TEACHING PRONUNCIATION WITH VISUAL FEEDBACK

Kathleen Albertson

Abstract

Sound discrimination often becomes a frustrating stumbling block for adult language students unable to hear the differences between two contrasting sounds. The use of a modified oscilloscope, in this case the Visipitch, enables students to see the differences in pronunciation patterns they may be unable to hear. With the aid of immediate feedback, students are able to monitor and mosify the activity of their speech organs and compare their productions of specific sounds to those produced by the instructor. This instrument is useful in teaching intonation and rhythm patterns as well as most phonemes and allophones.

Key Words: Language instruction; oscilloscope; pronunciation; TESOL; visual feedback.

TEACHING PRONUNCIATION WITH AN OSCILLOSCOPE

Pronunciation training is difficult. Students are often unable to differentiate differences in the target language that were not distinctive in their native language. Whether these differences are phonemic (individual contrasting sounds such as /b/ or /d/), allophonic (slight variations of the individual sounds), or prosodic (having to do with emphasis or accent), they still affect the perceptability of what is being said or meant and are part of the total package of language learning and communication in the target language.

Instructors also may be unable to hear allophonic or prosodic features that are nondistinctive (or insignificant) in English, but are distinctive in the native language of the student. These sounds can confuse the student who can hear differences of which the instructor is unaware. When asked "Do I say it this way or that?" the instructor is often at a loss to give an answer as the same sound is heard twice. Such is the case with the aspirated and unaspirated bs and ts of Hindi. Only an instructor with some knowledge of Hindi, or with a very sensitive ear, will be able to respond appropriately to "Do I say /ball/ or /b all/?

For these reasons, coupled with a firm belief that the stimulation of multiple senses will result in better retention, a visual feedback instrument has been used as a teaching aid in English as a second language classes. This paper describes the use of the Kay Elemetrics Visipitch which enables students to see important features of their speech they are unable to hear and helps them to make appropriate modifications to errors by giving virtually immediate feedback to their attempts at duplicating the utterance of the instructor.

WHAT IS A VISI-PITCH?

The Visi-Pitch is a pitch and intensity (volume) extraction device which is mated to a Tektronix storage oscilloscope. The speech energy is received through a hand-held microphone and converted into an electrical trace which is displayed on a 4½" by 5½" cathode ray tube. (Fig. 1) The pitch and intensity values are displayed on the vertical axis, over time, which is displayed on the horizontal axis. These values can be displayed either singly, or simultaneously as seen in Figure 2. In the latter case, the pitch values appear as a solid line (Figure 2-lower trace) while the intensity values appear as a dotted line (Figure 2-upper trace). The length of the trace time can be varied from one to eight seconds so that it is possible to display a single word, or an utterance of several sentences on the width of the screen.

Because the oscilloscope has not only a split screen but also storage capabilities, the trace of the instructor's speech can be stored on one section of the screen while the student tries to match it on the other, with each new attempt erasing the former trace. Since the utterance is displayed within a fraction of a second from the time it was uttered, the feedback is virtually immediate. With practice, the students can learn to manipulate the trace with their speech mechanisms (vocal cords, lips, tongue, etc.) and thereby improve their pronunciation.

Because the Visi-Pitch is a new device and not widely used for second language training, what follows is intended as an introduction to some of the possibilities for its use in the field, and not as a total summary of the device's capabilities.

First to be discussed will be the use of the Visi-Pitch in working with the prosodic features of language commonly referred to as the accenting of syllables or words, and the intonation patterns. Next, the use of the instrument in teaching the individual sounds or phonemes will be explained. Finally, some general comments regarding the teaching of language in general using visual feedback will be made.

PROSODIC PARAMETERS

Because non-phonemic errors often cause the second language learner to be "not understood" as often as phonemic variations cause him to be "misunderstood", (Prator, 1972) it is important to deal with the parameters of phonetic duration, average speech power, and average fundamental frequency (pitch).

It is these parameters that combine to give languages their distinctive rhythms and melodies which are referred to as the prosodic aspects of language. In English, all three, duration, power and frequency, combine to form the distinctive stress pattern whereby we emphasize some syllables in the information bearing words and reduce the emphasis on other syllables within the same word, or other words such as articles and prepositions within the same utterance. As will be seen, this emphasis or stress is achieved by varying the phonetic duration, average speech power, or average fundamental frequency either singly, or in combination with each other.

PHONETIC DURATION

Phonetic duration refers to the length of time a sound is held and is easily seen when the Visi-Pitch is used with the intensity display as a lengthing of the trace on the horizontal axis.

In Figure 3, the trace for the word danger /deyn dʒər/ , shows that when this speaker produces the word, the first syllable is stressed by holding it longer than the second, thus increasing the phonetic duration. Stress from this parameter is seen frequently in words with tense vowels, often referred to as the long vowel sounds.

Displaying phonetic duration has been useful in dispelling the frequently held notion that syllables are stressed only by increasing the volume, or average speech power, with no mention of phonetic duration.

AVERAGE SPEECH POWER

Average speech power refers to the intensity with which a sound is produced and includes the power of the breath stream as well as what is commonly referred to as volume. An increase in average speech power is seen as an increase in the height of the trace on the vertical axis when the Visi-Pitch is used with the intensity display.

In the word "automobile" /ɔˊtəməbiyˋl/ , in Figure 4, the increase in average speech power can be seen on the first syllable which receives primary stress, as the trace rises significantly higher for this syllable than it does for the others. Secondary stress can be seen on the fourth syllable by the increase in phonetic duration or the lengthening of the trace on the horizontal axis.

In Figure 2 (lower trace), the independent breath stream power of the consonants is seen as small hills on either side of the vowel nucleus in the word "this".

On the sentence level, the pattern of English is to stress the information bearing words, especially those which carry new information or are used to show contrast. This rhythm can be seen in Figure 5, where the words "yours" and "mine" receive primary and secondary sentence stress. The absence of this stress pattern would cause a sentence to be perceived as little more than a string of words because American listeners are accustomed to having the more important words emphasized.

AVERAGE FUNDAMENTAL FREQUENCY

Stress

The contribution of the parameter of average fundamental frequency, or pitch, can be seen while using the pitch display either singly or simultaneously with the intensity display.

Although not much attention has been given to the contribution of average fundamental frequency to stress for instructional purposes, it is useful to note that some speakers achieve the emphasis of stress in some words, by raising the fundamental frequency rather than the average speech power. In figure 6, in this speaker's production of the word "constitution", primary stress is placed on the third syllable by raising the fundamental frequency or pitch. The average speech power and phonetic duration for the syllables is about equal. While the dual trace is sometimes difficult to read, it is often helpful to see the change in fundamental frequency in relation to the change in intensity, or average speech power.

Intonation Patterns

In addition to the function of average fundamental frequency in stress or rhythm, this same parameter is the essence of the melody of a language or its intonation patterns. When the Visi-Pitch is set in the pitch display mode, the various patterns can be seen easily. A difference can be seen between the questions which require a yes or no answer and have the rising pitch on the end usually attributed to interrogatives, in contrast to the questions which start with Wh-words and have the rising-falling pattern on the end.

In Figure 5, the first part of the sentence "Is this drink yours, or mine?" can be seen to have the rising intonation pattern usually attributed to Yes-No questions, while the second part shows the falling pattern of questions.

Of importance also is the fact that intonation patterns are perceived as indicators of attitude. When these patterns vary cross-culturally, misunderstandings can arise. A steadily falling pattern indicates to Americans that the speaker is unhappy, which in turn can indicate depression or even anger. However, this same pattern is used in French simply for declarative sentences. A happy Frenchman could therefore be falsely misunderstood as being unhappy if he carries his French intonation pattern over into English.

PHONEMIC DISPLAYS

Wave Types

In order to understand the display of the phonemes or the individual sounds that make up words, it is important to have some knowledge of the four speech wave types produced by the vocal mechanisms. The quiescent, burst, quasi-random and quasi-periodic wave types, either singly or in combination with each other, can be related to the production of the different phonemes and their allophones.

When the Visi-Pitch is used with the pitch display and in frequency setting of D, traces appear which correspond to these four wave types.

The *quiescent wave type* is displayed as a total lack of energy as the trace falls to the bottom-most position on the screen. This wave type is seen in the first portion of a stop when no energy is permitted to pass the point of maximum closure as is displayed in Figure 7 when the voiceless stop /p/ is produced in the word "speculate". The quiescent wave occurs when the lips are closed prior to the pronunciation of the following vowel sound.

The *burst wave type* appears as a small spike showing the release of a stop, as can be seen in Figure 7 when the final stop /t/ is released. As soon as the tongue is removed from the tooth area, a small amount of air escapes with a burst.

The *quasi-random wave type* is shown as a random distribution of frequency values as the trace dot moves rapidly up and down, unable to arrive at a constant value. This wave type is seen in the production of any hissing or sibilant sound such as /s/ or /f/, as well as in the production of /h/. In figure 7, a quasi-random wave type can be seen in the production of the sound of /s/.

The *quasi-periodic wave type* is displayed as a consistent frequency distributional pattern, the height of which corresponds proportionally to the distance between the tongue and the roof of the mouth. This wave type is seen in the production of all vowels and in the sonorant consonants such as /1/ and /r/. In figure 7, it can be seen as the /kyuw ley t/ portions of speculate are produced. This particular display is interesting as it shows not only the complex character of u as /kyuw/ but also the dipthongization of the sound of /a/ or /ei/.

In regard to the quasi-periodic wave type, some mention must be made of the uniqueness of each person's vocal mechanisms. Because the size and shape of tongues, oral cavities etc. vary, so too will the wave forms produced by these mechanisms. Thus, the wave form for one speaker's vowel sound in "cat" may look a little different from that of another speaker but the sound can still be correct. However, the relative variation from one sound to the other across speakers should remain somewhat constant. In other words, each person should show a difference between the various vowel sounds in proportion to the differences shown by other speakers who produce the sounds correctly, although each speaker's traces for the same sound will probably not be the same as another's.

It is for this reason that it is recommended that each student be assisted in finding a word or environment in which he can produce the sound appropriately and place that sound on the screen, so that he can see what this production of the sound looks like. He can then try to duplicate it in isolation feeling what the correct sound feels like, and later try to duplicate the same sound in other environments. In this way the student is not asked to match another person's production with equipment that is different.

CONSONANTS

Obstruants

Stops. The feature stop, or non-continuant, involves a total closure somewhere in the vocal tract. The result of this obstruction, a cessation of energy flow, will be seen on the oscilloscope as a quiescent wave form followed by a burst wave as the captured energy is released. In the production of the voiceless stops, /p/, /t/, /k/, the trace will drop to the absolute bottom of the display area during the quiescent wave, to be followed immediately by a burst wave if the aspiration is released. In the production of the voiced stops, /b/, /d/, /g/, following the quiescent wave, the trace line may rise slightly indicating that voicing has begun. The burst wave will also be smaller after the voiced consonants, or may not be present at all.

In conclusion, what is seen in the stop displays, is the manner of articulation, or the fact that the speech power is momentarily blocked, as well as whether or not voicing is continuing, but not a difference in place, or the location of the blockage. As shown in Figure 8, there is a difference to be seen between pad and dad, but not between dad and bad, due to the fact that /p/ is voiceless while /b/ and /d/ are voiced.

Fricatives. The trace for the fricative phonemes or sounds made by a partial closure somewhere in the vocal tract, allowing some energy to escape with friction, such as /f/, shows a quasi-random wave type which varies in the amount of trace concentration on the screen according to the amount of energy which is permitted to escape. Hence the trace for /f/ (Figure 9-upper trace) shows a greater concentration of energy than does the trace for /s/ (Figure 9-lower trace) as seen in Figure 9. The voiced fricatives such as /v/ show either a quasi-periodic wave type alone or in speaker. See the word "vine" in Figure 19.

Affricates. As would be expected, the affricates, or sounds which are made by combining a stop with a fricative, such as /ts/ which is the first and last sound in the word "church", are seen as a combination of a quiescent wave for the stop portion of the phoneme, followed by a quasi-random wave type for the fricative portion. As with the voiced stops, the voiced affricates show the quasi-periodic wave type in addition, for the voiced portion of the stop, although this is subject to allophonic variation.

In figure 10 the difference between the voiceless affricate /ts/ (Figure 10-upper trace) as the first and last sound in church and the voiced affricate /dʒ/ (Figure 10-lower trace) as the first and last sound in judge, can be seen.

Nasals. The trace for the nasals, /n/, /m/, and /j/ or sounds made by closing the oral cavity and forcing the speech power to come through the nasal passage, appears as a quasi-periodic wave type which is very narrow when measured on the vertical axis, due to the maximum closure of the vocal tract. This wave type indicates little more than voicing as it is just slightly above the quiescent line or absolute bottoms of the screen. Once again, the Visi-Pitch is showing manner rather than place of articulation and all three nasals appear to be the same, as can be seen in Figure 11 where the words din, dim, and ding are displayed.

Liquids. The liquids of /l/ and /r/ appear as quasi-periodic wave types. Owing to the fact that the opening of the oral cavity is narrower in the production of the /l/ than it is in the production of the /r/, the trace for the /l/ is also narrower than the trace for the /r/ when measured on the verticle axis. This difference can be seen in Figures 12 and 13 when comparing "stealing" with "steering".

The presence of the schwa, or the sound of /ə/ in these displays is significant and helpful to students. Because most American speakers put the schwa between front vowels and the liquids, students are often more successful if they are conscious of doing the same thing.

However it must be added that there must be allowances for tongue variation in these sounds. It is possible to produce a good /l/ sound with only the tip of the tongue at the alveolar ridge, thereby allowing enough speech power around the sides of the tongue to produce a trace that looks like the trace for an /r/. In addition, it is possible to produce an /r/ with enough tongue height so as to make a trace that looks like that of the /l/. It is for this reason that it was stated earlier that it is advisable to isolate each person's appropriate productions of the target sound and attempt to repeat the sound in other environments.

Glides. Because the nature of the glides of /j/ and /w/ is to be articulated in close relationship to the vowels preceeding and following, (pronounced while gliding from one sound to the other), they are greatly affected by the environment in which they occur.

The phoneme /j/, (pronounced as the first sound in the word "yes"), usually appears as a narrow band close to the quiescent line when in the word initial or word final position as in the words "yes" change, as seen in Figure 14. However, when the sound occurs in the medial position of the surrounding vowels. Hence in the word "weighing", as seen in Figure 15, the trace for /j/ appears closer to the middle of the trace area due to the effect of the /e/ and /I/ which are non-high, non-low vowels although these differences are not great.

The glide /w/ is affected more by its environment than is /j/. Not only is the position and width of the trace for the /w/ on the screen determined by the position of the adjacent vowels but also whether it will appear with or without aspiration is determined by the environment. In Figures 16 and 17 a difference can be seen in the trace for the /w/ in the words "coward" and "reweed", although these differences are not always consistent.

It must be added that although it is difficult to describe the trace for /w/ because of the variations that occur, it is safe to say that the difference between the target trace for /w/ and its frequent substitutions such as /v/ is significant as can be seen in Figures 18 and 19, in which the words "wine" and "vine" are displayed.

It is also worth noting that for the sake of pronunciation training, a small hand held mirror is of great value in teaching the production of /w/. The oscilloscope however, is a good method of checking whether or not the sound has been produced successfully when compared to an earlier correct attempt. Self-confidence then comes from knowing that one is correct.

Vowels. Because the vocal cords vibrate regularly in the production of vowels, the wave type produced is quasi-periodic, with the traces varying in height on the vertical axis according to the placement of the tongue on the high to low scale. The traces for the higher vowels of /i/ and /u/ are narrow on the vertical axis while the traces for /æ/ and /a/ are wide on the vertical axis as seen in Figure 20.

Because of the variation with the position of the tongue, there is a dramatic difference in the traces for the vowels of /i/ and /ae/ in this display, but not much in the traces of /ae/ and /a/. However, if a minor adjustment is made in the intensity control of the oscilloscope, traces appear which show a difference in the internal structure of these vowels. Thus in Figure 21 and 22, we can see a difference in the internal structures of /æ/ and /a/ which are the vowel sounds in the words "hat" and "hot".

CONCLUSION

In conclusion, the Visi-Pitch can be a useful instrument in the pronunciation training of English as a second language students. While it may not replace the traditional training methods such as small mirrors for articulation training, or tape recorders for audio feedback, it does complement both by introducing the added dimension of being able to see the results of what is felt and heard.

As students watch their mouths with the mirrors, they can also speak into the microphone of the Visi-Pitch and see the results of the various positions of the articulators. When a tape recording is played back, this can also be done into the microphone of the Visi-Pitch to show students whether they are right or wrong. This can be especially dramatic with students who have a faulty stress pattern and may produce a stream of speech with syllables of equal duration and intensity. It is also revealing to students who omit final consonants and can see that their traces show a heavy concentration of vowels rather than a frequent drop to the quiescent line as would be the case if they had included some of the final stop consonants.

Finally, the Visi-Pitch has been especially helpful with older language learners in professional positions who do not have a lot of time to spend on the subject of pronunciation. They are able to see what they are doing and begin seeing which words they can use safely and knowing which words they must improve or for which they must find a suitable substitute.

Kathleen Albertson holds a B.S. in Education from the University of Dayton and a M.A. in Linguistics from Ohio University where she majored in TESOL.

Kathleen Albertson, Director
Ohio University Cambridge Program
POB 1202
Cambridge, OH 43725
(614) 432-4404

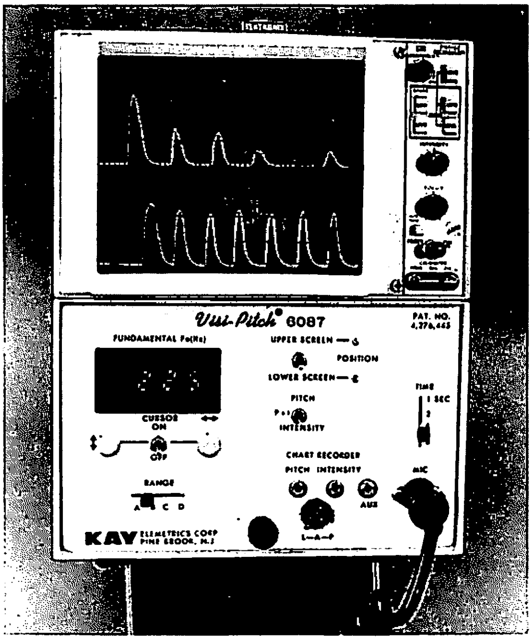

FIGURE 1 The VISI-PITCH from Kay Elemetrics with a Tektronics oscilloscope mounted on top.

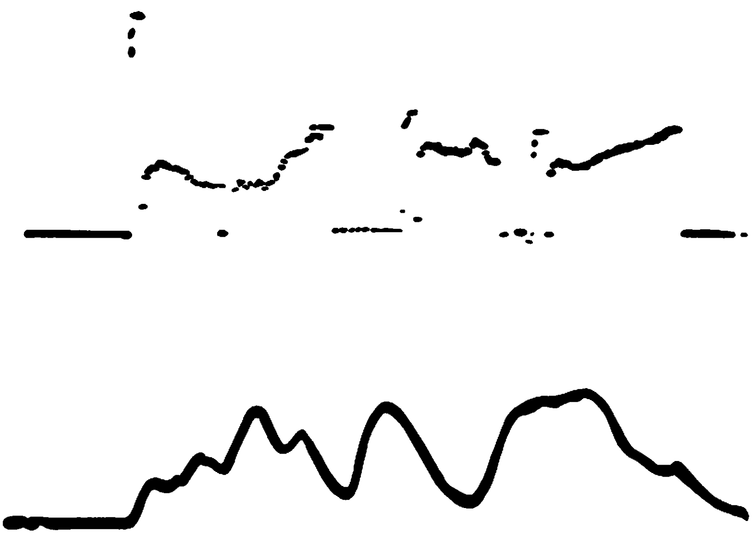

FIGURE 2 Is this drink yours?

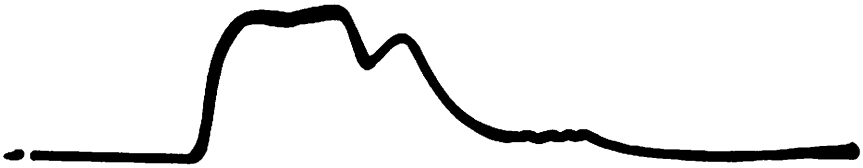

FIGURE 3 danger /deyn dʒer/

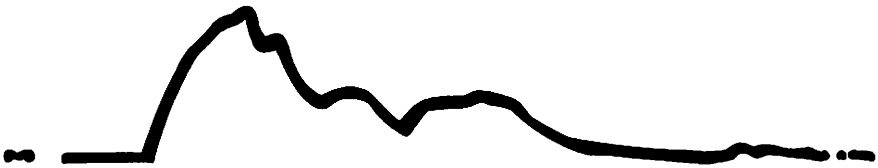

FIGURE 4 automobile /ɔʼtə mə biyˋl/

FIGURE 5 Is this drink yours or mine?

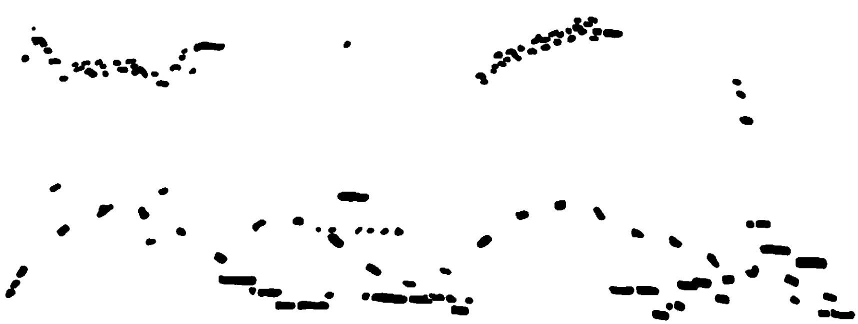

FIGURE 6 constitution /kan stə tu ʃən/

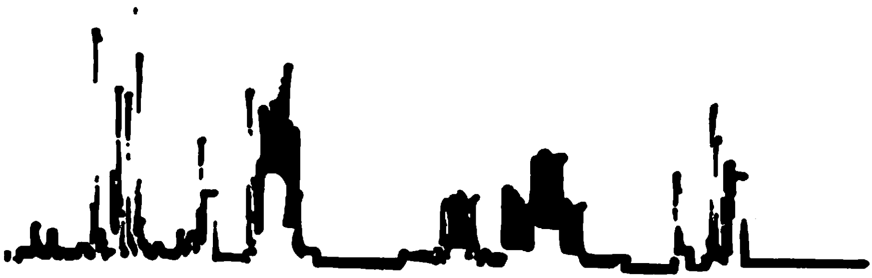

FIGURE 7 speculate /sp e′ kyuw ley t/

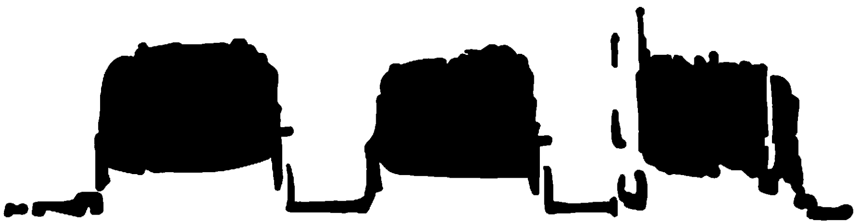

FIGURE 8 bad, dad, pad /bæd/ /dæd/ /pæd/

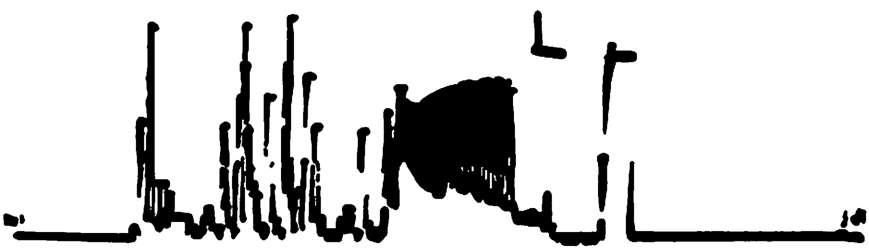

FIGURE 9 (upper trace) fat /fæt/

FIGURE 9 (lower trace) sat /sæt/

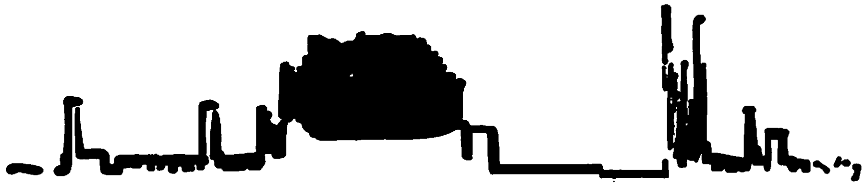

FIGURE 10 (lower trace(judge /dʒ ʌ̀ dʒ/

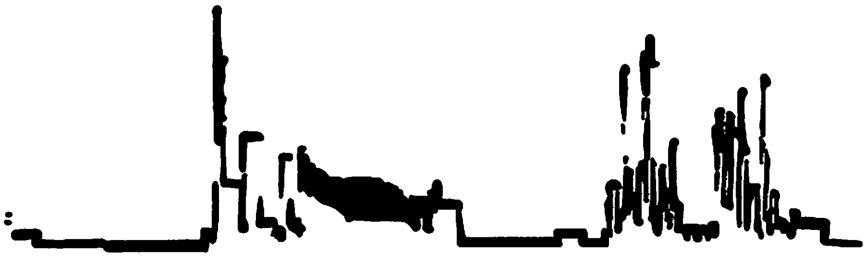

FIGURE 10 (upper trace) church /tʃ ər tʃ?

FIGURE 11 din, dim, ding /din/ /dim/ /diŋ/

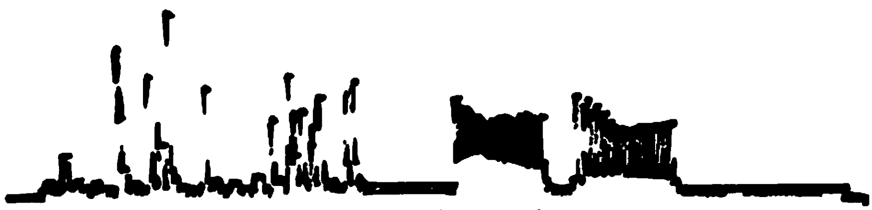

FIGURE 12 stealing /st iy ə liŋ/

FIGURE 13 steering /st iy ə r iŋ/

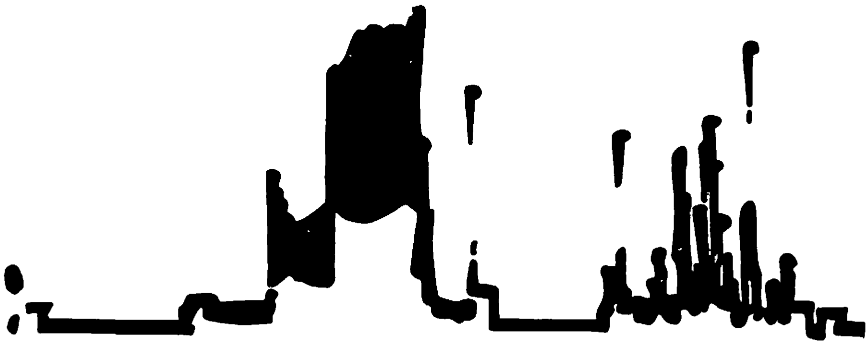

FIGURE 14 yes /y e s/

FIGURE 15 weighing /w ey iŋ/

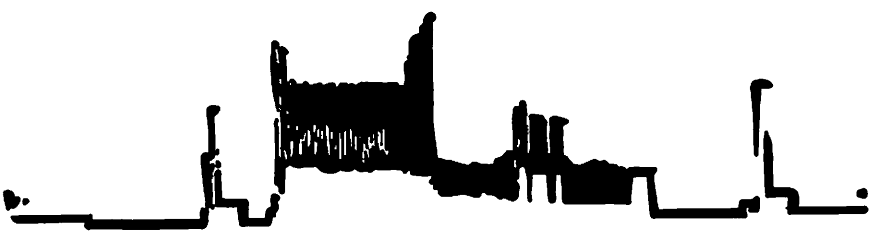

FIGURE 16 coward /k aw ər d/

FIGURE 17 reweed /riy wiyd.

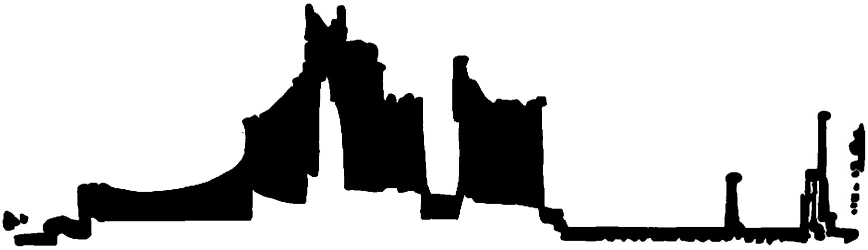

FIGURE 18 wine /w a y n/

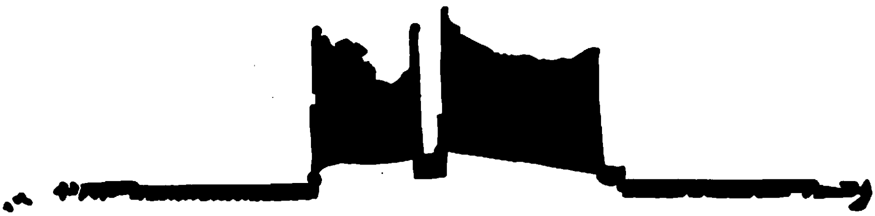

FIGURE 19 vine /v a y n/

FIGURE 20

heed hid head hat hot hut hood whoed
/hiyd/ /hid/ hɛd/ /hæt/ /hat/ hət/ /hud/ /huwd/

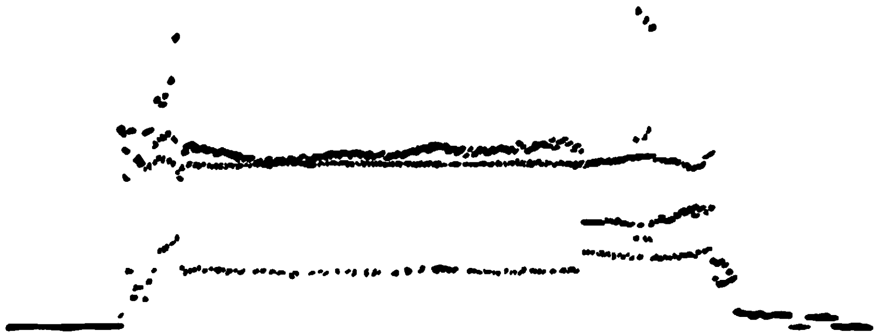

FIGURE 21 hat /hæt/

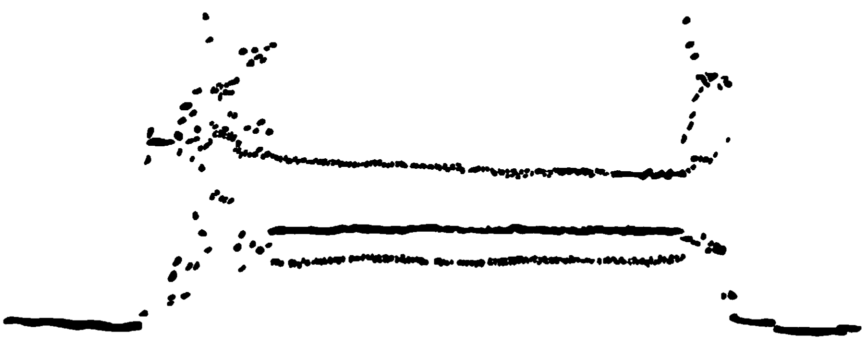

FIGURE 22 hot /hat/