Spectrographic Analysis and Patterns in Pronunciation

Garry Molholt

Coordinator of Computer Assisted Instruction, English Department, 547 Main Hall, West Chester University, West Chester, PA 19383, USA

Abstract: Though there are good reasons to improve instruction in pronunciation, the teaching of pronunciation has lost popularity among language teachers. This is because the traditional indirect analyses of sounds according to places and manners of articulation are clumsy when applied to classroom teaching. By shifting the focus of instruction to the direct feedback of real-time acoustic analysis in the visual mode, instructors are free from the complex and often unproductive terminology of articulatory phonetics, and students are free from the burden of translating instructors' general comments such as "try again" or "repeat after me" into plans for specific changes.

Key Words: acoustic phonetics, applied linguistics, computer-assisted instruction, phonology, pronunciation, second language acquisition, spectrographic analysis, speech, visual feedback.

Over the past decade, students, parents and legislators have applied increasing pressure on universities to improve the spoken English of international graduate students who are paid to teach introductory level classes (Bailey et al., 1984, pp. 3—15). This problem has received more public attention than what may be even more acute problems: the need to improve the spoken language of students in foreign language departments, and the need for individuals to improve their spoken language in jobs which require a second language. The teaching of spoken language, especially pronunciation, has therefore not received sufficient attention to meet these needs. According to Pennington (1988, p. 203):

> Advocates of modern comprehension-based or communicatively oriented language curricula generally take the view that pronunciation should not be taught explicitly but should rather be allowed to develop naturally as a by-product of attempts by students to communicate.

This "trickle-down" theory of the acquisition of second language phonology is unnecessarily slow and may lead to premature fossilization: cessation of improvement before adequate sophistication is reached (Molholt et al., 1988, p. 74).

Though it is relatively common to find spectrograms in linguistics textbooks, neither linguistics nor language education are typically the primary focus of current research in digital signal processing, which produces facilities for advanced analysis of speech. Progress in technology has been due, therefore, to factors outside of language education. For example, electrical engineers are more concerned with voice input for computer systems because speech (in one's native dialect) requires little training and speech is the human's highest capacity output channel (Chapanis, 1977, pp. 101—26). As educators, we may assume that when we can automate the analysis of normal conversation, computers will then be able to

---

Garry Molholt is Assistant Professor of Linguistics and English as a Second Language, and Coordinator of Computer Assisted Instruction. His research interests are the applications of speech processing to instruction in the acquisition of second language phonology. He has published "Computer Assisted Instruction in Pronunciation for Chinese Speakers of American English," in TESOL Quarterly, *and (with Ari Presler), "Correlation between Human and Machine Ratings of Test of Spoken English Reading Passages," in* Technology and Language Testing.

Computers and the Humanities **24**: 81—92, 1990.

compare patterns of pronunciation of different speakers, and programs can be developed to monitor students' progress in conforming to desired standards.

Despite the facts that language education is not a primary focus in speech processing research and that engineers are still far from producing machines which automatically analyse and compare speech from a variety of speakers (White, 1984, pp. 213–23), there are still many benefits for language educators in the applications of speech processing research already completed (Molholt and Presler, 1985, 1986; Molholt et al., 1988; Molholt, 1988). A major breakthrough has been the development of real-time displays of speech patterns such as the Kay Elemetrics Speech Spectrographic Display (SSD 8800) and the Visi-Pitch 6095 for use with the Apple IIe. Kay's latest, the DSP Sona-Graph, model 5500, has multiprocessor and dual bus architecture, which provide high speed processing and perform simultaneous real-time acquisition, filtering, analysis, storage and display functions. In conjunction with IBM hardware and speech processing software, the DSP Sona-Graph 5500 is a sophisticated workstation with a wide variety of applications.

Before these high-speed real-time machines were available, language researchers had to wait 90 to 120 seconds to receive paper output displays which showed only a few features of one to two seconds of speech. The analysis was therefore too slow for effective teaching of pronunciation. Now that students are able to see immediate displays of their speech patterns and match them on a split screen with target patterns, they can quickly learn to recognize the location, type, size and significance of their errors and monitor their progress with the aid of reliable and precise feedback (Molholt, 1988). In so doing they gain the confidence needed to improve speaking skills.

The purpose of this paper is to show how speech processing equipment can be used to help instructors communicate with students about patterns of pronunciation. Though the equipment can be used for any level of sophistication for any language or dialect, I will concentrate on comparing features of English spoken by non-natives with English spoken by natives. Of course, the equipment could also be used for more sophisticated tasks, such as helping a newscaster or actor who wants to sound like Winston Churchill.

Rather than being viewed as a tool in language education, acoustic analysis has been used to study both the extent of the applicability of several hypotheses of the acquisition of second language phonology and the feasibility of using computers to rate automatically the accuracy of pronunciation of non-native speakers. For example, according to Flege and Hillenbrand (1987, p. 199):

> The accuracy with which a learner produces the sounds of a foreign language can be objectively assessed in a variety of ways:
> (1) through the use of rating scale judgments by native speakers of the target language,
> (2) by calculating the frequency with which L2 [target language] phones [sounds] are correctly identified, and
> (3) through acoustic analysis.

These researchers show (p. 196) that both perceptual and acoustic criteria agree to support their hypothesis that inaccurate perceptual targets are "factors that might limit the accuracy with which adult learners produce phones found in a foreign language" (p. 176).

In an earlier study, Molholt and Presler (1986) studied the correlation of human and machine ratings of the pronunciation of 20 subjects reading a disclosed form of the reading passage of the Test of Spoken English administered by the Educational Testing Service. They found (p. 121) the rank order correlation coefficient (Spearman's Rho) to be 0.9318, which compares quite favorably to the 0.88 intercorrelation between pronunciation and fluency reported in the *Test of Spoken English Manual for Score Users* (1982, p. 23).

This paper will not deal directly with the questions of why certain sounds are more difficult to acquire than others or how certain rating systems compare. Therefore, this is not a test of any form of the Contrastive Analysis Hypothesis, Language Universals Hypothesis, or Markedness Differential Hypothesis (Anderson, 1987, p. 279). Several studies in Ioup and Weinberger (1987) refer to the utilization of speech processing equipment to study such hypotheses, especially with respect to interlanguage phonology.

Whether we look at pronunciation from the bottom up, starting with the production of phonemes and going to suprasegmentals or from the top down and consider general properties such as rhythm and voice quality (Pennington, 1988, p. 204), there still remain quantifiable ranges of acoustic features which correlate with our perceptions of the level of comprehensibility (Rabiner and Schafer, 1978, especially pp. 38—60). From spectrograms, we can easily identify patterns of features, including the basic acoustic features of frequency, duration, and intensity, from which we can see additional linguistic patterns of voicing, aspiration, addition and deletion, and articulatory patterns of turbulence, tight closure and loose closure.

Phonological systems are built on sensitive patterns which easily yield wide variations in interpretation if small mistakes are made. For example, the following sentence has many different interpretations if spoken by people with different accents:

She's taking the gold.

A Spanish speaker might use tight closure and say *cheese* instead of *She's*. A Chinese speaker might say *tacking* for *taking*, because of an error in vowel frequency. It is common for Koreans to add a short vowel at a word boundary when the first word ends in a voiced sound and the second word starts with a voiceless stop. Thus *attacking* would be substituted for *taking* when this error is added to an error in vowel frequency. Deletion of final consonants, devoicing of voiced sounds — as in German final consonants — and simplification of clusters are all quite common. Hence *gold* could sound like *goal, cold, coal, colt, coat*, and even *goat*. In a worst case rendition combining several types of errors, the sentence could sound like:

Cheese attacking the coat.

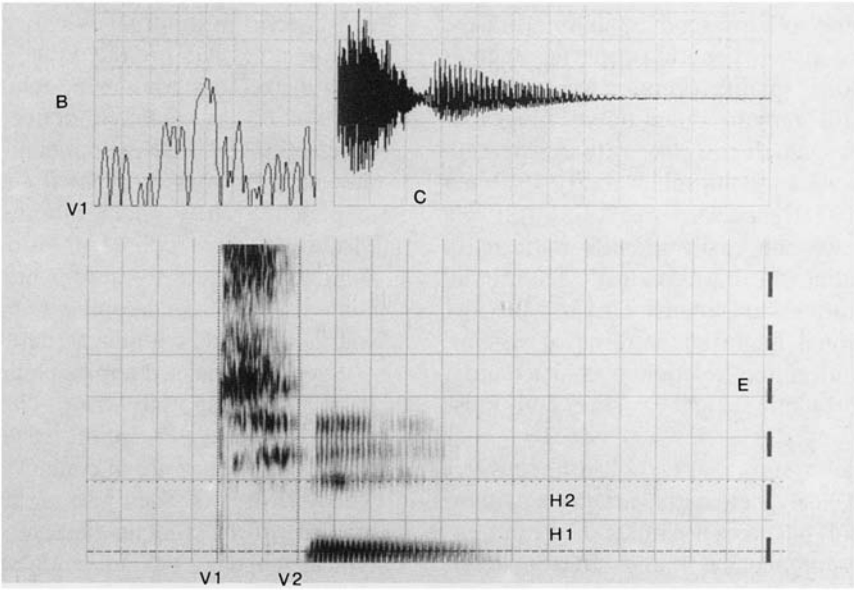
Combining errors of grammar and word usage with these errors further reduces the chances of a message being interpreted as intended.

Traditional methods of teaching students to correct these errors rely heavily on providing students with opportunities to notice and correct them on their own. Students read descriptions about how to pronounce sounds, they hear lectures, and go to a language lab to listen and repeat and then listen to the tapes they make so that they can check the differences between their pronunciation and that on the tape. Typical feedback from instructors consists of general terms such as *very good* or *try again*. Students do not get precise descriptions of their pronunciation or the patterns they need to work on. They do not receive precise feedback to help them monitor their progress. Moreover, the feedback they do receive is substantially delayed, so there is little chance that the students will associate the feeling of producing sounds correctly with the actual sounds.

Such a haphazard approach limits the effectiveness of language lab work. Those students who have fine aural perception learn correct pronunciation when they are in contact with the language, whether or not they use a language lab. The majority of the students need more help. Since the typical language lab does not provide them with sufficiently precise feedback regarding the differences between their pronunciation patterns and the targets, their progress is impeded. Thus, pronunciation teaching has been viewed by many language teachers as frustrating, and even detrimental to language learning.

With spectrographic displays of pronunciation on a split screen to provide immediate and precise feedback, teachers and students can become sensitive to the location, size, type, and significance of error patterns, without worrying about taking so much time from language learning that fluency will be diminished. Because students are immediately able to see patterns of their speech, they seem to be able to associate the kinesthetic feelings of production with the patterns on the screen. Eventually many seem to be able to associate these feelings with the target sounds. Then they no longer need the intermediary step of seeing the patterns. This can be accomplished without using complex linguistic vocabulary about the manner or place of articulation, since it is being handled on the acoustic level in a visual mode. For busy students interested only in improving their fluency, not in learning about teaching or linguistics, such as approach seems to be quite attractive, since it is quite direct, reliable and quantifiable.

Figure 1 is a sample display from a Kay Elemetrics DSP Sona-Graph 5500, and demonstrates only one of many ways to display and analyse data with this machine. Here the screen is




```
KAY ELEMETRICS CORP. MODEL 5500
SIGNAL ANALYSIS WORKSTATION

Date: JULY      3F 2027    Time: 12:20:57 AM

Analysis by:

INPUT SETTINGS      Channel 1            Channel 2
Source              LEFT CONNECTORS      LEFT CONNECTORS
Frequency Range     DC -  8 KHz.         DC -  8 KHz.
Input Shaping       HI-SHAPE             FLAT
Buffer Size         2.0  SECONDS         2.0  SECONDS

ANALYSIS SETTINGS   Lower Screen         Upper Screen
Signal Analyzed     CHANNEL 1            CHANNEL 1
Analysis Format     SPECTROGRAPHIC       POWER AT CURSORS
Transform Size      100 pts. ( 300 Hz)   256 pts. ( 117 Hz)
Time Axis           1.0  SECOND          1.0  SECOND
Frequency Axis      FULL SCALE           FULL SCALE
Analysis Window     HAMMING              HAMMING
LOFAR Set Up        NO AVERAGING         NO AVERAGING

DISPLAY SETTINGS    Lower Screen         Upper Screen
Time Divisions      0.05000 SEC.         0.05000 SEC.
Freq. Divisions      1000.00 HZ.             0.00 HZ.
Dynamic Range       42 dB                42 dB
Analysis Atten.     15 dB                5 dB
Set Up Options Set  to:  #00

CURSOR READINGS:
FC1:  1240.00 Hz., FC2:   400.00 Hz.,  ^F:   840.00 Hz.
^T: 0.1031 SEC.

SUBJECT MATTER
```

Figure 1. Display of the word *tea* on a DSP Sona-Graph 5500.

divided into three parts. Part A is a spectrogram with grids and cursors. The vertical axis shows frequency, with grid markers every 1000 Hz., from 0 to 8000 Hz. The horizontal axis shows time, with grid markers every 0.05 seconds, from 0 to 1 second. The two vertical cursors, V1 and V2, are set 0.1125 seconds apart. They are marking the duration of the explosion and the aspiration of the initial /t/ of *tea*. The two horizontal cursors, H1 and H2, are set at 280 Hz. and 1960 Hz. They mark the first and second vowel formants of the /i/ of *tea*.

The third axis, shown by the gray scale, represents amplitude, the strength of the signal at a particular frequency. In the vowel /i/, for example, the lower formant at H1 has higher amplitude than the formant at H2; thus the lower one is darker. We can see that the amplitude across the middle of the /t/ is high, because it is darker. This can be measured precisely by placing V1 at the exact place to be measured and then looking at the upper left display, display B. The lower horizontal line of B is labeled as V1, and corresponds to the vertical cursor V1 in A. V1 in B shows the amplitude of the slice of time taken from A at the location of vertical cursor V1. The vertical grid markers in B show frequency from 0 to 8000 Hz. marked every 1000 Hz. Thus we can see in B that the highest level of amplitude occurred at exactly 5000 Hz. The settings and display type used for this sample may be changed for other applications.

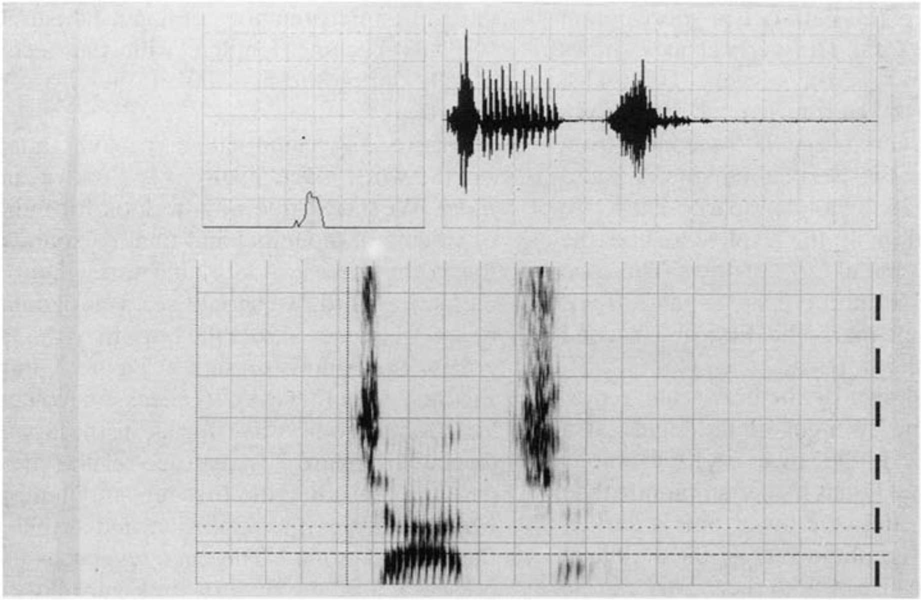
The third display area, C, is a wave form display similar to that generated by an oscilloscope. It shows the random noise of the /t/ followed by the patterned (periodic) sound of the /i/. For this particular configuration, the waveform display shows everything to the right of the vertical cursor V1 in display A. The horizontal grid of C shows levels of amplitude, and the vertical grid shows duration.

The relationships between frequency, amplitude and duration within a speech signal provide the information we need to communicate with students about their patterns and help them learn to recognize errors and monitor their own progress as they attempt to bring their patterns within a desired range. The following examples of student attempts to match teacher patterns show how frequency, duration and amplitude may be translated into interpretations of linguistic patterns. As students become familiar with the techniques, usually during the first lesson, they learn to self-correct.

Figure 2 is a non-native speaker's attempt to say the word *judge*. Figure 3 is a native speaker's *judge*. We have three ways to look for indications of voicing of the initial and final /j/ sounds. Since the vocal cords are vibrating throughout almost the entire word, we should see a horizontal band of low frequency across the bottom of the spectrograms. The band is present in Figure 3, but not in Figure 2. Another way to check for voicing is to look for periodic wave forms in the upper right display. In Figure 3 we can see regular prevoicing patterns preceding the first big amplitude pattern which indicates the explosion and turbulence of the initial /j/ sound. The same regular patterns are present before the final /j/. In Figure 2 we do not find such patterns. The third way we can check to see if a sound is voiced or not is to place the vertical cursor in the spectrogram at the desired location and check the amplitude indication of the power spectrum in the upper left display. In Figure 3 the indication is that we have a low frequency signal at the start of the word. In Figure 2 it is a mid frequency signal. Voiced sounds have a low frequency component, so we know that the initial sound in Figure 3 is voiced and in Figure 2 it is voiceless.

These are excellent tools of analysis. We do not need to use them all to communicate with the students, however. It is usually sufficient to show patterns on the spectrogram, explain how the patterns are made, give examples, and show why they are important. The class or individual lessons can thereby move at a rapid pace. Exercises in Bowen (1975) are helpful for demonstrating to the students why the distinctions are important.

In Figure 5 we see a native speaker's *big*. Again we have voicing across the bottom of the spectrogram, periodic patterns in the waveform display preceding the explosions of the /b/ and the /g/, and a low frequency marker in the power spectrum indicating voicing. In Figure 4 we have a different story. Neither the /b/ nor the /g/ are voiced. In addition, we can see from the grids and cursors that the vowel in Figure 4 is /i/ not /I/, since the first and second formants in 4 are at 360



KAY ELEMETRICS CORP. MODEL 5500
SIGNAL ANALYSIS WORKSTATION

Date: JULY 3F 2027 Time: 12:20:02 AM

Analysis by:

INPUT SETTINGS	Channel 1	Channel 2
Source	LEFT CONNECTORS	LEFT CONNECTORS
Frequency Range	DC - 8 KHz.	DC - 8 KHz.
Input Shaping	HI-SHAPE	FLAT
Buffer Size	2.0 SECONDS	2.0 SECONDS

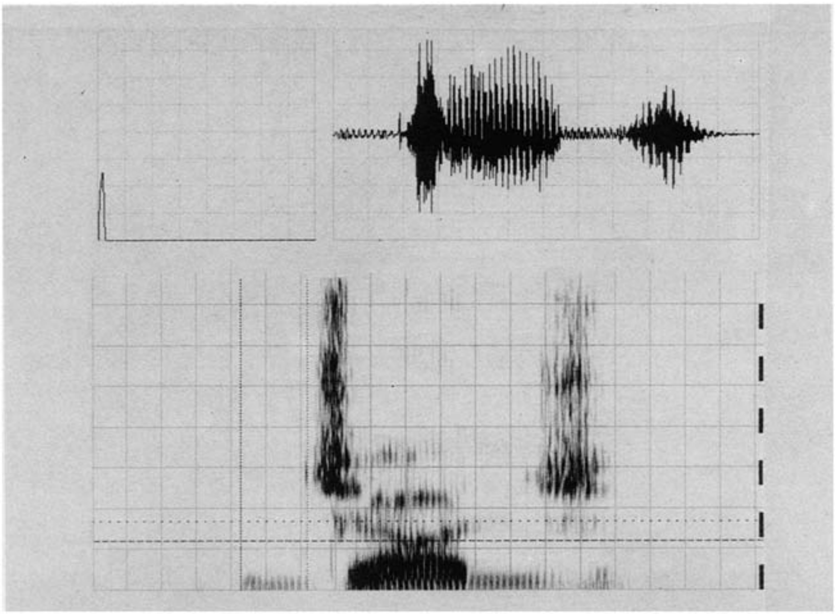
ANALYSIS SETTINGS	Lower Screen	Upper Screen
Signal Analyzed	CHANNEL 1	CHANNEL 1
Analysis Format	SPECTROGRAPHIC	POWER AT CURSORS
Transform Size	100 pts. (300 Hz)	256 pts. (117 Hz)
Time Axis	1.0 SECOND	1.0 SECOND
Frequency Axis	FULL SCALE	FULL SCALE
Analysis Window	HAMMING	HAMMING
LOFAR Set Up	NO AVERAGING	NO AVERAGING

DISPLAY SETTINGS	Lower Screen	Upper Screen
Time Divisions	0.05000 SEC.	0.05000 SEC.
Freq. Divisions	1000.00 HZ.	0.00 HZ.
Dynamic Range	42 dB	42 dB
Analysis Atten.	15 dB	5 dB
Set Up Options Set to:	#00	

CURSOR READINGS:
FC1: 4440.00 Hz., FC2: 6680.00 Hz., ^F: 2240.00 Hz.
^T:

SUBJECT MATTER

Figure 2. Non-native speaker's attempt at *judge*.

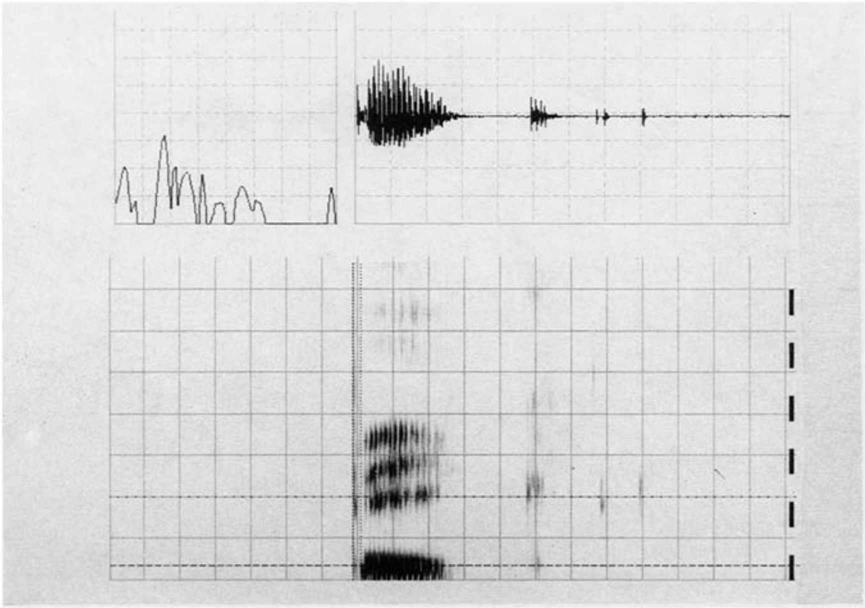


```
KAY ELEMETRICS CORP. MODEL 5500
SIGNAL ANALYSIS WORKSTATION

Date: JULY        3F 2027    Time: 12:20:25 AM

Analysis by:

INPUT SETTINGS      Channel 1              Channel 2
Source              LEFT CONNECTORS        LEFT CONNECTORS
Frequency Range     DC -  8 KHz.           DC -  8 KHz.
Input Shaping       HI-SHAPE               FLAT
Buffer Size         2.0  SECONDS           2.0  SECONDS

ANALYSIS SETTINGS   Lower Screen           Upper Screen
Signal Analyzed     CHANNEL 1              CHANNEL 1
Analysis Format     SPECTROGRAPHIC         POWER AT CURSORS
Transform Size      100 pts. ( 300 Hz)     256 pts. ( 117 Hz)
Time Axis           1.0  SECOND            1.0  SECOND
Frequency Axis      FULL SCALE             FULL SCALE
Analysis Window     HAMMING                HAMMING
LOFAR Set Up        NO AVERAGING           NO AVERAGING

DISPLAY SETTINGS    Lower Screen           Upper Screen
Time Divisions      0.05000 SEC.           0.05000 SEC.
Freq. Divisions      1000.00 HZ.               0.00 HZ.
Dynamic Range       42 dB                  42 dB
Analysis Atten.     15 dB                  5 dB
Set Up Options Set to:  #00

CURSOR READINGS:
FC1:  3600.00 Hz., FC2:  1880.00 Hz., ^F:  1720.00 Hz.
^T: 0.2046 SEC.

SUBJECT MATTER
```

Figure 3. Native speaker's *judge*.

Hz. and 2000 Hz. In 5 they are at 480 Hz. and 1640 Hz. Thus, the word in Figure 4 sounds more like *peak* than *big*. This is a very common problem, especially for east Asians. (See Molholt [1988] for more details.)

Figures 6 and 7 show a non-native and native speaker's attempts at pronouncing the word *zero*. Since the non-native speaker uses tight closure and no voicing for the /z/, it sounds and looks more like the /ts/ at the end of the word *cats*.



KAY ELEMETRICS CORP. MODEL 5500
SIGNAL ANALYSIS WORKSTATION

Date: JULY 3F 2027 Time: 12:20:47 AM

Analysis by: ' ' '

INPUT SETTINGS	Channel 1	Channel 2
Source	LEFT CONNECTORS	LEFT CONNECTORS
Frequency Range	DC - 8 KHz.	DC - 8 KHz.
Input Shaping	HI-SHAPE	FLAT
Buffer Size	2.0 SECONDS	2.0 SECONDS

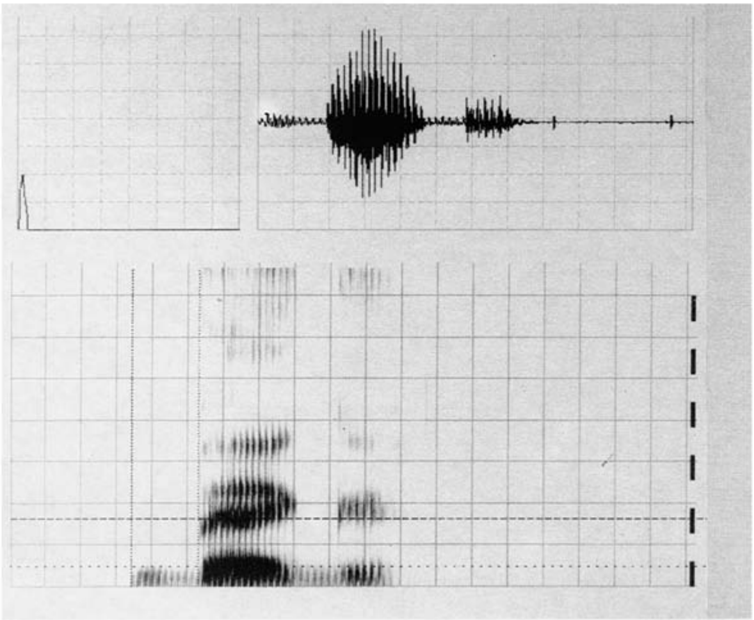
ANALYSIS SETTINGS	Lower Screen	Upper Screen
Signal Analyzed	CHANNEL 1	CHANNEL 1
Analysis Format	SPECTROGRAPHIC	POWER AT CURSORS
Transform Size	100 pts. (300 Hz)	256 pts. (117 Hz)
Time Axis	1.0 SECOND	1.0 SECOND
Frequency Axis	FULL SCALE	FULL SCALE
Analysis Window	HAMMING	HAMMING
LOFAR Set Up	NO AVERAGING	NO AVERAGING

DISPLAY SETTINGS	Lower Screen	Upper Screen
Time Divisions	0.05000 SEC.	0.05000 SEC.
Freq. Divisions	1000.00 HZ.	0.00 HZ.
Dynamic Range	42 dB	42 dB
Analysis Atten.	15 dB	5 dB
Set Up Options Set	to: #00	

CURSOR READINGS:
FC1: 1920.00 Hz., FC2: 1600.00 Hz., ^F: 320.00 Hz.
^T:

SUBJECT MATTER

Figure 4. Non-native speaker's attempt at *big*.



KAY ELEMETRICS CORP. MODEL 5500
SIGNAL ANALYSIS WORKSTATION

Date: JULY 3F 2027 Time: 12:20:06 AM

Analysis by: ' ' '

INPUT SETTINGS	Channel 1	Channel 2
Source	LEFT CONNECTORS	LEFT CONNECTORS
Frequency Range	DC - 8 KHz.	DC - 8 KHz.
Input Shaping	HI-SHAPE	FLAT
Buffer Size	2.0 SECONDS	2.0 SECONDS

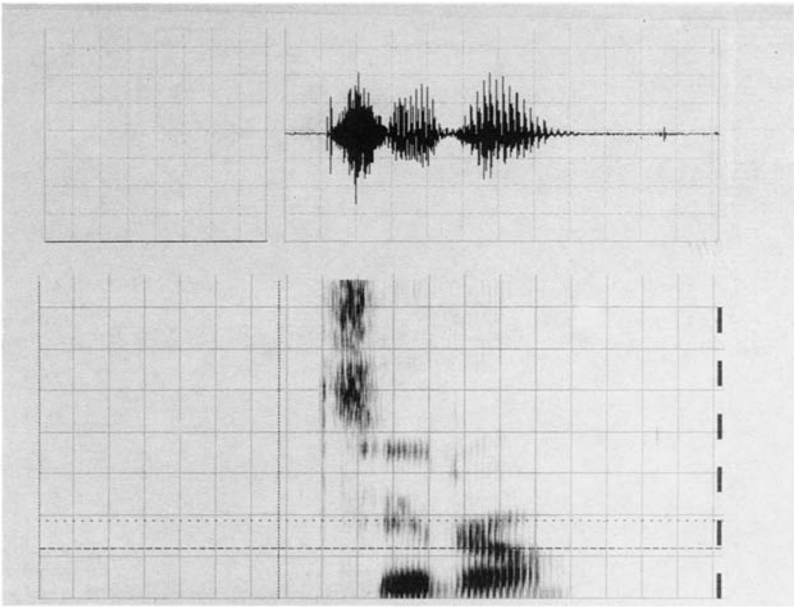
ANALYSIS SETTINGS	Lower Screen	Upper Screen
Signal Analyzed	CHANNEL 1	CHANNEL 1
Analysis Format	SPECTROGRAPHIC	POWER AT CURSORS
Transform Size	100 pts. (300 Hz)	256 pts. (117 Hz)
Time Axis	1.0 SECOND	1.0 SECOND
Frequency Axis	FULL SCALE	FULL SCALE
Analysis Window	HAMMING	HAMMING
LOFAR Set Up	NO AVERAGING	NO AVERAGING

DISPLAY SETTINGS	Lower Screen	Upper Screen
Time Divisions	0.05000 SEC.	0.05000 SEC.
Freq. Divisions	1000.00 HZ.	0.00 HZ.
Dynamic Range	42 dB	42 dB
Analysis Atten.	15 dB	5 dB
Set Up Options Set	to: #00	

CURSOR READINGS:
FC1: 1240.00 Hz., FC2: 3480.00 Hz., ^F: 2240.00 Hz.
^T:

SUBJECT MATTER

Figure 5. Native speaker's *big*.



KAY ELEMETRICS CORP. MODEL 5500
SIGNAL ANALYSIS WORKSTATION

Date: JULY 3F 2027 Time: 12:20:06 AM

Analysis by: ' ' '

INPUT SETTINGS	Channel 1	Channel 2
Source	LEFT CONNECTORS	LEFT CONNECTORS
Frequency Range	DC - 8 KHz.	DC - 8 KHz.
Input Shaping	HI-SHAPE	FLAT
Buffer Size	2.0 SECONDS	2.0 SECONDS

ANALYSIS SETTINGS	Lower Screen	Upper Screen
Signal Analyzed	CHANNEL 1	CHANNEL 1
Analysis Format	SPECTROGRAPHIC	POWER AT CURSORS
Transform Size	100 pts. (300 Hz)	256 pts. (117 Hz)
Time Axis	1.0 SECOND	1.0 SECOND
Frequency Axis	FULL SCALE	FULL SCALE
Analysis Window	HAMMING	HAMMING
LOFAR Set Up	NO AVERAGING	NO AVERAGING

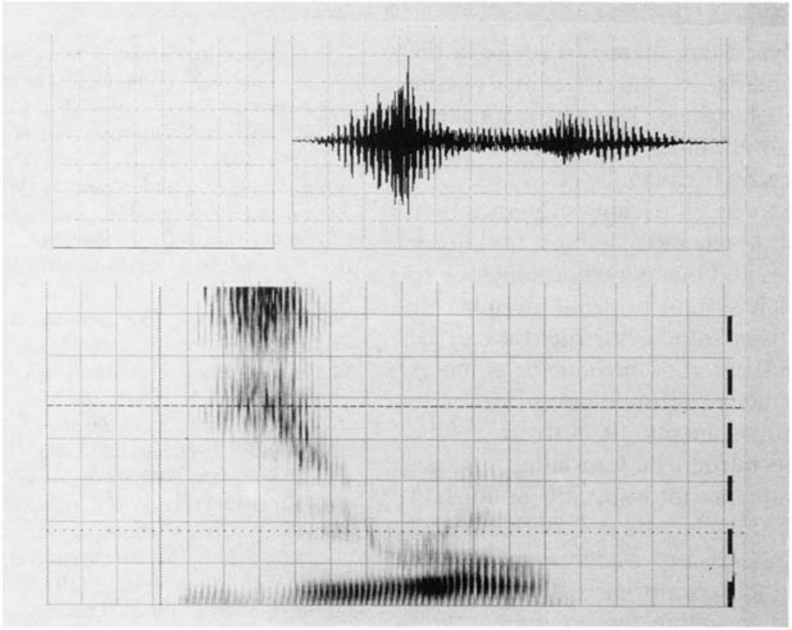
DISPLAY SETTINGS	Lower Screen	Upper Screen
Time Divisions	0.05000 SEC.	0.05000 SEC.
Freq. Divisions	1000.00 HZ.	0.00 HZ.
Dynamic Range	42 dB	42 dB
Analysis Atten.	15 dB	5 dB

Set Up Options Set to: #00

CURSOR READINGS:
FC1: 1200.00 Hz., FC2: 1880.00 Hz., ^F: 680.00 Hz.
^T: 0.02343 SEC.

SUBJECT MATTER

Figure 6. Non-native speaker's attempt at *zero*.



```
KAY ELEMETRICS CORP. MODEL 5500
SIGNAL ANALYSIS WORKSTATION

Date: JULY      3F 2027    Time: 12:20:37 AM

Analysis by:    '      ' '

INPUT SETTINGS      Channel 1              Channel 2
Source              LEFT CONNECTORS        LEFT CONNECTORS
Frequency Range     DC - 8 KHz.            DC - 8 KHz.
Input Shaping       HI-SHAPE               FLAT
Buffer Size         2.0 SECONDS            2.0 SECONDS

ANALYSIS SETTINGS   Lower Screen           Upper Screen
Signal Analyzed     CHANNEL 1              CHANNEL 1
Analysis Format     SPECTROGRAPHIC         POWER AT CURSORS
Transform Size      100 pts. ( 300 Hz)     256 pts. ( 117 Hz)
Time Axis           1.0 SECOND             1.0 SECOND
Frequency Axis      FULL SCALE             FULL SCALE
Analysis Window     HAMMING                HAMMING
LOFAR Set Up        NO AVERAGING           NO AVERAGING

DISPLAY SETTINGS    Lower Screen           Upper Screen
Time Divisions      0.05000 SEC.           0.05000 SEC.
Freq. Divisions     1000.00 HZ.            0.00 HZ.
Dynamic Range       42 dB                  42 dB
Analysis Atten.     15 dB                  5 dB
Set Up Options Set to:  #00

CURSOR READINGS:
FC1:  2080.00 Hz., FC2:   280.00 Hz., ^F:  1800.00 Hz.
^T: 0.1328 SEC.

SUBJECT MATTER
```

Figure 7. Native speaker's *zero.*

Because the /r/ has a tap, there is a break in the display corresponding to the touching of the tongue to the alveolar ridge. The native speaker's patterns in Figure 7 are much smoother. The /z/ is voiced, and there is no break in the /r/.

By shifting the focus of teaching pronunciation, away from the indirect and complex traditional emphasis on places and manners of articulation to the direct feedback system of real-time acoustic analysis in the visual mode, instructors can be freed from the frustration of introducing complex and often unproductive terminology in order to communicate with students. It also frees the students from the burden of translating instructors' general comments into specific notions of what worked well and what did not. With minimal instruction and supervision, students are able to learn to monitor their own progress in mastering the patterns. By connecting the equipment to a VCR, students can make a video tape so they can see and hear their work between scheduled lessons and practice sessions.

References

Anderson, Janet I. "The Markedness Differential Hypothesis and Syllable Structure Difficulty." In *Interlanguage Phonology: The Acquisition of a Second Language Sound System.* Ed. Georgette Ioup and Steven Weinberger. Cambridge, MA: Newbury House, 1987, pp. 279—91.

Bailey, Kathleen, Frank Pialorsi, and Jean Zukowski-Faust. *Foreign Teaching Assistants in U.S. Universities.* Washington, DC: National Association for Foreign Student Affairs, 1984.

Bowen, J. Donald. *Patterns in English Pronunciation.* Rowley, MA: Newbury House, 1975.

Chapanis, A. "Studies in Interactive Communications II." *Human Factors,* 19, 2 (1977), 101—26.

Educational Testing Service. *Test of Spoken English Manual for Score Users.* Princeton, NJ, 1982.

Flege, James E., and James Hillenbrand. "Limits on Phonetic Accuracy in Foreign Language Speech Production." In *Interlanguage Phonology: The Acquisition of a Second Language Sound System.* Ed. Georgette Ioup and Steven Weinberger. Cambridge, MA: Newbury House, 1987, pp. 176—203.

Ioup, Georgette, and Steven H. Weinberger, eds. *Interlanguage Phonology: The Acquisition of a Second Language Sound System.* Cambridge, MA: Newbury House, 1987.

KAY Elemetrics. *User Manual for SSD 8800.* Pine Brook, NJ, 1988.

KAY Elemetrics. *User Manual for Visi-Pitch 6095.* Pine Brook, NJ, 1988.

KAY Elemetrics. *User Manual for DSP Sona-Graph 5500.* Pine Brook, NJ, 1988.

Lea, W. A., ed. *Trends in Speech Recognition.* Englewood Cliffs, NJ: Prentice Hall, 1980.

Maddieson, Ian. *Patterns of Sounds.* Cambridge: Cambridge University Press, 1984.

Molholt, Garry. "Computer Assisted Instruction in Pronunciation for Chinese Speakers of American English." *TESOL Quarterly,* 22, 1 (1988), 91—111.

Molholt, Garry, Linda Lane, Joyce Tanner, and Laurie Fischer. "Computer Graphics in the Language Lab." *Technological Horizons in Education Journal,* 15, 6 (1988), 74—78.

Molholt, Garry, and Ari Presler. "Computer Assisted Evaluation of Pronunciation: A Progress Report." In *Proceedings of the Harris Users' Exchange Annual Symposium.* Ed. Terry Ward. Fort Lauderdale: The Harris Users' Exchange, 1985, pp. 200—20.

Molholt, Garry, and Ari Presler. "Correlation between Human and Machine Ratings of Test of Spoken English Reading Passages." In *Technology and Language Testing.* Ed. Charles Stansfield. Washington, DC: Teachers of English to Speakers of Other Languages, 1986, pp. 111—28.

Pennington, Martha. "Teaching Pronunciation from the Top Down." In *University of Hawai'i Working Papers in English as a Second Language.* Ed. Charlene Sato. University of Hawai'i at Manoa: Department of English as a Second Language, 1988, pp. 203—27.

Rabiner, Lawrence, and Ronald Schafer. *Digital Processing of Speech Signals.* Englewood Cliffs, NJ: Prentice Hall, 1978.

White, G. M. "Speech Recognition: An Idea whose Time is Coming." *Byte* (1984), 213—23.