, a rapid closure and/or opening is effected at some point in the oral cavity. Behind the point of closure a pressure is built up which is suddenly released when the closure is released.

During the period of closure the vocal cords may or may not vibrate; if they do, we have a voiced stop; if they do not, we have a voiceless stop. Although in many instances the presence or absence of voicing serves to distinguish /b/ /d/ /g/ from /p/ /t/ /k/, in English voicing is not crucial to this distinction. The essential difference between these two classes of stops lies in the fact that in the production of the latter more pressure is built up behind the closure than in the production of the former. This difference in pressure results in higher intensity bursts and accounts for the well-known fact that /p/ /t/ /k/ bursts are often followed by an aspiration, which is not present in the case of /b/ /d/ /g/. Differences in the spectra of the bursts of these two classes of stops and in the duration of the preceding vowel can also be observed (see below). Since the role of the vocal-cord vibrations is thus relatively less important, the traditional terms "voiced" and "voiceless" seem somewhat inappropriate and will not be used here. Instead we shall refer to /p/ /t/ /k/ as "tense" and to /b/ /d/ /g/ as "lax" stops.[1]

The acoustic correlates of the complex of movements involved in the production of stops are rapid changes in the short-time energy spectrum preceded or followed by a fairly long period (of the order of at least several centiseconds) during which there is no energy in all bands above the voicing component (above 300 cps). This "silence" is a necessary cue for the perception of a stop sound: if the "silence" is filled by any other type of sound except voicing, a stop is not perceived.

When the stop is adjacent to a vowel, the movement in the oral cavity to and/or from the closure results in rapid changes in the formant frequencies. These rapid changes in the vowel formants adjacent to a silence are known as *transitions* and they are important cues for the perception of the different classes of stops.[2]

The rapid opening of the oral cavity is commonly accompanied by a short burst of noise. The spectral properties of the burst constitute another set of cues for the perception of the different classes of stops.

When a stop sound is adjacent to a vowel, we usually have all three cues: silence, burst and transition, or transition, silence and burst, as in "tack" in Fig. 1. Of these three, however, only the silence is a necessary cue—the silence with either transition or burst is a sufficient cue for identifying a stop. Thus, for example, in words like "task" (see Fig. 1) the identification of the final stop must evidently be attributed to the spectral properties of the stop burst; since the stop is not adjacent to any vowel, there can be no transition cue. On the other hand, in the ordinary pronounciation of words like "tact" (see Fig. 1), there is only a single silence followed by a single burst, although two stops /k/ and /t/ are perceived. The cue for the stop /k/ must, therefore, be contained in the transitions of the vowel formants preceding the silence.

* This work was supported in part by the U. S. Army (Signal Corps), the U. S. Air Force (Office of Scientific Research, Air Research and Development Command), and the U. S. Navy (Office of Naval Research); and the National Science Foundation.

† Also Department of Modern Languages, Massachusetts Institute of Technology.

‡ Also Department of Electrical Engineering, Massachusetts Institute of Technology.

§ Now at Air Force Cambridge Research Center, Hanscom Field, Bedford, Massachusetts.

[1] Jakobson, Fant, and Halle, Preliminaries to Speech Analysis, Technical Report 13, Acoustics Laboratory, Massachusetts Institute of Technology, May, 1952, pp. 36–39. The phonetic symbols used in this article are those of the International Phonetics Association (IPA).

[2] Ever since it became clear that vowels are products of the resonances of different configurations of the vocal tract, it has also been obvious that these resonances had to change as the geometrical configuration of the vocal tract changed. Not until the development of the sonagraph was it possible to follow these changes easily, although single investigators with unusually acute ears, like the Russian phonetician, A. Thomson, drew attention to these changes more than half a century ago. Thomson wrote: "Depending on the pitch of the proper tone of the mouth cavity ('pitch' refers to the second formant—M. H.) in its articulation of the preceding consonant, the vowel often begins considerably higher or lower, and then continuously and rapidly moves to its characteristic pitch on which it is held for a relatively long time. Towards the end, as it approaches the following consonant, the vowel again rapidly rises or drops, depending on the shape of the resonator characteristic of that consonant . . . even in the central part of the vowel there is no complete constancy in pitch. The same movement from the characteristic pitch of the preceding consonant to that of the following consonant continues here too." A. I. Thomson, Russ. Filol. Vestnik **54,** 231 (1905).

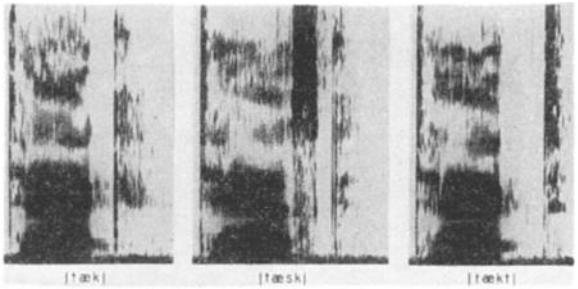

FIG. 1. Sonagrams of the words "tack" "task" "tact" (male speaker G) illustrating the role that transitions and bursts play in the perception of stops. In the /k/ of "tack" both transition and burst are present; in that of "task" only the burst is present; while in that of "tact" the transition alone is present.

The objective of the research reported below was to study the burst and transition separately in order to establish ways in which they could be used in a mechanical identification procedure.

SPECTRAL PROPERTIES OF STOP BURSTS

For this phase of the study our corpus consisted of monosyllabic words spoken in isolation by two males and one female. The list of words contained the six stops of English in position before and after the vowels /i/ /ɪ/ /ʌ/ /ɑ/ /u/. In addition, the list contained the voiceless stops in nonvocalic contexts, e.g., in the word "whisk," as well as words ending in the vowels /i/ /u/ /ɑ/.

Since we propose to study the bursts in isolation, the question might well be raised: Is a listener able to identify stop bursts isolated from their context as /p/ /t/ or /k/? We tried to answer this question by performing the following experiment. Words ending in stops produced with a burst and without vocal-cord vibration ("leap," for example) were selected from the corpus. The first 20 msec of each stop burst was gated out and rerecorded. Care was taken not to introduce any perceptible gating transients. The gated stop bursts were presented to listeners with instructions to judge them as /p/ /t/ or /k/. In the initial experiment, we experienced great difficulty in obtaining a reasonable response from our subjects, but with a certain amount of training it was possible to elicit fairly consistent responses. Our five best subjects gave the following percentages of correct responses: 65, 70, 75, 80, and 96. The last three subjects had had a considerable amount of experience with bursts in isolation; the first two subjects had received only a few minutes of instruction before the test. Since the percentages of correct responses of these five subjects were at least twice the percentages that might be obtained by guessing, we concluded that the bursts in isolation are identifiable as particular stops by listeners.

We hypothesized that the clues that make possible the identification of the bursts as different stops, reside in the spectrum. Consequently, we prepared detailed spectra of the stop bursts of all the words in our corpus. The first 20 msec of each stop burst, or the interval from the onset of the burst to the onset of the vowel—whichever was shorter—was fed into a filter of fixed band width whose center frequency was continuously variable over the range from 250 cps to 10 000 cps. For our filter we used a Hewlett Packard 300 A wave analyzer modified so that its band width was approximately 150 cps. The output of the wave analyzer was amplified, full-wave rectified, and integrated, and the resultant dc voltage was fed to a holding circuit and meter.[3] Samples of spectra are shown in Figs. 2, 3, and 4.

In examining the spectra, we note that the three classes of stops associated with different points of articulation differ from each other as follows:

/p/ and /b/, the labial stops, have a primary concentration of energy in the low frequencies (500–1500 cps).

/t/ and /d/, the postdental stops, have either a flat spectrum or one in which the higher frequencies (above 4000 cps) predominate, aside from an energy concentration in the region of 500 cps.

/k/ and /g/, the palatal and velar stops, show strong concentrations of energy in intermediate frequency regions (1500–4000 cps).

The differences among the various speakers were not very regular or marked. Much greater were the differences in the spectra of tense and lax stops. Since most of our lax stops were pronounced with vocal-cord vibration their spectra contained a strong low-frequency component. This component does not appear in the examples in Figs. 2, 3, and 4, because we passed all lax stops through a 300 cps high-pass filter before measuring them. The lax stops also show a significant drop in level in the high frequencies. This high-frequency loss is a consequence of the lower pressure associated with the production of lax stops and is therefore a crucial cue for this class of stops.

The most striking differences, however, were found in spectra of /k/ and /g/ in position before different vowels. Before front vowels (i.e., vowels having a second formant above 1200–1500 cps) the spectral peak of the bursts was in the region between 2000 and 4000 cps; before back vowels (second formant below 1200 cps) the spectral peaks were at much lower frequencies. These differences are not surprising, since it is well known that in English the phonemes /k/ and /g/ have two distinct contextual variants; one, before front vowels, produced with a closure nearer the front of the vocal cavity, and the other, before back vowels, produced with a closure more to the rear of the oral cavity. In position after vowels these contextual

[3] G. W. Hughes and M. Halle, J. Acoust. Soc. Am. 28, 303–305 (1956), gives a detailed description of measuring procedure and equipment characteristics.

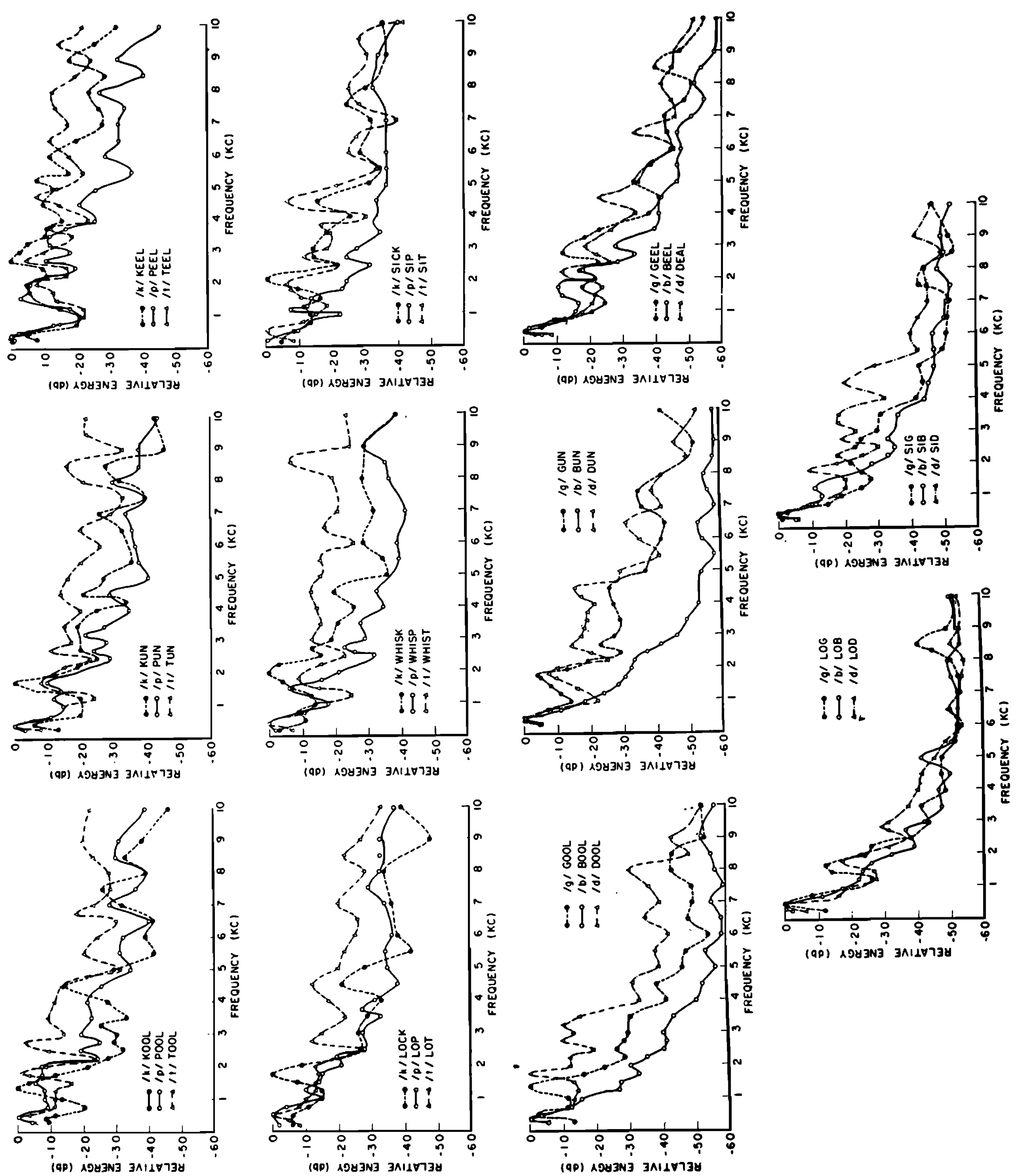

FIG. 2. Energy density spectra of stop bursts as spoken by speaker *G* (male).

differences were much less marked, which is to be expected since the "silence" between the end of the vowel and the burst was of the order of 100 msec.

A number of spectra deviated from the norms described above. Two particularly striking examples are given in Fig. 5; others can be found in Figs. 2, 3, and 4.

In spite of these divergences, the spectra possessed enough uniformity to make possible a statement of criteria that separate the spectra into three classes which are associated with the different points of articulation: First, the intensity in the 700–10 000 cps and 2700–10 000 cps bands was measured for all bursts. When the burst possessed significant energy in

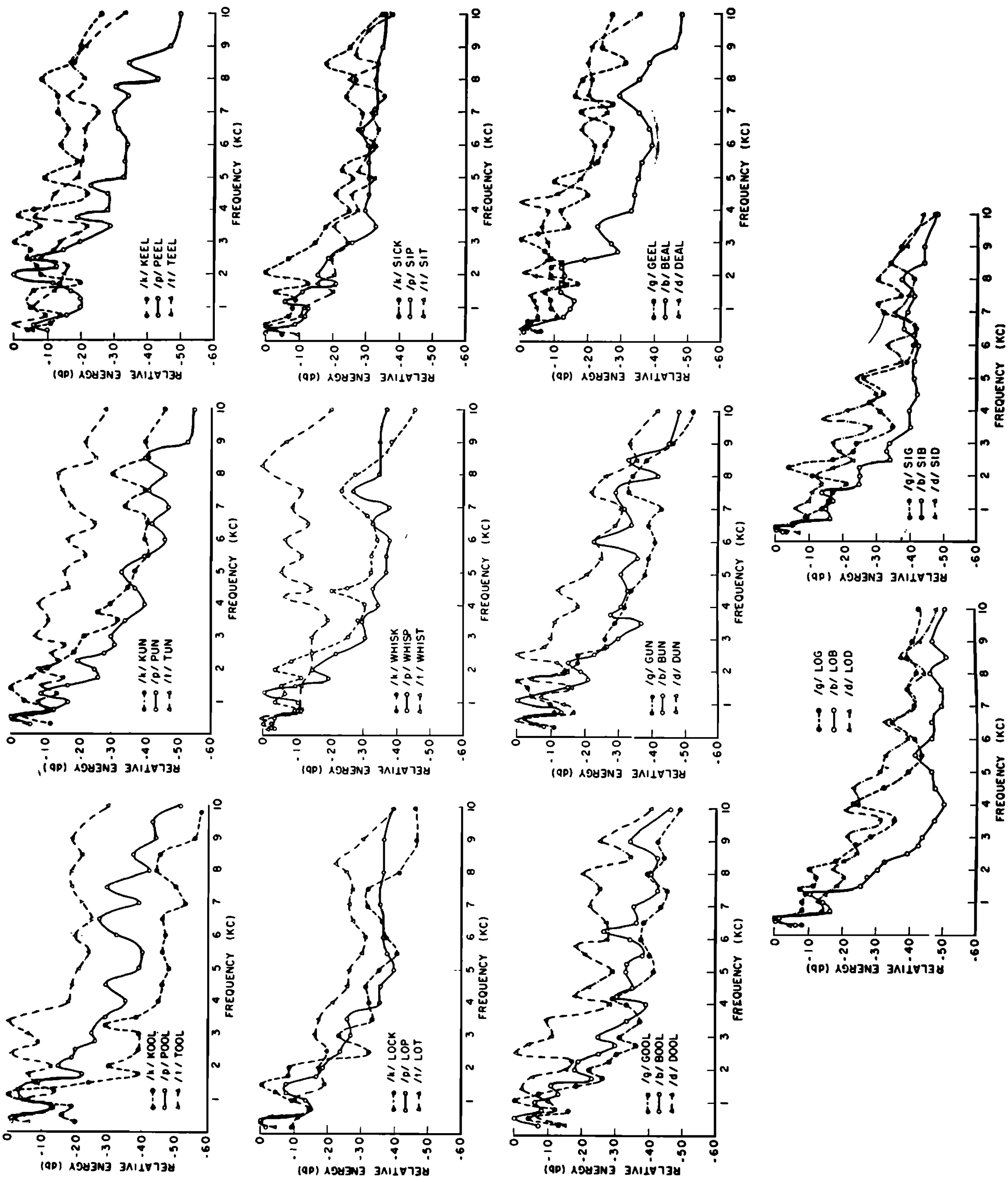

FIG. 3. Energy density spectra of stop bursts as spoken by speaker *A* (male).

the upper frequencies, these two values differed little (5 db for tense stops, 8 db for lax stops). When there was no significant energy in the higher frequencies, the two measurements differed greatly. As we have remarked, significant energy in the high frequencies is a characteristic of /t/ and /d/ and of the front variants of /k/ and /g/, while /p/ and /b/ and the back variants of /k/ and /g/ are characteristically weak in the high frequencies. We subjected our entire catalog of sounds to these two measurements and obtained correct classifications in 95% of the cases.

This step classified all the sounds into two classes; one, which we shall call the *acute* class, contained /t/ /d/ and the front variants of /k/ and /g/; the

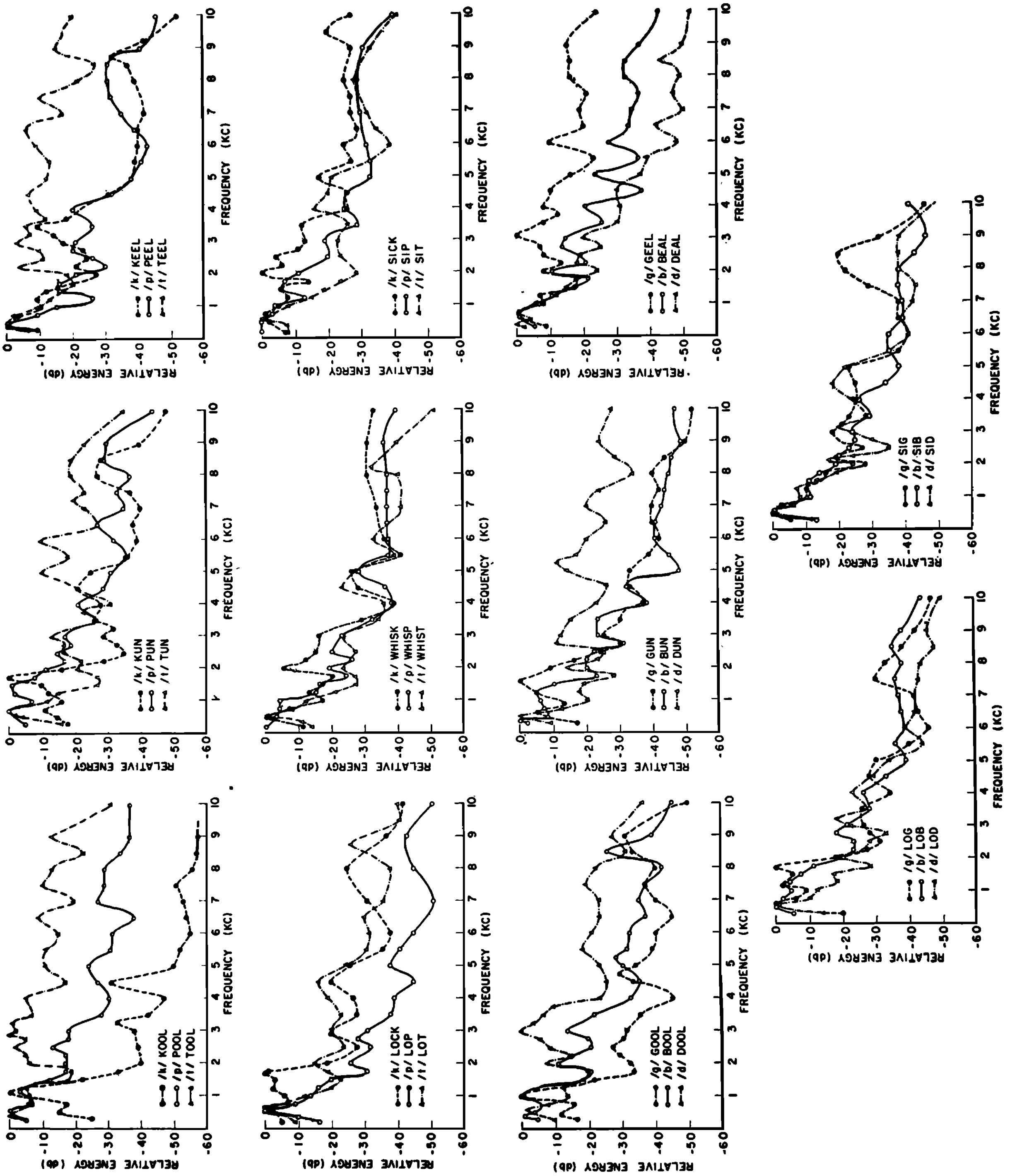

FIG. 4. Energy density spectra of stop bursts as spoken by speaker *R* (female).

other, which we shall call the *grave* class, contained /p/ and /b/ and the back variant of /k/ and /g/.[4] Acute /k/ and /g/ variants were rare in final position due to the "silence" that separates the burst from the preceding vowel. The next task involves subdividing these two classes again into two; thus obtaining the correct identification of the stops. In order to do this we found it necessary to apply a different procedure for the grave consonants than for the acute.

For the grave consonants, we noted the difference in levels between the two most intense spectral maxima and plotted it as a function of the frequency position

[4] See R. Jakobson *et al.*, reference 1, pp. 26–36, for an explanation of the terms "grave" and "acute."

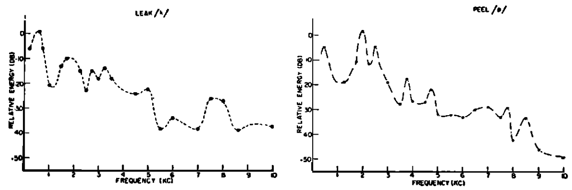

FIG. 5. Deviant spectra of stop bursts /k/ in "leak" (male speaker G) and /p/ in "peal" (male speaker A).

of the maximum of highest frequency. Such a plot is shown in Fig. 6. By the simple expedient of drawing straight lines across these graphs (the higher line for tense than for lax stops can easily be justified on the grounds of correcting the effects of the high-frequency drop that is characteristic of lax stops), we obtained correct separation of 85% of the tense bursts and approximately 78% of the lax bursts. The only really bad cases are the four low /g/ in final position. No special significance is to be attached to the shape or the position of the line other than that it separates what we know to be different.

We separated the acute stops into /t/ /d/ and /k/ /g/ by measuring graphically the average level of the spectrum between 300 cps and 10 000 cps and comparing it with the average level of the spectrum between 2 kc and 4 kc. Since /k/ and /g/ had an energy concentration in the central frequencies, the spectrum level in these frequencies considerably exceeded (5 db for tense stops, 8 db for lax stops) the average level of the entire spectrum. These values yielded the correct classification in approximately 85% of the cases. This procedure was very efficient in the case of the tense stops; for the lax stops it was not quite as reliable.[5]

TRANSITIONS

The other major class of cue that is important in the identification of the stop consonants is the transitions in the formants of the adjacent vowel. In recent years, a great deal of valuable information concerning this class of cue has been gathered by the researchers at the Haskins Laboratories in their work on the perception of synthetic speech stimuli in which certain features were systematically varied.[6] The problem, as we saw it, was to determine to what extent the uniformities observed with synthetic stimuli were applicable to natural speech.

We approached this problem from two directions: (1) by studying a large number of spectrograms and attempting to correlate the observed formant transitions with the stop that was uttered by the speaker; and (2) by conducting perceptual tests that were similar to those carried out by the Haskins group except that natural speech was used in place of synthetic stimuli.

Spectrographic Studies[7]

For examining the transitions, the principal tool was a Kay Electric sonagraph. The sonagraph has certain limitations for a study of this kind, the most serious of which are the restricted range of automatic gain control and the fixed band widths of its analyzing filter which make it impossible to obtain clearly defined formants for certain speakers. Nevertheless, in the majority of our samples the presentation was quite satisfactory.

The corpus in this study consisted of words, primarily monosyllables, in which the English vowels /i/ /ɪ/ /e/ /ɛ/ /æ/ /ɑ/ /ʌ/ /ɔ/ /o/ /u/ and /ʊ/ were combined with the six stops in initial and final position. The list was read twice by three males and one female whose dialects varied somewhat.

Our first problem, how to define a transition, illustrates well the difficulties that are encountered in the study of natural speech, but can easily be avoided when one has control over the stimulus.[8] In the sonagrams, shown in Fig. 7, there are a number of stop transitions that differ greatly in duration, rate of change, and their terminal (beginning and end) points. In Fig. 7(a), the final transition is considerably longer than 50 msec. The stop transition cannot be distinguished here from the change in the position of the vowel formants arising from the diphthongal pronounciation of the vowel. In Fig. 7(b), the vowel lasts only 100 msec and it is difficult to find any steady-state

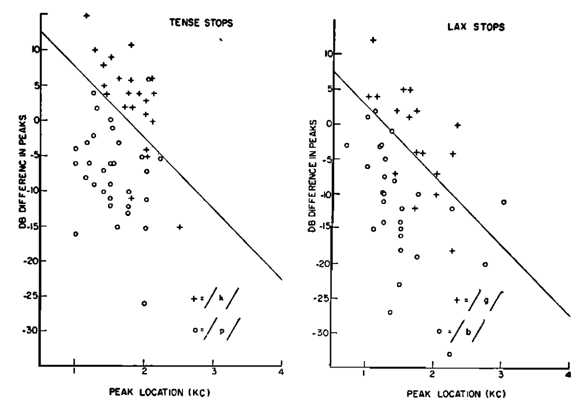

FIG. 6. Plots of the intensity differences of the two most intense spectral maxima as a function of the frequency position of the maximum of highest frequency. The straight lines in each plot divide /k/ and /g/ bursts from /p/ and /b/ bursts.

[5] Attention is drawn to a similar scheme proposed for the identification of Russian stop sounds in a forthcoming book: M. Halle and L. G. Jones, *The Russian Consonants* (Mouton and Company, 'S-Gravenhage, Netherlands, to be published).

[6] Liberman, Delattre, Cooper, and Gerstman, "The role of consonant-vowel transitions in the perception of stop and nasal consonants," Psychological Monographs No. 379, 1954; see also publications cited therein.

[7] For the material (including more than 500 sonagrams) on which the following discussion is based, see J.-P. A. Radley, "The role of transitions in the identification of English stops," S.M. thesis, Department of Electrical Engineering, Massachusetts Institute of Technology, 1955.

[8] Delattre, Liberman, and Cooper, J. Acoust. Soc. Am. **27**, 769–773 (1955).

portion or to decide where the transition begins. In Fig. 7(c), the final transitions of the different formants start at different times. Note that here the final $F2$ transition lasts only 20 msec. The initial $F2$ and $F3$ transition can only be discerned with the greatest difficulty even if the aspiration is included with the vowel transition.

All of these difficulties show that it is not easy to decide what segment on a given sonagram constitutes a transition. Specifically, it is often impossible to identify a transition by examining on a single sonagram a fixed time interval before or after a silence or by looking for certain rates of change in the formant center frequency or the formant band width.

We find, however, that by looking at sets of sonagrams of minimally different words, e.g., "seep," "seat," "seek," differences in the formant transitions, if they exist, can be easily pointed out. The transition, like so many other linguistic concepts, must be defined with respect to a set of entities that are otherwise identical.

The regularities that we observed were considerably more complex than the elegant "locus" rules that summarize the results of the Haskins experiments. We found dependencies not only on the steady-state position of the adjacent vowel, but also on the position of the stop with respect to the vowel (preceding or following), and of the feature tense-lax in both the consonant and the vowel.[9]

The least satisfactory group was that of the tense stops in initial position. For these stops, generalizations are only possible about transitions in the lax vowels /ɪ/, /ʌ/, and /ʊ/ and in the tense vowel /u/. Before /ʌ/, $F2$ and $F3$ have neutral or negative transitions for /p/; $F2$ has a positive transition and $F3$ has a negative or neutral transition for /t/; and $F2$ and $F3$ converge for /k/. Before /ɪ/, negative transitions in both $F2$ and $F3$ are associated with /p/, neutral or moderately positive transitions in both $F2$ and $F3$ are correlated with /t/, while /k/ is associated with a positive $F2$ and a neutral $F3$ transition, which can be thought of as a case of convergence. Before /ʊ/ and /u/, the $F3$ transition could not be seen in quite a few instances; in those cases in which it was visible it seemed to us to be neutral. $F2$ had positive transitions for /t/ and neutral or slightly negative transitions for both /p/ and /k/, which, however, could not be separated on the basis of their transitions.

In the case of tense stops in final position, the transitions are considerably more uniform. A partial explanation for this may be that imploded stops are quite common in final position. In these cases, the transition cue is not supplemented by a burst cue, as it is in initial position. The transitional cues fall into two classes: after /i/ /ɪ/ /ɛ/ /e/ /æ/ /ɑ/ /ɔ/ /ʌ/ and after the rounded back vowels /u/ /ʊ/ /o/. In the former class /p/ induces a markedly negative transition in $F2$, which is absent for /t/ and /k/. The latter two can be distinguished by noting that /k/ has a convergence of $F2$ and $F3$. In the case /u/ /ʊ/ /o/ the cue for /t/ is a markedly positive $F2$ transition, which sometimes meets a descending $F3$, while the absence of such an $F2$ transition is the cue for /p/ and /k/, which do not differ significantly from each other.

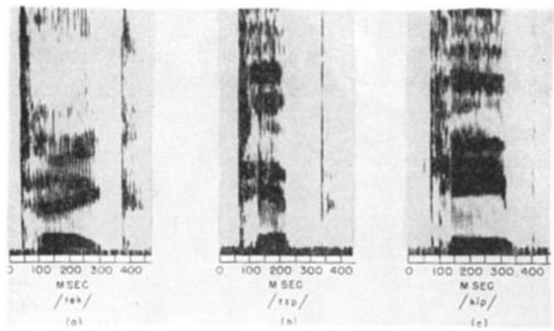

FIG. 7. Sonagrams illustrating the great variety of transitions encountered in actual speech. (a)—"take" (male speaker R); (b)—"tip" (male speaker B); (c)—"keep" (male speaker H).

For lax stops in initial position, /b/ has a more negative or less positive $F2$ and $F3$ transition than /d/ for every vowel except /i/, where the transitions may be similar. As we move along the vowel triangle from /i/ to /u/, /d/ gives progressively more positive $F2$ transitions, in conformity with the results obtained in the Haskins perceptual tests; but it was not possible to specify a "locus" frequency. Before the front vowels, /i/ /ɪ/ /e/ /ɛ/ /æ/, /g/ has a more positive $F2$ transition than the other two stops, while before the back vowels, /g/ has a less positive transition than /d/. Convergence of $F2$ and $F3$ is common, though not universal, for /g/ transitions, particularly before back vowels.

In the case of the lax stops in final position, the rules are fairly similar to those stated for tense stops in final position. As with /p/, there is the very marked negative transition before /b/ in all vowels except those with the lowest $F2$: /u/ /o/. Again we find something like a "locus" phenomenon in the behavior of /d/ transitions, which are slightly negative or neutral with vowels having a high second formant and become progressively more positive as $F2$ of the vowels is lower. After front vowels, /g/ has a more positive transition than /d/, and is further distinguished from /d/ by a negative $F3$, thus giving convergence. For the back vowels, however, the $F2$ transition is considerably more positive for /d/ than for /g/. In these instances, /g/ transitions cannot be differentiated consistently from /b/ transitions.

[9] See K. N. Stevens and A. S. House, J. Acoust. Soc. Am. 28, 578–585 (1956). Following common usage, we call a transition "negative" if its terminal (beginning or end) point is of lower frequency than the steady-state or average position of the formant in the vowel; "positive" if it is of higher frequency; and "neutral," if it is of the same frequency.

TABLE I. Responses of the listeners to syllables from which the final stop bursts were gated out. Each box contains the results of tests with syllables beginning with the consonant plus vowel sequence indicated. The vertical columns show the number of different consonant judgments made with respect to a single stimulus.

		Final stop in stimulus syllable /la-/							
		/p/	/t/	/k/	/#/	/b/	/d/	/g/	Total
	/p/	5	0	1	0	3	0	0	9
Number	/t/	9	25	21	0	0	0	0	45
of listener	/k/	5	1	2	0	0	0	0	8
judgments	/#/	9	1	6	25	3	1	0	45
	/b/	2	0	0	4	22	0	3	31
	/d/	0	2	0	0	2	29	20	53
	/g/	0	0	0	1	0	0	7	8
	Total	30	29	30	30	30	30	30	

		Final stop in stimulus syllable /li-/							
		/p/	/t/	/k/	/#/	/b/	/d/	/g/	Total
	/p/	26	14	20	0	3	0	0	63
Number	/t/	3	12	9	0	0	1	1	26
of listener	/k/	0	1	1	0	0	0	5	7
judgments	/#/	0	3	0	20	8	13	19	63
	/b/	0	0	0	6	15	1	5	27
	/d/	1	0	0	3	4	24	4	36
	/g/	0	0	0	1	0	1	1	3
	Total	30	30	30	30	30	40	35	

		Final stop in stimulus syllable /lu-/							
		/p/	/t/	/k/	/#/	/b/	/d/	/g/	Total
	/p/	18	2	11	0	2	0	0	33
Number	/t/	5	24	4	0	0	0	0	33
of listener	/k/	1	1	4	0	0	0	0	6
judgments	/#/	6	1	9	27	30	4	17	94
	/b/	0	2	2	2	7	0	12	25
	/d/	0	0	0	1	1	26	0	28
	/g/	0	0	0	0	0	0	1	1
	Total	30	30	30	30	40	30	30	

		Final stop in stimulus syllable /sʌ-/						
		/p/	/t/	/k/	/b/	/d/	/g/	Total
	/p/	21	0	2	9	0	0	32
Number	/t/	7	27	16	0	0	10	50
of listener	/k/	1	1	11	0	0	0	13
judgments	/b/	1	0	0	11	0	4	16
	/d/	0	2	0	7	19	7	35
	/g/	0	0	1	3	1	19	24
	Total	30	30	30	30	20	40	

		Final stop in stimulus syllable /sɪ-/						
		/p/	/t/	/k/	/b/	/d/	/g/	Total
	/p/	23	5	0	5	0	0	33
Number	/t/	6	23	1	1	7	1	39
of listener	/k/	0	0	27	0	0	2	29
judgments	/b/	1	0	0	24	1	2	28
	/d/	0	2	0	0	22	1	25
	/g/	0	0	2	0	0	23	25
	Total	30	30	30	30	30	29	

Perceptual Tests

It was, unfortunately, impossible to carry out large-scale perceptual tests. The data reported below are, therefore, to be taken as preliminary results.

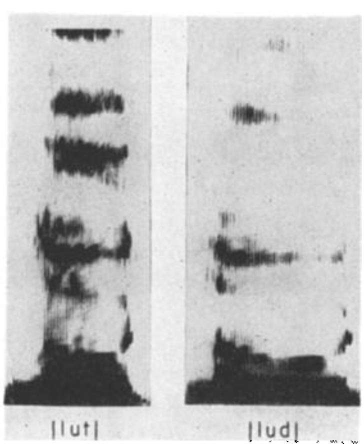

FIG. 8. Sonagrams of the words "lute" "lewd" (male speaker G) illustrating the differences in duration between the vowels preceding tense and lax stops.

In the tests, listeners were presented with monosyllables recorded by one female and two male speakers. The syllables contained the vowels /i/ /ɑ/ /u/ /ɪ/ /ʌ/ followed by each of the six stops. The first three vowels were also contained in open syllables. Thus, a representative paradigm contained the following stimuli: /lip/ /lit/ /lik/ /lib/ /lid/ /lig/ /li/. By means of an electronic gate each syllable including the open one (i.e., the one ending in a vowel) was terminated immediately after the vowel so that no stop burst appeared in the test. The subjects were asked to identify the end of the syllable as one of the six stops or as "nothing" (#). After the lax vowels, where in English an open syllable is impossible, the judgment "nothing" was omitted. The results of the tests are given in Table I.

Because of the very marked difference in duration of vowels before tense and before lax stops (see Fig. 8), the subjects had little difficulty in distinguishing tense from lax stops. Subjects never confused an open syllable with a syllable ending in a tense stop, but occassionally an open syllable was thought to end in a lax consonant. On the other hand, closed syllables, particularly those ending in lax stops, were rather frequently judged to be open. The "mistakes" are correlated with the position of first formant of the preceding vowel. They seem to indicate that negative transitions of the first formant played a particularly significant role in the identification of stops as class. Where $F1$ was high and, consequently, free to move downward, closed syllables were relatively rarely thought to be open; where $F1$ was low and its downward movement restricted, closed syllables were frequently judged to be open.

Somewhat unexpectedly, in the light of previous studies, but not quite so surprisingly, in view of the transition data reviewed in the preceding section, the greatest differences in response were correlated with the differences between tense and lax vowels. After lax vowels, the transitions seemed, in general, easier to judge correctly than after tense vowels. Especially great difficulties were experienced with stimuli ending in /k/ and /g/ preceded by tense vowels. Here /k/ and /g/ judgments were avoided to such an extent that they constituted only 5% of the responses, although syllables ending in /k/ and /g/ accounted for 28.5% of the stimuli. This compares with 25.5% /k/ and /g/ responses *vs* 36% of the stimuli for the transitions in the lax vowels.

As for individual vowels, we must note the extreme reliability of the judgments for the vowel /ɪ/, which is explained by the very clear transitions of this vowel; see Fig. 9. The other lax vowel in our corpus /ʌ/ was also judged with fair reliability, although not nearly as reliably as /ɪ/. It is significant that /k/ and /g/ judgments are again the least reliable. After /u/ there was a very marked tendency to consider syllables ending in a lax stop as open, which is explained by the total absence of transitions (see Fig. 10). In view of the restricted size of the sample, the tendencies in the /i/ and /a/ transition judgments are not sufficiently marked to warrant specific conclusions.

CONCLUDING REMARKS

The data just reviewed, though fragmentary and in need of further elaboration, give support to the common view that, in the perception of stop sounds, information from the burst as well as from the transition is normally required. This view, however, presents us with the paradoxical situation that what appear to be the two most disparate acoustical phenomena, formant movements and bursts of sound, are perceptually equated. We shall try to account for this by considering some acoustical properties of bursts and transitions and their relation to the perception of speech.

Physically speaking, a formant[10] reflects organization of the acoustic energy in the frequency domain and no organization in the time domain; while a burst represents organization of acoustic energy in the time domain and no organization in the frequency domain. In the case of an ideal formant, there is no energy outside the infinitely narrow resonance frequency, but the sound lasts forever. An ideal burst (impulse), on the other hand, has an infinitely short duration; i.e., the energy is present only at a given instant, but its band width is infinite.

When a resonance is changing in frequency, the formant band width increases. The more rapid the movement, the broader the band width. In the limiting case of instantaneous movement, the band width is infinite; i.e., we have a phenomenon of strictly limited duration and infinite band width. But these are exactly the terms in which we previously described the burst. The burst can, therefore, be considered as an extreme

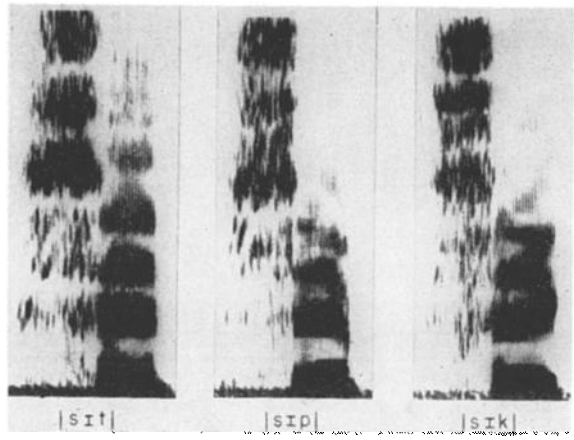

FIG. 9. Sonagrams of the words "sit" "sip" "sick" (male speaker *G*) illustrating the exceptionally unambiguous transitions in the end of the vowel /ɪ/. Note that in "sit" $F2$ and $F3$ transitions are neutral; in "sip" both transitions are negative; while in "sick" $F2$ has a positive transition and $F3$, a negative transition, and the two formants converge.

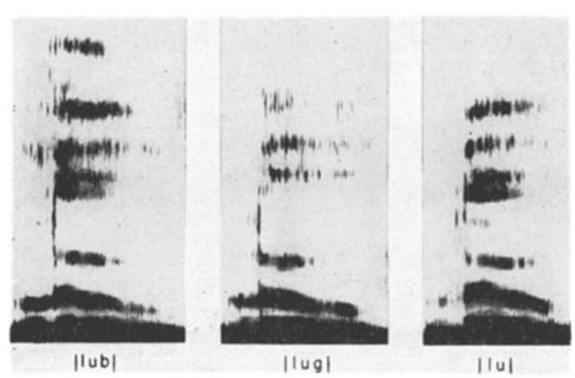

FIG. 10. Sonagrams of the syllables /lub/ /lug/ /lu/ (male speaker *A*) showing almost no differences in the transitions.

[10] We distinguish between "resonance" and "formant." The former refers to a maximum in the frequency response of a resonator that is computed by taking into consideration only the geometrical properties of the resonator and by specifically neglecting the special effects that are produced when the resonator is excited in a particular manner and its output measured by a particular device and procedure. The term "formant" takes all these factors into consideration; thus it refers to a frequency region of maximal intensity in an actually measured spectrum.

case of transition in which the changes in the short-time energy-density spectrum are very rapid and the organization of the energy in the frequency domain is replaced by organization in the time domain.

It is usually assumed that moving formants are perceived in a manner similar to that of stationary formants. This view seems to be an oversimplification, for no reasons have been advanced that would justify ignoring the fact that when a resonance is changing in frequency the problem of measuring this frequency is complicated by increases in the formant band width. It is obvious that in the case of instantaneous change of resonance frequency, it is meaningless to specify a formant frequency, or formant movement. In intermediate cases, the rate of change (or the formant band width) at which it becomes meaningless to specify a formant frequency depends upon the properties of the measuring apparatus,[11] which, in the case of interest here, are those of the human auditory system.

In experiments with damped sine waves, K. N. Stevens[12] has shown that, as the resonance band width is increased by increasing the damping, it becomes increasingly difficult to match the pitch of such a resonance to that of a pure tone. Although Stevens' experiments were not carried out with enough values of damping to provide conclusive evidence on this point, they indicate that, at some critical band width, which we estimate to be 300 cps, the train of damped sine waves can no longer be matched in pitch to a pure tone. This does not mean that all pitch judgments become impossible; for example, different trains of damped sine waves can still be ordered from high to low with some consistency.

It is interesting that we observe similar facts in judging the pitch of vowels and consonants. On the one hand, the vowels can be matched in pitch to pure tones without much difficulty; such matches have been performed with amazing accuracy ever since the seventeenth century, when Samuel Reyherr, in his book *Mathesis mosaica* (Cologne, 1679), gave in musical notation the "characteristic pitches" (second formant) of German and French vowels.[13] On the other hand, consonants cannot be matched easily in pitch to pure tones; and all "pitch" determinations of consonants are completely unconvincing and usually refer to the terminal stage of the second formant in the adjacent vowel.[14]

One might, therefore, suggest as an hypothesis that the distinction between vowels and consonants, which is the fundamental dichotomy of phonetics, is based on the organism's ability to perform much more elaborate pitch judgments with respect to certain classes of physical stimuli (vowels and undamped or moderately damped sine waves) than with respect to other classes (consonants, tonal masses, noises, and highly damped sine waves). Formant transitions would then be intermediate structures whose assignment to the vowels or to the consonants is a function of their band width, which in turn is dependent on their rate of change.

[11] L. L. Beranek, *Acoustic Measurements* (John Wiley and Sons, Inc., New York, 1949), pp. 538–542.

[12] K. N. Stevens, "The perception of sounds shaped by resonance circuits," Sc.D. thesis, Department of Electrical Engineering, Massachusetts Institute of Technology, 1952.

[13] Compare C. Stumpf, *Die Sprachlaute* (Berlin, 1926), p. 148.

[14] Compare the following remark of a nineteenth century English phonetician who attempted to determine the "pitch" of consonants: "When freed from connection with any vowel, the resonance of *f* can be carried a long way both up and down in pitch without at all spoiling the *f* itself. . . . It becomes clear that the essential quality of *f* is but vaguely linked with the actual pitch of its resonance." R. J. Lloyd, Proc. Roy. Soc. (Edinburgh) 22, 224 (1898).