COMMUNICATION SCIENCES
AND
ENGINEERING

XII. SPEECH COMMUNICATION*

Academic and Research Staff

Prof. K. N. Stevens	Prof. D. H. Klatt	Dr. Margaret Bullowa
Prof. M. Halle	Dr. A. W. F. Huggins	Dr. Paula Menyuk
Prof. W. L. Henke		Dr. H. Suzuki†

Graduate Students

Kay Atkinson	M. F. Medress	R. M. Sachs
A. J. Goldberg	J. E. Richards	J. J. Wolf

A. ON THE FEATURE "ADVANCED TONGUE ROOT"

The traditional features that are used to describe different vowel qualities specify tongue position in terms of tongue height and backness, together with lip rounding. This description is adequate for classifying the vowels in many languages, but it does not account for vowel systems which are said to have tense and lax vowel classes. Of particular interest with regard to the feature tense-lax are those West African languages that display vowel harmony based on this opposition. The feature is also used to describe the oppositions /i - ɪ /, /u - ʊ /, and others in English. In their formulation of phonetic features, Chomsky and Halle[1] tentatively make a distinction between the feature tense-lax, which applies, for example, in English, and a feature covered-uncovered, which applies to the West African vowels that exhibit harmony.

We would like to re-examine the hypothesis first suggested by Melville Bell that the lower pharynx or tongue root plays a decisive role in the tense-lax distinction. Moreover, as the recent work of Stewart[2] has suggested, the tongue root plays also the same basic role in the African languages that have the characteristic type of vowel harmony. It appears, therefore, that the features tense-lax and covered-uncovered have in common one and the same phonetic mechanism and should, therefore, be regarded as a single feature in the phonetic framework. This conclusion was already arrived at on other grounds by Jakobson (cf. Jakobson and Halle[3]), and the remarks that follow can be read as providing evidence in support of Jakobson's conception of the nature of the tense-lax feature.

Basing himself in part on radiograms published by Ladefoged,[4] Stewart has noted that in languages like Igbo the two classes of harmonizing vowels are distinguished by movements of the root of the tongue in the vicinity of the lower pharynx, the epiglottis, and the hyoid bone. As shown in Fig. XII-1, the phonetic difference between the so-called tense-lax pairs of English vowels (/i/ ~ /ɪ/; /u/ ~ /ʊ/) is characterized by a similar

*This work was supported in part by the U.S. Air Force Cambridge Research Laboratories, Office of Aerospace Research, under Contract F19628-69-C-0044; and in part by the National Institutes of Health (Grant 2 ROl NB-04332-06).

†On leave from Tohoku University, Sendai, Japan.

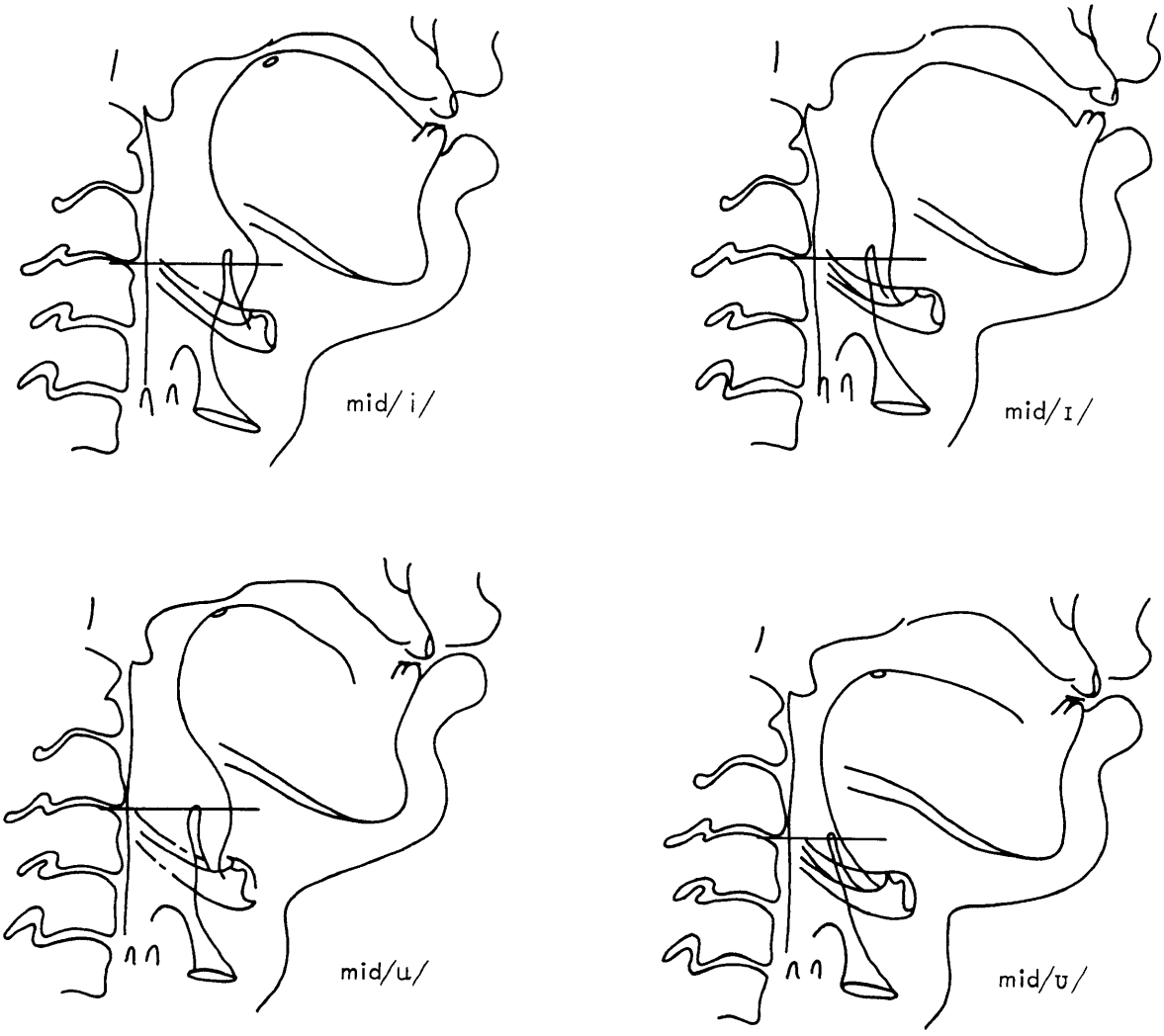

Fig. XII-1. Midsagittal sections (obtained from lateral cineradiographs) for 4 American English vowels. The upper pair illustrates the contrast in tongue-root position for the vowels /i/ and /ɪ/, and the lower pair illustrates the same contrast for /u/ and /ʊ/. The horizontal line drawn between the third and fourth cervical vertebrae indicates the region where advancing the tongue root has the greatest effect on vocal-tract shape.

distinction in the tongue-root position. The tense vowel of each pair has a much wider cavity in the vicinity of the hyoid bone and lower pharynx than does the corresponding lax vowel, and the root of the tongue has a concave shape. For the tense vowel, the tongue is somewhat higher, presumably as a consequence of drawing the root forward, tending to bunch the body of the tongue, and thereby to raise it.

The clearest and most consistent acoustic consequence of widening the vocal tract in the vicinity of the tongue root is a lowering of the first-formant frequency. This shift in F_1 can be predicted on a theoretical basis and can also be observed in the acoustic data. Independent of the particular vowel configuration there is always a maximum in the sound-pressure distribution in the vicinity of the glottis for all natural frequencies. We recall that expansion of the cross-sectional area of an acoustic tube in the vicinity of a maximum in the sound-pressure distribution in the standing wave for a particular natural frequency tends to lower that natural frequency (Chiba and Kajiyama[5]). The region in the vocal tract over which an expansion in cross-sectional area occurs appears to be centered 2-4 cm above the glottis. The maximum in sound-pressure distribution for the first formant always extends over at least the lowest 4 cm of the vocal-tract length, and hence expansion of the vocal tract in this region always causes a lowering of F_1. Lowering of the glottis, which often accompanies tongue-root advancing, would tend to accentuate further the downward shift in F_1.

An expansion of the vocal-tract cross-sectional area in the vicinity of the tongue root can also cause a change in the second-formant frequency. The region over which the maximum in the sound-pressure distribution near the glottis extends for the second-formant frequency is roughly 4 cm for back vowels and 2 cm for front vowels. In the case of front vowels (particularly high front vowels), there is a pressure minimum in the region 2-6 cm above the glottis. Consequently, one would expect tongue-root advancing to cause a downward shift of F_2 for back vowels and an upward shift for front vowels.

These changes in the first- and second-formant frequencies are in the direction that is observed in acoustic data for the pairs /i - ɪ/ and /u - ʊ/ in English (Peterson and Barney[6]). The shifts in F_1 are also consistent with formant-frequency measurements on tense-lax pairs of Igbo vowels reported by Ladefoged,[4] but Ladefoged's data on F_2 do not show the expected downward shift for the back vowels with tongue-root advancing.

It is probable that the physiological activity that gives rise to tongue-root advancing is contraction of the mylohyoid muscle and of the geniohyoid and lower fibers of the genioglossus muscles. As Hockett has pointed out, the "tense" vowels in English are characterized by a tension of muscles "above and in front of the glottis within the frame of the lower jaw".[7] Stewart notes that these same muscles are "pushed markedly downwards for ... the raised vowels [of Twi] and for none of the unraised vowels".[8]

In the case of the high vowels, advancing of the tongue root creates a maximally large

cavity volume posterior to the major vowel constriction. As has been noted, this increased cavity volume causes the first-formant frequency to become low. When the cavity volume is made larger, F_1 becomes less sensitive to linear changes in cavity volume and constriction size. Furthermore, for a large cavity volume the wall impedance places a lower limit on F_1 independent of the size of the cavity or the narrowness of the constriction (Fant and Sonesson[9]), and, as a consequence, the sensitivity of F_1 to changes in vocal-tract shape is further reduced. Thus high vowels that are generated with an advanced tongue root have the desirable property that F_1 is relatively insensitive to perturbations in articulatory shape (Stevens[10]), i.e., the demands on precision of articulation are not stringent.

When the first-formant frequency becomes low, vocal-tract excitation by glottal pulses causes large fluctuations in sound pressure immediately above the glottis. These fluctuations give rise to marked interaction between the vibrating vocal cords and the supraglottal system (Halle and Stevens[11]), and hence to irregular glottal vibrations. The amount of this interaction can be minimized by increasing the cavity volume immediately above the glottis; it can be shown that the sound-pressure fluctuations in this region are inversely proportional to the cavity size. It can be argued, therefore, that for high vowels tongue-root advancing is essential if the first-formant frequency is to be as low as possible.

Such considerations lead to the conclusion that unmarked or "natural" high vowels are produced with tongue-root advancing, in much the same way that the unmarked non-low back vowels are rounded. One would assume that unmarked low vowels do not have tongue-root advancing, since they are characterized by a maximally high F_1. The unmarked versions of midvowels probably also do not have tongue-root advancing, but this point needs further examination. In many languages, therefore, tongue-root advancing does not operate independently, but is concomitant with tongue height. For other languages, particularly the West African languages displaying vowel harmony, tongue-root advancing is clearly a feature that distinguishes between vowel pairs. It might be expected that languages in which both rounding and tongue-root advancing play independent roles are either rare or nonexistent. The acoustic consequences of both of these articulatory activities are similar (at least as far as the effect on F_1 is concerned), although they are not identical.

Many investigators have reported a dull or even breathy character for the vowels with advanced tongue root (cf. Sapir[12]). There appears to be a shift in the mode of vibration of the vocal cords for these vowels: the waveform of glottal vibration becomes less rich in high frequencies, presumably as a consequence of a broadening and smoothing out of the glottal pulses. Figure XII-2 shows sketches of the kinds of glottal waveforms that might be associated with vowels characterized by advanced tongue root as opposed to vowels that do not possess this feature. The acoustic effect of this change in the

waveform of the glottal pulses can often be observed in tense-lax pairs of vowels in English. Shown in Fig. XII-3 are acoustic spectra sampled within the vowels /i/ and /ɪ/

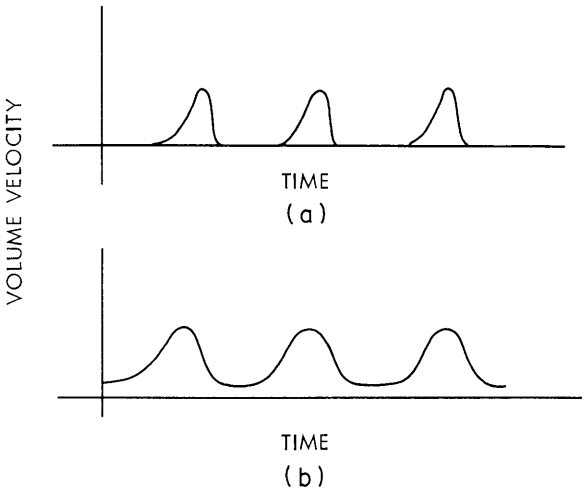

Fig. XII-2. Schematized representation of pulses of glottal volume velocity (a) for clear vowels, and (b) for breathy vowels. It is hypothesized that vowels characterized by advanced tongue root often have the breathy property.

uttered in the context /b - b/. The spectra were obtained with a filter bank having relatively wide filters (bandwidth 360 Hz). The spectra show that the high-frequency part of the spectrum, in the vicinity of the second and third formants, is considerably lower in

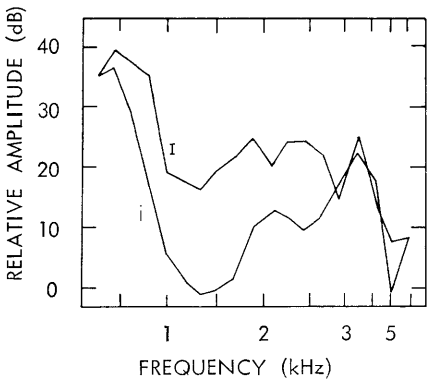

Fig. XII-3.

Acoustic spectra sampled within the vowels /i/ and /ɪ/. The lax vowel has relatively more high-frequency energy in the F_2-F_3 region than does the tense vowel.

energy for the [i] than for the [ɪ].[13] Some decrease in high-frequency energy would be expected in the tense vowel as a consequence of the lower F_1, but the reduction in the high frequencies is greater than would be predicted on the basis of the drop of 100-odd Hz in F_1.

Advancing of the tongue root appears, then, to have two acoustic consequences. The

first, and probably primary, effect is a downward shift in the first formant. This lowering of F_1 causes a decrease in high-frequency energy, simply as a consequence of the normal relationship between formant frequencies and formant amplitudes (Fant[14]). A second, and probably incidental, effect of tongue-root advancing is to broaden the glottal pulse, which results in an additional drop in high-frequency energy. It is significant that both of these effects – the downward shift in F_1, and the decrease in high-frequency energy of the glottal pulse – cause a reduction in the amount of high-frequency energy in the spectrum of the vowel, and consequently give a "dull" quality to the vowel. That is, both effects produce a shift in vowel quality that is in the same direction.[15]

One can at present only speculate on the anatomical and physiological mechanisms that give rise to the change in glottal waveform when the tongue root is advanced. There is a direct connection consisting of ligament and muscle tissue from the lateral edges of the epiglottis to the arytenoid cartilages, which are, of course, responsible for positioning the vocal folds. There are also ligamentous connections between the hyoid bone and the epiglottis. Contraction of the mylohyoid and other muscles causes a forward movement of the hyoid bone, to which the root of the tongue is anchored. Through the ligamentous connections the epiglottis is pulled forward, and this motion in turn may give rise to a lateral displacement of the arytenoid cartilages. This movement would tend to position the vocal cords less tightly together, possibly leaving a small chink between the cords in the static condition when there is no air flow. In other words, the lateral displacement of the arytenoids causes a static force that tends to keep the vocal cords slightly apart. Thus under conditions of air flow, the vocal cords come together less rapidly during the adduction phase of the vibratory cycle. The vocal cords may, in fact, not become approximated at any time during the vibratory cycle, thereby giving rise to a waveform like that shown in Fig. XII-2b, with the waveform displaced upward slightly from the baseline, even during maximum vocal-cord adduction.

Up to this point we have considered the effects of tongue-root advancing on non-consonantal articulations, i.e., on articulations that do not include a radical obstruction in the midsagittal region of the vocal tract. We must now inquire into the effects that tongue-root advancing will produce on consonantal articulations, i.e., on articulations that include a radical obstruction in the midsagittal region of the vocal tract. The principal acoustic effect here would be to lower the first-formant frequency of a vowel immediately preceding or following the consonant. The so-called heavy consonants of Javanese have this acoustic characteristic, and are presumably characterized by the feature advanced tongue root. The lower F_1, particularly in the region of the vowels adjacent to the consonant, can be clearly observed. The "slightly aspirated" Korean stop consonants may also have this feature, particularly in intervocalic position. The voicing that occurs through the stop gap of these consonants

when they appear in this position (Kim[16]) would be a result of the widened pharyngeal maneuver, which increases the vocal-tract volume behind the constriction and allows air to flow through the glottis, thereby causing vocal-cord vibration. At this point we are unable to attribute the increase in cavity volume observed along the entire pharynx in, e.g., English /d/ vs /t/ (Perkell[17]) to an advanced tongue root. It seems to us that a different mechanism may well be at work in these consonants. These questions with regard to the consonants are now being considered in greater detail, and we hope that future work will be able to establish more clearly the attributes of the feature advanced tongue root for consonantal articulations.

M. Halle, K. N. Stevens

References and Footnotes

1. N. Chomsky and M. Halle, The Sound Pattern of English (New York: Harper and Row, 1968).
2. J. M. Stewart, "Tongue Root Position in Akan Vowel Harmony," Phonetica 16, 185-204 (1967).
3. R. Jakobson and M. Halle, "Tenseness and Laxness," in D. Abercrombie et al. (eds.), In Honour of Daniel Jones (London: Longmans, Green and Co., 1964), pp. 96-101.
4. P. Ladefoged, A Phonetic Study of West African Languages (London: Cambridge University Press, 1964).
5. T. Chiba and M. Kajiyama, The Vowel: Its Nature and Structure (Tokyo: Tokyo-Kaiseikan Publishing Company, 1941).
6. G. E. Peterson and H. L. Barney, "Control Methods Used in a Study of the Vowels," J. Acoust. Soc. Am. 24, 175-184 (1952).
7. C.F. Hockett, A Course in Modern Linguistics (New York: Macmillan Co., 1958), see pp. 78-79.
8. J. M. Stewart, op. cit., p. 197.
9. C. G. M. Fant and B. Sonesson, "Speech and High Ambient Air-pressure," Speech Transmission Laboratory QPSR 2/1964, Royal Institute of Technology, Stockholm, pp. 9-21.
10. K. N. Stevens, "The Quantal Nature of Speech: Evidence from Articulatory-Acoustic Data," in E. E. David, Jr. and P. B. Denes (eds.), Human Communication: A Unified View (New York: McGraw-Hill Publishing Co., in press).
11. M. Halle and K. N. Stevens, "On the Mechanism of Glottal Vibration for Vowels and Consonants," Quarterly Progress Report No. 85, Research Laboratory of Electronics, M.I.T., April 15, 1967, pp. 267-271.
12. E. Sapir, "Notes on the Gweabo Language of Liberia," Language 7, 30-41 (1931).
13. The broadening of the glottal pulse for vowels or sonorant consonants with a low first formant is probably a necessary requirement if regular voicing is to be maintained (Halle and Stevens[11]).
14. C. G. M. Fant, "On the Predictability of Formant Levels and Spectrum Envelope from Formant Frequencies," in For Roman Jakobson (The Hague: Mouton and Co., 1956), pp. 109-120.
15. While the dull quality may be an incidental consequence of tongue-root advancing, it should be noted that a separate feature that provides a distinction between dull and clear vowels may operate in some languages. This feature may be actualized without concurrent adjustment of tongue-root position.
16. C.-W. Kim, "On the Autonomy of the Tensity Feature in Stop Classification (with Special Reference to Korean Stops)," Word 21, 339-359 (1965).
17. J. S. Perkell, Physiology of Speech Production: Results and Implications of a Quantitative Cineradiographic Study, Research Monograph No. 53 (Cambridge, Mass.: The M.I.T. Press, 1969).

B. ACOUSTIC MEASUREMENTS FOR SPEAKER RECOGNITION

The problem of speaker recognition, like most problems in pattern recognition, may be divided in 2 parts: measurement and classification. In the first part, the pattern under test (a voice signal from an unknown speaker, in this case) is subjected to measurements that result in a set of values characterizing the pattern. These values, in turn, act as inputs to a classification scheme, which compares them with stored information on previously labeled patterns and makes a decision about their class membership. This report concerns the importance and nature of the measurement phase.

Well-chosen characterizing measurements are important to pattern-recognition problems in several respects. First, they must adequately characterize the patterns under test. No amount of decision-making sophistication can compensate for a basic lack of information. Efficient measurements also lead to economy in storage of the data concerning the pattern classes and permit the use of simpler classification techniques. In speaker recognition, we should aim for acoustic measures that are related in as direct a manner as possible to the voice characteristics of the unknown speaker and are minimally affected by irrelevant factors.

As others have pointed out, the reasons why a person's voice has characteristic qualities that make it possible to recognize it stem from two fundamental bases: the structural characteristics of the individual vocal mechanism, and the habitual patterns of neural commands to the vocal tract muscles that are learned by each individual. We may thus understand the motivation for the measurements chosen in some prior speaker-recognition experiments such as the following.

1. Sampled intensity-frequency-time patterns over an utterance.[1, 2]
2. Formant frequencies and time derivatives over an utterance.[3, 4]
3. Pitch frequency and its time derivative over an utterance.[4, 5]
4. Long-time averaged spectra of voiced portions of an utterance.[1]

These are straightforward measurements encompassing aspects of the acoustic signal which are dependent on organic and learned differences among speakers. The measurements describe a point or a contour in an n-dimensional space. If the same linguistic content is used in tests, the differences in the points or contours measured for different utterances are due not only to speaker differences but also to other normal factors, such as emphasis, emotion, intentional modification and natural imprecision of articulation permitted by the phonology of the language. Because of the general nature of the measurements, it is essentially left to the classification scheme to separate the speaker-selective effects from the others.

We cannot deny the value and utility of statistical classification procedures, but there are limits to the complexity of the probability distributions that a given procedure can handle. If the results of speaker-recognition experiments are limited by the difficulties

in classification, then it would seem that progress will depend on devising more effective ways of characterizing the effects of speaker differences in the speech signal. More of the burden of the separation of speaker-dependent effects from irrelevant factors should be shifted from the classification process to the measurement process. The measurement phase should be selective and efficient, rather than merely systematic and sufficient.

Not every measurement will be significant in every part of an utterance. For example, low-frequency spectral data during /s̆/ and /s/ is irrelevant, so we need not clutter up the stored data and classifier computations with it. Rather, one could adapt the measurement strategy to suit the individual segments of an utterance.

If possible, measurements related to structural constraints should be closely related to or constrained by a specific part of the vocal tract, rather than be related to structure in only a general way. This would help to reduce the sources of variation in the measurement.

Independence of measurements is a desirable attribute. It can simplify statistical procedures, and it should lead to efficient representation in terms of information per measurement. For example, once you measure a short-time spectrum in the center of a tense vowel, another spectrum 10 or 20 msec later will contain little new information, yet it requires as much data storage as the first one. To be sure, temporal patterns should be important, but if so, why not measure them explicitly?

We have recorded under good conditions 10 repetitions of 6 short sentences from each of 21 American, adult, male speakers. The use of cooperative speakers and a prespecified context may be considered a simulation of a realistic speaker authentication situation, in which the speakers desire to be recognized.

Our principal analysis tools are a PDP-9 computer and a 36-channel filter bank covering 150-7000 Hz.[6] A system of programs for analysis and display of the recorded data has been written. Figure XII-4a is an example of the type of computer display that is used. The spectrum at the top is /m/ in the sentence, "I cannot remember it." The horizontal axis is frequency (linear to 1650 Hz, then logarithmic to 7 kHz), and the vertical axis is amplitude, in dB. The two graphs below are time functions of the utterance. The upper one is a measure of low-frequency energy (sum of certain filter outputs), useful as a syllable "map" of the utterance, and the lower one is pitch frequency during voiced portions, which is obtained by extracting the first harmonic with an analog low-pass filter and measuring the interval between zero crossings. (The figure "124" is the pitch value in Hz at the time of the spectrum shown.) The spectra are taken at 10-msec intervals. The vertical cursor shows the point in the utterance at which the displayed spectrum occurs.

We have been studying these data with respect to several parameters felt to be particularly suited to speaker recognition, and present our results thus far.

The nasal consonants should be suited to speaker recognition, because of the relatively fixed influence of the nasal cavity. Glenn and Kleiner,[7] for example, have used heavily averaged spectra of /n/ as measurements. The analysis and experiments by Fant[8] and by Fujimura[9] suggest that certain poles of the transfer functions of /m/ and /n/ are closely tied to resonances of the nasal cavity alone, and the interplay between mouth and nasal cavities can produce considerable variability in the 700-1600 Hz portion of the spectrum, depending on the nature of those cavities.

Figure XII-4a shows a clear example of an /m/ spectrum. The zero attributable to the mouth in this case has cancelled a pole between F_1 and 1 kHz, thereby leaving the pole at 1.3 kHz in the clear. The two upper formants occur around 2 and 3 kHz, and have been described as being roughly tied to nasal resonances.

Figure XII-4b illustrates the variation in /m/ spectra among speakers. In all of these multiple-spectrum display photographs, each row contains examples by a single speaker. The spectra were selected from the utterance manually, but according to definite rules. The top row is by the same speaker as on the previous slide, and shows the same characteristics.

The 2nd row shows a speaker whose F_2 is not as completely cancelled by the zero, thereby resulting in a small peak around 800 Hz.

The 3rd row shows a speaker whose F_2 and F_3 are both considerably cancelled by the effect of the zero, thereby resulting in a lack of consistent spectral peaks in that region.

The 4th row shows a speaker whose F_5 around 3 kHz is absent from the spectrum. This is probably due to the coincidence of F_5 and the second zero of the mouth cavity.

Figure XII-4c illustrates two speakers who make the job of characterizing nasal spectra even more difficult. Some speakers just do not seem to exhibit in their nasal spectra the invariability hoped for, and some exhibit nasal spectra in which the poles and zeros are probably unusually heavily damped, thereby resulting in few clearly defined prominences. We have found that suitably normalized filter outputs in the frequency regions corresponding with expected nasal consonant poles and zeros are more suitable features for speaker recognition than those between these regions. This may fall short of the criterion of tying a measurement directly to a structural constraint, but in the absence of the ability to characterize all speakers by the location of certain spectral features, it takes cognizance of the fact that the shape of the spectrum is due to certain underlying poles and zeros. Speakers may also be characterized by the prominence or absence of nasal-related spectral peaks.

The fricative /š/ depends mainly on the details of the region around and forward of the alveolar ridge. It has been found that the frequency of the spectral peaks for individuals varies over a rather wide range, but that the shape of the high-frequency region remains much more constant. That is, the number and relative position of uncancelled

poles in the transfer function make it possible and useful to classify examples of /š/ in terms of gross shape. Figure XII-4d illustrates the four shapes that we have defined: narrow peak, wide or double peak, flat, and exceptionally low major peak. These classes are defined so that they are mutually exclusive. To be sure, some of the speakers varied between two of these classes, but rarely across three. The appropriate characterizing measurement for the phoneme would be the estimated probabilities of occurrence of these 4 shapes. The fricative /s/ shows similar characteristics, higher in frequency.

In voiced stops following an unvoiced segment, the onset of voicing before the stop release is not uncommon. (This is the "voice bar" observed in spectrograms.) Although it is not consistent enough to be used as a cue for voiced stops in speech recognition, it is speaker-selective. For our purposes, a stop is termed "prevoiced" if voicing begins more than 20 msec before the burst. In the single instance of this recorded in our data, 6 out of 21 speakers prevoiced better than 5 out of 10 times, 4 speakers 1 or 2 times, and 11 speakers never. Again, the appropriate statistic is the estimated probability of this event. This measurement is particularly appealing because it concerns a rapid event that should be hard to modify consciously, and is an event of such specificity that it is probably independent of most other measurements.

The determination of vowel formant frequencies from a filter-bank analysis of this kind is often not possible if the formants are so close that their spectral peaks merge, as is frequently the case with the vowel /i/. As in the case of the nasals, however, it has proved profitable to measure spectral properties that reflect the underlying pole locations, specifically the second and third central moments of the high-frequency peak in /i/, which comprises three formants.

Atal[5] has shown that the pitch frequency contour over an utterance can be used as a set of measurements for speaker recognition. The pitch contours for a given sentence and a given speaker do show variations that cannot be characterized in a simple way, evidently because of nuances of expression and emphasis, and hence a powerful classification procedure is required. We have found that the values of pitch frequency in most mid-syllable positions are also useful measurements, provided the stress pattern of the utterance is unambiguous or controlled. For example, by using only range overlap to determine discriminability between two speakers, measurements of pitch frequency during three unstressed syllables produced discriminations of 40%, 37%, and 48%. If all three results were pooled, discrimination of 60% resulted, since the three measurements were not uncorrelated. If estimates of the probability distributions from these measurements were used in the decision procedure, the individual measurement scores would certainly be improved.

The strategy of measurements suggested here implies that speaker-recognition schemes require a degree of speech recognition capability. Since in many applications,

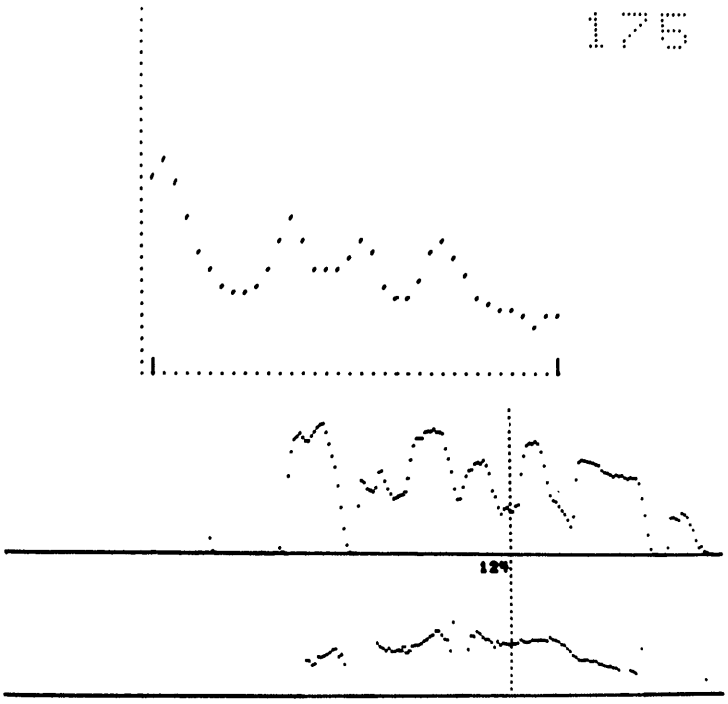

Fig. XII-4(a). Computer display of a spectrum of /m/. Below the spectrum the low-frequency energy and pitch functions of "I cannot remember it" are displayed.

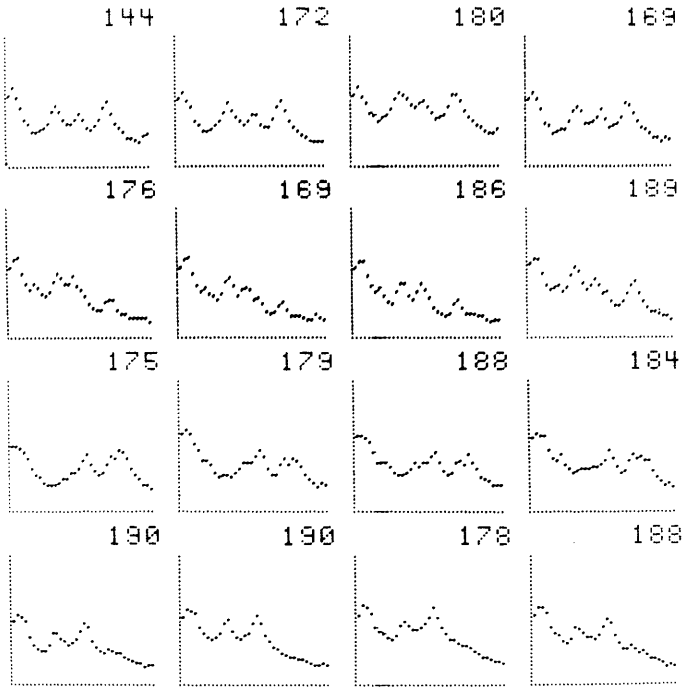

Fig. XII-4(b). Spectra of /m/. Each row contains four examples by one speaker.

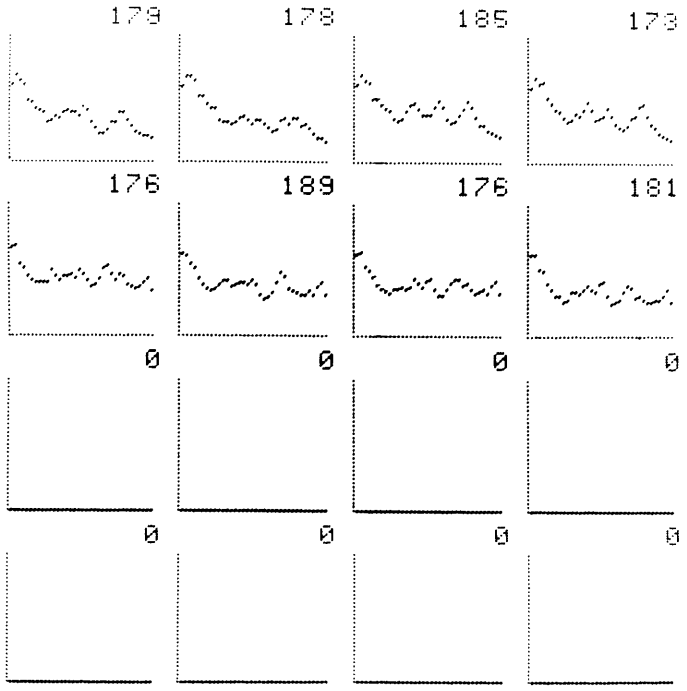

Fig. XII-4(c). Spectra of /m/. Each row contains four examples by one speaker.

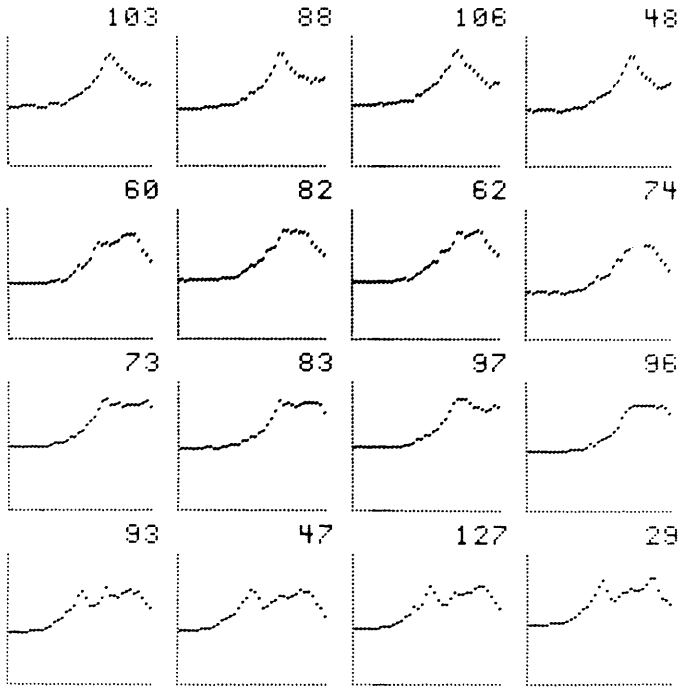

Fig. XII-4(d). Spectra of /š/. Each row contains four examples by one speaker.

the use of a known or prespecified context is a good assumption, this should present no practical difficulty for well-constructed utterances. Naturally, the number of measurements must be expanded in order to yield sufficient information for the operation of a practical speaker-recognition system. The areas of source spectrum characteristics, systematic vowel formant differences, and articulation dynamics should yield further specific measurements of the type discussed here.

J. J. Wolf

References

1. Sandra Pruzansky, "Pattern-Matching Procedure for Automatic Talker Recognition," J. Acoust. Soc. Am. 35, 354-358 (1963).
2. Sandra Pruzansky and M. V. Mathews, "Talker-Recognition Procedure Based on Analysis of Variance," J. Acoust. Soc. Am. 36, 2041-2047 (1964).
3. J. Edie and G. Sebestyen, "Voice Identification General Criteria," RADC-TDR-62-278, 16 May 1962.
4. W. Floyd, "Voice Identification Techniques," RADC-TDR-64-312, September 1964.
5. B. S. Atal, "Automatic Speaker Recognition Based on Pitch Contours," Ph. D. Thesis, Department of Electrical Engineering, Polytechnic Institute of Brooklyn, New York, 1964.
6. W. L. Henke, "Speech Computer Facility," Quarterly Progress Report No. 90, Research Laboratory of Electronics, M.I.T., July 15, 1968, pp. 217-219.
7. J. W. Glenn and N. Kleiner, "Speaker Identification Based on Nasal Phonation," J. Acoust. Soc. Am. 43, 368-372 (1968).
8. C. G. M. Fant, Acoustic Theory of Speech Production (Mouton and Co., The Hague, 1960).
9. O. Fujimura, "Analysis of Nasal Consonants," J. Acoust. Soc. Am. 34, 1865-1875 (1962).