VIII. SPEECH COMMUNICATION*

Academic and Research Staff

Prof. K. N. Stevens	Dr. Mary C. Bateson	Dr. D. H. Klatt
Prof. M. Halle	Dr. Margaret Bullowa	Dr. Paula Menyuk
Prof. W. L. Henke	Dr. A. W. F. Huggins	Dr. J. S. Perkell
Prof. A. V. Oppenheim	Dr. R. D. Kent	A. R. Kessler

Graduate Students

R. W. Boberg	R. M. Mersereau	H. A. Sunkenberg
D. E. Dudgeon	B. Mezrich	R. N. Weinreb
R. W. Hankins	M. R. Sambur	M. L. Wood, Jr.
Emily F. Kirstein	J. S. Siegel	V. W. Zue

A. PHYSIOLOGY OF SPEECH PRODUCTION: A PRELIMINARY STUDY OF TWO SUGGESTED REVISIONS OF THE FEATURES SPECIFYING VOWELS

1. Introduction

The feature 'tense' (as applied to vowels) has been widely discussed (cf. Chomsky and Halle,[1] Jakobson and Halle,[2] and Stewart[3]), particularly with respect to its articulatory correlates and relationship to vowel duration. Motivated apparently by persistent questions about the articulatory correlates of 'tense' and certain acoustic considerations, Halle and Stevens[4] have suggested a revision of the feature specification of vowels in which 'tense' is replaced by 'advanced tongue root.' They hypothesize that the features 'tense-lax' (which accounts for the oppositions /i-ɪ/, /u-ʊ/ and others in English) and 'covered-uncovered' (which applies to vowel harmony in West African languages) "have in common one and the same phonetic mechanism and should, therefore, be regarded as a single feature in the phonetic framework." Their suggestion of the feature 'advanced tongue root' is based on a study of vowel harmony in Asante Twi by Stewart[3] in which he observes that the vowels /ɪ, ɛ, a, ɔ, ʊ/ (unraised) are "raised" to /i, e, ɜ, o, u/ by advancing the root of the tongue.

Halle and Stevens[4] support their argument with the tracings of the vowel pairs /i-ɪ/ and /u-ʊ/ from one speaker of English in which it is obvious that for /i/ and /u/, the tongue root is drawn relatively forward, thereby causing an enlargement of the lower pharynx and a raising of the tongue body in the oral cavity. Their acoustic analysis shows that this gesture would cause the changes in F_1 and F_2 for front and back vowels that are commonly observed. They also argue that the gesture of advancing the tongue

*This work was supported in part by the U.S. Air Force Cambridge Research Laboratories under Contract F19628-69-C-0044; and in part by the National Institutes of Health (Grant 5 RO1 NS04332-08) and M.I.T. Lincoln Laboratory Purchase Order CC-570.

root is essential for F_1 to be as low as possible to produce the unmarked or "natural" high vowels. It follows then that "unmarked" low vowels do not have tongue-root advancing, since they are characterized by a maximally high F_1.

Since the unmarked, low vowel /a/ in English is specified as +tense, it is obviously impossible to characterize all +tense vowels with 'advanced tongue root.' Halle and Stevens[5] have recently suggested a second revision of the vowel features which would account for this discrepancy. They suggest deleting the feature 'low' and adding 'constricted pharynx' (or 'retracted tongue root' or 'constricted tongue root').

'Constricted pharynx' presumably corresponds to a narrowing of the lower pharynx past the neutral position in the region of the tongue root. It could be accomplished by the action of the middle and lower pharyngeal constrictors and contraction of the hyoglossi (which would cause a backward bulging of the pharyngeal tongue dorsum). The perturbation corresponding to 'constricted pharynx' is acoustically antagonistic to that of 'advanced tongue root,' so a "++" specification is precluded.[6]

This change, along with the addition of 'advanced tongue root' and elimination of 'tense' causes a pronounced rearrangement in the feature specification of the vowels. The resulting changes for the vowels of English are shown in Table VIII-1. Note that (i) there are only two possible tongue height specifications, + and - high, and (ii) former tense-lax distinctions are accounted for by the two new features.

The revised feature specification of the vowels is potentially interesting and appealing for a number of reasons (a few of which have been discussed by Halle and Stevens,[4]

Table VIII-1. The hypothesized feature specification of the vowels of English (excluding the feature 'round') is shown above the double line. 'Advanced tongue root' and 'constricted pharynx' would replace 'low' and 'tense' (shown below the double line).

Feature \ Vowel	u	ʊ	i	ɪ	o	ɔ	e	ɛ	æ	a	ʌ
High	+	+	+	+	-	-	-	-	-	-	-
Back	+	+	-	-	+	+	-	-	-	+	+
Advanced Tongue Root	+	-	+	-	+	-	+	-	-	-	-
Constricted Pharynx	-	-	-	-	-	+	-	-	+	+	-
Low	-	-	-	-	-	+	-	-	+	+	-
Tense	+	-	+	-	+	+	+	-	+	+	-

but only for 'advanced tongue root'). At this point, then, the linguistic, acoustic, and physiologic implications of the suggested system should be examined.

The purpose of this study is to take an initial look at the physiological implications of the suggested revisions. The experimental approach is based on the criterion of Chomsky and Halle,[1] that "the phonetic features can be characterized as physical scales describing independently controllable aspects of the speech event ...".

In order to test this criterion for the suggested system, it would be desirable to look at the overall behavior of the vocal tract as it correlates with vowel production. From the point of view of studying the organized function of the end organs, an ideal approach might lead to an expression of the features in terms of neural commands to muscle groups whose actions are identifiable with the feature attributes. Knowing the magnitude, organization and timing of these commands to the vocal tract would give us enormous insight into both physiological and linguistic mechanisms; however, this kind of information is not available for present methods of investigation. The closest that we can get to information of this kind is in the form of electromyographic data. These data can be very useful, but unfortunately, it is impossible to reach many of the muscles of interest with electromyographic probes. This limitation constricts our attempts to look at the organized behavior of muscle groups in a comprehensive manner. Also, it is often difficult to identify precisely the specific muscle(s) from which an electromyographic signal is being obtained, and the interdigitation of often opposing or orthogonal muscle fiber makes interpretation difficult. In order to obtain a comprehensive picture it can be useful to look at tracings of the articulator contours made from a lateral cineradiograph.[7] In this study, mid-vowel tracings of 11 vowels of English are examined for an overview of the manner in which articulatory configurations correspond to the hypothesized features.

2. Actions of the Muscles

To interpret tracings of the midsaggital vocal-tract contour, it would be helpful to review briefly the suggested function of the musculature responsible for positioning the tongue body and determining the size of the pharyngeal cavity (cf. Goss[9] and MacNeilage and Scholes[10]). It is assumed that the extrinsic tongue musculature is primarily responsible for positioning the tongue body. It is recognized that the intrinsic musculature plays a role, but this role is assumed to be secondary, particularly for vowel production.[11] The actions of the muscles are indicated schematically in Fig. VIII-1.

Actions of the genioglossi. Contraction of any portion of the genioglossi pulls the corresponding segment of the tongue dorsum toward the mandibular symphysis, and most likely causes a spatially compensating displacement of the remainder of the tongue body.

Actions of the hyoglossi. The hyoglossi pull the tongue body down and back toward the hyoid bone. If the hyoid position is stabilized, this will (i) cause the

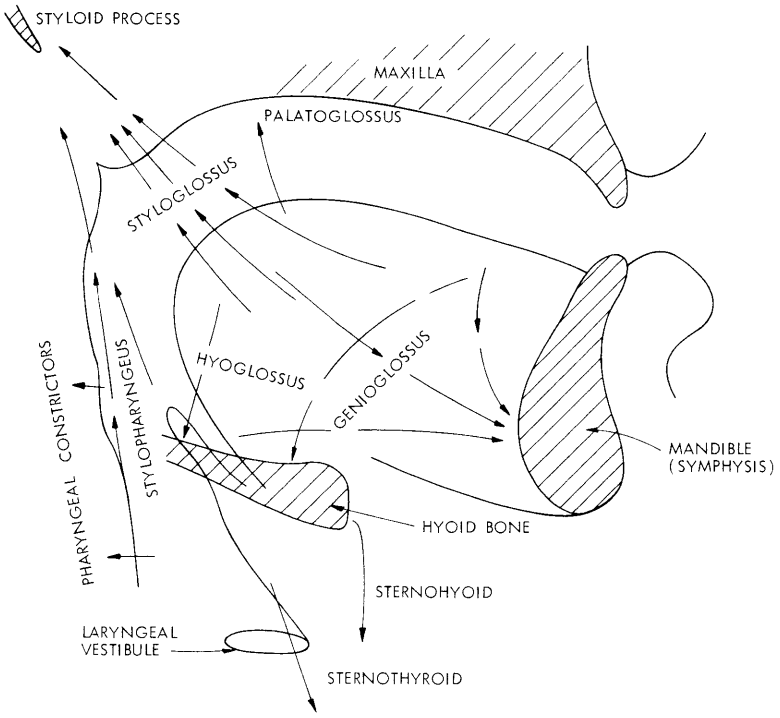

Fig. VIII-1. Schematic representation of the actions of some of the vocal-tract musculature responsible for positioning the tongue body and determining the volume of the pharyngeal cavity.

pharyngeal half of the tongue dorsum to bulge down and back toward the rear pharyngeal wall, and (ii) depress the oral half of the tongue dorsum.

Actions of the styloglossi. The styloglossi pull the middle part of the tongue body upward and backward toward the styloid processes.

Actions of the palatoglossi. The palatoglossi pull the tongue body upward, although these muscles are small and seem to act as much to lower the velum as they do to raise the tongue.[12]

Actions of the pharyngeal constrictors. The superior, middle, and inferior pharyngeal constrictors act to narrow the pharyngeal lumen. By virtue of their respective origins on the mandible, hyoid bone, and thyroid cartilage, they could act to pull these structures backward slightly.

Actions of the stylopharyngei. The stylopharyngei draw the sides of the lower pharynx upward and lateralward to increase its width.

Actions of the sternothyroidei. These muscles exert a downward pull on the thyroid cartilage and in so doing either lower or stabilize the position of the larynx, depending on the antagonistic pull of supralaryngeal musculature. Since the insertion on the thyroid laminae is anterior to the crico-thyroid articulation (see Fig. VIII-2) contraction

of the sternothyroidei might have a secondary effect of rocking the thyroid cartilage forward, thereby increasing the tension on the vocal folds.

Actions of the sternohyoidei. The sternohyoidei exert a downward pull on the body of the hyoid bone either lowering or stabilizing its position.

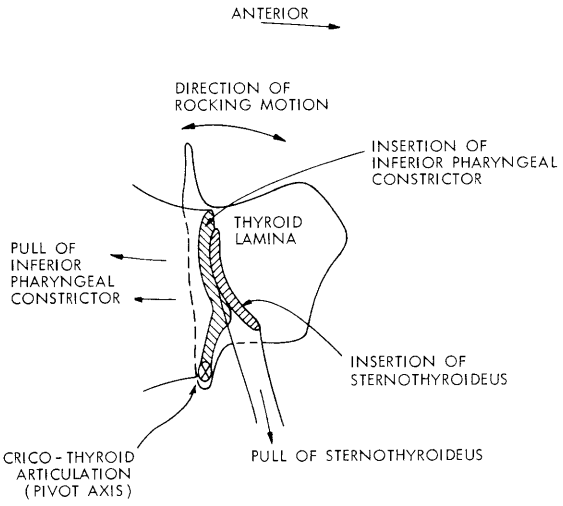

Fig. VIII-2. Outline of the right thyroid lamina showing the insertions and directions of pull of the inferior pharyngeal constrictor and the sternothyroideus. The axis and direction of the hypothesized resulting rotation is also shown.

3. Procedure

Mid-vowel (halfway between onset and offset of voicing) tracings were made from a lateral cineradiograph[13] of the vowels /i, ɪ, e, ɛ, æ, o, u, ʊ, a, ʌ, ɔ/ spoken in the environment /hɔ́bvb/. The subject, Speaker A, is a speaker of General American English.[14] Approximate values of the vowel formants and durations were obtained from spectrograms and are listed in Table VIII-2. It appears from the spectrograms that the /e/ is somewhat diphthongized, but the /o/ does not appear to be diphthongized significantly. Data were obtained for a second subject, Speaker B (General American English) in the form of mid-vowel tracings of /i, ɪ, ɛ, æ, u, ʊ, a/ spoken in the environment /hɔ́tv/.[15, 16]

To examine the articulatory differences supposedly governed by a particular feature, mid-vowel tracings were superimposed for vowels for which all features except 'rounding' and the feature in question are the same. This process yields several sets of overlapping vocal-tract contours in which contrasting contours represent "+" vs "−" the value of the feature. For each of the speakers, the tracings were superimposed with respect to the maxilla.[17]

Table VIII-2. Formant values (to the nearest 50 Hz) and durations (to the nearest 5 ms) of the vowels as measured from spectrograms.

Vowel	F_1 (Hz)	F_2 (Hz)	F_3 (Hz)	Duration (ms)
i	400	2300	2800	310
ɪ	500	1750	2500	225
e	500	1900	2600	310
ɛ	550	1700	2500	280
æ	750	1600	2400	310
u	450	1200	2250	295
ʊ	500	1200	2300	250
o	600	1200	2350	320
ɔ	700	1100	2300	330
a	800	1200	2350	315
ʌ	800	1300	2500	285

4. Relationship of Mid-vowel Vocal Tract Contours to the Proposed Vowel Features

These relationships are expressed in terms of simplified complexes of muscle contractions with the goal of providing a crude physiological framework corresponding to the features. We hope that this framework will contribute to a model that might be useful in testing cross-linguistic applications of the features, timing of commands, and coarticulatory effects.

1. The Feature High. Figure VIII-3 contains overlapping vocal-tract contours for the vowel pairs /u, o/, /ɪ, ɛ/, /ʊ, ʌ/, /i, e/ for Speaker A and /ɪ, ɛ/ for Speaker B. For all pairs the tongue body is higher and farther forward for the +high vowel than for the -high one. Also, the mandible is higher for the +high cognates. The hyoid bone and laryngeal vestibule are all lower for the +high vowels. This confirms the data of Perkell[15] which showed an inverse relationship between tongue and larynx height (for Speaker B). In all cases except /ɪ, ɛ/ for Speaker A the pharyngeal part of the tongue dorsum is farther forward for the +high vowels.

The relative displacements of the tongue and larynx for + vs -high tend to increase the ratio of posterior to anterior cavity volume. The primary acoustic

effect of this increase is to lower F_1. This is reflected in the spectral measurements shown in Table VIII-2.

These observation suggest that +high correlates with a complex of several muscle actions. The posterior third of the genioglossi contract to displace the tongue body upward and forward, and the styloglossi contract to pull upward and backward, producing a net displacement that is upward and slightly forward. Also, +high also seems to correlate with contraction of the sternohyoid and sternothyroid muscles to lower the hyoid bone and larynx. The raised mandible most likely helps to raise the tongue body.

The gestures of raising the mandible and lowering the larynx involve phonetically governed movements of large structures. Since the movements would tend to be sluggish they should diminish considerably in continuous speech.

2. The Feature Back. Figure VIII-4 contains overlapping vocal-tract contours for the vowels /u, i/, /ʊ, ɪ/, /o, e/, /ʌ, ɛ/, /a, ɔ, æ/ (Speaker A) and /u, i/, /ʊ, ɪ/, /ɑ, æ/ (Speaker B). In all cases the tongue body is farther back in the pharynx for the +back cognates. For Speaker A there seems to be a direct relationship between the antero-posterior positions of the tongue body and the mandible and hyoid bone.

The tongue body movement corresponding to +back could be accomplished by a combination of the backward and upward pull of the styloglossi and the downward and backward pull of the hyoglossi (which would cause the pharyngeal portion of the tongue dorsum to bulge out in a posterior direction). The anterior fibers of the genioglossi, as well as intrinsic musculature, could be active in keeping the tongue tip touching the floor of the mouth for +back and +high.[11]

The posterior movement of the mandible, hyoid bone, and larynx for Speaker A suggests that the pharyngeal constrictors play a role in +back, drawing back the entire framework of the vocal tract, as well as narrowing the pharyngeal cavity.

3. Advanced Tongue Root. Figure VIII-5 contains overlapping vocal-tract contours for the vowel sets /u, ʊ/, /i, ɪ/, /o, ʌ/, /e, ɛ/ (Speaker A) and /u, ʊ/, /i, ɪ/ (Speaker B). In all cases the posterior half of the tongue dorsum is farther forward for +advanced tongue root vowels. The epiglottis is also in a relatively anterior position for these vowels, but since the epiglottis was difficult to trace (Speaker A), this observation should be confirmed. At this point it is suggested that 'advanced tongue root' corresponds to a contraction of a small segment of the genioglossi at the tongue root (somewhat less of the muscles than for +high).

For the four examples of the [+high, +advanced tongue root] vowels, /u, i/ there is a concavity in the tongue contour at the root. This concavity is presumably the result of an additional effect of overlapping commands to the posterior segment of the genioglossi to contract. For the vowel /o/, only 'advanced tongue root' is influencing the command to the posterior genioglossi, so a slight forward displacement without the

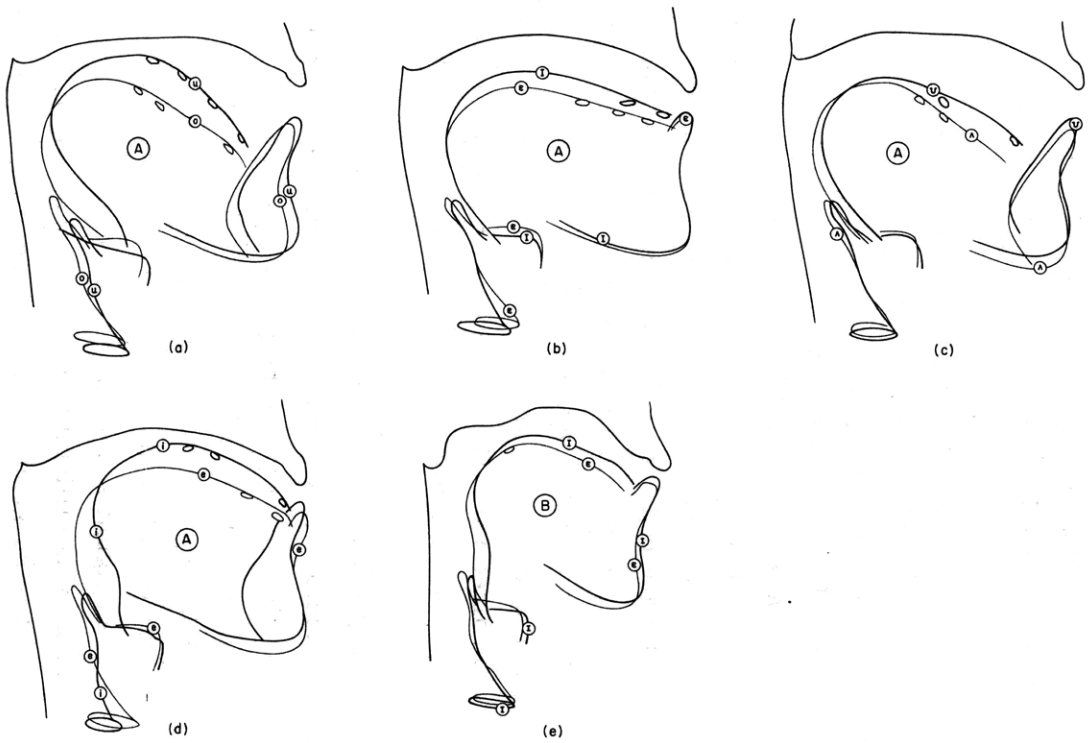

Fig. VIII-3. The feature high. Vocal-tract contours: (a) for the vowel pairs /u, o/; (b) and (e) for /ɪ, ɛ/; (c) for /ʊ, ʌ/; and (d) for /i, e/. The dark contours represent the +high cognates; the light contours, the -high cognates. The speakers are identified by encircled Roman letters.

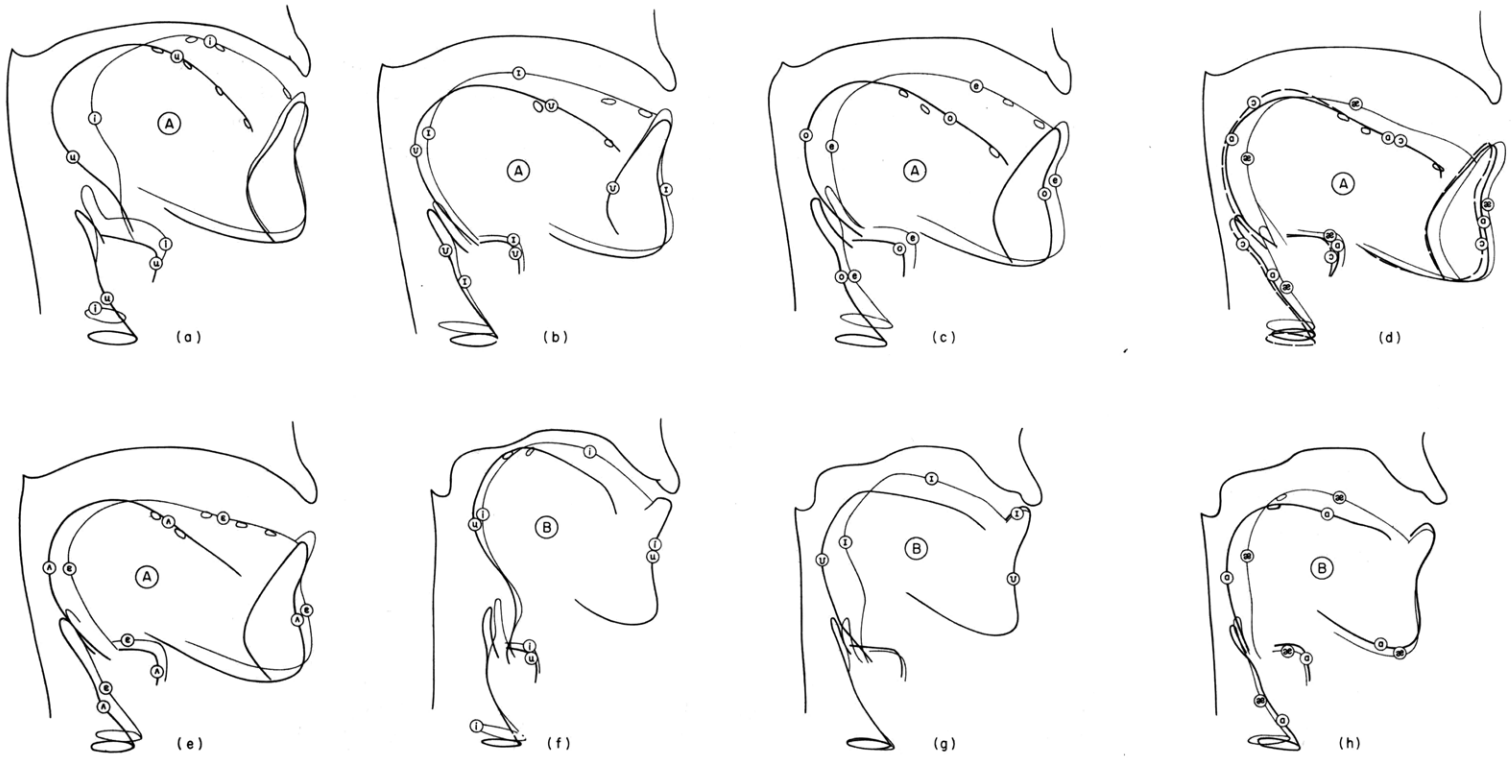

Fig. VIII-4. The feature back. Vocal-tract contours: (a) and (f) for the vowel sets /u, i/; (b) and (g) for /ʊ, ɪ/; (c) for /o, e/; (d) for /a, ɔ, æ/; (e) for /ʌ, ɛ/; and (h) for /a, æ/. The dark contours represent the +back cognates; the light contours, the -high cognates. The speakers are identified by encircled Roman letters.

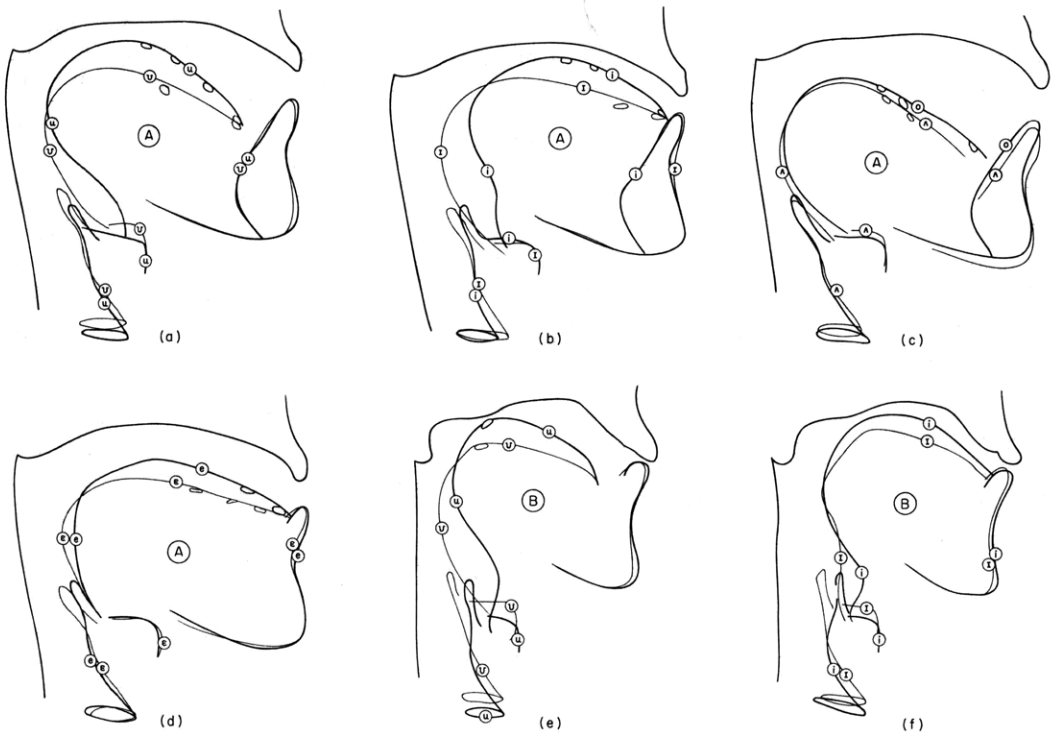

Fig. VIII-5. The feature advanced tongue root. Vocal-tract contours: (a) and (e) for the vowel pairs /u, ɪɾ/; (b) and (f) for (i, ɪ/; (c) for /o, ʌ/; and (d) for /e, ɛ/. The dark contours represent the +advanced tongue root cognates; the light contours, the -advanced tongue root cognates. The speakers are identified by encircled Roman letters.

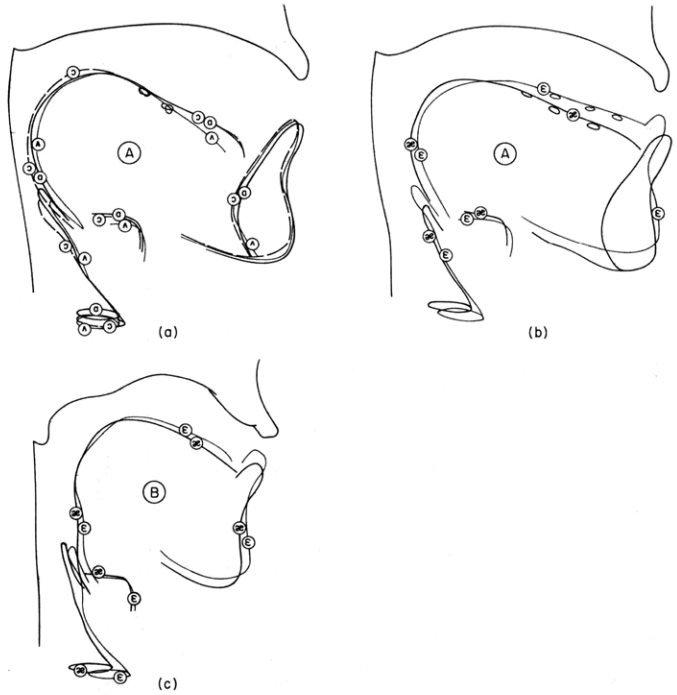

Fig. VIII-6. The feature constricted pharynx. Vocal-tract contours: (a) for the vowel sets /ɔ, a, ʌ/; (b) and (c) for /æ, ɛ/. The dark contours represent the +constricted pharynx cognates and the light contours, the -constricted pharynx cognates. The speakers are identified by encircled Roman letters.

extreme effect of a concavity is observed. The relative forward displacement of the tongue contour for /e/ is not most pronounced at the root, but slightly higher up. This could be because the vowel is diphthongized and what is being observed is a shape that represents the transition from /e/ to /i/.

4. Constricted Pharynx. Figure VIII-6 contains overlapping vocal-tract contours for the vowel sets /ɔ, a, ʌ/ (Speaker A) and /æ, ɛ/ for Speakers A and B. In all cases, the contour of the posterior half of the tongue dorsum, the epiglottis, and the hyoid bone are farther back for the +constricted pharynx vowels.

It is suggested that +constricted pharynx corresponds to contraction of the middle and lower pharyngeal constrictors and the hyoglossi. The action of the constrictors would be to narrow the lower pharyngeal lumen and pull back on the hyoid bone and thyroid cartilage, while the contraction of the hyoglossi would bulge the posterior tongue body downward and backward.

5. Other Results

Some of the functions suggested by the tracings are corroborated by results of two other studies.

Using ultrasonic measurements, Minifie, Hixon, Kelsey, and Woodhouse[18] have shown there is both inward and outward active movement of the lateral pharyngeal wall (from a neutral position) associated with vowel production. They used utterances of the form VCVCV in which the consonants were both /b/, /d/, or /g/ and the vowels all /u/, /i/, /ʌ/, /æ/, or /a/. For /u/ and /i/ they found outward movement of one wall ranging between 0 and 2 1/2 mm. Inward movement was 2-4 mm for /ʌ/ and 2 1/2-4 mm for /a/ and /æ/, with the movement for /a/ and /æ/ being consistently greater than for /ʌ/. This result tends to confirm not only that the pharyngeal constrictors must play a phonetically determined role in narrowing the pharyngeal lumen, but that there is also some action, probably by the stylopharyngei, which tends to widen the pharynx.

In an electromyographic study, Smith and Hirano[19] found that the anterior and posterior portions of the genioglossus muscle do act independently, and "that the genioglossus muscle is, in functional terms, not one muscle at all, but at least two, perhaps more, differently innervated units." They also report that the posterior portion is consistently and reliably more active for the tense, "high" vowels [i], [e], and [u] than for their lax counterparts [ɪ], [ɛ], and [ʊ], "and that it is "inhibited" for the '[a] vowel'." This result verifies our deductions about the function of the posterior genioglossus from tracings, and it tends to substantiate the method of making inferences about muscle activity from a knowledge of the anatomy and articulator displacement.

6. Implications and Conclusions

It is useful to summarize and slightly revise the hypothesized relationships between each muscle-group action and its underlying feature. The muscle actions can be visualized more readily by referring to the numbered diagram in Fig. VIII-7. Implementation of the feature 'high' causes raising of the tongue body by contraction of the posterior

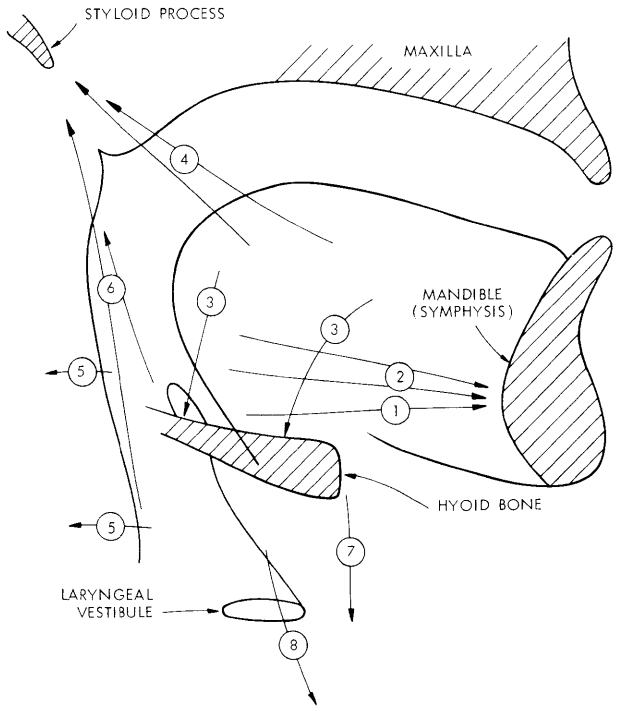

Fig. VIII-7. Schematic representation of the muscle function corresponding to the features in the suggested model. The muscles are: 1, genioglossus, small segment at the root; 1 and 2, genioglossus, posterior one-third; 3, hyoglossus; 4, styloglossus; 5, pharyngeal constrictors; 6, stylopharyngeus; 7, sternohyoideus; 8, sternothyroideus.

third of the genioglossi (1 and 2) and the styloglossi (4). The pharynx is further widened slightly by contraction of the stylopharyngei (6), and the larynx is lowered by the sternohyoidei (7) and sternothyroidei (8). Implementation of 'back' causes retraction of the tongue body through contraction of the styloglossi (4) and hyoglossi (3). The pharynx is further narrowed by contraction of the pharyngeal constrictors (5). 'Advanced tongue root' is implemented by contraction of a small posterior segment of the genioglossi (1) and a widening of the lower pharynx by the stylopharyngei (6). 'Constricted pharynx'

corresponds to contraction of the hyoglossi (3), causing the tongue body to bulge backward, and further narrowing of the lower pharynx by the middle and inferior constrictors (5). Presumably, overlapping commands to the same structures have additive effects.[20] It follows that the muscle actions corresponding to 'advanced tongue root' are a subset of those corresponding to 'high,' and the actions corresponding to 'constricted pharynx' are a subset of those corresponding to 'back'.[21] The hypothesized "physiological system" as it relates to the features is shown in matrix form in Table VIII-3.

Table VIII-3. The hypothesized "physiological" specification of the features. Numbers refer to the schematic representations of muscle action shown in Fig. VIII-7.

Functional Unit		Feature			
Muscle(s)	Number in Fig. VIII-7	High	Back	Advanced Tongue Root	Constricted Pharynx
Genioglossus, Root Segment	1	+	-	+	-
Genioglossus, Posterior Third	2	+	-	-	-
Hyoglossus	3	-	+	-	+
Styloglossus	4	+	+	-	-
Pharyngeal Constrictors	5	-	+	-	+
Stylopharyngeus	6	+	-	+	-
Sternohyoideus and Sternothyroideus	7 and 8	+	-	+	-

This system offers a rather simple possible explanation of the relationship between F_o and tongue height (cf. Peterson and Barney[22]). We have observed an inverse relationship between larynx and tongue height. Since the sternothyroidei would seem to rock the thyroid cartilage forward as well as lower it, their action would also cause an increase in vocal-fold tension and a resulting rise in F_o for high vowels. Acting in opposition, the inferior pharyngeal constrictors' pulling on the posterior edges of the thyroid laminae would tend to rock the cartilage backward, thereby decreasing the vocal-fold tension (see Fig. VIII-2).

As more observations are made of the vocal-tract function a picture of complete synergy continues to emerge. In these examples it may be seen that all possible mechanisms seem to be operating to achieve the phonetic and acoustic goals. Thus the scope of the simplified physiological mechanisms needed to account for the features becomes

broader and the systems more complex. It now seems that a satisfactory model of the tongue behavior can no longer be constructed by using only a cylinder, the position of which is specified by two numbers. A model utilizing the mechanisms suggested by these observations should be much more difficult to construct, but we hope that it would do a better job of accounting for the actual behavior.

These preliminary results seem to give physiological support to the proposed feature revisions of Halle and Stevens[4] and the underlying assumptions of Chomsky and Halle.[1] Thus it appears to be worthwhile to attempt to confirm these observations for additional speakers of English and to try to look at other languages by using cineradiographs. It should also be interesting to study more complete supporting linguistic and acoustic arguments.

J. S. Perkell

Footnotes and References

1. N. Chomsky and M. Halle, The Sound Pattern of English (Harper and Row Publishers, Inc., New York, 1968).
2. R. Jakobson and M. Halle, "Tenseness and Laxness," in D. Abercrombie et al. (Eds.), In Honour of Daniel Jones (Longmans, Green and Company, London, 1964), pp. 96-101.
3. J. M. Stewart, "Tongue Root Position in Akan Vowel Harmony," Phonetica 16, 185-204 (1967).
4. M. Halle and K. N. Stevens, "On the Feature 'Advanced Tongue Root'," Quarterly Progress Report No. 94, Research Laboratory of Electronics, M.I.T., July 15, 1969, pp. 209-215.
5. M. Halle and K. N. Stevens, Personal communication, 1971.
6. It will be seen that the muscle actions that produce 'advanced tongue root' and 'constricted pharynx' are not completely antagonistic in the usual sense of muscle antagonism. This will have implications in accounting for overlapping physiological domains of the features.
7. Because a major source of feedback may be in the form of spatial information (cf. MacNeilage[8]), data in this form may prove to be extremely useful in constructing articulatory models.
8. P. F. MacNeilage, "The Motor Control of Serial Ordering in Speech," Psychol. Rev. 77, 182-196 (1970).
9. C. M. Goss (Ed.), Gray's Anatomy (Lea and Febiger, Philadelphia, 1959).
10. P. F. MacNeilage and G. N. Scholes, "An Electromyographic Study of the Tongue during Vowel Production," J. Speech and Hearing Res. 7, 209-232 (1964).
11. MacNeilage and Scholes,[10] through measurements with surface electrodes have demonstrated differences in activity of intrinsic tongue musculature for different vowels. It is important to recognize that any articulatory gesture usually involves the activity of all the musculature that could act in either a synergistic or antagonistic manner to perform the gesture with the required degree of control. For understanding the behavior in terms of a hypothesized, simplified scheme of commands like a feature matrix, however, it is very helpful to limit the controllable parameters to a small group in a manner that could have some physiological significance. Thus it seems reasonable to allow vowel features determining

tongue-body position to be implemented only through the action of the extrinsic musculature. Hence, for purposes of future modeling of the behavior of the tongue, it may be assumed that the activity of the intrinsic musculature is determined by spatial and physiological constraints, such as keeping the tongue tip within a certain distance from the floor of the mouth for vowels. This particular activity might then be a physiological property of all vowels, but it would be expressed automatically, especially for vowels with tongue body positions farther away from the lower alveolar ridge. It would not, however, be associated with any feature that differentiates vowels from one another.

12. J. Lubker, B. Fritzell and J. Lindquist, "Velo-pharyngeal Function: An Electromyographic Study," Speech Transmission Laboratory, Quarterly Progress and Status Report 4/1970, Royal Institute of Technology, Stockholm, pp. 9-20.

13. I am grateful to Mr. Joseph DeClerk, Research Scientist, U.S. Army Electronics Command, for taking the cineradiograph and supplying me with a copy, a dubbing of the tape, and spectrograms containing the sound-synchronizing information.

14. For Speaker A, the dorsal tongue contour, especially in the epiglottis was difficult to trace, so it is possible that some of the values are not quite as accurate as would be desired. In any case the observations made in this study should be repeated with more subjects.

15. J. S. Perkell, Physiology of Speech Production: Results and Implications of a Quantitative Cineradiographic Study, Research Monograph No. 53 (The M.I.T. Press, Cambridge, Mass., 1969).

16. The vowels /ɪ, ɛ, æ, ʊ/ do not normally occur in open syllables in English, so these utterances are somewhat unnatural for Speaker B.

17. The tracings for Speaker B all include the outline of a single radio-opaque marker fixed to the tongue dorsum. The four markers fixed to the tongue of Speaker A could not always be seen clearly, and they are of much less value in determining the position of a specific flesh point.

18. F. D. Minifie, T. J. Hixon, C. A. Kelsey, and R. J. Woodhouse, "Lateral Pharyngeal Wall Movement during Speech Production," J. Speech and Hearing Res. 13, 584-594 (1970).

19. T. Smith and M. Hirano, "Experimental Investigations of the Muscular Control of the Tongue in Speech," Working Papers in Phonetics 10, University of California at Los Angeles, 1968, pp. 145-155.

20. Several assumptions are implied by modeling the vocal tract function in this manner. The muscle actions that correlate with a "+" value of a feature are assumed to operate against a restoring force which corresponds to a "vowel tonus" of the musculature. This "tonus" is presumably associated with the feature combination specifying "vowel", and if unperturbed by "+" values of the feature under discussion, it will produce the neutral, or /ɛ/ configuration.

The interaction of acoustically or physiologically antagonistic elements of a combination of commands (such as +back, +high) must be accounted for by some kind of cancellation. The cancellation usually takes place in more than one way. It can be in the form of an actual muscle antagonism, as might be the case of the opposing action of the pharyngeal constrictors and the stylopharyngei on pharynx width. The interaction of the effects of overlapping commands to the posterior genioglossi and hyoglossi could be accounted for by having the contractions take place, resulting in some grooving of the tongue with little net displacement. Alternatively, the commands could cancel at a higher neural level. Presumably, a variety of such mechanisms operate simultaneously in a highly complex manner. Any attempt at modeling vocal-tract behavior based on a feature system will have to consider these interactions, particularly with respect to the efficiency of computation and naturalness of the model.

21. The main difference between 'high' and 'advanced tongue root' and between 'back' and 'constricted pharynx' is whether or not the stylopharyngei contract. There is, however, an important additional difference between the overlapping domains of these features. By virtue of the anatomy of the pharyngeal constrictors and the hyoglossi, the effects of +back or +constricted pharynx on the configuration of the lower pharynx cannot be highly localized, and are probably indistinguishable from each other. This would suggest that if the stylopharyngei were caused to contract by +high, there might be only a small observable difference between [+high, +back] and [+high, +constricted pharynx]. It would be useful to observe an example of this contrast, if such an example exists.

    On the other hand, there is a hypothetical difference between the localized effect of 'high' and 'advanced tongue root' on the posterior genioglossus, in that 'advanced tongue root' involves a smaller part of the muscle. Thus we would expect to see a difference between [+back, +advanced tongue root] and [+back, +high].

22. G. E. Peterson and H. L. Barney, "Control Methods Used in a Study of the Vowels," J. Acoust. Soc. Am. 24, 175-184 (1952).

B. COMPUTER-GENERATED SPECTROGRAMS AND CEPSTROGRAMS

Computer generation of spectrograms offers great flexibility and permits interesting on-line analysis. An objective of the work presented in this report was to produce high-quality spectrograms on a PDP-9 computer using the fast Fourier transform (FFT) algorithm for spectral analysis. The theoretical techniques have been described previously.[1,2] The results that we present here indicate that spectrograms obtained digitally, using the FFT, are comparable to those obtained by conventional analog methods, and have a potential advantage in terms of increased flexibility.

The programming package used to display spectra three-dimensionally has also been applied to displaying cepstra.[3] Since the cepstrum has an energy concentration at an interval corresponding to the short-time pitch period, the three-dimensional "cepstrogram" yields contours of intensity that give a visual indication of pitch period behavior.

1. Implementation

Digital spectrograms are obtained by computing the discrete Fourier transform of sampled speech multiplied by a finite-duration window w(n). As the window advances over the speech waveform, new spectral cross sections are computed. It can be shown that the magnitude of each spectral sample corresponds to the output of a full quadrature filter into which the speech samples are played. The frequency response of the lowpass prototype of the quadrature filter is the Fourier transform of the window w(n) used in the evaluation of the discrete Fourier transform.

In general, it is desirable for the window w(n) to be of short duration and also for its transform to have low sidelobes past a specified cutoff frequency. Often, a convenient choice for w(n) is a Hanning window defined as

$$w(n) = \begin{cases} 1/2\left[1+\cos\frac{\pi n}{N}\right]; & |n| \leq N \\ 0 & \text{otherwise.} \end{cases}$$

More generally the impulse response of a frequency selective nonrecursive digital filter with good frequency selection characteristics can be used as a window for the digital spectral analysis.

Analog spectrograms are generated by analyzing one frequency band for all time and then incrementing the frequency, whereas digital spectrograms are typically generated by analyzing all frequencies for one finite-length time segment and then advancing the window over the speech waveform. After an FFT is computed on a time segment, the magnitude is formed from the real and imaginary parts. For the examples presented in this report it is then raised to the 0.8 power which serves to enhance lower energy points.

Speech energy tends to fall off in frequency at a rate of 6-12 dB/octave from 300-3000 Hz with a total dynamic range of approximately 40-50 dB.[4] In a typical analog spectrogram machine like the Voice Print (VP) Laboratories model, the marking paper has a dynamic range of ~12 dB. To fit the speech range into this dynamic range, the VP playback amplifier is designed with a 12 dB/octave boost from 300-3000 Hz, above which it is flat and below which it falls off rapidly (see Fig. VIII-8). To generate digital spectrograms similar to VP spectrograms this frequency shaping was applied by multiplying the enhanced DFT magnitude points by the playback amplifier frequency response.

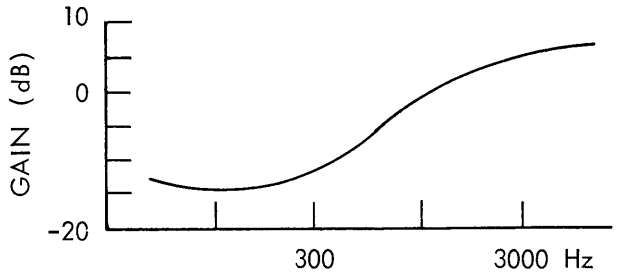

Fig. VIII-8. Voice-print playback amplifier frequency response.

The three-dimensional picture is formed by duration-modulating points in a two-dimensional CRT raster with 256 points in the vertical (frequency) dimension and 512 points in the horizontal (time) dimension. The duration is proportional to the amplitude at the spectral point to be displayed. The duration modulation is accomplished by displaying each point in the raster a number of times proportional to the amplitude of the spectral sample. At present, 32 different intensity levels are recorded, corresponding to a dynamic range of 30 dB. The integrating property of the Polaroid film

used for hard copy yields an intensity proportional to the amplitude.

When the speech sampling rate is 10 kHz, the frequency dimension spans 5 kHz and the time dimension approximately 1 s, yielding the same aspect ratio as in VP spectrograms. The frequency points are spaced at ~20 Hz and the time points at ~2 ms.

In the examples presented here narrow-band analyses are performed by a 512 point (51.2 ms) DFT using a Hanning window. The equivalent bandwidth is 28 Hz (compared with 50 Hz in the VP machine) with an output point every $5000/256 \cong 20$ Hz. The successive 51.2 ms time segments are advanced by ~8 ms increments and thus correspond to every fourth raster column. Intervening columns are obtained by linear interpolation.

Wideband analyses are performed by a 128 point (12.8 ms) DFT using a window obtained by the frequency sampling method.[5] The equivalent bandwidth is 250 Hz (compared with 300 Hz in the VP machine) with an output point every $5000/64 \cong 80$ Hz. These 64 points are expanded to the 256 required by the raster by using linear interpolation. The successive 12.8 ms time segments are advanced ~2 ms and thus correspond to every raster column.

All of the above-mentioned parameters are easily adjusted in the computer so that various frequency and time ranges, frequency shapings, filter bandwidths, and so forth can be chosen. In effect, a spectrogram can be tailor-made to a waveform that is under analysis.

To make cepstrograms, the three-dimensional display package was used to display cepstra instead of transforms. The cepstra of 51.2 ms segments are computed and the portion from 3.2-12.8 ms (pseudo-time or quefrency) is displayed. This permits display of fundamental frequencies from 80-310 Hz. The output points are squared to provide peak enhancement, and a linear emphasis from 1 at 3.2 ms to 4 at 12.8 ms is applied.

2. Results

Figure VIII-9 shows conventional wideband and narrow-band analog spectrograms of "Joe took Father's shoe bench out," spoken by a male speaker. The digital spectrograms are shown in Fig. VIII-10. The formants, pitch striations, and frications all are evident with similar dimensions and relative intensities. Figures VIII-11 and VIII-12 show time expansions. Figures VIII-13 through VIII-17 are spectrograms (digital and analog) and cepstrograms of several sentences. Because of limitations in the printing process the contrast and dynamic range of these examples may appear somewhat less than in the original pictures.

3. Conclusions

Computer-generated spectrograms of quality comparable to that of analog

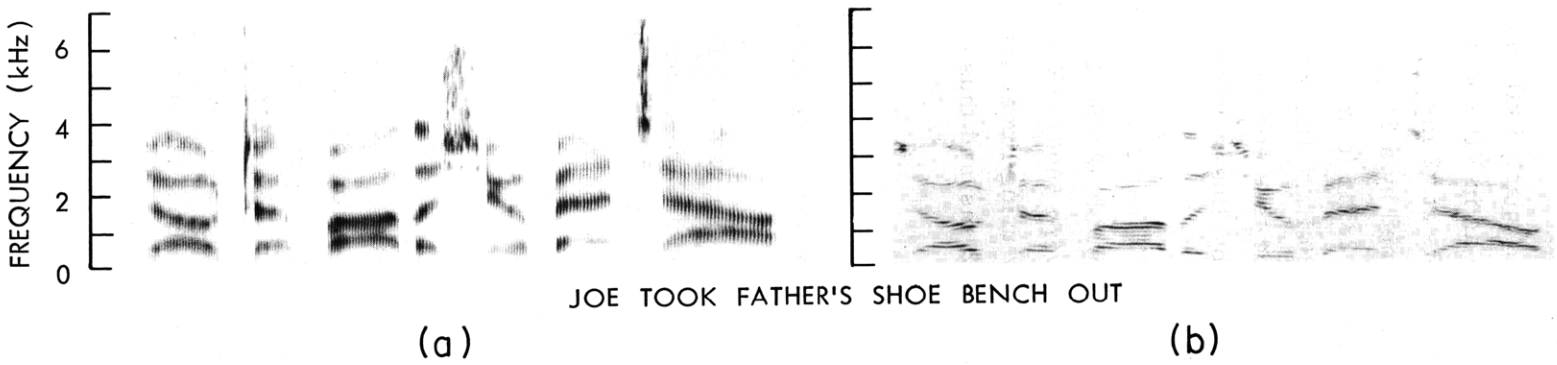

Fig. VIII-9. Analog spectrograms. (a) Wideband. (b) Narrow-band.

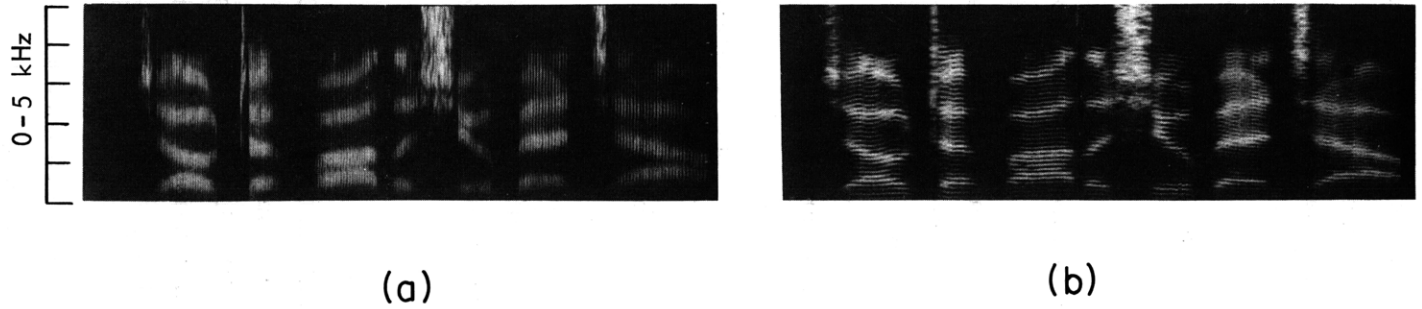

Fig. VIII-10. Digital spectrograms. (a) Wideband. (b) Narrow-band.

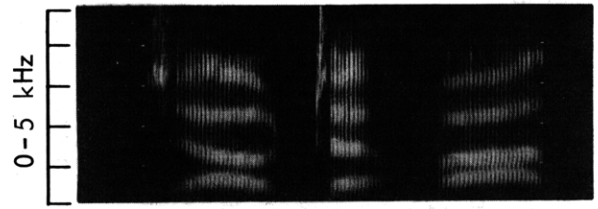

Fig. VIII-11. Wideband spectrogram. Time expansion by 1.5.

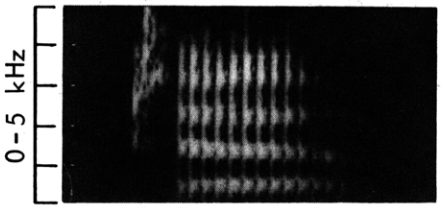

Fig. VIII-12. Wideband spectrogram. Time expansion by 4.

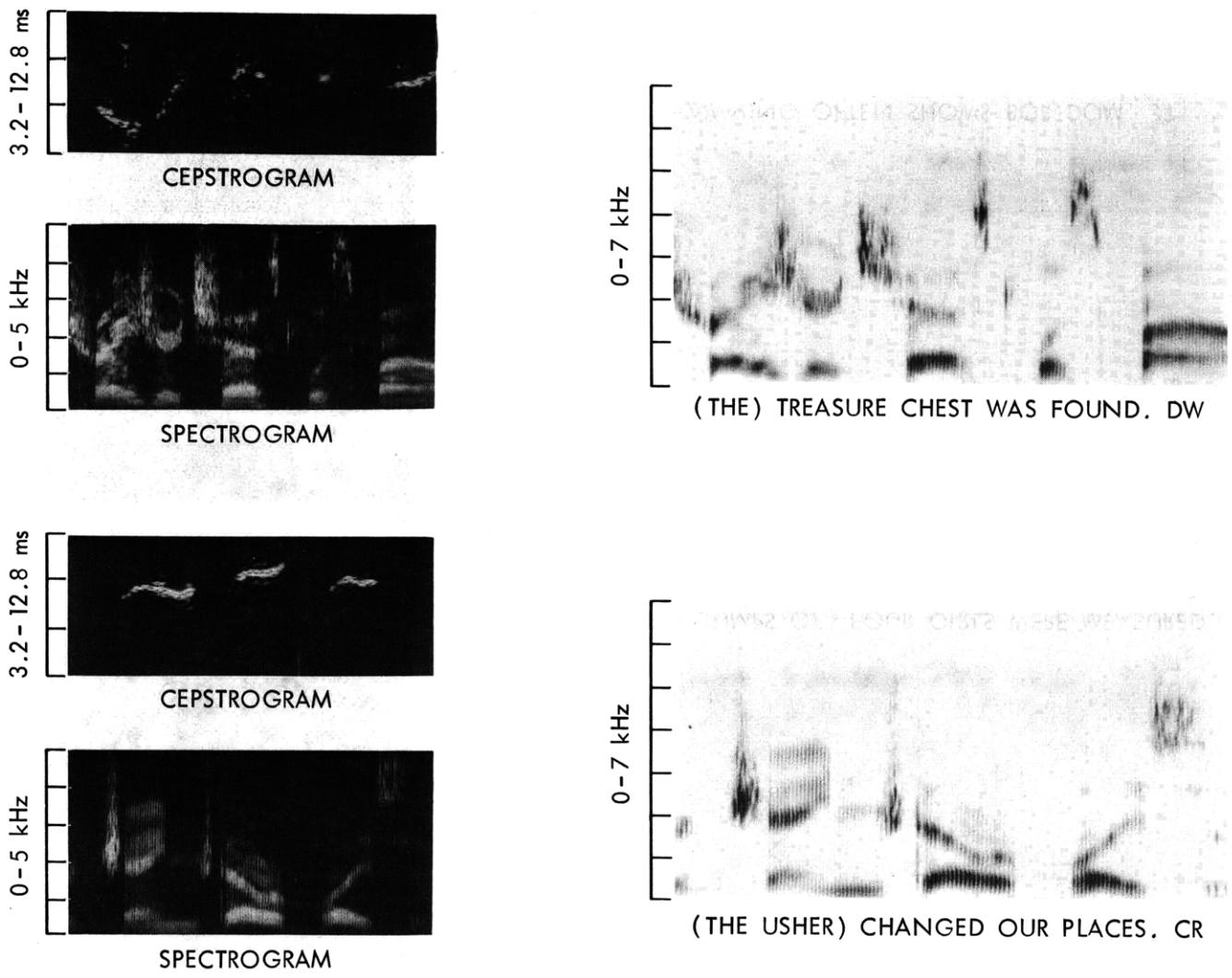

Fig. VIII-13. Spectrograms and cepstrograms.

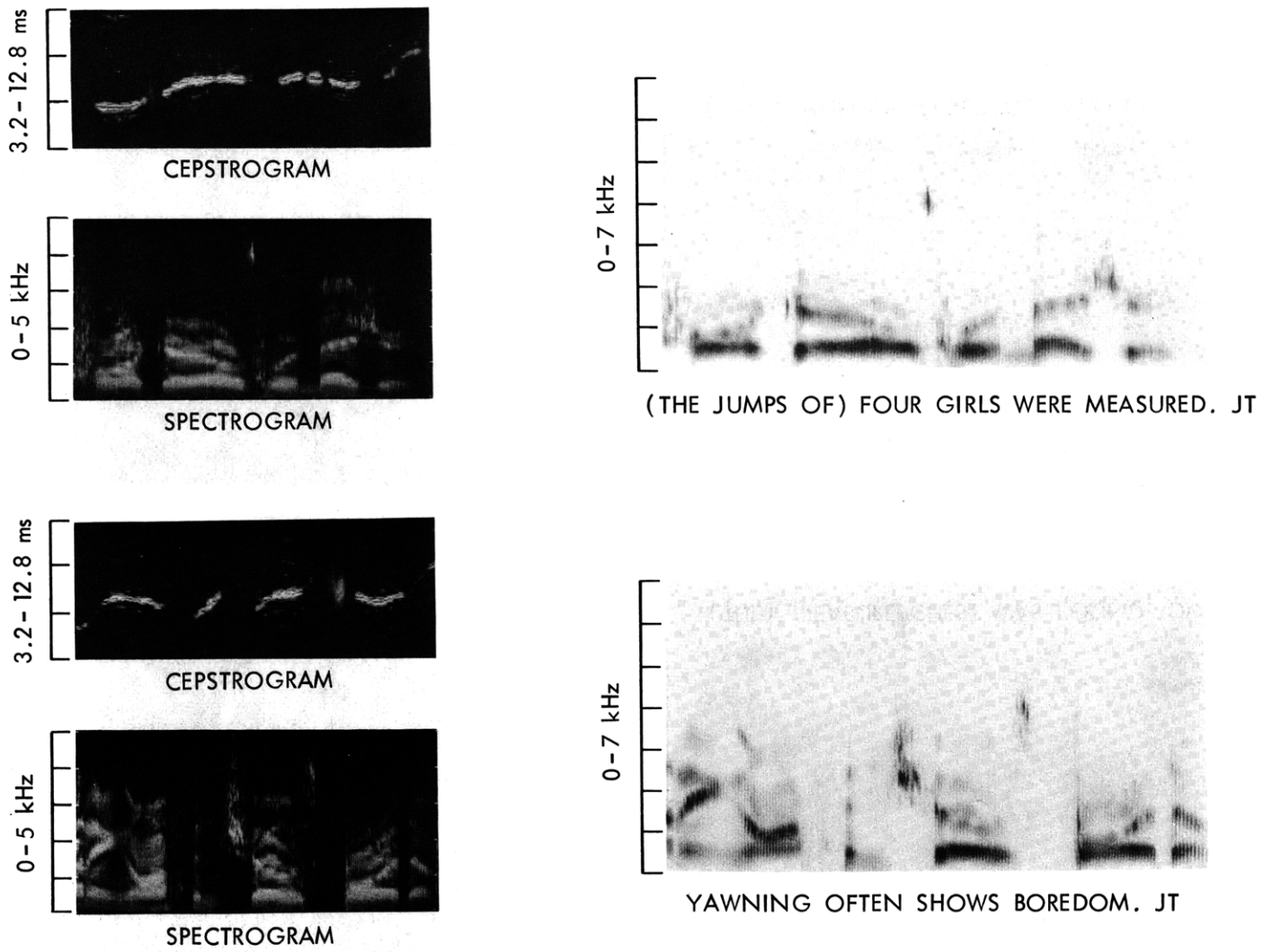

Fig. VIII-14. Spectrograms and cepstrograms.

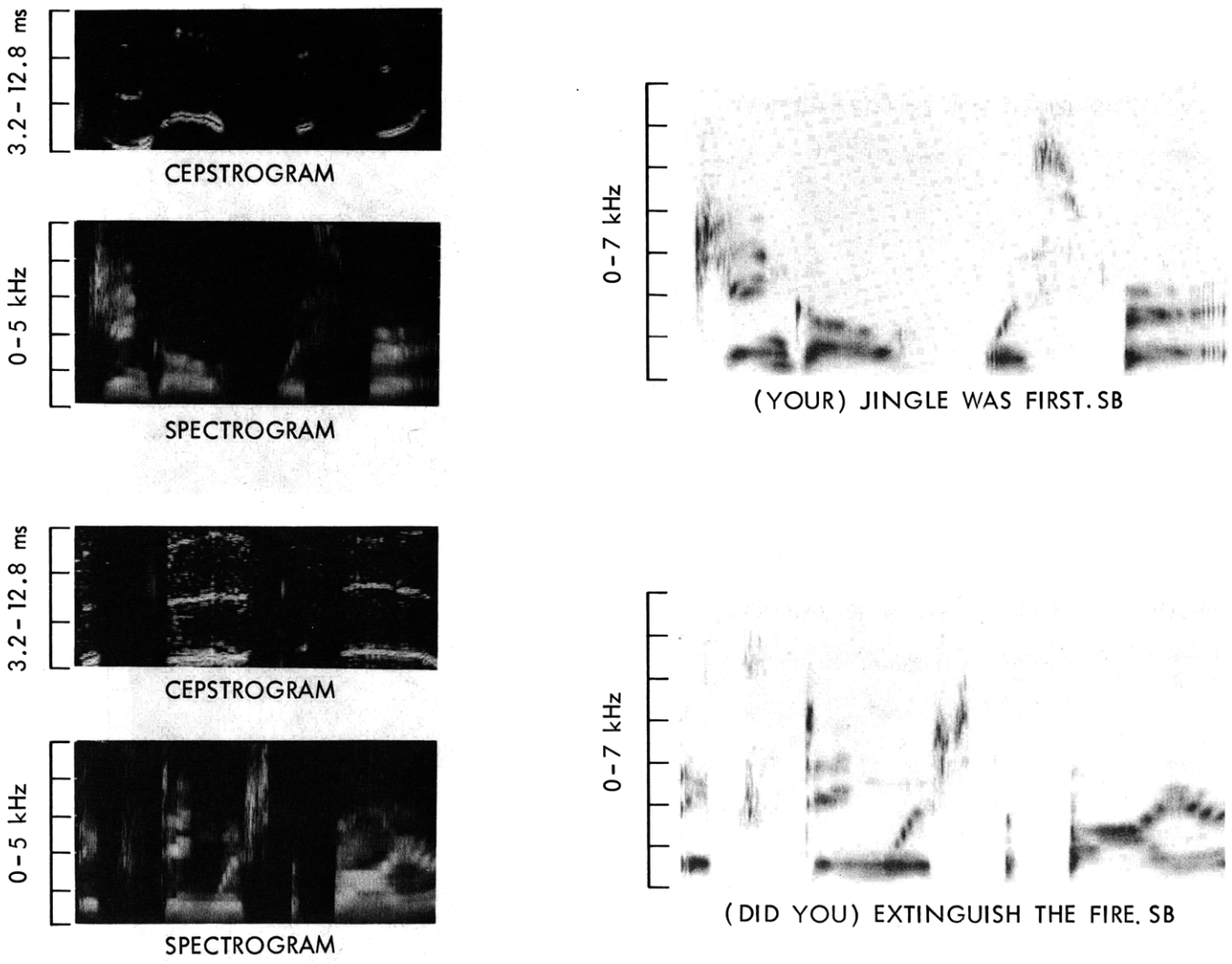

Fig. VIII-15. Spectrograms and cepstrograms.

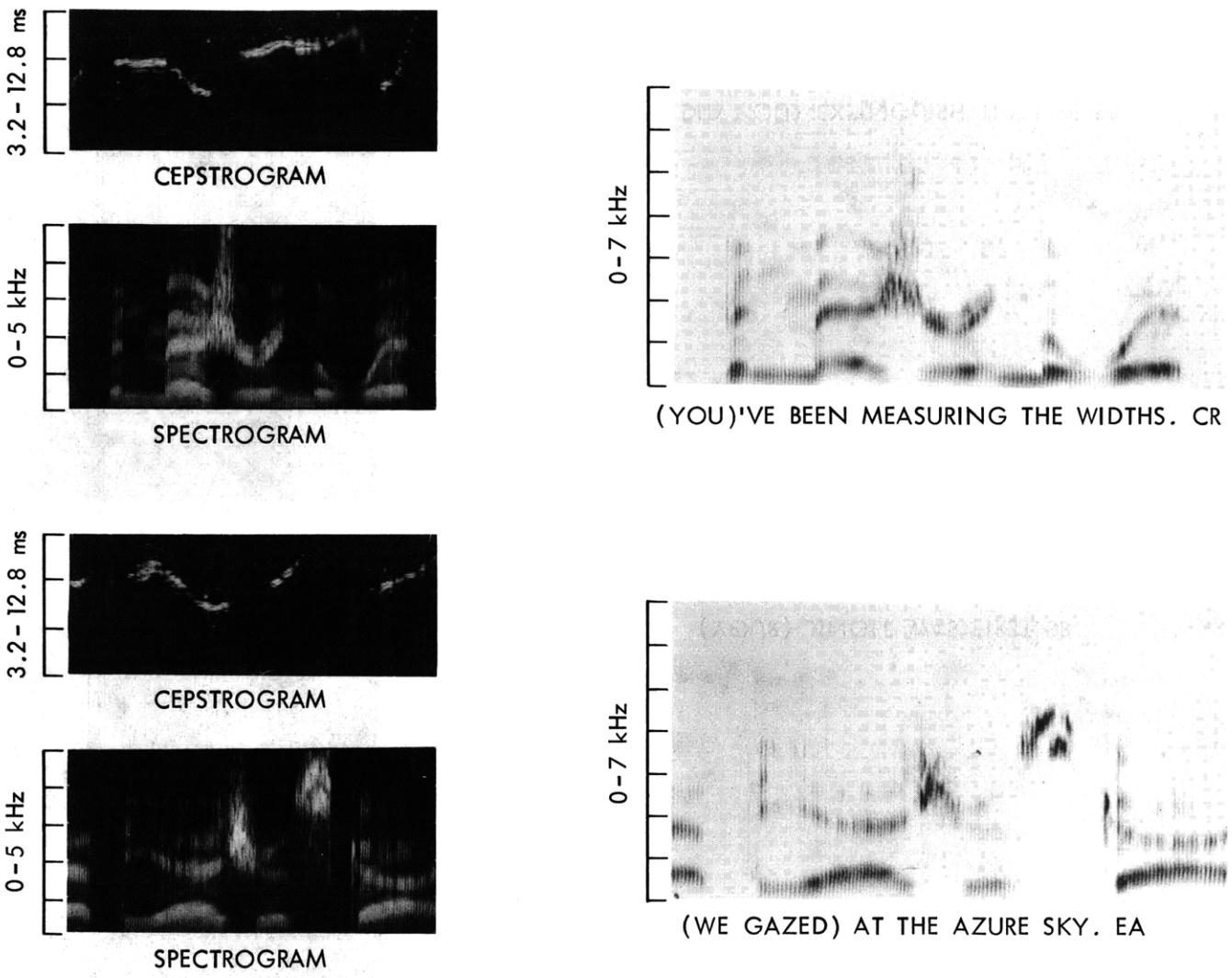

Fig. VIII-16. Spectrograms and cepstrograms.

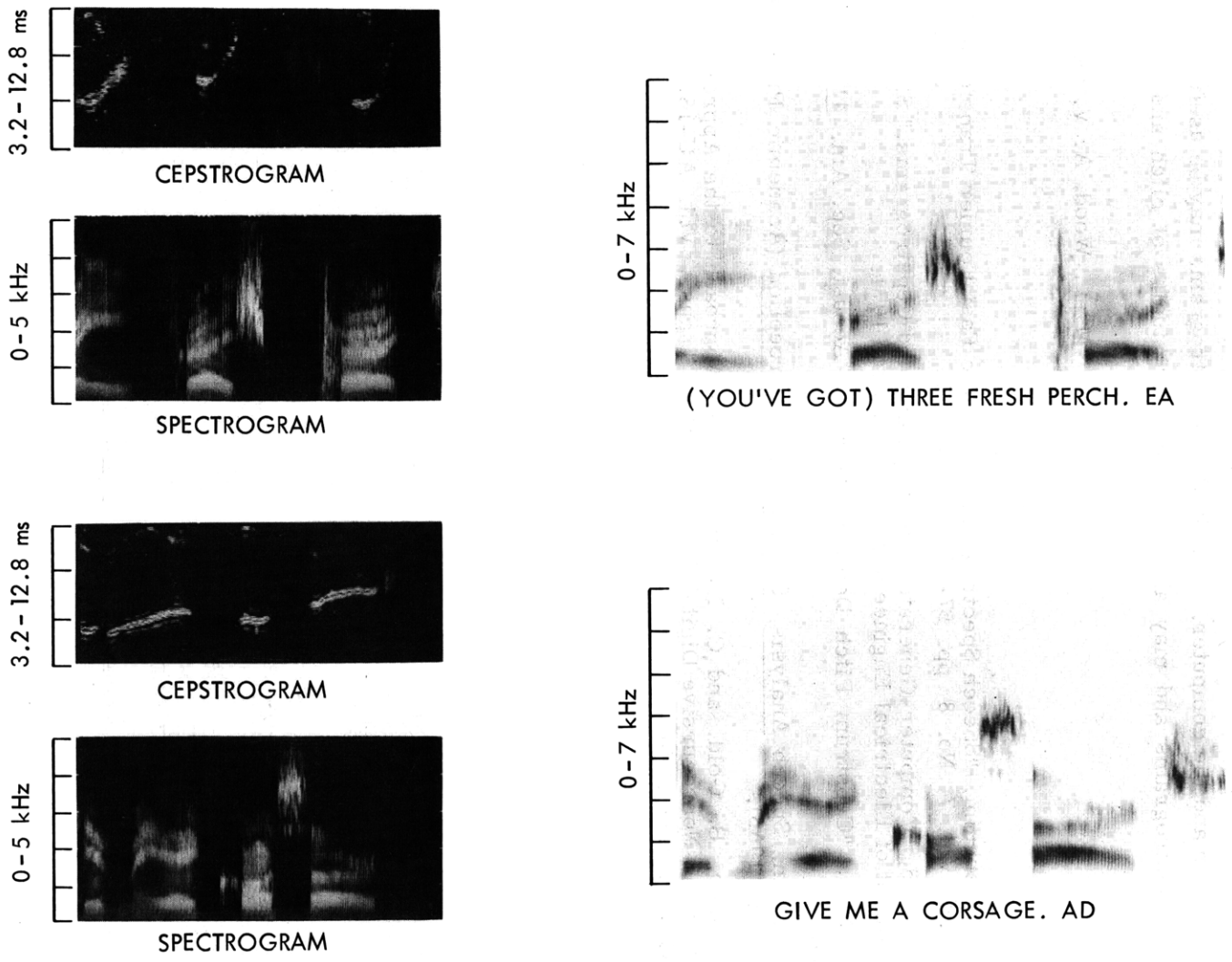

Fig. VIII-17. Spectrograms and cepstrograms.

spectrograms can be made, and the computer flexibility is an advantage over analog processes. With faster computers or special-purpose digital hardware and higher speed displays, real-time generation is possible. This could greatly increase the interaction between the user and the computer. Potentially, cepstrograms may be useful as an aid in reading spectrograms and may also be a tool for studies of pitch and voicing in language.

M. L. Wood, A. V. Oppenheim

References

1. A. V. Oppenheim, "Speech Spectrograms Using the Fast Fourier Transform," IEEE Spectrum, Vol. 7, No. 8, pp. 57-62, August 1970.
2. M. L. Wood, "Computer-Generated Spectrograms and Cepstrograms," S.M. Thesis, Department of Electrical Engineering, M.I.T., June 1971.
3. A. M. Noll, "Cepstrum Pitch Determination," J. Acoust. Soc. Am. 41, 293-309 (1967).
4. J. Flanagan, Speech Analysis, Synthesis and Perception (Academic Press, Inc., New York, 1965).
5. L. Rabiner, B. Gold, and C. McGonegal, "An Approach to the Approximation Problem for Nonrecursive Digital Filters," IEEE Trans., Vol. AU-18, No. 2, pp. 83-106, June 1970.