

PAC3 at JALT 2001

Conference
Proceedings



International
Conference
Centre

Kitakyushu
JAPAN

November
22-25, 2001

The Taming of English Vowels

Stephen Lambacher
The University of Aizu

Electronic visual feedback (EVF) is a type of computerized training for accent reduction that has received a great deal of attention. This paper introduces some of the basic features of EVF software and pedagogy. It proposes a training plan for helping Japanese learners improve their pronunciation of difficult English vowels. Pronunciation patterns utilizing acoustic data (EVF) are used to assist Japanese learners to reduce their Japanese-flavored accent of English vowels. EVF instruction is based on a learner's imitation of model patterns with an EVF displaying acoustic characteristics of the model and its imitation. Students visualize the differences between the sound features on their monitors as produced by their teacher, or from files stored in a database, with which they compare their own production. One of the main objectives of EVF software is for students to be able to associate the frequency patterns of the segmentals with the sounds they are producing.

この論文はEVFの基本的な特徴と教授法を紹介する。日本人学習者が、難しい英語の母音を習得していく上でこれを助ける練習計画をこの論文は提案する。

Learning how to pronounce a second language (L2) without an accent can be a formidable task for even the most talented and motivated language learner. One reason for this is that L2 learners often retain the flavor of their native (L1)

language in their pronunciation of nonnative sounds, a process by which a learner's phonological filter acts like a "sieve" and passes through information useful to categorizing sounds in the L1 (Trubetzkoy, 1969). Japanese "accented" pronunciation of English has been well documented (e.g., Riney and Anderson-Hsieh, 1993). One of the most notoriously difficult sounds for Japanese is the English /r/ and /l/. However, Japanese learners frequently have trouble pronouncing certain English vowel sounds as well. One problem is that there are about three times more vowels in English than in Japanese. While some English vowels are relatively easy for Japanese to pronounce, e.g. the vowels in the words *beat* and *boot*, others are more difficult and require special attention in the language classroom. One type of computerized training for accent reduction that has received a great deal of attention is electronic visual feedback (EVF) (See, for example, Molholt, 1990; Anderson-Hsieh, 1992; and Lambacher, 1999). EVF instruction is based on a learner's imitation of model patterns with an EVF displaying acoustic characteristics of the model and its imitation. Students visualize the differences between the sound features on their monitors as produced by their teacher, or from files stored in a database, with which they compare their own production. Many EVF programs include a dual display with top and bottom screens, which helps students to objectively evaluate their pronunciation errors and

progress through analyzing and visually comparing their own pronunciation with a model pattern. One of the main objectives of EVF software is for students to be able to associate the frequency pattern of the segmental with the sound they are producing.

Researchers agree that EVF can be an effective tool for accent reduction for learners of various cultural backgrounds and learning levels (e.g., Molholt, 1990; Anderson-Hsieh, 1992; Lambacher, 1999). Some excellent EVF programs currently on the market for a PC include Kay Elemetric's Multispeech and Visipitch, IBM's Speechviewer, and Signalize (developed for a Macintosh). There are also some EVF programs, such as Cool Edit Pro and Cool Edit 2000 by Syntrillium Software, OGRI Speechoolkit, Wavesurfer, and Praat, which can be downloaded from the web for free. The relatively large selection of software currently available with speech analysis functions catering to a wide variety of needs and interests comes as little surprise (See Anderson-Hsieh, 1998 for a survey of different pronunciation software/hardware programs on the market).

The main purpose of this paper is to introduce some of the basic characteristics of EVF software and pedagogy. Specifically, an EVF training plan is proposed for helping Japanese learners improve their pronunciation of difficult English vowels. Pronunciation patterns utilizing acoustic data (EVF) are presented as a

way to assist Japanese learners to overcome the negative influences of L1 interference in order to help them reduce their Japanese-flavored accent of English vowels. EVF provides L2 learners with a deeper sense of their own pronunciation by enabling them to graphically compare their own pronunciation with their teacher or with sound data samples that have been copied and stored into a database.

EVF Software Features

With EVF, users can record their voice and perform an acoustic analysis of their speech with functions for showing and measuring intonation, intensity, duration and frequency range. Figures 1-3 below show the different types of the acoustic analysis that can be performed using EVF. Real-time input of the teacher's own speech via a microphone, pre-recorded speech samples, as well as wave files from databases and other sources can be used as the target model. During the initial stages of training, the teacher selects the models for students, although it is possible to let experienced learners select their own models. The model speech sample is visualized on the student's computer screen utilizing the signal analysis procedure, which is best suited for the specific training task. For example, a waveform and a spectrogram of the speech signal are normally used for teaching vowels and consonants (Figure 3), while a combination of a waveform, pitch

contour and intensity curve are required for the acquisition of pitch and intonation (Figures 1 and 2).

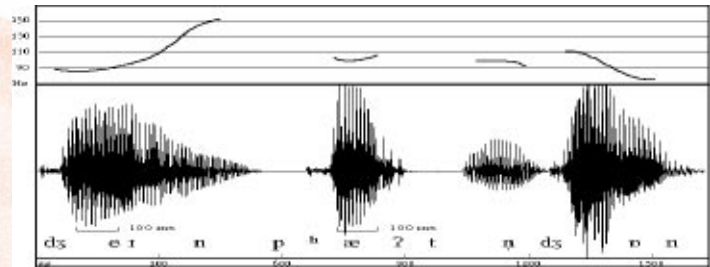


Figure 1: *Waveform--Pitch display*

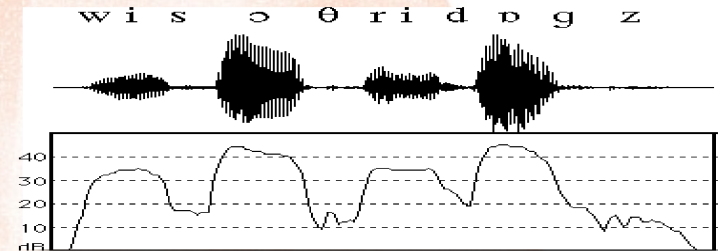


Figure 2: *Waveform--Intensity display*

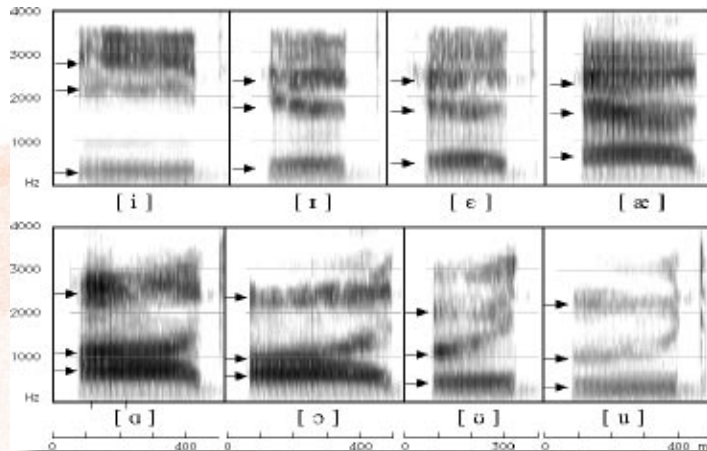


Figure 3: *Spectrographic display*

Students are initially given a small tutorial on how to analyze the speech signal by learning some basic information about acoustic phonetics, i.e., what a sound wave, spectrogram, pitch contour and intensity display all are. Students learn that aperiodic sounds, such as vowels, contain a frequency pattern in which acoustic energy radiates at many different frequencies but in a repeated pattern of change. Energy concentrated at certain frequency levels shows up as dark markings on a spectrogram that are measured in Hertz (Hz). Vowels have three formants or overtones that pertain to the resonating frequencies of the air in the vocal tract (Ladefoged, 2001: 31-35). This acoustic phonetic

instruction is briefly repeated when appropriate at the beginning of each EVF session until it is clearly understood by students. Then, students are presented with the model pattern of a word or sentence to imitate and instructed in how to interpret the target sound feature(s), provided with practice opportunities, and guided along by the teacher until they are able to imitate the target pattern. After learning to recognize the patterns of target sounds, students learn how the adjustments they make in their vocal tract affect the visual display of the sounds on their monitors. The focus of this practice is to help students understand the relation between their articulatory activity and acoustic output. With minimal training using EVF, students begin to recognize the acoustic patterns of the sound features and begin mastering them on their own. For training in discrimination of similar sounds, students practice minimal pair exercises to improve their awareness of the difference between the sounds. For building fluency, practice the target sounds within words, sentences, or short dialogues that are stored in files in the computer. Finally, students record, analyze (acoustically) and compare their productions of the sounds with the teacher's model patterns and with their earlier attempts. Students typically begin to notice improvement in their pronunciation, which is determined by the degree to which their productions match the teacher's model pattern with regard to the

targeted feature(s). With many EVF programs, the students' results can be printed out for homework and grading purposes. The model speech signal or its selected part can also be played back an unlimited number of times. Initially, training sessions are conducted in class. Afterwards, students benefit from having self-access to EVF for additional practice. By showing a more precise feature, a visual display provides a more accurate and objective measurement by which students and teachers can evaluate students' pronunciation errors and progress.

Accent Reduction of Vowels

Most, if not all, teachers do have not sufficient time to address all the sounds of English in the pronunciation classroom. The number of vowels in English Depending on the dialect, the English language can have upwards of 19 distinct vowels including diphthongs. The number of vowels in American English is shown below in Table 1. Of these, there are some that cause more difficulty for Japanese speakers than others. Therefore, prioritizing training objectives is essential.

Table 1: The 15 vowels of American English, each presented within an example word. The phonetic symbols of the International Phonetic Association (IPA) are used (adapted from Ladefoged 2001).

	b_d	IPA		b_d	IPA
1	bead	i:	9	boode	oo
2	bid	ɪ	10	boood	u:
3	bayed	eɪ	11	bud	ʌ
4	bed	e	12	bird	ɜ:
5	bad	æ	13	bide	aɪ
6	bod(y)	ɑ:	14	bowed	aʊ
7	bawd	ɔ:	15	Boyd	ɔɪ
8	budd(hist)	ʊ			

Due to their small inventory of the five vowels /i, e, a, o u/, Japanese learners typically have difficulty with English vowels that either completely differ from any Japanese vowel (/æ/ as in bad, for example) or are only slightly different from one of the five Japanese vowels (English /ɑ/ as in cot). Prior research has, in fact, shown that Japanese learners have difficulty perceiving and pronouncing the English vowels /æ/ /ɑ/ /ʌ/ /ɔ/ and /ə/ (see Strange et. al., 2001; Lambacher et. al., 2000). Since the Japanese language has only one vowel (/a/) that occupies a space within area of the English central and low vowels /ɑ/ /ʌ/ /ɔ/ and /ə/, Japanese speakers

typically substitute /a/ for these vowels in spoken English (See Figure 5 below). Japanese also commonly substitute a long /a/ vowel in place of the English /ə/ or any number of vowel + /r/ combinations, as in [baa:d] for [bird].

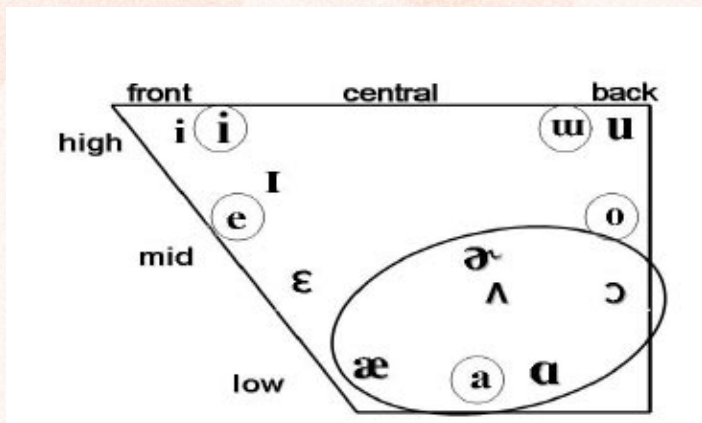


Figure 4: A vowel chart showing the five point vowels of Japanese in small circles and the location of the five AE vowels that are particularly difficult for Japanese speakers within the larger circle. The descriptors front, central, back, high, mid, and low correspond roughly to the location of articulation of each vowel within the vocal tract.

In the classroom, the teacher can give priority to the English mid and low vowels or to any other vowel(s) the students are struggling with. One of the objectives

of EVF in teaching English vowels is to help Japanese learners to distinguish the phonetic differences between similar vowels. Initially, familiarizing students with the patterns of the target vowel sounds achieves this. Using EVF, the vowel formant patterns clearly show up on the monitor, enabling students to easily and objectively observe and measure the formants of the vowels they produce. Students can record minimal pairs that contrast words containing mid and low vowels, for example, [bad]—[bud], [pat]—[pot], [taught]—[Tut], etc., analyze the first three formants of the target vowels using the frequency measuring bars of the EVF software, and then compare their patterns with those of the teacher's. Figure 5 shows the first three formants for the words *bad* and *bud*. Notice the F2 of /æ/ is about 500 Hz. higher than that of /ʌ/. Of course, vowel formant frequencies will vary based on the speaker's gender and vocal tract length, so it is necessary for the teacher to set target ranges for students to imitate. The main objective is for students to approximate an acceptable target and not to imitate a model pattern precisely.

Another important phonetic feature of vowels is duration. Vowels are classified as long or short depending on their quality and the context they occur in. A vowel is shorter when it occurs before a voiceless consonant than before a voiced consonant, as in the words [bat] and [bad], respectively.

Japanese vowels also have this long—short distinction,

so it usually takes just a few attempts for students to imitate the vowel durational pattern. English vowels can also vary in length depending on their vowel quality. For example, the vowel in the word [bed] is shorter than the vowel in the word [bad].

Other examples include the vowels in [beat]—[bit] and in [cook]—[kook]. The teacher can produce these vowel patterns for students to visualize and then have them measure the vowel length on their monitors. In addition, the English diphthongs [ai], [ei], [oi], and [ou] can also be presented in isolation and within words to

help students work on vowel duration.

Students also learn how vocal tract adjustments influence the pattern of the target sounds on their monitor, which helps them understand the relation between their articulatory activity and acoustic output. One difference between Japanese and English vowels may be related to more subtle articulatory gestures of Japanese while they are speaking. For example, it takes large gestures to create vowel qualities that extend into the back and lower ranges. Americans tend to round their lips to produce the mid-back vowel /ɔ/

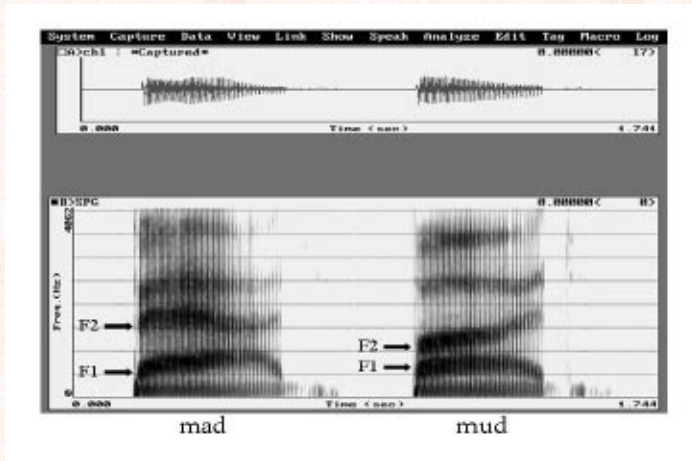


Figure 5: Spectrograms of the words *mad* and *mud* as recorded by an English native speaker. The vertical line is the frequency from 0 to 4,000 in (Hz). The horizontal line is duration (ms)

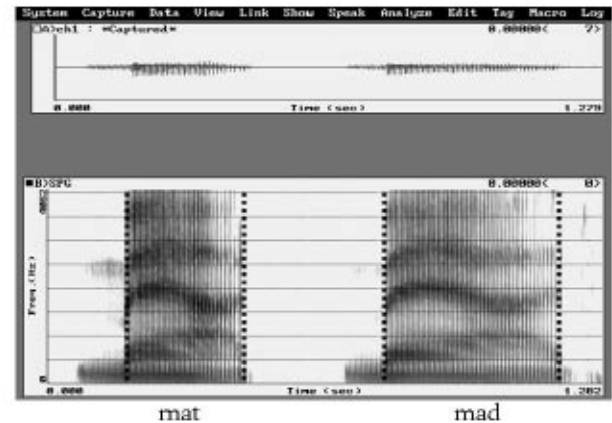


Figure 6: Spectrograms of the words *mat* and *mad* as recorded by an English native speaker. The vowel sounds are can be seen within the dotted line. The vertical line is the frequency from 0 to 4,000 in (Hz). The horizontal line is duration (ms).

and drop their jaw to produce the low vowels /æ/ and /ɑ/. For more appropriate pronunciation, students alter their vocal tract gestures for more appropriate vowel production by moving their articulators (e.g., tongue, jaw, lips, etc.) to more closely approximate the English vowels as modelled by the teacher. Students can then return to using EVF to see if these articulatory adjustments result in their pattern coming closer to the target range. It usually takes only a few attempts during a single class period before students begin to notice improvement in their vowel production.

Conclusion and Future Directions

This paper has introduced some of the basic features of EVF software and methodology for accent reduction. Specifically, an EVF training plan was proposed for helping Japanese learners improve their pronunciation of difficult English vowels. One benefit of EVF is it is motivating to students because it appeals to more senses. This is important to Japanese learners, in particular, because they tend to exhibit a learning style that responds well to visual stimuli. By showing the exact features that need changing, EVF provides an objective measurement by which students and teachers can evaluate and assess learners' mistakes and progress. EVF can greatly assist teachers and students in identifying speech errors and progress than by just listening to students' production in class or from recorded tapes.

Because teachers can provide feedback to students in real-time, students can correct their mistakes right away.

Even with its superior technological functions and positive results, some have questioned the promises of EVF as an effective pronunciation-training tool (e.g., Pennington, 1999: 431-32). A common problem for some users is that EVF displays are not particularly user-friendly. Because EVF displays were not originally intended for language learning, they are sometimes too complicated to operate. To help alleviate this problem, teachers should first spend sufficient time becoming familiar with the EVF equipment beforehand using it in the classroom. A basic understanding of acoustic phonetics is very useful. *The acoustic analysis of speech* by Kent (1992) is an excellent resource for learning the acoustic properties of English segmentals and suprasegmentals. Another area of focus should be on the development of pedagogy that facilitates the transfer of EVF training to communicative situations outside of class. Japanese learners should be provided with opportunities in the classroom to transfer their knowledge to natural settings, for example, role-play and oral presentations. Training should also focus on teaching self-monitoring skills to enable students to apply the skills and knowledge they learn to situations outside the classroom. Even so, the benefits of EVF far outweigh any shortcomings. Students exposed to both audio feedback and EVF tend to repeat sentences

more often and make more effort to correct their mistakes than when exposed to only audio feedback. EVF is not intimidating since the goals are more objective than those of traditional methods of speech and pronunciation instruction and students can work independently on their pronunciation outside of class.

Students correct a pronunciation error when it was pointed out to them with EVF than without it. Even less motivated students are easily excited by the graphical display of voice patterns and pitches, which encourages them to keep practicing by imitating the native speaker model.

References

- Abberton, E. & Fourcin, A. (1975). Visual feedback and the acquisition of intonation. In Lenneberg & Lenneberg, *Foundations of Language Development 2*, 157-165. New York: Academic Press.
- Anderson-Hsieh, J. (1992). Using electronic visual feedback to teach suprasegmentals, *System*, 20, 51-62.
- Anderson-Hsieh, J. (1998). *TCIS Colloquium on the uses and limitations of pronunciation technology: Considerations in selecting and using pronunciation technology*. Paper presented at the 32nd Annual TESOL Convention, Seattle, WA.
- Best, C. (1995). A direct realist view of cross-language speech perception. In Strange, *Speech perception and linguistic experience: Issues in cross-language research*, pp. 171-204, York Press: Baltimore.
- Kent, R. D. & Read, C. (1992). *The Acoustic Analysis of Speech*. San Diego: Singular Publishing Group, Inc.
- Ladefoged, P. (2001). *Vowels and Consonants: An Introduction to the Sounds of Languages*. New York: Harcourt Brace Jovanovich College Publishers.
- Lambacher, S. (1999). A CALL tool for improving second language acquisition of English consonant sounds by Japanese learners. *Computer Assisted Language Learning*, 12 (2): 137-156.
- Lambacher, S., Martens, W., G. Molholt. (2000). A comparison of identification of American English vowels by native speakers of Japanese and English. *Proceedings of Meeting of the Phonetic Society of Japan*. Chiba, Japan: 213-218.

- Molholt, G. (1990). Spectrographic analysis and patterns in pronunciation. *Computers and the Humanities*, 24: 81-92.
- Pennington, M. (1999). Computer-Aided pronunciation pedagogy: Promises, limitations, directions. *Computer Assisted Language Learning*, 12 (5): 427-440.
- Riney, T. & Anderson-Hsieh, J. (1993). Japanese pronunciation of English. *JALT Journal*, 15 (1): 21-36.
- Strange, W., Yamada, R., Kubo, R., Trent, S., Nishi, K. & Jenkins, J. (1998). Perceptual assimilation of American English vowels by Japanese listeners. *Journal of Phonetics*, 26: 311-344.
- Trubetzkoy, N. (1969). *Principles of Phonology*. Translated by C. A. Baltaxe, Berkley: University of California Press.