DOCUMENT RESUME

ED 030 859 AL 001 898

By-Chase, Richard Allen; And Others

Teaching New Vowel Sounds Using Real-Time Spectral Displays.

Johns Hopkins Univ., Baltimore, Md. Neurocommunications Lab.
Spons Agency-National Inst. of Child Health and Human Development, Bethesda, Md.
Pub Date 68
Note-20p.; Paper published in the 1968 Annual Report, Neurocommunication Laboratory, Dept. of Psychiatry, Johns Hopkins University, Baltimore, Md. 21218.
EDRS Price MF-$0.25 HC-$1.10
Descriptors-Acoustic Phonetics, Age Differences, Articulation (Speech), *Pronunciation Instruction, *Spectrograms, *Visible Speech, *Visual Learning, Vowels
Identifiers-*Visible Speech Translator, VST

The primary objective of this study was to find out if young (normal-hearing) children could be taught a novel vowel sound by means of visual information alone, that is, without benefit of auditory presentation of the sound or instructions on the shaping of the vocal tract for its production. A second objective was to find out whether the rate and nature of learning by means of visual information were different for children and young adults. A modified version of the Visible Speech Translator developed by Bell Telephone Laboratories was used in the study. The subjects participating consisted of ten young adult males and ten children, males and females, between the ages of four and five years old. The vowel /y/ which occurs in the phonological systems of French and German but not in American English was selected for learning. All subjects showed evidence of learning in the context of the experiment. In every case, marked modification of vocalization patterns in the direction of the novel vowel was obtained in the first 100 trials. Most subjects continued to demonstrate improvement through the remainder of the training sequence. No striking differences in the performance of adults and children were obtained. (DO)

TEACHING NEW VOWEL SOUNDS USING REAL-TIME SPECTRAL DISPLAYS

Richard Allen Chase

Richard L. Mobley

Rachel E. Stark

The preparation of this paper was supported by Contract PH 43-65-637 from the National Institute of Child Health and Human Development, National Institutes of Health.

TEACHING NEW VOWEL SOUNDS USING REAL-TIME SPECTRAL DISPLAYS

Introduction

The recent growth in information about the acoustic structure of speech, and in technology for analysis and display of speech signals, has led to an increased interest in the use of visual displays for purposes of teaching speech motor gestures. It has been suggested that the auditory system has evolved unique capabilities for processing the speech signal (2). However, the possibility that speech information can be processed by the visual system has yet to be investigated fully.

Some attempts have been made to evaluate the efficiency of visual displays in teaching speech to the deaf. The Bell Telephone Laboratories have developed a Visible Speech Translator (VST) capable of generating real-time spectrographic displays of speech. The work of Kopp and Kopp with this instrumentation (1) and studies in this laboratory using a modified form of the VST (5) have demonstrated that such displays are moderately useful in modifying the speech patterns of deaf children. The use of other types of display for this purpose has been covered in recent reviews (3, 4).

Objectives

The primary objective of this study was to find out if young, normal-hearing children could be taught a novel vowel sound by means of visual information alone, that is, without benefit of auditory presentation of the sound or instructions on the shaping of the vocal tract for its production.

A second objective was to find out whether the rate and nature of learning by means of visual information were different for children and young adults.

Instrumentation

The modified Visible Speech Translator used in this study employs 18 parallel filter channels to obtain the amplitude spectrum of speech sounds. These filters are arranged on a semi-compressed scale spanning a frequency range from approximately 125 to 5000 Hz. Filter bandwidths vary from 125 Hz at the lower end of the scale to 600 Hz for the uppermost channels. The display of spectral information is achieved by means of the cathode ray tube of a Tektonix RM 564 Storage Oscilloscope. This provides a rectangular spectrogram of approximately 5 X 4 inches, with time represented along the abscissa and frequency along the ordinate in a conventional manner. Signal amplitude at any particular time-frequency coordinate is registered by presence or absence of the trace. This binary representation of intensity is accomplished by comparing the output of a particular filter channel with a fixed reference, and unblanking the oscilloscope writing beam for all signal levels above the reference. Thus, a simplified dot-pattern representation of the spectrographic information is achieved. The anaylsis and control functions of the Visible Speech Translator are shown in the flow diagram of Figure 1.

Pilot Studies

The vowel /y/ was selected for all of the studies to be reported. This vowel occurs in the phonological systems of French and German but not in American English. The simplicity of the spectrographic representation of the vowel made it suitable for preliminary experiments. However, it was not

at first clear whether the subject should be required to produce the vowel in isolation or in C-V-C monosyllables. Control of the duration of the vowel would be imposed if a C-V-C syllable were chosen. In addition, this would be closer to real speech.

Work with a few subjects demonstrated that the task associated with use of a C-V-C syllable was extremely difficult. When the subject was concentrating on achieving a match with certain features of the spectral pattern, he would lose control over other features. In addition, it was difficult for the subject to distinguish those disparities between the spectrographic display of his utterance and the model that were of real significance, and those which represented minor variations in consecutive vocalizations. We therefore decided to work with the vowel in isolation.

For the vowel /y/, the model initially consisted of three horizontal regions corresponding to the first three formants. The areas were defined by inspection of the correct production of the vowel by young adult males. These studies indicated that if subjects were asked to vocalize the vowel /i/ and the vowel /u/, the horizontal space bounded by the lower limit of the F-2 region for /i/ and the upper limit of the F-2 region for /u/ would define an approximate F-2 region for the vowel /y/. Since there is not a great deal of difference between the F-1 and F-3 regions for the three vowels, these regions were covered with an opaque sheet of paper, leaving only the F-2 regions visible.

Plan of Experiment and Procedures

The subjects participating in this experiment consisted of ten young adult males and ten children, males and females, between the ages of four and

five years. The ages and sex of each subject are shown in Tables 1 and 2. The young adult subjects were paid volunteers, and the children were obtained, with the consent of their parents, from the population of The Johns Hopkins Hospital Nursery School. All subjects were in good general health, and gave no indication, by history, of impaired auditory or visual acuity. In addition, historical information was obtained that would allow us to screen out prospective subjects who had marked language disabilities.

All subjects were brought into a sound treated room, and seated on a stool adjusted in height so that the display was at eye level. They were told that, as they made different speech sounds, different patterns of light would appear on the screen. Each subject was encouraged to experiment to verify this relationship for himself. The subjects were then asked to provide five to ten consecutive samples of the vowel sound /i/ followed by five to ten consecutive samples of the vowel sound /u/. The experimenter allowed the spectrograms of consecutive vocalizations to be overlaid one on another.

The experimenter placed a clear plastic screen over the cathode ray tube, and outlined the F-2 regions for /i/ and /u/. Strips of translucent blue plastic were then glued over these regions. The intervening horizontal space was clear. The F-1 and F-3 regions for the three vowels were masked with opaque paper. Thus, the clear space of the training templet defined the F-2 region for /y/ (see Figure 2).

The adult subjects were told that they were to experiment with the production of vowel sounds in order to discover one which would put most of the dots in the middle clear region, while at the same time removing them

completely from the lower blue region, and as much as possible from the upper blue region. For a few of the subjects it was necessary to give these instructions in three stages, i.e., first they were instructed to put most of the dots in the middle clear region, secondly to continue to do so and at the same time remove the dots from the lower blue region, and thirdly to continue as in the second stage but at the same time to remove the dots from the upper blue region. An attempt was made to obtain five hundred vocalizations from each of these subjects. All vocalizations were recorded, utilizing a Bruel and Kjaer 4131 microphone and a Tandberg 74B tape recorder. The tape recorder was located in an adjoining control room, and operated by an assistant who was able to view the subject through a window, and to hear the subject's performance and the experimenter's instructions through an intercommunication system. The assistant announced the count of the subject's vocalizations through this system. The subject was instructed to wait a few seconds after each count and then to produce his next vocalization, which remained on the screen for 3 to 5 seconds before being erased. In this manner, a new vocalization was obtained approximately every 30 seconds. When a total of fifty vocalizations had been produced, the subject was allowed a two-minute rest period before continuing with the next block of fifty vocalizations.

The child subjects were told that they were to make "little sounds" like /i/ and /u/ but different in the sense that these would not be sounds that they would ordinarily make. They were told simply that these were to be sounds that would place light behind the clear region of the screen. After they accomplished this, they were told to change the sound they were making in

order to remove the lights from the bottom blue region while at the same time keeping them in the middle clear region. If progress could be made with respect to both of these requirements, the child was then told to try to remove some of the lights from the upper blue region.

The children were allowed to work at their own pace, and to play when they complained of fatigue or were obviously distracted. The number counting system of cueing vocalizations was not used with the children because they ignored the procedure almost completely. The experimenter encouraged the child to look at the visual display corresponding to each vocalization. The displays were left on the screen for 3 to 5 seconds as in the case of the adults. It usually was not possible to obtain five hundred vocalizations from the children. The children's vocalizations were recorded in the same manner as the adults' vocalizations.

Analysis of Data

The Visible Speech Translator was used for analysis of the vocalization data. Every fifth vocalization was played from a magnetic tape through the VST filters and the resulting outputs were examined. Each of the filter channels was represented on the horizontal axis of the cathode ray tube, and amplitude was represented on the vertical axis. The maximum filter excitation corresponding to the second formant was identified and plotted for each sample vocalization. In many instances the vocalizations were non-steady state and therefore did not lend themselves to this type of analysis. In these cases, the complete range of filters exhibiting substantial excitation was noted. An

attempt was made to decide whether the vocalization in question represented a diphthong or a change in lip rounding.

Results

Data has been analyzed for the first three children described in Table 1 and the first three young adults described in Table 2. The graphs shown in Figure 3 display the peak-energy F-2 center frequency for every fifth successive vocalization for each of the three children. Comparable data displays for each of the three young adult subjects are shown in Figure 4. The following are notes describing the performance of each of these subjects.

Subject C-1

A total of 420 vocalizations was obtained in the course of three test sessions with interposed play sessions. Approximately equal numbers of vocalizations were obtained in each of the three test sessions. During the first 150 vocalizations most of the peak-energy F-2 frequencies were in the F-2 region for /u/. During the next 150 vocalizations a non-steady state vowel characterized by changes in lip rounding was used. This was progressively differentiated until a new non-steady state vowel was stabilized. This fulfilled the visual pattern match requirements of the experiment extremely well. Most of the vocalizations obtained during the last session sounded like excellent approximations to the /y/ vowel. It was possible for the subject to vocalize the novel vowel without benefit of the visual displays immediately following the termination of the experiment.

Subject C-2

A total of 125 vocalizations was obtained in three test sessions with interposed play sessions. Approximately equal numbers of vocalizations were

obtained in each test session. This subject showed the most rapid pattern of learning of all the subjects studied in this experiment. Within the first five vocalizations he obtained excellent control over his ability to put energy into the novel F-2 region defined by the training templet. Thereafter, he maintained a fairly stable pattern of vocalization that fulfilled the visual pattern match criteria extremely well, and sounded like a good approximation to the novel vowel /y/. This subject produced a non-steady state vowel by changing lip rounding. His experimentation with changes in lip rounding was very obvious during the course of the experiment, despite the fact that no instructions were given about the usefulness of this procedure. This subject frequently showed evidence of boredom, and required a fair amount of verbal reinforcement to produce a relatively small sample of vocalizations. On a few occasions, the subject did not monitor his vocalizations by inspecting the visual displays. This behavior was associated with marked departure of the vocalization pattern from the criteria.

Subject C-3

A total of 275 vocalizations was obtained in two successive test sessions with an interposed play session. During the first 150 vocalizations the subject experimented with a wide range of speech sounds. There was progressive increase in the frequency with which signal appeared in the novel F-2 region. When told to try to keep the lights in the middle clear region while at the same time removing them from the lower blue region, the subject produced increasing numbers of sounds that fulfilled the visual pattern match criteria quite well and sounded like good approximations to the /y/ vowel. The peak-

energy F-2 frequencies showed the widest scatter during early stages of the experiment, when it was more difficult for the subject to re-establish the visual pattern on departing from this, than in the later stages of the experiment. As the experiment progressed, longer sequences of vocalizations were produced which met the visual pattern match criteria satisfactorily.

Subject A-1

During the first 50 trials the subject searched very widely through the English vowels and minor variations of the English vowels. Early in the second series of 50 vocalizations he discovered a sound that fulfilled the visual pattern match criteria quite well, and he continued to make minor modifications of this sound until, by the end of the second series of 50 vocalizations, he had fulfilled the visual pattern match criteria extremely well. In subsequent trials, there was progressive stabilization of his ability to do so. However, even by the end of 500 trials there were occasional departures from the target position, primarily in the direction of the F-2 region for /u/. During the early portions of the experiment the subject experimented actively with changes with vocal tract shape which he referred to as "making nasal sounds" and "pursing the lips". Acoustic analysis of this subject's vocalizations demonstrated increasing frequency of steady state vowels with peak-energy F-2 frequencies exactly centered in the novel F-2 region. It should be noted that this subject had one-half year of French in the eighth grade. He claims, however, that his experience with spoken French was minimal.

Subject A-2

During the first series of 50 vocalizations the subject experimented with a wide variety of English vowel sounds and minor modifications of English vowel sounds. During the course of this exploration, the subject found a vowel sound that partially satisfied the visual pattern match criteria. He proceeded to effect minor modifications of this vocalization and succeeded in fulfilling the visual pattern match criteria more satisfactorily as the experiment progressed. The subject stated that he concentrated on the movements of lips and tongue that resulted in good visual pattern matches. He observed that when he fulfilled the visual pattern match criteria well, his tongue was usually "raised in the back and lowered in the front", and that the lips were "open and puckered". It was only during the later stages of the experiment that the subject was aware of having attended to the sound of his vocalizations. This subject felt that he relied most heavily on visual pattern information for learning the new vowel sound.

Subject A-3

During the first 25 vocalizations this subject learned how to place energy in the novel F-2 territory with fair reliability. This was preceded by brief experimentation with modifications of the English vowels. During the course of the experiment increasingly longer strings of vocalizations were produced which fulfilled the visual pattern match criteria extremely well. The departures from the desired visual pattern became progressively less marked, and were accompanied by increasing ease of return to the correct visual pattern. This subject felt that he made maximal use of the

visual displays during the course of learning the new vowel sound. Auditory feedback from the vocalizations that resulted in good visual target approximations was also found to be useful, but to a lesser degree. This subject frequently experimented with lip rounding, but was for the most part unaware of this.

Discussion

All of the subjects for whom we have data demonstrate evidence of learning in the context of the experiment. In every case, marked modification of vocalization patterns in the direction of the novel vowel was obtained during the first 100 trials. Most subjects continued to demonstrate improvement through the remainder of the training sequence. No striking differences in the performance of adults and children were obtained.

We feel that these observations are indicative of the effectiveness of a spectrographic display in learning a new vowel sound without benefit of an auditory model. It is to be noted, however, that the visual pattern displayed by the training templet constitutes only a portion of the total spectrum for the speech sound being taught. The subject is provided information about the F-2 region of the novel vowel (target) and about F-2 of /i/ and /u/ (error regions). In contrast, "identification" of /i/, /u/, and /y/ spectrographically requires, as a minimum, the display of both F-1 and F-2. This implies that the informational requirements for efficient learning and identification of speech sounds via spectrographic displays may be quite different.

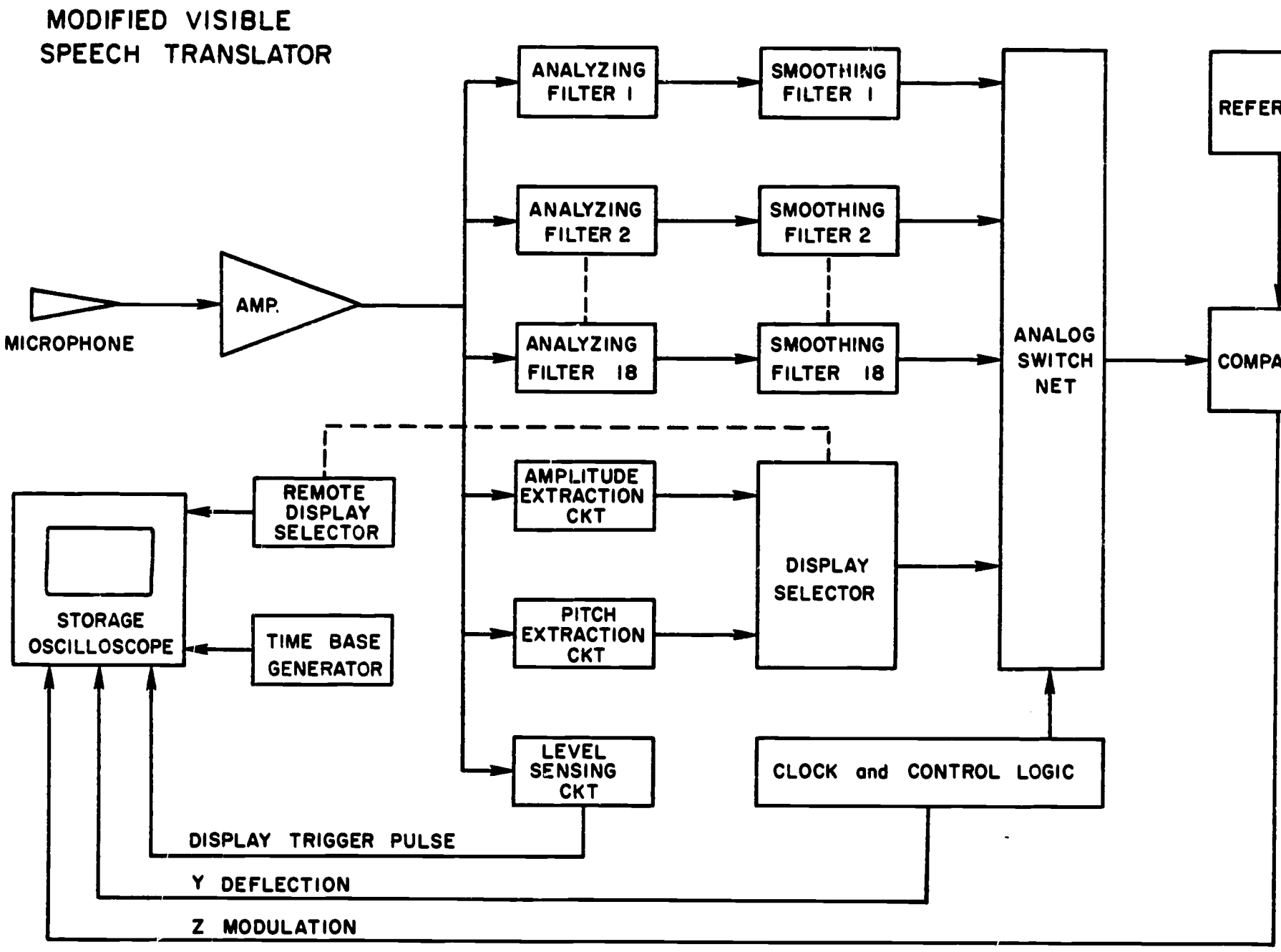

Figure 1. Flow diagram of the modified visible speech translator.

Subject	Age	Sex	Nationality and Race	Languages spoken By Parents or Siblings	Languages Studied	Evidence of Language Disability	Subjective Appraisal of New Learning
C-1	4	F	American - W	English		No	+++
C-2	4	M	American - W	English		No	+++
C-3	5	F	American - W	English, French		No	+++
C-4	5	M	American - W	English		No	+++
C-5	5	M	American - W	English		No	+
C-6	4	M	American - W	English		No	+
C-7	4	F	American - N	English		No	++
C-8	5	M	American - N	English		No	+
C-9	4	F	American - W	English		No	++
C-10	4	M	American - N	English		No	+

Table 1. Background information about the children studied in this experiment. The subjective appraisals of new learning were made by one experimenter in the course of the actual experimental session. Minimal evidence of new learning is represented by (+), and moderate and marked evidence by (++) and (+++) respectively.

Subject	Age	Sex	Nationality and Race	Languages Spoken by Parents or Siblings	Languages Studied	Evidence of Language Disability	Subjective Appraisal of New Learning
A-1	21	M	American - W	English	French (1/2 yr. 8th grade)	No	+++
A-2	23	M	American - W	English	Spanish (2 yrs. H.S., 2 yrs., Col.), Latin (2 yrs. H.S.)	No	+++
A-3	20	M	American - W	English, Yiddish	Hebrew (5 yrs. grade school) Spanish (5 yrs. H.S.)	No	+++
A-4	25	M	American - W	English	Spanish (2 yrs. H.S.)	Dyslexic	+
A-5	30	M	Irish - W	Gaelic	English, Latin (through H.S.)	No	++
A-6	19	M	American - W	English	Spanish (4 yrs. H.S.) German (6 mos. Col. no conversation)	Occasional Stuttering	++
A-7	25	M	American - N	English	Spanish (1-1/2 yrs. H.S.)	Past History of Stuttering	+
A-8	22	M	American - W	English	Latin (5 yrs. H.S., 1 yr. Col.)	No	++
A-9	30	M	Chinese	Cantonese	English (10 yrs. H.S. and Col.), German (1 yr. Col.)	No	+++
A-10	23	M	American - N	English	None	Past History of articulation disorder	+

Table 2. Background information about the young adults studied in this experiment. The subjective appraisals of new learning were made by one experimenter in the course of the actual experimental session. Minimal evidence of new learning is represented by (+), and moderate and marked evidence by (++) and (+++) respectively.

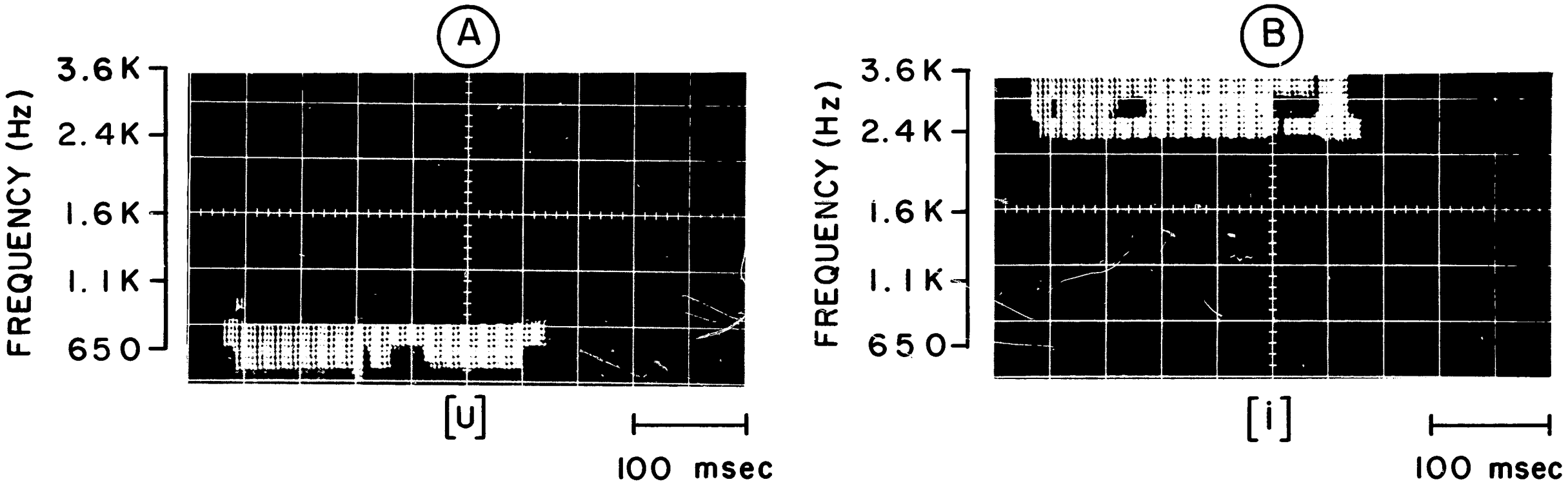

Figure 2. Photographs of the storage oscilloscope screen showing the appearance of F-2 for the vowels /i/(A), /u/(B), /y/(C) and the vowel diphthong /iu/(D). The training templet is shown over the oscilloscope screen, and the F-2 regions for /i/ and /u/ have been overlaid with horizontal strips of blue plastic. The vertical bars to the right of each photograph show the height of each of these blue regions. All of the speech samples displayed were obtained from the same young adult speaker.

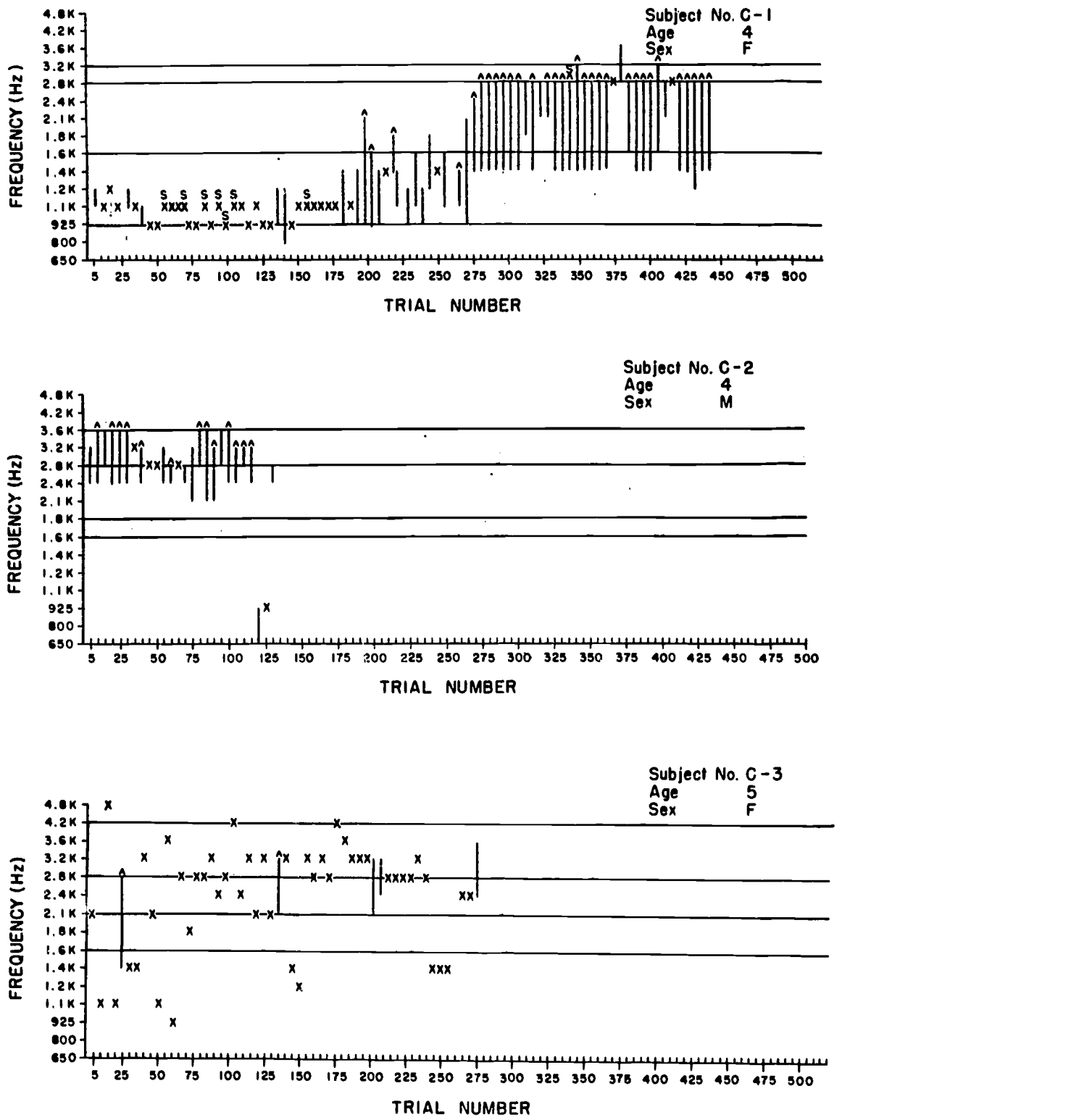

Figure 3. Graphs showing the vocoder channel center frequency of F-2 of every fifth vocalization for each of three children. In the case of vocalizations that are not steady state vowels, the F-2 range is indicated by a vertical bar. The dense horizontal bars define the regions of the training templet: the upper stippled area shows the F-2 region for /i/, the lower stippled area shows the F-2 region for /u/, and the center zone shows the F-2 target area for /y/.

KEY

X peak energy F-2 frequency

| vowel diphthong

↑ change in lip rounding

S shouting

N nasality

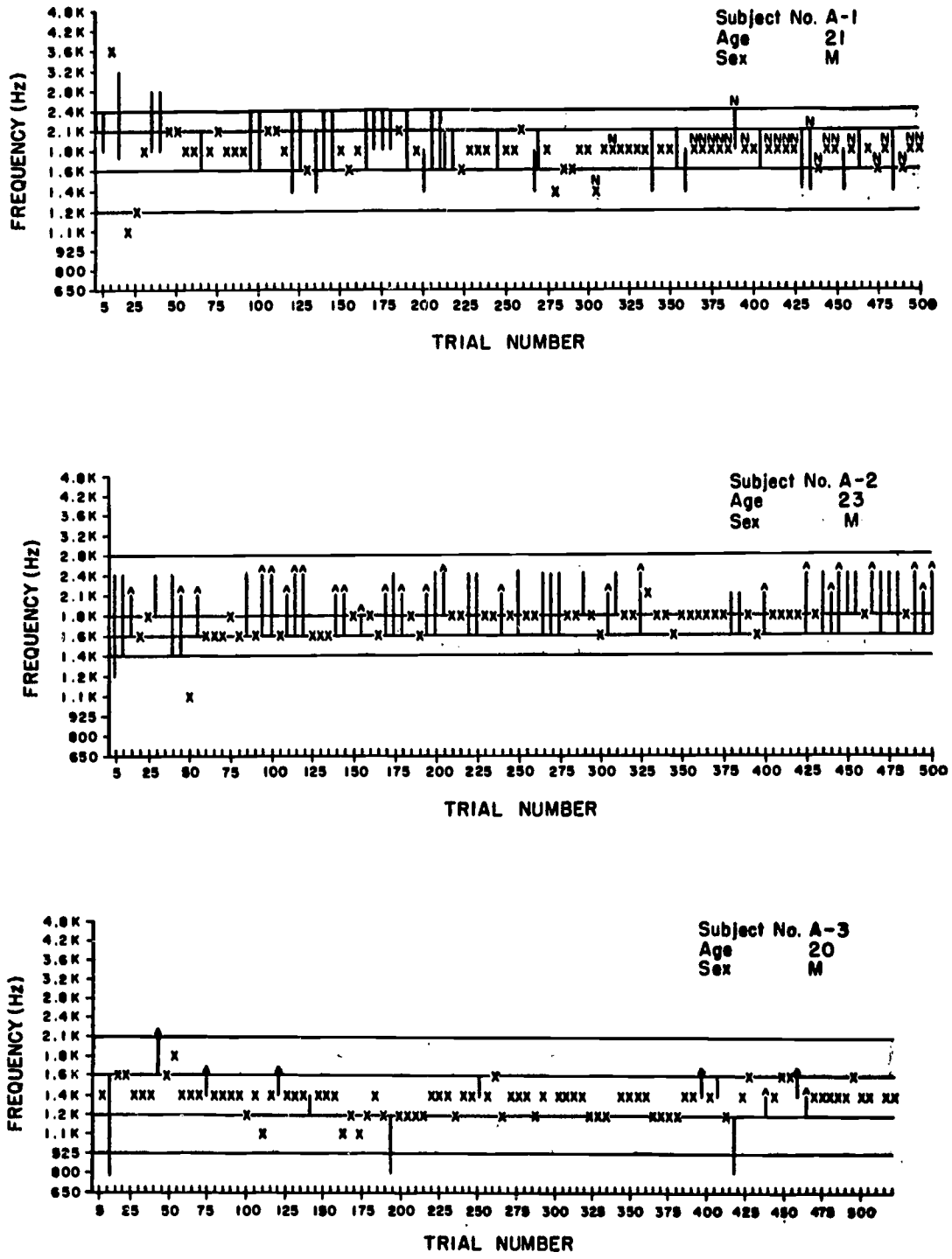

Figure 4. Graphs showing the vocoder channel center frequency of F-2 of every fifth vocalization for each of three young adults. In the case of vocalizations that are not steady state vowels, the F-2 range is indicated by a vertical bar. The dense horizontal bars define the regions of the training templet: the upper stippled area shows the F-2 region for /i/, the lower stippled area shows the F-2 region for /u/, and the center zone shows the F-2 target area for /y/.

References

1. Kopp, G. A. & Kopp, Harriet G. An investigation to evaluate usefulness of the visible speech cathode ray tube translator as a supplement to the oral method of teaching speech to deaf and severely-deafened children. Final Report, 1963, Grant Number RD-526, Office of Vocational Rehabilitation.

2. Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. Why are speech spectrograms hard to read? Amer. Ann. Deaf, 1968, 113, 127-133.

3. Pickett, J. M. Recent research on speech-analyzing aids for the deaf. IEEE Trans., 1968, AU-16, 227-234.

4. Pronovost, W. Developments in visual displays of speech information. Volta Review, 1967, 69, 365-373.

5. Stark, Rachel E., Cullen, J. K., Jr., & Chase, R. A. Preliminary work with the new Bell Telephone Visible Speech Translator. Amer. Ann. Deaf, 1968, 113, 205-214.