

On the front cavity resonance and its possible role in speech perception

G. M. Kuhn

Haskins Laboratories, New Haven, Connecticut 06510
(Received 10 March 1975; revised 24 April 1975)

Spectrographic data are presented which suggest that it may be possible to estimate the frequency of the fundamental resonance of the cavity behind the mouth opening, the "front cavity resonance," from information in the speech signal. It is shown that place of articulation information in the steady states, transitions, and bursts of F_2 (or sometimes F_3) can be reinterpreted to be information from the front cavity resonance. Furthermore, a number of synthesis results that have appeared anomalous when described in terms of numbered formants seem to find a coherent explanation in terms of the front cavity resonance. Implications for theories of speech perception include the possibility that an estimate of front cavity resonance frequency may serve for continuous articulatory reference.

Subject Classification: 70.20; 70.30.

INTRODUCTION

According to the acoustic theory of speech production, the fundamental resonance of the cavity next to the mouth opening, the "front cavity resonance," may be associated with any of the first four formants (Fant, 1960, p. 72). But, as tongue constriction is relaxed, there is less dependence of any formant on one subpart of the vocal system, so little emphasis has been placed on cavity affiliations when describing the speech signal.¹ Instead, the description of acoustic cues for place of articulation remains largely in terms of numbered formants, with particular emphasis on F_2 .

It is of interest, therefore, that the spectrographic data presented below suggest that it may be possible to estimate the front cavity resonance frequency from information in the speech signal. As a result, it appears that a more articulatory description of the acoustic cues can be provided, and that several anomalous results of experiments on acoustic cues can be explained.

The spectrographic data come from analysis of two types of speech. The first type is normal speech, and the second type is speech produced with a fricated source, or "fricative speech." In fricative speech, palatal frication is substituted for laryngeal voicing, and the nasal port is kept closed. The position of the palatal frication adjusts with the articulation until it feels more nearly velar in backed environments. It should be noted that the frication constriction is maintained even for speech sounds that are not normally characterized by significant constriction of the vocal tract (e.g., central vowels). Two interesting properties of fricative speech are, first, that it seems highly intelligible, and second, that the fundamental resonance of the front cavity appears as a prominent spectral component.² The acoustic similarities between fricative and normal speech suggest that a front cavity resonance frequency estimate can be made for normal speech.

1. ON THE POSSIBILITY OF ESTIMATING THE FRONT CAVITY RESONANCE FREQUENCY

Figure 1 shows spectrographic analyses of the phrase "Where were you a year ago?" spoken under two condi-

tions of excitation: fricative speech (top) and normal speech (bottom). Visual inspection of the top spectrogram indicates the presence of two components in the fricative speech token. The most obvious component varies in frequency from 700 to 3000 Hz and is visible in all excited portions of the token. Another component is fixed above 3500 Hz and is less visible when lip-rounding increases. While the fixed component may be due to the fricative constriction, the variable component can be interpreted to be the fundamental, quarter-wave resonance of the front cavity. The variations in front cavity resonance frequency appear to reflect changes in the position of fricative constriction (from velar to prepalatal) and changes in lip opening (from rounded to retracted). Using the formula $l = c/4f$, and setting $c = 353$ m/sec (for 35 °C), a quarter-wave resonance at 700 Hz would indicate that the front cavity has a functional length of about 12.6 cm; at 3000 Hz, a length of about 2.9 cm.

It comes as no surprise that the front cavity resonance should vary so continuously in fricative speech, since tongue constriction is extreme. What is interesting, however, is that this resonance can be traced so easily in the normal speech token. A comparison of the two spectrograms shows that this is the case. The comparison also illustrates the point that the fundamental resonance of the front cavity cannot always be associated with the same numbered formant: it may be associated with F_2 in /*ɜ/* and /*u/*, but it is more strongly associated with F_3 in /*i/*.³

Figure 2 shows spectral cross sections of eight vowels, all spoken by the same adult male. There are two sections per vowel, one each from fricative speech (left) and normal speech (right).

It may not be inappropriate, at this point, to insert a comment about the ease of production of these fricative speech vowels. Fant (1960, p. 115) reports vocal-tract cross-sectional areas for /*ieaou/*. In the region of the tongue constriction, the cross-sectional area appears to fall to 1 cm² or less for /*iaou/*, but to no less than 2 cm² for /*e/*. Similarly, it seems easy to make the constriction for a satisfactory fricative speech close

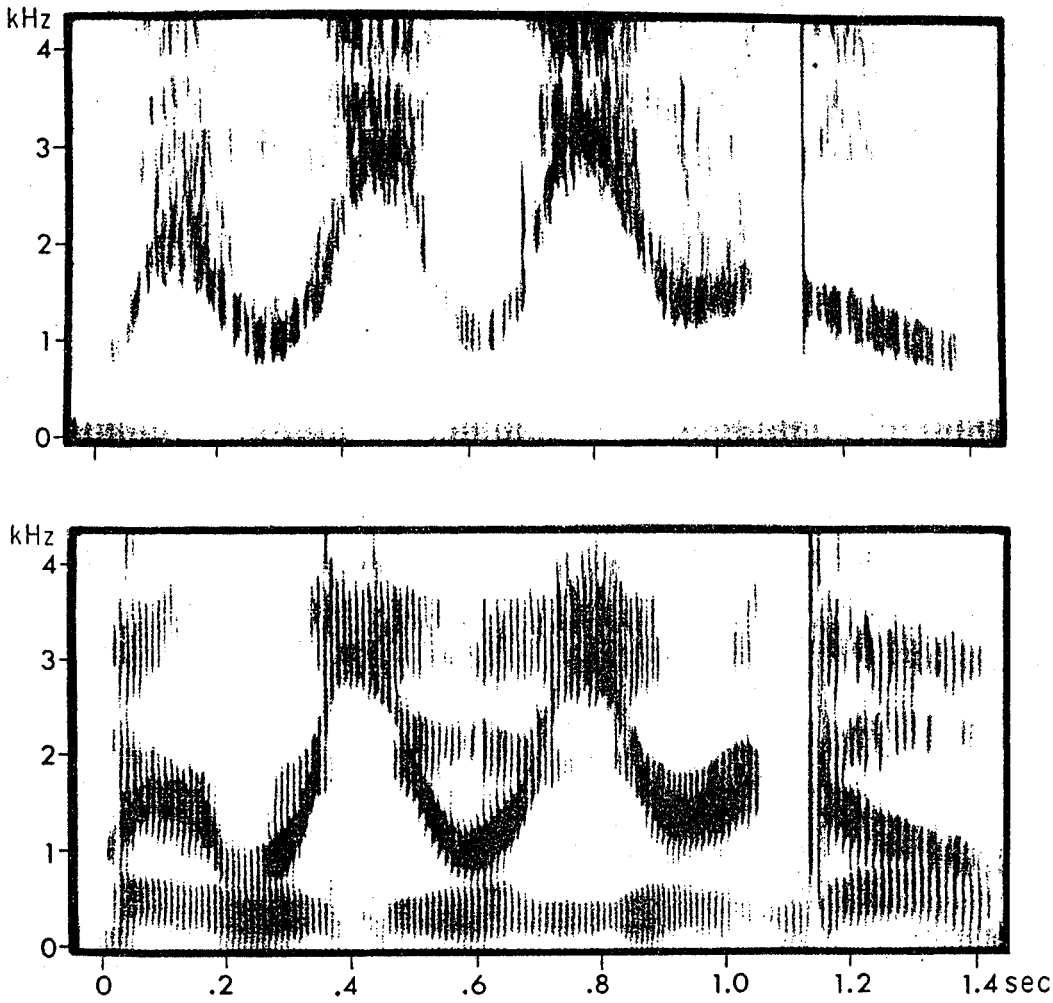


FIG. 1. Spectrographic comparison of the phrase "Where were you a year ago?" for two conditions of excitation: fricative speech (top) and normal speech (bottom).

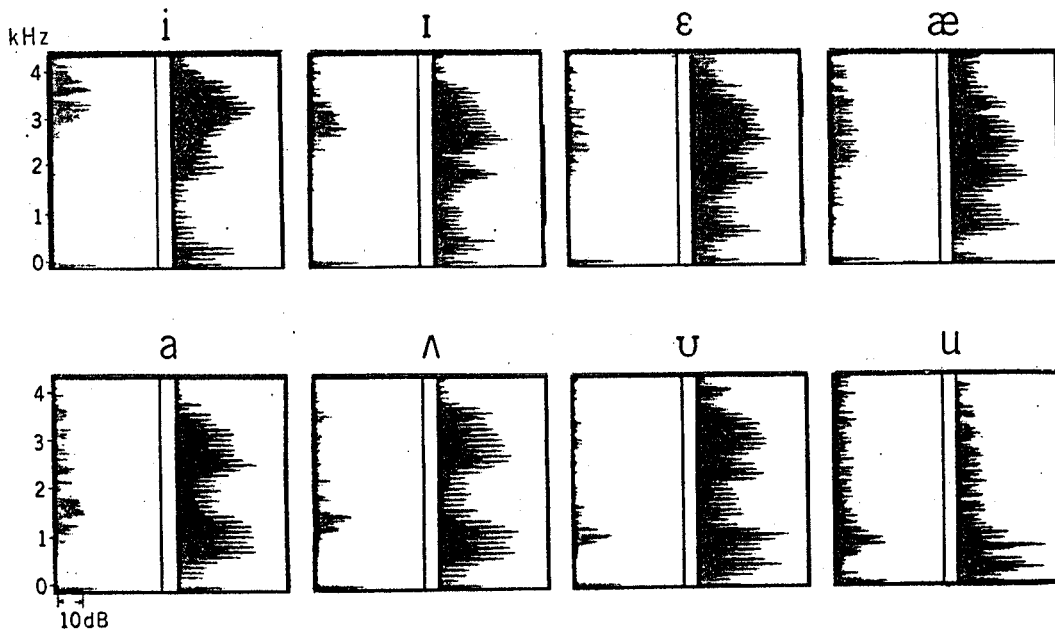


FIG. 2. Spectral cross sections of eight vowels. There are two sections per vowel, one each from fricative speech (left) and normal speech (right).

front vowel (here, /i/ and /I/). It also seems easy to make the constriction for the vowels with a backed tongue position (/aΔUu/), where we were more aware of manipulating the lip opening when trying to adjust the perceived color. However, it seems less easy to lower the jaw and produce convincing fronted palatal constriction for the more open front vowels /ε/ and /æ/.

These cross sections give further indication that a front cavity resonance frequency estimate can be made for normal speech. In these sections, the length of the front cavity seems to have an important effect on the overall spectral shape. The fricative and normal speech spectra seem to be shaped toward the high frequencies when the front cavity is short, as for /i/, and toward progressively lower frequencies as the front cavity is apparently lengthened for each successive vowel. In addition to the effect of the length of the front cavity, there also seems to be an effect due to the amount of tongue constriction involved. The greater the constriction, the more the front cavity resonance in the fricative speech seems to correspond to a formant in the normal speech. This correspondence seems very close for F_3 of /i/ and /I/, and for F_2 of /aΔUu/. For all eight vowels, however, the front cavity seems to be associated with what is perhaps the most intense group of formants: with the F_3 group for /iIεæ/, and with the F_2 group for /aΔUu/.⁴ Notice in particular the change in overall spectral shape from /æ/ to /a/, where the front cavity shifts its strong association from F_3 to F_2 and the weight of the spectrum shifts to frequencies below 2000 Hz.⁵ This change occurs despite the fact that frequencies of F_1 , F_3 , and F_4 are essentially unchanged. These comparisons with fricative speech seem to lead us to an observation about spectral shape that is substantially the same as that made by Fant (1960, p. 123), namely, that the front cavity can have an important effect on F_2 (and thus on the mean of F_1 and F_2), or on the mean of F_2 and all higher formants.

Figure 3 shows spectrograms of 12 consonant-vowel syllables, the consonants /bdg/ followed by the vowels /iæa u/. There are two spectrograms per syllable, one each from fricative speech (left) and normal speech (right). These spectrograms indicate that a front cavity resonance frequency estimate can be made for highly constricted normal speech consonants. They show the remarkable similarity of burst and transition information in fricative and normal speech. Notice again the shift in spectral weight toward the lower frequencies, this time as the vowel goes from /æ/ to /a/.

These observations suggest a general effect of the front cavity: that it is a determiner of the overall spectral shape. Nevertheless, it appears possible to construct a formula to estimate the front cavity resonance frequency from formant frequency data. For constricted vowels, this formula should place the front cavity resonance frequency estimate somewhere between the low values for F_2 as in back vowels and the high values of F_3 as in front unrounded vowels like /I/.

Carlson *et al.* (1973) have expressed exactly these concerns in designing a formula for predicting a per-

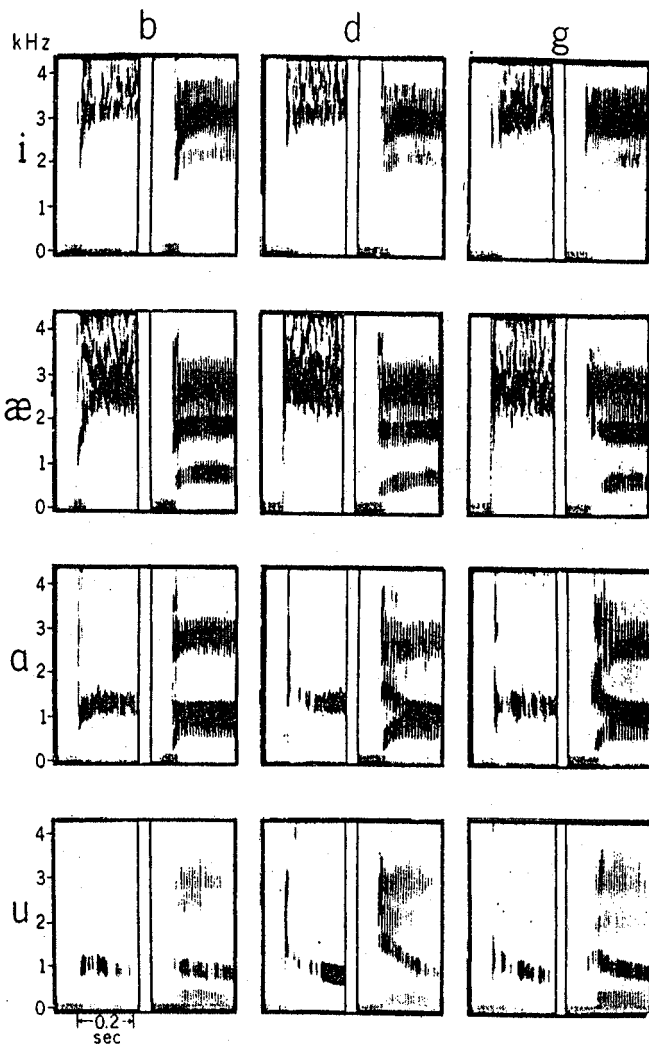


FIG. 3. Spectrograms of 12 CV syllables, the consonants /bdg/ followed by the vowels /iæa u/. There are two spectrograms per syllable, one each from fricative speech (left) and normal speech (right).

ceptual " F_2 prime" for vowels. The notion of F_2' arises from a desire to represent natural vowels in a perceptually equivalent two-formant space (see, e.g., Delattre *et al.*, 1951; Fant, 1959). The F_1 of the natural vowel is replaced by the F_1 of the two-formant equivalent, while all higher formants of the natural vowel are replaced by the F_2 (the so-called F_2') of the two-formant equivalent. From the data of a matching experiment in which techniques for two-formant, parallel resonance synthesis were used, Carlson *et al.* (1970) report values of F_2' for several Swedish vowels. The matching experiment values of F_2' range from about 700 Hz for /u/ to about 3000 Hz for /I/. These limiting values, and the other, intermediate values reported, appear to lie close to the front cavity resonance frequency as estimated from fricative speech. The thought arises, then, that the front cavity resonance frequency may be what F_2' predicts. If this is so, then it might be appropriate to estimate the front cavity resonance frequency using the formula proposed by Carlson *et al.* (1973). That formula is

$$F_2' = \frac{F_2 + c(F_3 F_4)^{1/2}}{1 + c},$$

where

$$c = \left(\frac{F_1}{500}\right)^2 \left(\frac{F_2 - F_1}{F_4 - F_3}\right)^4 \left(\frac{F_3 - F_2}{F_3 - F_1}\right)^2.$$

The formula apparently generates the results of the matching experiment to within 65 Hz, on the average. Carlson *et al.* (1973) report that the values of F_2' predicted by the formula are also within 75 Hz, on the average, of values predicted by a model of the cochlea. When the reference vowels from the matching experiment were the input to their cochlear model, then the two most prominent peaks in the output "were found to correspond closely to F_1 and the F_2' of the two-formant matching." Thus, there is some rather indirect evidence that the front cavity resonance frequency could also be estimated from speech data that is in a form perhaps more like that found in the auditory system. The authors attribute the close agreement between the formant equation and those of the cochlear model, at least in part, to "single component prominence." This explanation appears to be consistent with an emphasis on the front cavity as a determiner of the overall spectral shape.

The methods for predicting F_2' suggest how the front cavity resonance might possibly be estimated for vocalic sounds. This estimate might be expected to be more consistent for the more constricted sounds, where the formant frequencies can move more continuously. In return, the front cavity resonance may provide an articulatory rationale for F_2' , which, heretofore, has been motivated mainly by perceptual considerations.

II. A REINTERPRETATION OF CUES FROM F_2 AND F_3

Since it appears that the front cavity resonance could be estimated from information that is intense in the speech signal, one may ask for indications that this happens, in fact, during speech perception. What follows is an attempt to reinterpret familiar data from studies of the perception of synthetic speech, in a fashion consistent with a possible role for the front cavity resonance.

The front cavity resonance appears to play a role in vowel perception. Single-formant equivalents of two-formant vowels have been reported for /i u o ə a/ (Delattre *et al.*, 1952). These single-formant equivalents lie close to the frequency of the front cavity resonance as estimated from fricative speech. In two-formant vowel synthesis, weighted averaging of the natural F_2 and F_3 has been used for front vowels, where the front cavity resonance in the natural case may be more strongly associated with F_3 . For example, the two-formant /i/ of Delattre *et al.* (1952) had an F_2 at 2880 Hz, and that of Liberman *et al.* (1954) had an F_2 at 2760 Hz, whereas the natural F_2 appears to be located at about 2300 Hz and the natural F_3 at about 3000 Hz (Peterson and Barney, 1952). Again, Carlson *et al.* (1970, 1973) are investigating a perceptual F_2' that, for both front and back vowels, may track the front cavity resonance.

The front cavity resonance appears to play a role in the perception of stop consonant formant transitions. The front cavity resonance in fricative speech is close to F_2 in /a/, and Liberman *et al.* (1954) found that changes in the F_2 transitions alone were sufficient to produce /ba/, /da/, and /ga/ responses. But the front cavity resonance is close to F_3 in /i/, and Harris *et al.* (1958) produced /bi/, /di/, and /gi/ responses by changing the F_3 transitions alone.⁶

Finally, the front cavity resonance appears to play a role in the perception of stop consonant bursts. Only the voiceless stop bursts are mentioned here, because of the relevance of the cited synthesis results. The descriptions appear to be applicable to the homorganic voiced stops.

For /t/ bursts, the cavity behind the mouth opening is small, extending back only to the alveolar constriction, regardless of the resonator configuration for a following vowel. Synthesis should therefore reveal the importance of a high frequency and relatively unchanging spectral component. It does: Liberman *et al.* (1952) obtained /t/ responses for bursts above 3000 Hz before the vowels /i e ε a o u/.

For /p/ bursts, the spectrum immediately following lip release can be broad and flat, because there is no resonator of significance in front of the constriction. Then, as the lips open further, the front cavity resonance can rise abruptly in frequency and amplitude. Before unrounded vowels these excursions may be quite salient, but before rounded vowels, if the lip opening remains small, they would be diminished. (Compare the spectrograms for /bi/ and /bu/ in Fig. 3). In synthesis, the excursions of the front cavity resonance might therefore be expected to play a more important role before unrounded vowels than before rounded ones. Indeed, Liberman *et al.* (1952) found that /p/ responses dominated when a schematic burst was positioned some 360 Hz below the formant closest in frequency to the front cavity resonance in a following /i e ε a/. But before /o u/, they found /p/ responses to dominate when the burst was neither near the frequency of the front cavity resonance nor in the /t/ region, but rather around 1500 Hz.

For /k/ bursts, the cavity behind the mouth opening extends back to the hump of the tongue, so that a front cavity resonance component of the burst should be affected at once by concomitant positioning of the tongue hump for a following vowel. The question is whether synthesis reveals a strong dependence of the burst upon the frequency of the formant closest in frequency to the front cavity resonance of a following vowel. In fact, the data of Liberman *et al.* (1952) show that /k/ responses predominated when a schematic burst was placed at, or slightly above, the formant closest in frequency to the front cavity resonance in a following /i e ε a o u/.

III. AN EXPLANATION OF ANOMALIES

Stevens and House (1956) suggested that some anomalies encountered in perceptual studies of transitional cues may be attributed to the changing cavity affiliations

of F_2 and F_3 . Since the possibility of explaining anomalies is at least as compelling as that of reinterpreting phenomena, it is interesting to note that if the role of the front cavity resonance is emphasized in describing the speech signal, several anomalies of the acoustic phonetic literature seem to find an explanation. Consider the explanations of the following three anomalies in terms of a possible role for the front cavity resonance.

One anomaly is the burst of noise at 1440 Hz that cued a /pi/, /ka/, or /pu/ response in Liberman *et al.* (1952). Before /i/ this burst appears to be interpreted as part of the rise in frequency of the front cavity resonance as it moves up to F_3 . Before /a/, the burst appears to be interpreted as part of the fall in frequency of the front cavity resonance as it moves to a slightly lower value in F_2 . Before /u/ it appears to be interpreted as part of a flat, lip-release spectrum, and was a somewhat weaker cue. The /pi/-/ka/-/pu/ result is then consistent with the suggestion that the front cavity resonance plays a role in the perception of /p/ and /k/.

A second anomaly is the F_3 transition that was important for the /d/ in /di/ but not for the /d/ in /du/ (Harris *et al.*, 1958). This result may be due to the fact that the fundamental resonance of the front cavity is strongly associated with F_3 of /i/, but not with F_3 of /u/. If so, the /di/-/du/ result suggests that transitions of the front cavity resonance can play a role in the perception of /d/.

A third anomaly is the F_2 transitions in two-formant synthesis of /g/: one could extrapolate the F_2 transitions to a virtual /g/ locus at 3000 Hz before /ie ε a/, but before /o u/ the locus would have to be much lower in frequency, if indeed it existed at all (Liberman, 1957). Figure 3, above, indicates that, like the velar bursts, velar transitions covary in frequency with the front cavity resonance of the following vowel. Indeed, the transitions that produced predominantly /g/ responses in Liberman *et al.* (1954) lie at the same frequency as bursts that produced predominantly /k/ responses in Liberman *et al.* (1952). These results suggest that the real, relative frequency of the transition of the front cavity resonance plays a role in the unchanging perception of the velar stop consonants.

IV. DISCUSSION

Acoustic anomalies like those above led Liberman *et al.* (1967) to express the belief that speech perception might involve a simplifying reference to articulation. The data and arguments of this paper suggest that such a simplifying reference may be available directly from the speech signal: despite the acoustic complexity of the anomalies mentioned, an interpretation in terms of the front cavity resonance seems to provide, in each case, a rational account.

One can try to show that an articulatory reference is available in the speech signal without arguing that this reference is interpreted by a process of analysis by synthesis. The task of synthesizing an acoustic pattern to subtract from the incoming signal now appears sim-

pler: the rules required to generate those curious speech acoustics do not seem anomalous when expressed in terms of the resonator system that produces them. But, at the same time, the task of directly perceiving the incoming signal appears simpler too: there appears to be an intense component of the signal that carries important information about place of articulation.

These observations suggest that a person who is perceiving speech might be described as one who is interpreting at least part of the signal as a contribution specifically of the front cavity. Given the quarter-wave resonator model, a front cavity resonance frequency estimate is also an estimate of the front cavity length. And for a given articulation, the front cavity length may not vary a great deal across individuals, not, for example, as much as the length of the pharyngeal cavity (Fant, 1966). Therefore, a front cavity resonance frequency estimate would be almost an estimate of place of articulation. It is necessary to say "almost" an estimate of place of articulation for at least two reasons: first, because of possible differences in front cavity length; and second, because similar lengths of the front cavity could arise in different combinations of fronted tongue constriction with lip rounding, or backed constriction without rounding. This last consideration indicates a possibly important use for continuous tracking of the front cavity resonance: spectra that are articulatorily ambiguous might be disambiguated if the preceding or following configuration of the slowly changing resonator system is unambiguous.

V. CONCLUSION

We have attempted to present a new technique (fricative speech) and articulatory rationalizations of some of the acoustic cues for speech. These have been used to emphasize a relationship which seems to deserve more attention, namely, the relationship between the fundamental resonance of the front cavity and the perceived place of articulation. This relationship would tend to arise to the extent that speech requires significant constriction of the vocal tract, as may be the case for consonants generally, and for many (though not all) vowels. We believe that such constriction contributes to the solution of the problem of deriving an articulatory description from the acoustics of speech. A front cavity resonance frequency estimate seems to be a useful way to represent part of that contribution.

ACKNOWLEDGMENTS

F. S. Cooper, C. G. M. Fant, O. Fujimura, M. Studdert-Kennedy, A. M. Liberman, R. McGuire, P. Mermelstein, K. N. Stevens, and the referee offered many helpful, substantive criticisms while this paper was in various stages of preparation. S. Koroluk and A. McKeon prepared the final manuscript and figures.

¹However, for a discussion of the effect of isolated articulatory movements on formant positions, see Delattre (1951).

²We know of no reference to fricative speech in the acoustic phonetics literature. However, for discussion relevant to fricative speech, see Fant (1960, p. 72). There it is sug-

gested that a static, three section model of the vocal-tract can show some of the essentials of velar and palatal articulation. Specifically, for the model of the articulation of /k/ or /g/ before /a/, /æ/, or /i/, it is suggested that the fundamental resonance of the front cavity can be associated with F_2 , F_3 , or F_4 , respectively.

³Sometimes the association of the front cavity with F_4 of /i/ is mentioned (Fant, 1960, see preceding footnote), sometimes its association with F_3 of the same vowel (Fant and Pauli, 1974). What appears to have been the emphasis of the earlier discussion, and what we attempt to emphasize again here, is not the affiliation of the front cavity with a given formant, but the ability of the front cavity resonance to move more or less continuously in frequency given significant vocal-tract constriction.

⁴Such phrases as "most obvious component" or "perhaps the most intense group of formants" should be accepted only with qualification. Figures 1-3 show speech spectra after lift has been applied (approximately 6 dB per octave between 300 and 3000 Hz). Also, in Figs. 1 and 3, automatic gain control and 300-Hz "broad-band" filtering have been applied. These operations have been made available on commercial sound spectrographs because they have been thought helpful for revealing perceptually relevant aspects of speech. This is not enough, of course, to make us want to assume that such operations make speech spectrograms look exactly like speech sounds.

⁵The front cavity resonance in fricative speech appears to be most clearly associated with F_3 of /i/ ϵ æ / and with F_2 of /a Δ α U υ u/. This association is consistent with the nomograms of Fig. 1.4-9 of Fant (1960), where the cavity affiliations of F_2 and F_3 appear to change at about 2000 Hz. For the model, this change has a constriction coordinate of approximately 11 cm from the glottis, which, in turn, is consistent with the estimate of "two-thirds of the total length of the vocal-tract" of Stevens and House (1956).

⁶This is our interpretation of their results. Harris *et al.* (1958) showed that a flat F_2 and different rising transitions of F_3 could cue /g/ and /d/ responses. They also showed that a sharp rise in both F_2 and F_3 could cue a /b/ response. We are interpreting this last case to be equivalent to an F_3 transition that starts below a flat F_2 .

Carlson, R., Granstrom, B., and Fant, G. (1970). "Some studies concerning perception of isolated vowels," STL-QPSR 2-3, 19-35.

Carlson, R., Fant, G., and Granstrom, B. (1973). "Two-formant models, pitch, and vowel perception," paper discussed at the Symp. Aud. Anal. Percept. Speech, Leningrad (21-24 Aug.).

Delattre, P. (1951). "The physiological interpretation of sound spectrograms," PMLA 66, 864-875.

Delattre, P., Liberman, A. M., and Cooper, F. S. (1951). "Two formant synthetic vowels and cardinal vowels," *Le Maitre Phonétique* (July-Dec.).

Delattre, P., Liberman, A. M., Cooper, F. S., and Gerstman, L. J. (1952). "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.

Fant, C. G. M. (1959). "Acoustic analysis and synthesis of speech with applications to Swedish," *Ericsson Tech.* 1, 3-108.

Fant, C. G. M. (1970). *Acoustic theory of speech production* (Mouton, The Hague), 2nd ed.

Fant, C. G. M. (1966). "A note on vocal tract size factors and nonuniform F -pattern scalings," STL-QPSR 4.

Fant, C. G. M., and Pauli, S. (1974). "Spatial characteristics on vocal-tract resonance modes," preprints of the Speech Commun. Sem., Stockholm 1, 121-132.

Harris, K. S., Hoffman, H., Liberman, A. M., Delattre, P., and Cooper, F. S. (1958). "Effect of third-formant transitions on the perception of the voiced stop consonants," *J. Acoust. Soc. Am.* 30, 122-126.

Liberman, A. M., Delattre, P., and Cooper, F. S. (1952). "The role of selected stimulus-variables in the perception of the unvoiced stop consonants," *Am. J. Psychol.* 65, 497-516.

Liberman, A. M., Delattre, P., Cooper, F. S., and Gerstman, L. J. (1954). "The role of consonant-vowel transitions in the perception of the stop and nasal consonants," *Psychol. Monogr.* 68, 1-13.

Liberman, A. M., (1957). "Some results of research on speech perception," *J. Acoust. Soc. Am.* 29, 117-123.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., and Studdert-Kennedy, M. (1967). "Perception of the speech code," *Psychol. Rev.* 74, 431-461.

Peterson, G. E., and Barney, H. L. (1952). "Control methods used in a study of the vowels," *J. Acoust. Soc. Am.* 24, 175-184.

Stevens, K. N., and House, A. S. (1956). "Studies of formant transitions using a vocal-tract analog," *J. Acoust. Soc. Am.* 28, 578-585.