DOCUMENT RESUME

ED 041 280 24 AL 002 490

AUTHOR Naeser, Margaret A.
TITLE Criteria for the Segmentation of Vowels on Duplex Oscillograms.
INSTITUTION Wisconsin Univ., Madison. Research and Development Center for Cognitive Learning.
SPONS AGENCY Department of Health, Education, and Welfare, Washington, D.C. National Center for Educational Research and Development.
REPORT NO TR-124
BUREAU NO BR-5-0216
PUB DATE Mar 70
CONTRACT OEC-5-10-154
NOTE 22p.; Report from the Project on Language Concepts and Cognitive Skills Related to the Acquisition of Literacy

EDRS PRICE EDRS Price MF-$0.25 HC-$1.20
DESCRIPTORS *Acoustic Phonetics, *Language Research, *Phonology, Spectrograms, Speech, *Vowels
IDENTIFIERS *Duplex Oscillograms

ABSTRACT

This paper develops criteria for the segmentation of vowels on duplex oscillograms. Previous vowel duration studies have primarily used sound spectrograms. The use of duplex oscillograms, rather than sound spectrograms, permits faster production (real time) at less expense (adding machine paper may be used). The speech signal can be more spread out on a duplex oscillogram than on a sound spectrogram, increasing ease of segmentation; duplex oscillograms provide equally clear displays for speech of high- or low-fundamental frequency. Segmentation criteria are developed for /p, b, s, z/ occurring in initial and final position from the sound spectrogram segmentation criteria reported by Peterson and Lehiste (1960). Vowel duration measurements segmented from 64 sound spectrograms and 64 duplex oscillograms of the same CVC utterances are compared. The two sets of measures correlate .97, indicating that the criteria for the segmentation of vowels on duplex oscillograms presented in this paper are reliable. (Author/JD)

CRITERIA FOR THE SEGMENTATION OF VOWELS ON DUPLEX OSCILLOGRAMS

WISCONSIN RESEARCH AND DEVELOPMENT
CENTER FOR
COGNITIVE LEARNING

RESEARCH
COGNITIVE
LEARNING
DEVELOPMENT

Technical Report No. 124

CRITERIA FOR THE SEGMENTATION OF VOWELS ON DUPLEX OSCILLOGRAMS

by Margaret A. Naeser

Report from the Project on Language Concepts and Cognitive Skills Related to the Acquisition of Literacy

Robert C. Calfee and Richard L. Venezky, Principal Investigators

Wisconsin Research and Development
Center for Cognitive Learning
The University of Wisconsin
Madison, Wisconsin

March 1970

Published by the Wisconsin Research and Development Center for Cognitive Learning, supported in part as a research and development center by funds from the United States Office of Education, Department of Health, Education, and Welfare. The opinions expressed herein do not necessarily reflect the position or policy of the Office of Education and no official endorsement by the Office of Education should be inferred.

Center No. C-03 / Contract OE 5-10-154

STATEMENT OF FOCUS

The Wisconsin Research and Development Center for Cognitive Learning focuses on contributing to a better understanding of cognitive learning by children and youth and to the improvement of related educational practices. The strategy for research and development is comprehensive. It includes basic research to generate new knowledge about the conditions and processes of learning and about the processes of instruction, and the subsequent development of research-based instructional materials, many of which are designed for use by teachers and others for use by students. These materials are tested and refined in school settings. Throughout these operations behavioral scientists, curriculum experts, academic scholars, and school people interact, insuring that the results of Center activities are based soundly on knowledge of subject matter and cognitive learning and that they are applied to the improvement of educational practice.

This Technical Report is from the Language Concepts and Cognitive Skills Related to the Acquisition of Literacy Project in Program 1. General objectives of the Program are to generate new knowledge about concept learning and cognitive skills, to synthesize existing knowledge, and to develop educational materials suggested by the prior activities. Contributing to these Program objectives, this project's basic goal is to determine the processes by which children aged four to seven learn to read, examining the development of related cognitive and language skills, and to identify the specific reasons why many children fail to learn to read. Later studies will be conducted to find experimental techniques and tests for optimizing the acquisition of skills needed for learning to read. By-products of this research program include methodological innovations in testing paradigms and measurement procedures; the present study is an example.

ACKNOWLEDGMENTS

The author wishes to thank the Principal Investigators of the Language Concepts Study (Project No. 104) Robert C. Calfee and Richard L. Venezky for their helpful suggestions in this study. Special thanks for inspiration and diligence go to Robin S. Chapman. I thank Professor Jeanette Johnson Harries for her guidance and the Linguistics Department for providing the instrumentation. The advice of Dr. Fred D. Minifie of the Communicative Disorders Department was most helpful and I thank Richard M. Merson, a graduate student of that department, for his help in the segmentation research. For the many correlation checks I am grateful to Edward H. Haertel and for the transcription checks, Donna H. Riddel. I would also like to thank the eight participants in the study from the Wisconsin Research and Development Center.

CONTENTS

LIST OF FIGURES

ABSTRACT

The paper develops criteria for the segmentation of vowels on duplex oscillograms. Previous vowel duration studies have primarily used sound spectrograms. The use of duplex oscillograms, rather than sound spectrograms, permits faster production (real time) at less expense (adding machine paper may be used). The speech signal can be more spread out on a duplex oscillogram than on a sound spectrogram, increasing ease of segmentation; duplex oscillograms provide equally clear displays for speech of high- or low-fundamental frequency. Segmentation criteria are developed for /p, b, s, z/ occurring in initial and final position from the sound spectrogram segmentation criteria reported by Peterson and Lehiste (1960). Vowel duration measurements segmented from 64 sound spectrograms and 64 duplex oscillograms of the same CVC utterances are compared. The two sets of measures correlate .97, indicating that the criteria for the segmentation of vowels on duplex oscillograms presented in this paper are reliable.

I
INTRODUCTION

This paper will present segmentation criteria for the measurement of vowel duration on duplex oscillograms. Previous studies have primarily used sound spectrograms in measuring vowel duration;[1, 2, 3] the use of duplex oscillograms has been suggested by Fant[4] as a more economic and efficient method of obtaining visual displays for duration measurement.

In particular, duplex oscillograms have the following desirable qualities. Since they are made by a direct-writing oscillograph on paper, the process takes place continuously in real time; sound spectrograms must be made in 2.4 second segments and require a reproduce time.[5] Duplex oscillograms are made on inexpensive adding machine paper, rather than the expensive photosensitive paper used in making sound spectrograms. These properties conserve both time and money for the investigator. In addition, duplex oscillograms may present more discriminable segmentation cues than the conventional sound spectrograms. For example, when maximum paper speed (20 cm/sec.) is used on the Oscillomink,* the speech signal is more spread out on a duplex oscillogram (1 mm = .005 sec.) than on a sound spectrogram (1 mm = .0075 sec.). Also, the duplex oscillograms of a female speaker can be read as easily as those of a male speaker, a statement which cannot be made for sound spectrograms, see pages 4-8.[6] Thus, the process of making duplex oscillograms is fast, inexpensive, spreads the speech signal out clearly on a fine time scale, and may be used with equal success on speakers with high- or low-fundamental frequencies.

One must still ask, however, what segmentation criteria are to be used for vowel duration and whether these criteria are as reliable as those developed for sound spectrograms.

*Siemens, 1966 model.

II
DUPLEX OSCILLOGRAMS

Because the Peterson and Lehiste[2] sound spectrogram segmentation criteria are the bases for the segmentation criteria in this study, a short explanation concerning the differences between sound spectrograms, plain oscillograms and duplex oscillograms is given below.

A sound spectrogram displays the speech signal on a frequency (ordinate)/time (abscissa) plot, where the frequency range is generally 85-8000 Hz. A plain oscillogram displayed on a conventional oscilloscope shows the speech signal on an amplitude (ordinate)/time (abscissa) plot. The positive and negative parts of the amplitude wave are separated by the zero line. A duplex oscillogram as written out by an Oscillomink also shows the speech signal on an amplitude (ordinate)/time (abscissa) plot. The duplex oscillogram, however, differs from the plain oscillogram in that it makes separate use of the positive and negative parts of the amplitude wave. To obtain a duplex oscillogram, the speech signal must first be filtered. A direct-writing Oscillomink has a frequency response of 1000 Hz. Thus, a filtering device must filter the speech signal so that the negative part of the amplitude wave is replaced by the rectified function of the speech signal above 1000 Hz. The high-frequency sounds (above 1000 Hz.) appear clearly on the duplex oscillogram as marked negative dips below the zero line. This additional difference in amplitude display makes segmentation of a duplex oscillogram easier than that for a plain oscillogram.[4] For comparison of duplex oscillograms and sound spectrograms, see Figures 1-5 on pages 4-8.

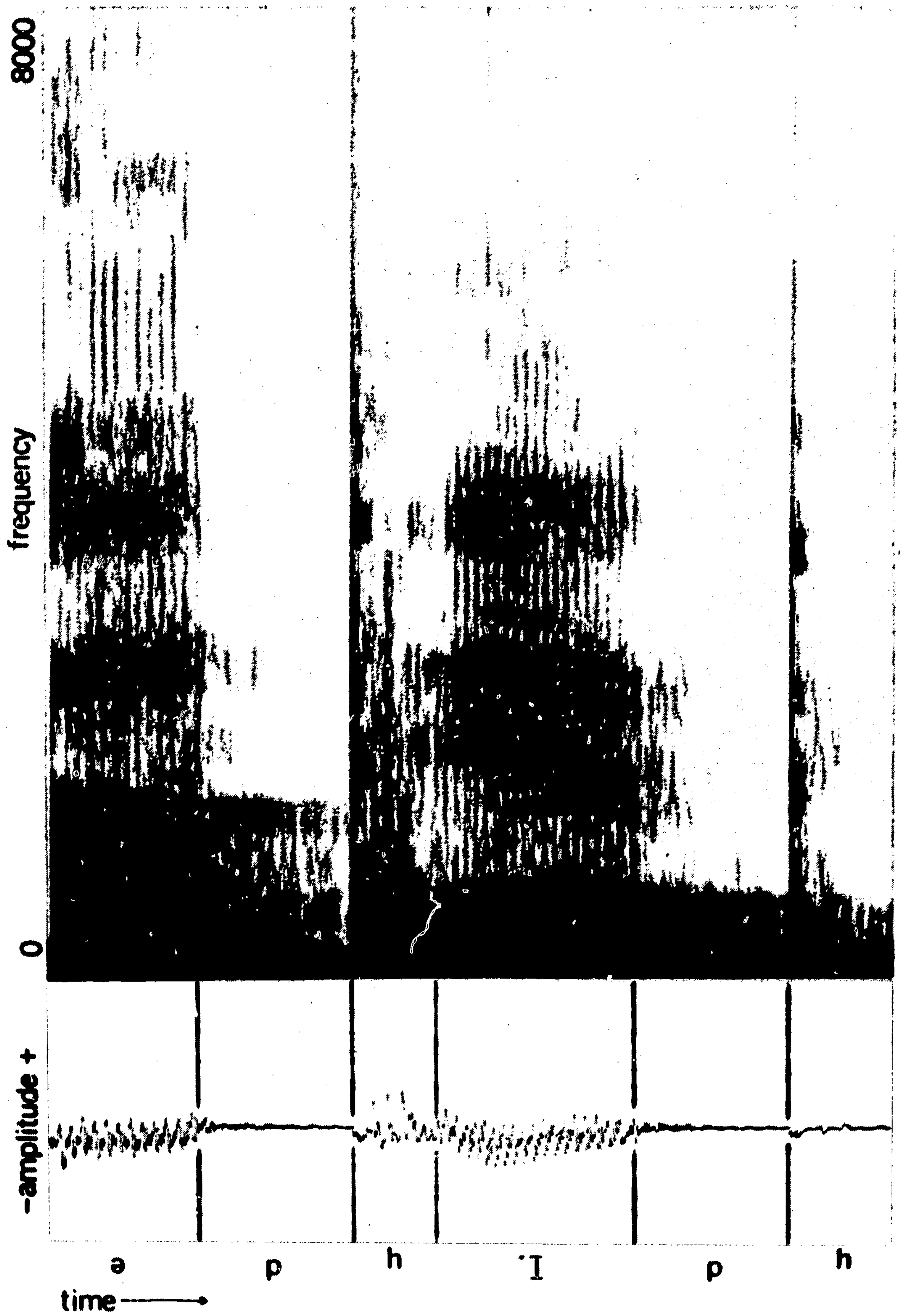

Figure 1. Male speaker. Sound spectrogram (enlarged) above, duplex oscillogram (real time 20 cm/sec) below.

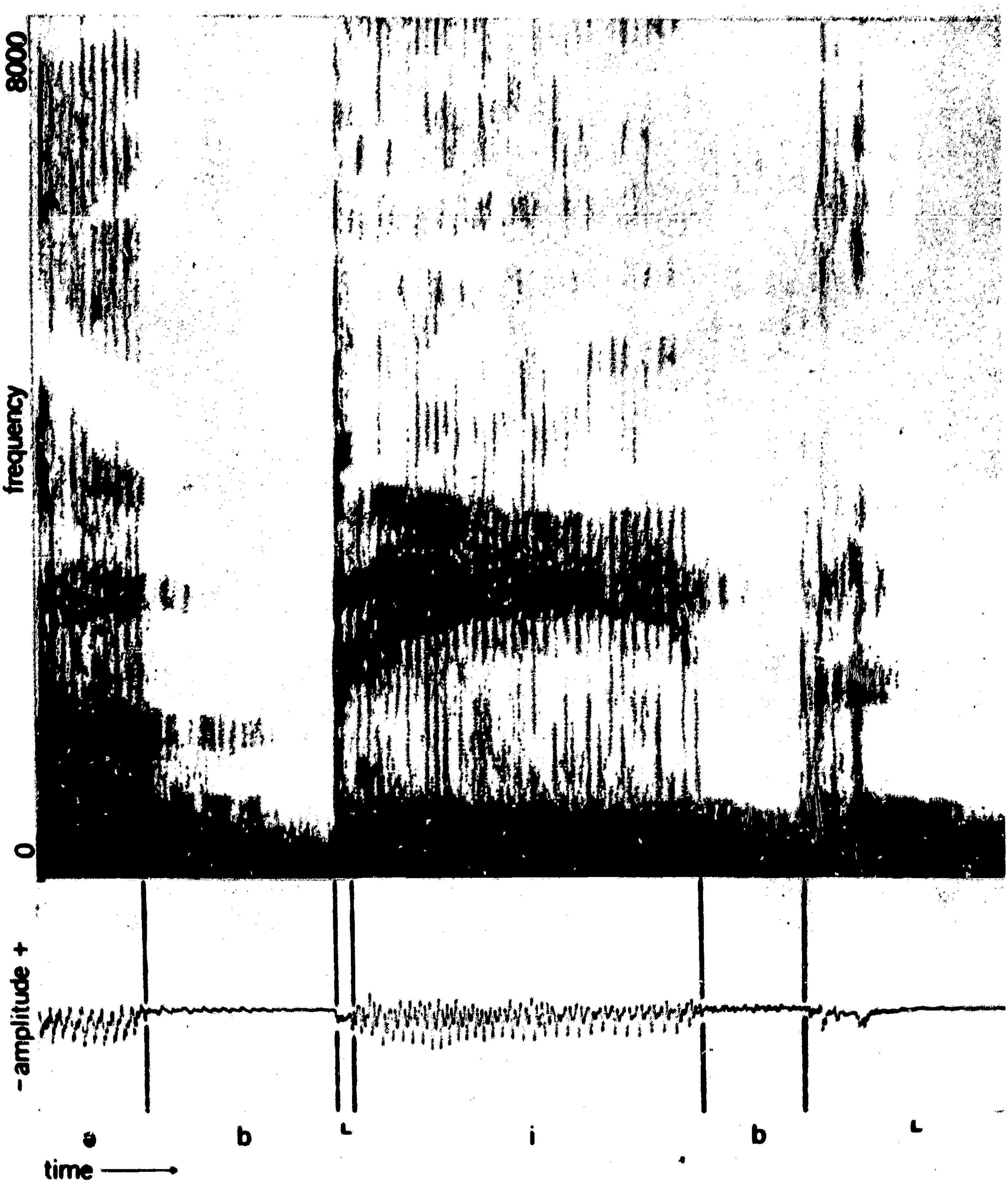

Figure 2. Male speaker. Sound spectrogram (enlarged) above, duplex oscillogram (real time 20 cm/sec) below. (L = release of voiced stop).

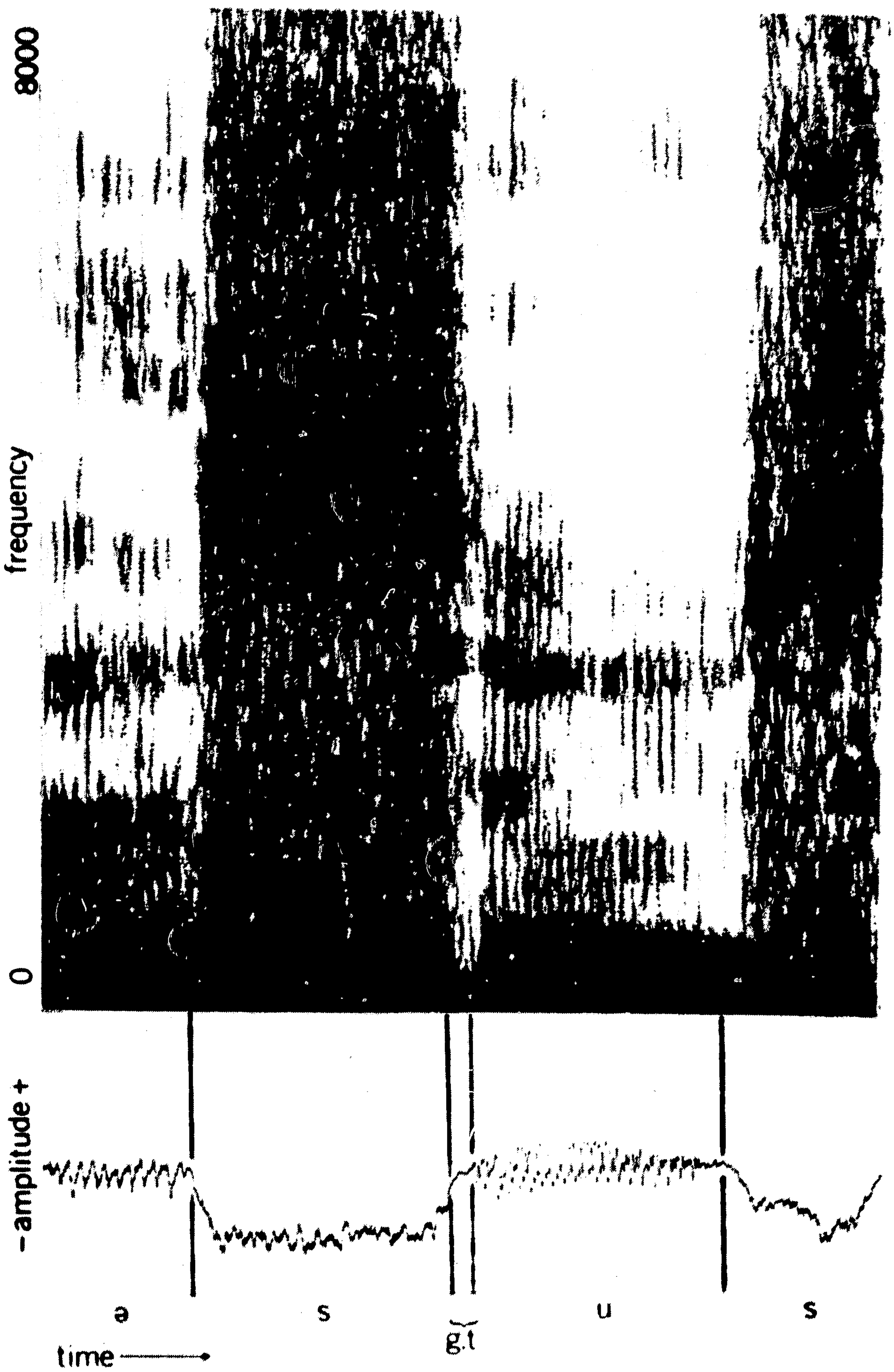

Figure 3. Male speaker. Sound spectrogram (enlarged) above, duplex oscillogram (real time 20cm/sec) below. (g.t. = glottal transition).

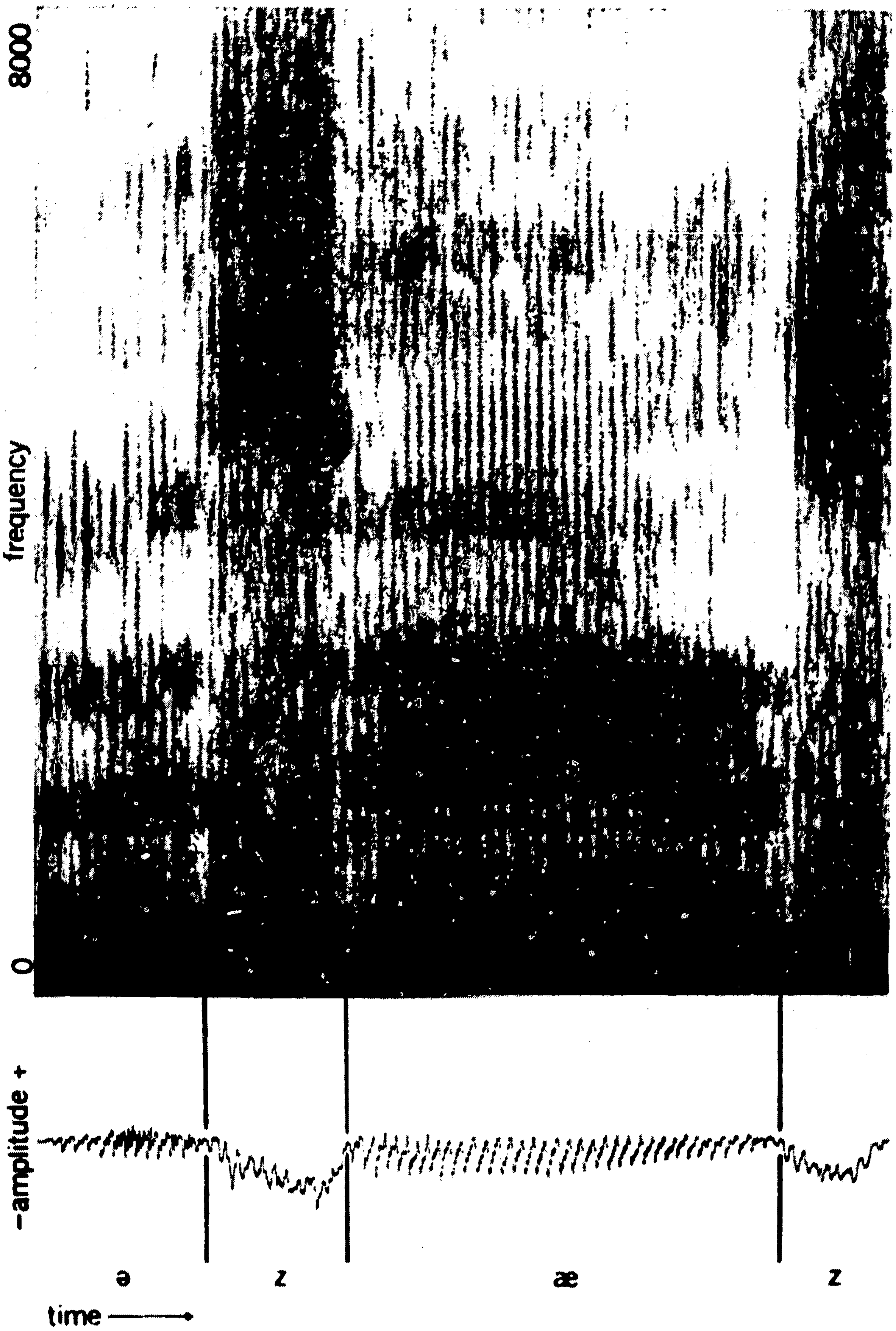

Figure 4. Male speaker. Sound spectrogram (enlarged) above, duplex oscillogram (real time 20 cm/sec) below.

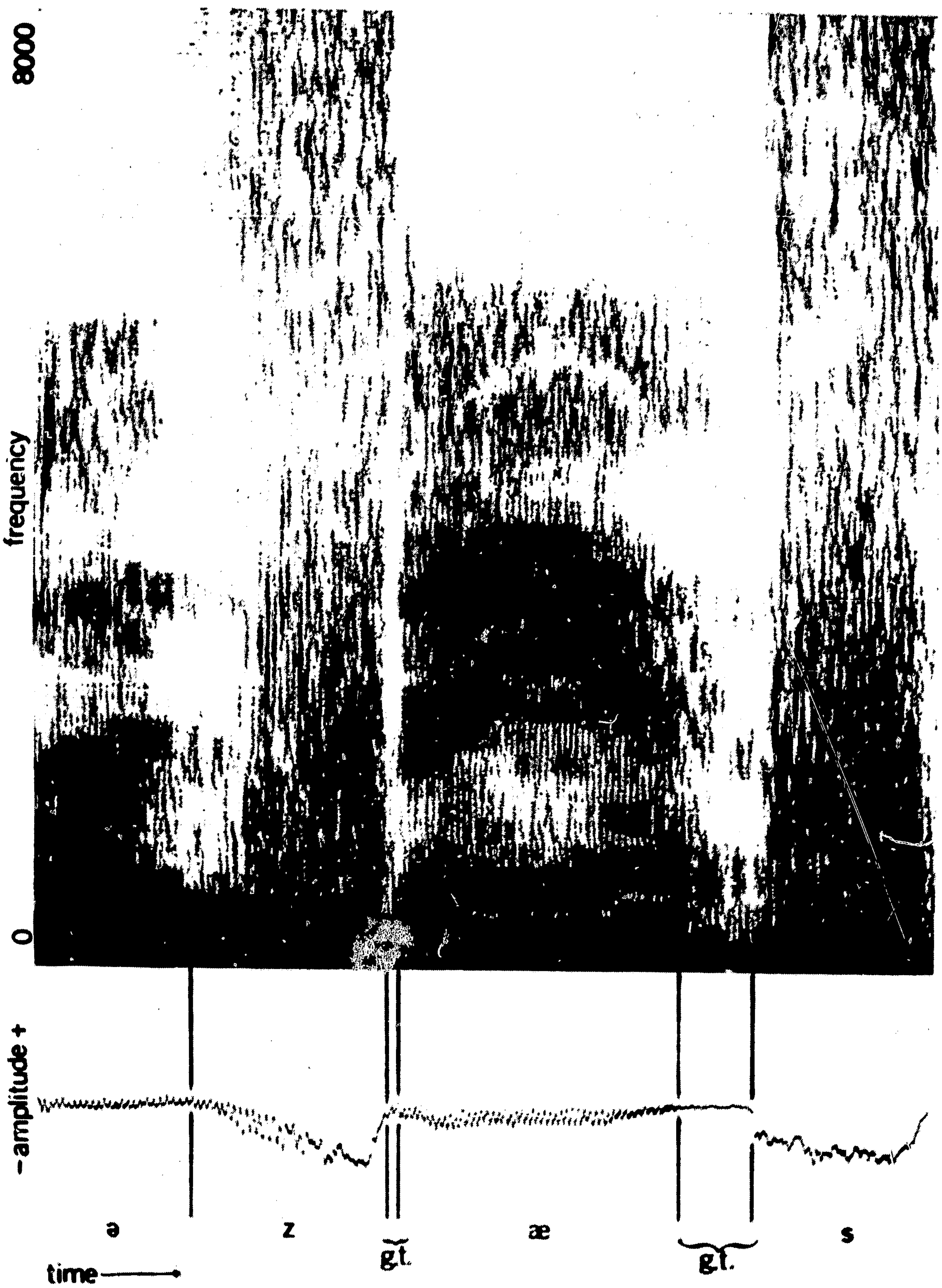

Figure 5. Female speaker. Sound spectrogram (enlarged) above, duplex oscillogram (real time 20 cm/sec) below. (g.t. = glottal transition).

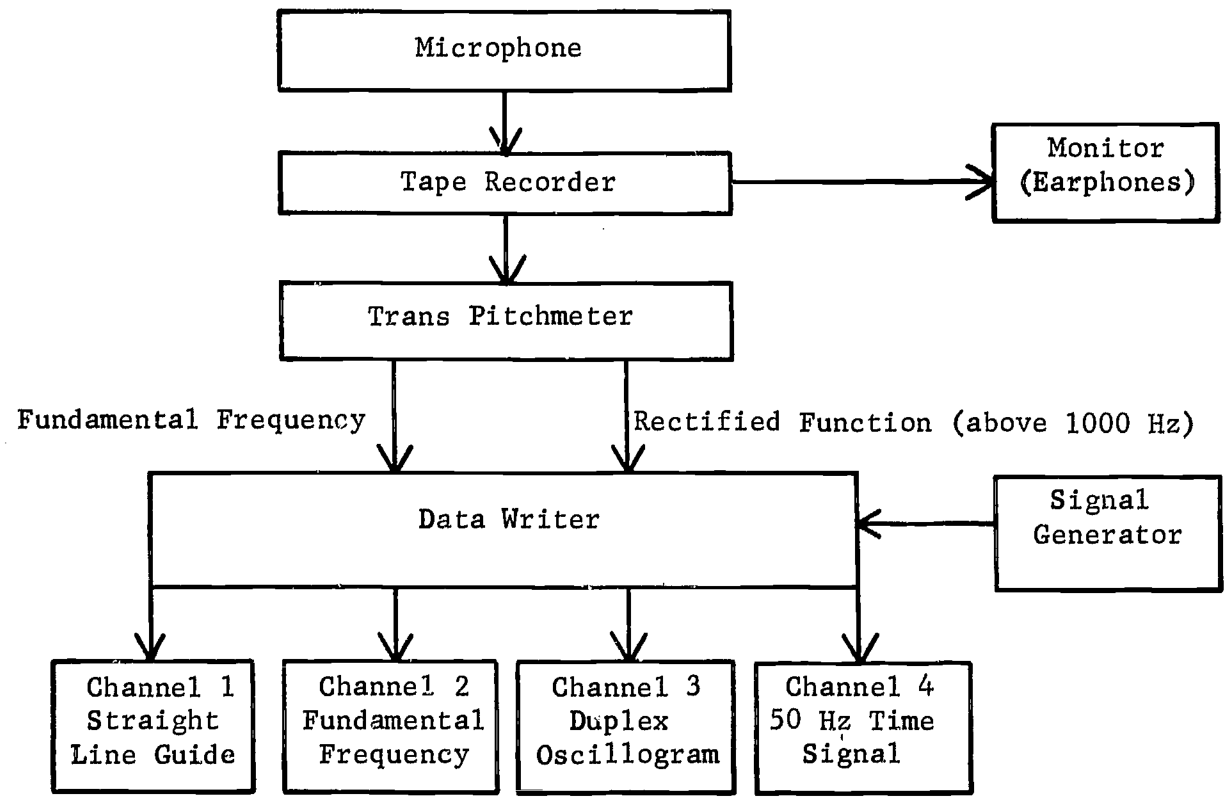

Figure 6. Generalized chart. Production of Duplex Oscillograms

III
SEGMENTATION CRITERIA[7]

VOWELS - GENERAL

The periodic vibrations of vowels on duplex oscillograms are characterized by sharp-pointed amplitude peaks. These points contrast with those of the aperiodic vibrations of consonants where the ends of the amplitude peaks are rounded. Each vowel charted in this study /ɪ/, /i/, /u/, /æ/ has its own characteristic pattern (Figures 1-4). The amplitude patterns for /æ/ for the two speakers, male and female, are quite different in their shapes (Figures 4 and 5).

VOWELS AFTER INITIAL CONSONANTS

Initial Voiceless Stop

An initial voiceless stop on the sound spectrogram (SS) is marked by a blank space (Figure 1). On the duplex oscillogram (DO) it is marked by no deflection from the zero line. The release is marked by a spike on the SS and by a short negative dip on the DO. The concentration of high-frequency fricative energy throughout the aspiration period, marked by a negative dip in the DO is not counted in the vowel duration. Vowel duration after a voiceless stop is measured from the first patterned[8] amplitude deflection.

Initial Voiced Stop

An initial voiced stop in a SS is marked by the presence of the voice bar in an otherwise blank space (Figure 2). In the DO it is marked by little or no deflection from the zero line; there may be ripples in the amplitude pattern, however, which distinguish the voiced stop from the voiceless stop. In SS the release of the stop is marked by a short frication period (energy present across all frequencies) and in DO by a short negative dip. The release period is not measured as a part of the vowel duration, which begins with the first patterned amplitude deflection.

Initial Voiceless Fricative

In SS the beginning of a vowel after an initial voiceless fricative is determined by the onset of voicing in the first formant of the vowel (Figure 3). Sometimes the end of the fricative noise does not coincide with the onset of the vowel and there is an interval which the author has termed glottal transition (Figure 3). The onset of vowel duration begins with a complete striation across two or more formants in the SS. The fricative noise in the higher frequencies of the SS, as mentioned with the aspirated release of stops above, shows up as a large negative dip in the DO. The first patterned deflection of the vowel amplitude after this negative dip marks the beginning of the vowel duration. If the glottal transition period is present, the deflection line hovers around zero before starting the vowel pattern; this transition is not counted as part of the vowel duration. (Peterson & Lehiste do not mention a glottal transition period. Apparently in their study the cessation of fricative noise coincided with onset of vowel formant activity.)

Initial Voiced Fricative

The segmentation criteria for the initial voiceless fricative applies also to the initial voiced fricative. In SS, the cessation of noise and beginning of vowel striations across two or more formants is considered to be the beginning of the vowel duration (Figure 4). In DO the onset of vowel duration is taken from the first patterned deflection of vowel amplitude registration. If a glottal transition is present before the vowel, it is not measured as part of the vowel duration.

VOWELS BEFORE FINAL CONSONANTS

Final Voiceless Stop

In SS the final voiceless stop is marked by the abrupt cessation of all formants (Figure 1). The final striation which can be found in two or more vowel formants is considered to be the end of the vowel; sometimes, however, the first formant does follow through the voiceless stop. In DO, this abrupt cessation of formant energy is manifested by a sudden leveling out of the deflection line. The vowel duration measurement terminates with the end of the patterned deflection, which is often before complete silence (shown by no deflection of the zero line on the DO). In DO the aspirated release of the final stop is marked by a negative dip from the zero line similar to, but of less distance, than the dip after the initial voiceless stop.

Final Voiced Stop

The same criteria for a final voiced stop are applied for a final voiceless stop. In SS, when there is a cessation of formant activity across most of the formants above the first, the vowel is considered to be terminated (Figure 2). In DO the termination of a vowel is marked by a sudden decrease in amplitude and absence of a patterned amplitude deflection as shown by a general leveling off of the deflection line. Some ripples of amplitude deflection around the zero line continue through the voiced stop. The release, if any, is marked by a short negative dip.

Final Voiceless Fricative

In SS the termination of a vowel before a voiceless fricative is marked by the sudden onset of random noise in the high frequencies (Figure 3). In DO the beginning of high frequency energy is marked by the sudden onset of a large negative dip in the deflection line. However, there are times when actual fricative onset is not sudden. Peterson and Lehiste[2] consider the vowel to be terminated at the point where the noise pattern begins, even though voicing in a few low harmonics often continues for a few centiseconds. It is this short continuation of periodic vibrations which sometimes makes vowel duration measurement before a final fricative difficult in DO. The most satisfactory cues for terminating the vowel in this position are 1) the onset of negative dip or 2) the end of the periodic pattern in the amplitude wave; i.e., when the amplitude wave becomes irregular and begins to dip into the negative half of the DO. (It may still retain a positive segment, however.) If a glottal transition is present here, it is counted as part of the vowel duration, not the fricative noise, as only the onset of fricative noise signals the end of the vowel and beginning of the fricative.

Final Voiced Fricative

The same criteria in segmenting vowels before a final voiceless fricative are applied to segmenting them before a final voiced fricative. In SS the onset of high frequency enery is considered to be the beginning of a voiced fricative (Figure 4). In DO the negative dip marks the onset of the fricative and the end of the vowel. If the aperiodic vibrations precede the negative dip, they signal the end of the vowel and beginning of the voiced fricative. If a glottal transition is present where the deflection line hovers over the zero line, before onset of the negative dip, it is considered to be a part of the vowel duration just as with final voiceless fricatives.

GPO 817-309-2

IV
RELIABILITY TESTING

The segmentation criteria given above were developed from 512 duplex oscillograms obtained in the following manner. Eight adult speakers (4 male, 4 female), each repeated 64 items (Appendix) from a pre-recorded tape. The items consisted of all possible CVC combinations of the 4 consonants /p/, /b/, /s/, /z/ and 4 vowels /ɪ/, /i/, /u/, /æ/. High-quality sound equipment was used for the recording.[9] Duplex oscillograms were made of these utterances on a 4 channel Siemens Oscillomink (1966 model). Channel 1 was used for a straight line which served as a guide for lining up a right triangle for segmentation purposes; Channel 2, for the fundamental frequency as filtered out by a fundamental frequency extractor, the Trans Pitchmeter (B. Frøkjaer-Jensen, Denmark); Channel 3, for the duplex oscillogram filtered out by the Trans Pitchmeter. A 50 Hz. sine wave time signal from a Hewlett Packard signal generator (200 CD Wide Range Oscillator) was displayed on Channel 4. The tape was played back on a Sony tape recorder (TC-777-4) at the recording speed (7 1/2 ips.). The utterances were monitored through earphones as they were processed by the data writer and phonetic transcriptions were entered on the output tape (see Figure 6).

The vowel duration on duplex oscillograms was measured to the nearest .5 mm. A clear plastic ruler with .5 mm markings and a .5 mm 4H lead pencil were used in marking the actual segmentation lines.

The reliability of the duplex oscillogram segmentation was tested against sound spectrograms. Sixty-four sound spectrograms (Set A) (8 items for each of the 8 speakers) were segmented by the author using the Peterson & Lehiste sound spectrogram segmentation criteria. A set of the same 64 sound spectrograms were segmented by two other persons (Sets B and C).[10] These segmentations correlated with Set A at .97 and .99. Sixty-four duplex oscillograms (Set D) of the same items were then segmented by the author using the duplex oscillogram segmentation criteria developed from sound spectrogram segmentation as reported above.

The vowel duration measurements taken from the duplex oscillograms (Set D) correlated with those vowel duration measurements taken from the sound spectrograms (Set A) at .97. The question asked in this study—Can duplex oscillograms be segmented as reliably as sound spectrograms?—is thus answered affirmatively.

V
SUMMARY

The reliable segmentation criteria of duplex oscillograms as developed in this study may be summarized as follows:

1. Vowels are marked by patterned periodicity and sharp points of the amplitude peaks. Vowel duration begins where a steady amplitude pattern begins and terminates, before stop consonants, where the patterned deflection stops. Vowel duration terminates before fricative consonants where the onset of the fricative is obvious.

2. Stop consonants are marked by little or no deflection around the zero line. The aspirated release of a voiceless stop is marked by the large negative dip while the release of a voiced stop is marked by a small negative dip.

3. Fricative consonants are marked by a long, deep negative dip of the deflection line.

4. Glottal transitions between a vowel and a fricative are marked by a hovering of the deflection line around zero. They are included in vowel duration measurements only when they occur after the vowel pattern and before the final consonant pattern.

APPENDIX

Each word repeated in the frame sentence: "A ________ is a word."

Randomized List of 64 CVCs

(C→/p, b, s, z/; V→/ɪ, i, u, æ/)

1. pɪp ____________
2. siz ____________
3. bus ____________
4. zæb ____________
5. bip ____________
6. sɪs ____________
7. zub ____________
8. pæz ____________
9. sup ____________
10. zɪs ____________
11. pib ____________
12. bæz ____________
13. sɪp ____________
14. pæb ____________
15. buz ____________
16. zis ____________
17. sæp ____________
18. puz ____________
19. zib ____________
20. bɪs ____________
21. pip ____________
22. zæs ____________
23. bɪb ____________
24. suz ____________
25. bæp ____________
26. zɪz ____________
27. nus ____________
28. sib ____________
29. zup ____________
30. biz ____________
31. sæb ____________
32. pɪs ____________
33. zip ____________
34. sæs ____________
35. bɪz ____________
36. pub ____________
37. bis ____________
38. pæp ____________
39. zuz ____________
40. sɪb ____________
41. bup ____________
42. zæz ____________
43. pɪb ____________
44. sis ____________
45. zɪp ____________
46. piz ____________
47. bæb ____________
48. sus ____________
49. sɪz ____________
50. pis ____________
51. bub ____________
52. zæp ____________
53. bib ____________
54. sæz ____________
55. pup ____________
56. zɪb ____________
57. bæs ____________
58. sub ____________
59. ziz ____________
60. bɪp ____________
61. pæs ____________
62. sip ____________
63. zus ____________
64. pɪz ____________

FOOTNOTES

1. House, A. S. & Fairbanks, G. "The Influence of Consonant Environment upon the Secondary Acoustical Characteristics of Vowels," Journal of the Acoustical Society of America, 1953, 25, 128-136.

2. Peterson, G. E. & Lehiste, I. "Duration of Syllable Nuclei in English," Journal of the Acoustical Society of America, 1960, 32, 693-703.

3. House, A. S. "On Vowel Duration in English," Journal of the Acoustical Society of America, 1961, 33, 1174-78.

4. Fant, C. G. M. "Modern Instruments and Methods for Acoustic Studies of Speech," in Proceedings of the Eighth International Congress of Linguists, Oslo University Press, Oslo, 1958, p. 326.

5. All references to sound spectrograms in this paper are to those made on a Kay Electric Sonagraph (6061A), a model widely used in speech research throughout the U.S.

6. Because of the limited band width of the sound spectrograph filters, the sound spectrograms of a female speaker with a high-fundamental frequency display less energy than those of a male speaker with a low-fundamental frequency. Because duplex oscillograms are displayed on an amplitude/time plot, a speech signal with a high-fundamental frequency presents no special problem in the production of clear duplex oscillograms.

7. Because the duplex oscillogram segmentation guidelines are based on the sound spectrogram criteria, both sets are listed together.

8. When looking at vowel amplitude, it is important to look at the whole general pattern of the vowel amplitude first and to begin the vowel duration measurement from where the general steady pattern is established. Thus, pattern in this paper refers to established pattern for that particular vowel segment and not to transition pattern which may occur briefly in the beginning of the vowel.

9. Ampex tape recorders (1100), Ampex microphone (2001).

10. A professor and a graduate student from the Communicative Disorders Department.

GPO 817-309-1