English Language & Literature Teaching, Vol. 13, No. 2 Summer 2007

# Teaching Pronunciation Using Sound Visualization Technology to EFL Learners

**Sujung Min**[*]
(Kongju National University)
**Hubert H. Pak**[**]
(Kongju National University)

**Min, Sujung & Hubert H. Pak. (2007). Teaching pronunciation using sound visualization technology to EFL learners. *English Language & Literature Teaching, 13*(2), 129-153.**

When English language teachers are deciding on their priorities for teaching pronunciation, it is imperative to know what kind of differences and errors are most likely to interfere with communication, and what special problems particular first-language speakers will have with English pronunciation. In other words, phoneme discrimination skill is an integral part of speech processing for the EFL learners' learning to converse in English. Training using sound visualization technique can be effective in improving second language learners' perceptions and productions of segmental and suprasegmental speech contrasts. This study assessed the efficacy of a pronunciation training that provided visual feedback for EFL learners acquiring pitch and durational contrasts to produce and perceive English phonemic distinctions. The subjects' ability to produce and to perceive novel English words was tested in two contexts before and after training; words in isolation and words in sentences. In comparison with an untrained control group, trainees showed improved perceptual and productive performance, transferred their knowledge to new contexts, and maintained their improvement three months after training. These findings support the feasibility of learner-centered programs using sound visualization technique for English language pronunciation instruction.

**[Pronunciation/Sound visualization technique/Phoneme discrimination/Pitch and durational contrasts]**

---

[*] 1st author
[**] 2nd author

## I. INTRODUCTION

There has been a large increase in the interest shown towards pronunciation instruction in the field of English language teaching. However, no one method or technique has been proved to be more beneficial than another (Pennington & Richards, 1986). Teachers have had to rely on their own intuitions or those of materials developers to decide on the emphasis a pronunciation course should take. Many language teachers are also realizing that pronunciation cannot be separated from communication since people use sounds to communicate and understand lexical, grammatical and socio-linguistic meaning (Morley, 1991; Pennington & Richards, 1986).

Phonetics and phonology have recently been revived in pronunciation teaching and the new approach is more balanced in focus, placing emphasis on both the segmental as well as prosodic features of pitch, rhythm and stress. In this way, phonetic and phonological structures are used only as a tool to assist learners in achieving communicative competence. Although attention is given to individual phonemes and allophones, more attention is placed on the suprasegmentals and how they are used to communicate meaning. The goal is not just competency at the segmental level (which can be counterproductive), but a focus on suprasegmentals with practice in stress, intonation, co‐articulation and voice quality settings; those qualities that help define the voice of an individual and the accent of their particular language dialect (Flege, 1995; Morley, 1994).

Along with this renewed interest in pronunciation, the availability of multimedia in computer assisted language learning (CALL) has led to a growing interest in ways of improving second language (L2) learners' pronunciation skills using computer‐based techniques. Over the last twenty years, the development of advanced visualization equipment has opened new approaches to teaching methods. The evolution of sophisticated technology such as the computer has led to the creation of software programs that can quickly and accurately extract pitch, frequency, and intensity of recorded speech and display the acoustic and spectrographic patterns on a computer screen. Some applications also include a dual display that can assist teachers and students in objectively evaluating speech errors and progress as well as analyze and compare phonetic data of English with other languages.

Two types of computer programs developed for teaching pronunciation include one model for measuring the sound source (vocal chords and articulators) and another model for measuring the acoustic signal (de Bot, 1980). Kalikow and Swets (1972) developed a program that provides information about the position of articulators from the acoustic signal without directly measuring them. They later designed spectrographic equipment that uses band - pass filters to measure frequency distribution, amplitude and durational distribution from isolated sounds or short words with applications for L2 learning. Leon and Martin (1972) and de Bot (1981) focused on building intonation awareness via fundamental frequency displays using pitch contour displays (Leather, 1983).

Past research findings have shown that a combination of audio and visual feedback can be effective in teaching all the aspects of English pronunciation (Lambacher, Martens & Kaeshi, 1994). Computer programs can provide the learner with real - time information about the salient properties of acoustic production. A visual display can provide an objective and instant measurement of speech production, helping teachers and students to identify the features that need changing. de Bot (1981) was successful in using visual displays to help learners in L2 intonation as long as students were first briefly exposed to theoretical instruction. Molholt (1988) helped training assistants overcome pronunciation problems in Chinese using visual display technology, and he reports improvement during a 4 - 6 week period, 1 hour per week, as long as correct instruction was given beforehand to reinforce the visual display.

This paper will show that computer-based training using sound visualization technique can be effectively implemented to Korean EFL learners to produce and perceive English pitch and durational contrast for words in sentences as well as for words in isolation.

## II. SOME KEY RESEARCH FINDINGS

In the development of CALL materials for pronunciation instruction, a number of valuable insights can be derived from previous work. Especially, EFL/ESL speakers' difficulty in using duration and pitch to produce and perceive English phonemic distinctions has been noted in various studies (Celce-Murcia, Brinton & Goodwin, 1996; Chun & Plass, 1996; Kang, 1999; Ladefoged, 1993). Here we describe three particularly important findings that will be put into practice in the present study.

### 1. The Advantages of Identification Training

As noted above, L2 speech research has used both discrimination training, in which trainees must say whether pairs of L2 segments (i.e., consonants and vowels) are the same or different, and identification training in which the trainees must specify the speech sound they have actually heard. Evidence indicates, however, that discrimination tasks are not optimal for training learners (Hewings, 2004; Jenkins, 2000; Long, 1996; Yavas, 2006). Rather, identification tasks have yielded better results in these studies, possibly because they lead trainees to direct their attention to the specific characteristics of a speech sound that make it differ from the other member of the contrastive pair.

### 2. The Perceptual Fading Technique

In everyday situations the speech that we hear varies in clarity. Noting this, Chapelle (1990) found improvement in French speakers' perceptions of English interdentals (/θ/ and /ð/) when they used a perceptual fading technique. In this approach the trainees began with very distinct exemplars of the two speech sounds, comparable to carefully articulated speech. Over time, less clear exemplars were introduced into the training to widen the stimulus range so that it would resemble the range of speech that trainees would hear outside the laboratory. Because careful control over stimulus properties is necessary, this approach requires synthetic speech or digitally modified natural speech.

### 3. Speaker Variability

Following perceptual training, participants are sometimes better able to perceive speech from a voice that they have been trained on than from unfamiliar voices (Larsen-Freeman & Long, 1991). For instance, if they have learned to hear the difference between lake and rake with voice 'A,' they may have trouble hearing the difference with voice 'B'. This problem can be partly reduced by using a variety of different voices during the training phase rather than a single voice (Morley, 1994).

## III. RESEARCH PURPOSE

### 1. Pronunciation Problems of Korean Students in Learning English

Before initiating into a research, it is necessary to discuss problems of Korean students in learning English as there are prevailing misconceptions about the nature of spoken language learning. Identification of the kinds of errors that is most likely to interfere with communication before deciding on the priorities for pronunciation teaching must be in order (Hewings, 2004). Furthermore, knowing how different the first language is crucial factor deciding on the priorities for pronunciation teaching as each different individual face difficulties in each different area of pronunciation. It is imperative to verify the needs of each different L1 speakers. The responsibility of the teacher is to identify the needs of learners when it comes down to the final stage of correction of the speech. It is not always viable to just differentiate the sound and pronunciation gap as the pronunciation teaching cannot always be relied on the learners' own awareness (Kim, 2004). Many researches have shown that shown that the 'vowel length', 'vowel shift' and 'consonant clusters (blends)' to be more problematic for Korean students (Hewings, 2004; Kim, 2004;Yavas 2006).

The following best summarizes the major trouble spots for Korean learners set out by Yavas (2006).

- missing target phonemes
- sounds existing as allophones
- salient phonetic differences
- insufficient separation of target vowel contrasts
- onset and coda clusters
- stress
- rhythm

The sound system is the biggest problem for Korean speakers to face when it comes to speak English more fluently. Four major problems are individual speech sounds due to transference from the first language, obtrusive vowel insertion due to differences in syllable formation between Korean and English, lexical stress due to the syllable-timed nature of Korean, and the use of weak forms and in the placement of sentence-stress due to unfamiliarity with the stress-timed nature of English. With the major trouble spots set out by Yavas (2006), dividing these into the problems of the sound system must be segregated and verified prior to the teaching of English.

## 2. Research Purpose

Like reading, effective listening requires the listener to be active and thinking.

Understanding what you hear depends on your background knowledge of the topic or situation and your knowledge of the language itself. While reading requires knowing the symbols of written language, listening requires recognizing the sound of knowledge. The information sources for comprehension in listening are therefore parallel to those skills for speaking.

Specially designed listening material is useful to provide practice for specific skills, but many of the best listening activities are those where listening is integrated with speaking, reading and writing. All the ideas for communicative and collaborative activities for listening lead naturally into pronunciation practice. As a general rule, unless a student's speech is interfering with communication, it is much better to spend time on general language development rather than on isolated speech activities aimed at improving pronunciation. However, where speech is a barrier to communication and the learner of EFL is no longer in the very early stages of English, it may be necessary to devise some more specific activities.

The software that is specifically developed to overcome this barrier providing a comprehensive pronunciation activity using audio-visual aid for listening and speaking skills is the Sound Visualization Technique. Sound Visualization Technique is the software that enables to turn a selected sound into a visualizing sound. It samples various information from a sound to visualize and express as an image, so that image respond to sound. The image becomes full of movement by producing image effects with selected sound which assists student to acquire accurate and proper English pronunciation visually, and ultimately improves listening skills. The technique initially concentrates on pronunciation practice, and later on general fluency and speech rhythms. Once a speaker has mastered the stress and intonation patterns of English through this technique of visualizing sound, pronunciation difficulties with specific sounds are often much less apparent.

Students are allowed to discriminate between target sounds visually and then to repeat words or phrases that contain these sounds, with the Technique audio-visually correcting where necessary. The activities using this Sound Visualization Technique leads students to predict, identify, sort, match, discover rules, and exchange information. The activities are intended to help students to become more aware of their own English pronunciation, stress and intonation, and ultimately speech rhythms, and those of native English speakers to analyze them to produce various features of pronunciation in relevant context, both visually and audibly.

## IV. RATIONALE FOR THE STUDY

### 1. Pronunciation Instruction Using Sound Visualization Technique System

The computerized Sound Visualization Technique enables users to perform an acoustic analysis of their speech. Included are functions for measuring amplitude, intonation, pitch, duration, and frequency, the results of which can be shown on a spectrographic display of voice patterns and pitches. The users' voice or voices from various (audio and visual) (AV) devices (laser disks, videos, TV, 8mm, tapes) can be recorded and two separate voices can be simultaneously displayed on dual screens. Voice data that is both visually and audibly analyzed by students can be called up on a workstation (WS) screen, and voice data analyzed by the teacher can be shown to students. It is also possible for a teacher to also control a set table displayed on the WS to guide students individually or in groups as well as monitor what each student sees, listens to, and says. The function of the Sound Visualization Technique system is very appealing and effective as a teaching tool since it allows students to visualize their pronunciation as they learn to associate the patterns on the display with the sounds.

**FIGURE 1**
**Unprocessed Fundamental Frequency Measurements**

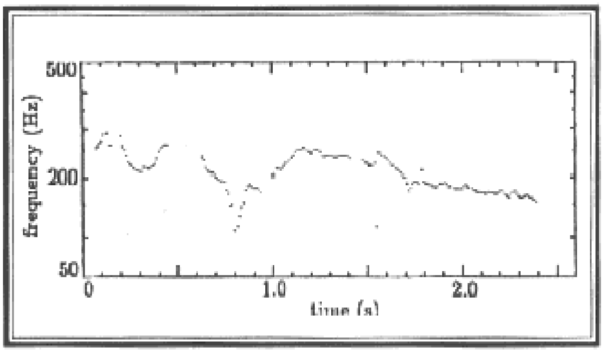
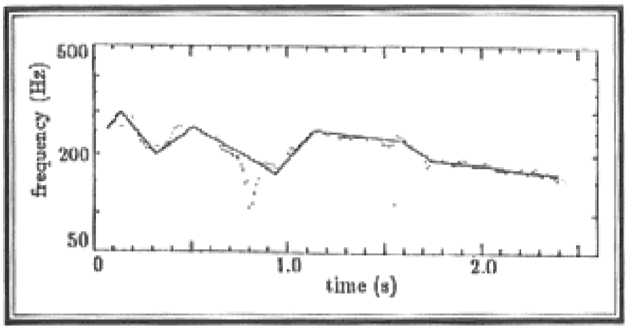


**FIGURE 2**
**"Unprocessed" Pitch Contour (dots) and Stylized Pitch Contour (straight line)**




## V. RESEARCH DESIGN

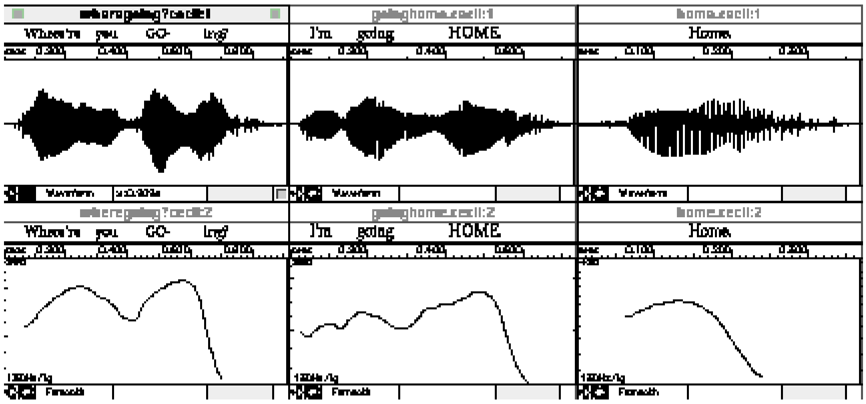
### 1. Trained versus Control Groups

Training techniques for the production of L2 contrasts have been developed and examined in several experimental studies (Chapelle, 1998; Chun & Plass, 1996; Doughty, 1991; Leather, 1983; Leon & Martin, 1972; Swain & Lapkins, 1995). These studies conducted experiments to show that certain aspects of L2 speech production can be trained, and that production training might also lead to improved perception. However, the development of training techniques for L2 pronunciation is in its infancy, and there is much to be explored in assessing whether various methods of pronunciation training are effective in enabling subjects to accurately produce L2 contrasts (see Pennington & Richards, 1986, for discussion). In particular, none of these experimental studies above compared trained subjects with control subjects who did not participate in training. Simply testing trained subjects before and after training does not reveal how much of the subjects' improvement could be attributed to the training and how much was from simply repeating the tests. In order to differentiate the training effect from the repetition effect, the present study compared trained and control (untrained) subjects in a pre-test/post-test design. An effect of training should be indicated by the difference between the two groups' improvements from the pre-test to the post-test. The first question explored in the present study, therefore, was whether native Korean subjects who have participated in production training improve their performance significantly more than the control subjects who only take a pre-test and a post-test.

**FIGURE 3**

**Examples of Fundamental Frequency Contours of Model Sentences**

"Neutral" question and two statements (replies)




*Where're you going?* *I'm going home.* *Home.*

## 2. Word-in-Isolation versus Word-in-Sentence Contexts

In seven out of the ten training sessions of the present training program, target words were embedded in phrases or sentences. The effectiveness of computer-based training on three English vowel contrasts – /i/-/I/, as in beat vs. bit, /u/-/ʊ/, as in Luke vs. look, and /ɛ/-/æ/, as in bet vs. bat, were tested. The participants are speakers of Korean, who are likely to conflate these pairs because they do not contrast phonemically. For at least the /i/-/I/ contrast, many learners tend to rely inappropriately on length as a means of distinguishing the two sounds. Korean speakers seem unaware that both vowels can be produced as long or short. This tendency may be reinforced by the common but largely incorrect pedagogical practice of teaching the /i/-/I/ contrast as “long” and “short” categories (Munro & Derwing, 1995). Therefore, in order to perceive and produce this pair correctly, learners must learn to ignore length differences and pay attention to how the vowel has been articulated by the speaker. In other words they most focus on vowel quality rather than length. This contrasts with the earlier studies (e.g., Kalikow &Swets, 1972; Krashen & Terrell, 1983; Leather, 1983; Munro & Derwing, 1995) that presented only isolated words of L2 contrasts. For example, the minimal pair words contrasting in pitch, /u/-/ʊ/, as in Luke vs. look, were not presented sequentially in isolation, but embedded in sentences. Similarly, the durational difference between /i/-/I/, as in beat vs. bit, was not made explicit, since the words were embedded in separate sentences. Thus, subjects' attention was directed not to how these words minimally differed from each other, but to imitating the pitch contours of the overall utterances.

The intent of this method was to provide subjects with minimal pair words in a variety of connected speech contexts. A number of studies have shown that perception and production of words or segments in isolation is different from that of words in sentence contexts (Kalikow &Swets, 1972; Krashen & Terrell, 1983; Leather, 1983; Munro & Derwing, 1995). The second question in this study, therefore, was whether production training that focuses on words embedded in phrases or sentences has an effect on trained subjects' performance in isolated word and sentence contexts.

### 3. Production versus Perception

Several studies agree that gains from perceptual training on L2 contrasts can transfer to production (Bradlow, Pisoni, Akahana-Yamada, & Tohkura, 1997; Jenkins, 2000; Lambacher, Martens, & Kelly, 1969; Munro & Derwing, 1995). However, there has been little agreement on the effects of production training relative to perceptual training (Bradlow, Pisoni, Akahana-Yamada & Tohkura, 1997; de Bot, 1981; Leather, 1990; Seidlhofer, 2003). Leather (1990) concluded, "training in one modality tended to be sufficient to enable a learner to perform in the other" (p. 201). On the other hand, Seidlhofer (2003) pointed out that the training of production or perception may dissociate the link between production and perception that the trained subjects initially had.

In the present study, the task for the native Korean participants during the production training was to match f0 (fundamental frequency) and duration patterns exhibited by English target words (either isolated or embedded in sentences). Participants were given as real-time visual feedback the f0 contours of their own speech and of models. They could also listen to model tokens whenever they wished during training. Thus, this method did involve perception to some extent. However, since subjects might attempt to imitate the models without attending to the actual sounds of the tokens, but only to the visual input, it is not clear whether this training method affects subjects' perception as well as production. The third question addressed in this study was whether the present training method had effects not only on production but also on perception. For this question, subjects were tested in these two domains before and after training.

## VI. RESEARCH QUESTIONS

The primary research question in this study is as follows: (1) Can computer-based training using procedures developed in laboratory contexts be effective in improving L2 learners' vowel perception? Our approach differs from many previous training studies in that we do not prescribe a rigid schedule or amount of training for the participants. Instead, we allow them some control over the content of their training, and on the time and amount of practice they receive. Moreover, the training is given over two months — a longer time frame than in any other study we know of. This approach might be compared with allowing EFL students to participate in computer-based practice sessions outside of classes at their own discretion. Improvement in perceptual performance under such conditions would suggest that laboratory training techniques may be applied successfully in pedagogical environments.

In addition to our main research question, we address three secondary questions: (2) Can training with synthetic speech stimuli help Korean listeners learn to ignore duration differences between vowel pairs and pay attention to vowel quality instead? (3) Will trainees generalize newly acquired knowledge about vowels to new speech tokens and new speakers? And (4) Will new knowledge about vowels acquired from training be retained after the training has stopped?

In summary, the questions addressed by this study were: (1) Is production training that provides f0 contours as visual feedback effective in assisting the overall acquisition of English pitch and duration contrasts? Does the overall improvement of the trained group from the pre- to the post-tests differ from the improvement of the control group? (2) Having participated in the training that provided isolated words, phrases, and sentences, does the trained group improve its test scores in both word-in-isolation and word-in-sentence contexts? (3) Does this production training have effects on both production and perception? If production learning occurs, is it accompanied by perceptual learning?

## VII. METHODOLOGY

The structure of the experiment was as follows: one group of native Korean speakers was trained to produce English pitch and durational contrasts in isolated words and in words embedded in phrases or sentences. This trained group took a pre-test (before training) and a post-test (after training) in both production and perception. A control group also took the same pre- and post-tests, but did not participate in training. Production and perception were assessed with isolated words (word-in-isolation context) and with those same words embedded in carrier sentences (word-in-sentence context).

### 1. Participants

Eight native speakers of Korean (5 males and 3 females) who were taking a second-year English Conversation course in the department of English Education at the Kongju National University (50 min a day, five times a week) participated. The participants were randomly assigned to a training and a control group ($n$ = 4 each). The participants' ages ranged from 19 to 21 years old (mean = 19.3 for each group). None of the participants reported hearing problems. Besides having taken the introductory English Conversation course at the Kongju National University in the previous year, all participants had studied English before college. The participants had had exposure to spoken English mainly by listening to their instructors and audiotapes that were required for the present and the past courses. Two of the subjects (one in the training and one in the control groups) had studied French in high school for less than 4 years, but the others had no experience of studying any other foreign languages. The training group participated in a pre-test, training, and a post-test, and the control group took the pre- and the post-tests at the same time as the training group. The control group, in addition, was given the opportunity to participate in the training after the post-test although their training data were not part of the present experiment.

### 2. Training Procedure

The training materials included 33 pairs and 2 triplets of English words contrasting in pitch and duration. These pairs and triplets of words were presented in isolation or embedded in phrases or sentences. The training consisted of 10 training sessions. Sessions 1-3 presented a total of 47 words in isolation. Session 4 presented words in

15 short phrases, and Sessions 5-10 presented words in a total of 57 sentences of different length. The speech materials were recorded by four native speakers of English (2 males and 2 females), as models for participants' production training. The materials were all recorded on DAT cassettes, and then digitized at 20 kHz on a PC.

An acoustic display of speech on a computer is not easily interpretable by non-phoneticians. The problem of dealing with raw acoustics of speech in computer assisted language learning has been pointed out by Chun and Plass (1996), Pennington and Richards (1986), and Swain (1985). For example, word boundaries do not correspond to breaks in f0 lines, and an f0 peak might not exactly correspond to the location of pitch accent perceived by native English speakers (Hewings, 2004; Yavas, 2006). It may not be clear to L2 learners whether a pitch rise or fall is phonetically relevant. To remedy such problems, a handout was made for each session, in which explanation was given on how to pronounce the utterances using prosody graphs. A prosody graph (Fig. 1; with circles and arrows) is a schematic description of how native English speakers perceive f0 contours. The length of the vertical lines indicated perceived relative pitch height, and the horizontal lines at the bottom corresponded to words or phrases. With the prosody graphs subjects were able to figure out on their own which part of the words or sentences corresponded to which part of the f0 contours, and roughly where the word boundaries would be in the f0 contours.

Subjects were trained individually in the language laboratory of the Kongju National University. In each of the 10 training sessions, the subjects opened each model audio file, listened to the model (isolated word, word in phrase, or word in sentence), and watched the pitch pattern of the model in real time. Then they produced the utterance on another empty window, and the program overlaid the model pitch contours onto the subjects' pitch contours. The subjects repeated this procedure, producing the same material until their pitch contours matched the model contours. The subjects were able to listen to model tokens whenever they wanted. Although the number of times subjects listened to models was not recorded, the experimenter observed in all the training sessions that it was approximately 1-4 times per token. It was also observed that the number of times they listened to models did not exceed the number of times they produced them.

Subjects were instructed that they move on to the next training set when they thought that the steepness and the duration of their f0 contours matched that of the model. The examiner was available at any time of the subjects' training sessions to answer their questions regarding the operation of CSL-Pitch or the materials. Each session took about 30 minutes. The 10 training sessions were completed in three and a half weeks.

## 3. Testing Procedure

### 1) Production Tests

Production and perception tests were conducted before and after training. Twenty-one words including pitch and durational contrasts were chosen for production tests, for example, /i/-/I/, as in beat vs. bit, /u/-/ʊ/, as in Luke vs. look, and /ɛ/-/æ/, as in bet vs. bat. Nine of the 21 words were used in training, and 12 words were not. The same sets of words were used in the pre- and the post-tests. Subjects' production of the 21 words was recorded three times in isolation (word-in-isolation context), and three times in the carrier sentence (word-in-sentence context). The order of the 21 words was randomized in each of the three repetitions.

The subjects' production was evaluated by two native speakers of English from Toronto, who were naive as to the purpose of the experiment and who had never heard the subjects' voices. Each judge identified a total of 2016 utterances produced by all subjects [21 words x 3 repetitions x 2 contexts (word-in-isolation and word-in-sentence) x 2 tests (pre- and post-tests) x 8 subjects]. Analog tape copies of the original DAT cassettes were prepared for the pre-test and the post-test and randomly labeled Test A and Test B. One judge listened to Test A first, and then proceeded to Test B, whereas the other judge listened to the two sets of tapes in the opposite order. The order of the subjects that each judge heard was randomized in each of Tests A and B, but production across different subjects was not mixed, that is, each judge identified all of one subject's production before moving to the next subject. The order of the 21 words was the same randomized order as they were originally recorded by the subjects. Each judge identified all the utterances produced by all subjects.

On the evaluation sheets, the judges were asked to choose one out of 3-5 alternatives, for example, "bɛt/bæt/ 'cannot-tell'" (three alternatives), "'b**I**t/b**i**t/b**i:**t/ 'cannot-tell'" (four alternatives), and "lo ʊ k/lu:k/l ʊ k/luk/ 'cannot-tell'" (five alternatives), which were all written in English orthography. For example, for the word /Luke/, which was randomly ordered among the 21 test words and recorded by subjects, the judges were asked which of the three words ("loʊk/lu:k/lʊk/luk/) they thought the subjects had intended to say. The number of the judges' identifications that matched the subjects' intent (production intelligibility) was then calculated. Since preliminary analysis on each judge's identification indicated no major discrepancy between the two judges, the present paper only reports the combined data.

2) Perception Tests

Perception tests also consisted of a pre-test and a post-test. Each test contained 30 novel words that did not appear in training. The perceptual stimuli were recorded by two female native speakers of English. Thirty English words were produced both in isolation (word-in-isolation context) and in carrier sentences (word-in-sentence context). Thus, there were a total of 60 utterances (30 words2 speakers) in each of the four tests (the word-in-isolation and the word-in-sentence tests in both the pre- and the post-tests), randomly ordered and divided into six blocks. For the perception tests of words-in-isolation, subjects first clicked a play button on a computer screen to listen to each word spoken in isolation. Subjects then chose by clicking one of the nine pitch patterns (HL, LH, HLL, LHL, LHH, HLLL, LHLL, LHHL, LHHH; H = high, L = low) of prosody graphs that matched the word they had just heard. This method required the ability to identify not only the pitch but also the duration of the words. For example, for the word /cutey/, the correct answer is LHH (HH: representing a long vowel), and not LH (H: representing a short vowel).

For the perception tests of words-in-sentences, a carrier sentence /she is a cutey/ was written on the computer screen. Subjects saw this carrier sentence, clicked the play button, and heard a sentence that included a target word variably. For this purpose, four different versions of each stimulus varying in pitch were created. Then the subjects were asked to choose the pitch pattern that matched the target word. Each of the pre- and the post-tests lasted about 30 min.

3) Analysis

The percent correct scores of the word-in-isolation and the word-in-sentence tests in pre- and post-tests were calculated. These perceptual scores and the production intelligibility scores (i.e., percent correct scores of the native English judges' identifications of the subjects' production) were examined in a mixed design analysis of variance (ANOVA). The factors were Group (trained vs. control), Test (pre-test vs. post-test), Context (word-in-isolation vs. word-in sentence tests), and Domain (production vs. perception). Group was a between subjects factor, and Test, Context, and Domain were within-subjects factors.

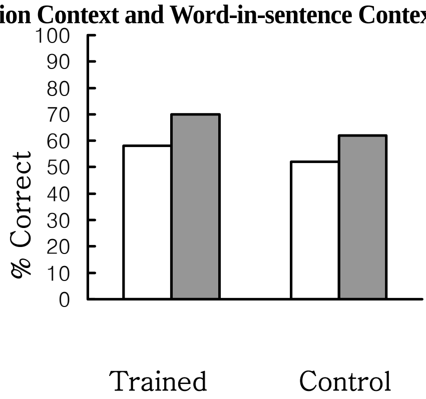
## VIII. RESULTS & DISCUSSION

### 1. Overall Effects of Training

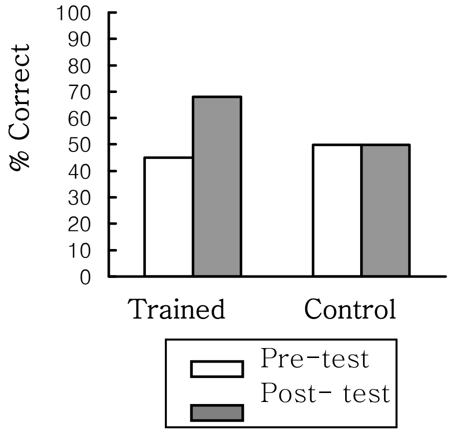
The first question of interest was whether the trained group improved its overall test scores from the pre- to the post-tests significantly more than the control group. The trained group of subjects showed a larger improvement in their overall test scores (including production and perception tests) from the pre-test (52.4%) to the post-test (69.5%) than the control group did (pre-test: 52.3%, post-test: 57.3%). The four-factor ANOVA revealed a significant Group x Test interaction [$F(1, 6)=11.972$, $p=.013$]. This indicated that the trained and the control groups showed significantly different amounts of improvement from the pre-test to the post-test. For the pre-test, the difference between the scores of the two groups was not significant [$t(6)=0.008$, $p=.190$]. However, for the post-test, the difference between the two groups was significant [$t(6)=1.364$, $p=.003$]. These results indicate that the training program was effective for improving the trained subjects' overall ability to distinguish English words contrasting in pitch and duration.

**FIGURE 4**
**Percent Correct Scores for Trained and Control Groups at Pre- and Post-test in Word-in-isolation Context and Word-in-sentence Context**

**Word-in-isolation**



**Word-in-sentence**




2. Training Effects on Word-in-Isolation versus Word-in-Sentence Contexts

The next question evaluated was whether the trained subjects showed equal improvement for the word-in-isolation and word-in-sentence contexts. The four-factor Group x Test x Context x Domain ANOVA revealed only a marginal Group x Test x Context interaction [$F(1, 6)=5.156$, $p=.064$]. Figure 2 shows the trained versus the control groups' pre- and post-test scores in the two contexts separately. In this figure, the subjects' production and perception scores in each context were combined and averaged. In the word-in-sentence context ((b) in Fig. 2), the trained group showed a significant improvement: the difference between the pre- and the post-tests for the trained group (21.4%) was significant [$t(3)=4.949$, $p=.016$]. In contrast, the difference for the control group (0%) was non-significant [$t(3)=0.006$, $p=.996$]. In the word-in-isolation context ((a) in Fig. 2), the difference between the two groups' improvement was much less than that in the word-in-sentence context. However, the trained group's improvement (12.8% from the pre- and the post-tests) reached significance [$t(3)=3.326$, $p=.045$] while the control group's improvement (10%) did not [$t(3)=2.132$, $p=.123$].

In summary, the trained group showed significant improvement in the two contexts, but it improved more in the word-in-sentence than in the word-in-isolation contexts. In contrast, the control group's improvement was non-significant in both the word-in-isolation and the word-in-sentence contexts.

### 3. Training Effects on Production and Perception

The above four-factor ANOVA revealed no interaction among Group, Test, and Domain [$F(1, 6)$=.034, $p$=.860]. This assures that the Group x Test interactions in the two domains (production and perception) was comparable to the Group x Test interaction reported earlier. Figure 3 shows the percent correct scores of the pre- and the post-tests by the trained versus the control groups separately for perception and production, averaging over the word-in-isolation and the word-in-sentence tests. For both perception and production, the trained group showed greater improvement from the pre- to the post-tests than the control group. For perception ((a) in Fig. 3), the difference between the pre-test scores of the two groups (0.9%) was not significant [$t(6)$=0.084, $p$=.101]. However, the difference between the post-test scores of the two groups (12.1%) was significant [$t(6)$=1.453, $p$=.004]. Comparable results were found for production ((b) in Fig. 3). In summary, the trained subjects significantly improved their ability not only to produce but also to perceive the pitch and durational contrasts.

An additional analysis examined production scores obtained just for novel words (i.e., scores on the nine words that appeared in training were excluded from the overall production scores reported earlier). The production scores on the 12 novel words by the trained versus control groups were calculated for both the pre- and the post-tests. The results were comparable to those on all production words. At the pre-test, the score of the trained group did not differ from that of the control group (69.3% vs. 69.3%) [$t(6)$=0, $p$=.413]. However, at the post-test, the score of the trained group was significantly higher than that of the control group (83.5% vs. 77.1%) [$t(6)$=.631, $p$=.040]. In summary, effects of training were found not only on the overall production scores, but also on the scores excluding the practiced production words.

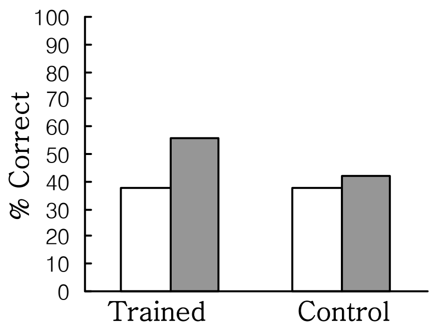
The scores of individual subjects were examined to evaluate changes in production and perception. Table 1 shows the individual subjects' percent correct pre- and post-test scores for production and perception. The trained subjects' improvement in production and perception was not consistently greater than the control subjects' improvement. The magnitude of training effects, if any, depended on each subject. For trained subjects #2 and #4, the effect of training was found more clearly in perception than in production. For example, trained subject #2 and control subject #1 both showed the same amount of improvement in production (10.7%), but trained subject #2's improvement was twice as great as control subject #1's in perception (20.0% vs. 10.0%). Similarly, trained subject #4's improvement in production was greater than that of control subjects #2 and #3 only by 1.9-2.0%. However, trained subject #4's perception improvement (25.8%) was approximately twice as much as that of control subject #2 (13.3%), and contrasted even more with that of control subject #3 (1.7%).

In contrast, the effect of training was greater for production than perception in the comparison of trained subject #1 and control subject #1. Although their improvement in perception was similar (9.1% vs. 10.0%), the improvement of trained subject #1 (18.7%) was superior to that of control subject #1 (10.7%) in production. Similarly, the production improvement of trained subject #3 (29.4%) was greater than that of control subject #2 (5.5%) (or greater than all control subjects) although the perception improvement of these two subjects was comparable (15.8% vs. 13.3%).

In summary, the group means indicated that the trained subjects improved both in production and perception. However, the comparison of individual subjects indicated that the degree of improvement in the trained subjects varied in production and perception. Two trained subjects (#2 and #4) showed greater perception improvement than the control subjects whose production improvement was comparable. The other trained subjects (#1 and #3), however, showed greater production improvement than the control subjects whose perception improvement was comparable.

**FIGURE 5**
**Percent Correct Scores of Perception and Production Test for Trained and Control Groups at Pre- and Post Tests**

**Perception**



**Production**

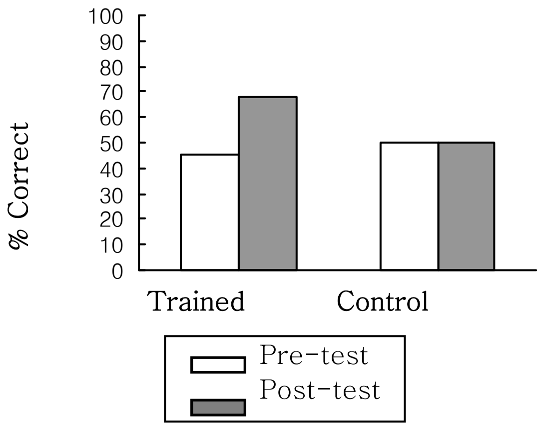



**TABLE 1**
**Individual Subjects' Improvement on Production vs. Perception Test Scores**

| Subjects | Production | | | Perception | | |
|---|---|---|---|---|---|---|
| | Pre-test | Post-test | Difference | Pre-test | Post-test | Difference |
| Trained 1 | 71.0 | 89.7 | 18.7 | 46.7 | 55.8 | 9.1 |
| Trained 2 | 64.7 | 75.4 | 10.7 | 33.3 | 53.3 | 20.0 |
| Trained 3 | 51.6 | 81.0 | 29.4 | 36.7 | 52.5 | 15.8 |
| Trained 4 | 80.2 | 87.7 | 7.5 | 35.0 | 60.8 | 25.8 |
| Control 1 | 71.4 | 82.1 | 10.7 | 42.5 | 52.5 | 10.0 |
| Control 2 | 41.7 | 47.2 | 5.5 | 19.2 | 32.5 | 13.3 |
| Control 3 | 85.3 | 90.9 | 5.6 | 63.3 | 61.7 | -1.7 |
| Control 4 | 65.1 | 63.9 | -1.2 | 30.0 | 27.5 | -2.5 |

## IX. CONCLUSIONS

The results obtained in this study support positive responses to all four of the research questions posed. The most important outcome was that the trainees' perception improved significantly on all of the phonetic contrasts on which they were trained, whereas a control group showed no such improvement. This study yielded three major findings.

First, the subjects who participated in the production training improved their overall test scores significantly more than the control subjects did. The trained subjects' attention was directed to both pitch and duration of the utterances they practiced producing with the Sound Visualization Technique. The focus of the training was for subjects to match the f0 contours of words, phrases, and sentences to those of the models. Minimal word pairs were only implicitly used in phrases and sentences in 7 out of 10 sessions.

It must be noted that the training method differed quite greatly from the testing method. In the perception tests, subjects were asked to identify pitch patterns, which they weren't during the training. While in the production training, they were trained to match f0 contours with those of the exemplars, they never received feedback on whether native speakers could perceive their attempt correctly.

The training resulted in the increased ability to perceive the pitch and durational patterns of the words, and the increased intelligibility of the subjects' production as evaluated by native English judges. The trained group improved on the practiced words but also shown a significant gain in their testing on novel words. We can thereby conclude that by training in the audio visualization program, the subjects acquired a general ability to produce and perceive English words that contrasted both in duration and pitch in contrast, the control subjects showed no significant improvement.

The second finding concerns the subjects' performance as tested in isolated word and sentence contexts. The present training method using words, phrases, and sentences was found to have an effect on the trained group's performance in both word-in-isolation and word-in-sentence contexts. The control group showed greater improvement in the word-in-isolation than in the word-in-sentence contexts, although neither one was significant. In contrast, the trained group made more significant improvements in their word-in-sentence performance than in their word-in-isolation performance. These results contrast with those of Carr (1999) in which trained L2 learners improved their scores for isolated words but not for sentences.

An implication of the present results is that EFL learners can learn from sentences to distinguish minimal pair words, even though their perception and production of L2 contrasts are consistently less accurate in sentence than in word contexts. Developing the ability to perceive word-in-sentence contexts is much more important for L2 learners when they encounter a real-world situation.

They must eventually learn to pay attention to multiple aspects of speech, that is, not only how a word is distinguished from another in a minimal pair but also, for example, whether the word has a contrastive pitch in a sentence. (In reality, L2 learners must also integrate the acquired phonetic and phonological knowledge into the use of language for meaningful communication, which is beyond the scope of this paper.) It is traditional to present minimal pair words of L2 contrasts in isolation in order to direct learners' attention to fine differences between those words. The finding that subjects could learn from sentences, however, suggests that it does not have to be that way, perhaps, for intermediate or advanced learners.

The third finding was that training subjects in producing English words and sentences improved not only their production but also their perceptive abilities. While training the group in productive methods, it can be inferred that it involved a certain inadvertent improvement in perception. This is not all that surprising when we consider that the exposure to perceiving the model and producing a replica would inherently improve the perceptive ability of the listener.

However, this finding must be interpreted with caution. Out of four trained subjects, only two subjects (#2 and #4) showed greater improvements in perception than production when they were compared to comparable control subjects. The opposite pattern of improvement was found for the other two trained subjects (subjects #1 and #3). Bradlow et al. (1997) concluded from their perception training study that perceptual learning generally transfers to production, but "learning in the perceptual domain is not a necessary or sufficient condition for learning in the production domain: the processes of learning in the two domains appear to be distinct within individual subjects" (p. 230). Since the number of subjects in the present study was small, more data is necessary to determine whether production training also yields great individual variation in the production-perception improvement as in Bradlow et al. (1997).

In conclusion, the findings in the present experiment suggest that the production training method developed and tested in this study was effective in enabling native Korean speakers to produce and perceive English pitch and durational contrasts. This exploratory experiment, however, motivates further studies to address various theoretical and pedagogical issues on the production and perception of L2 contrasts.

## REFERENCES

Bradlow, A., Pisoni, D., Akahana-Yamada, R., & Tokura, Y. (1997). Second language instruction. *Journal of Acoustical Society of America*, *13*, 431-469.

Carr, P. (1999). *English phonetics and phonology*. Oxford: Blakwell Publishing.

Celce-Murcia, M., Brinton, D. M., & Goodwin. (1996). *Teaching pronunciation: A reference for teachers of English to speakers of other language*s. Cambridge: Cambridge Univ. Press.

Chapelle, C. A. (1998). Multimedia CALL: *Lessons to be learned from research on instructed SLA. Language Learning & Technology, 2*(1), 22-34.

Chapelle, C. A. (1990). The discourse of computer-assisted language learning: Toward a context for descriptive research. *TESOL Quarterly*, *24*(2), 199-225.

Chun, D. M., & Plass, J. L. (1996). Effects of multimedia annotations on vocabulary acquisition. *The Modern Language Journal, 80*(2), 183-198.

de Bot, C. (1980). The role of feedback and feedforward in the teaching of pronunciation—an overview. *System, 8*(1), 35-45.

de Bot, C. (1981). Visual feedback of intonation, and experimental approach. In B. Sigurd & J. Svartvik (Eds.), *Proceedings of the Institute of Phonetics*, 3 (pp. 24-39).

Doughty, C. (1991). Second language instruction does make a difference: Evidence from an empirical study of SL relativization. *Studies in Second Language Acquisition, 13*(3), 431-469.

Flege, J. E. (1995). Second language speech learning: theory, findings, and problems. In W. Strange (Ed.), *Speech perception and linguistic experience* (pp. 233–277). York Press, Timonium, MD.

Hewings, M. (2004). *Pronunciation practice activities*. Cambridge Univ. Press.

Jenkins, J. (2000). *The phonology of English as an international language*. Oxford Univ. Press.

Kang, S. (1999). Relationships between the use of ESL learning strategies and English language proficiency of Asian students. *English Language & Literature Teaching, 5*, 1-25.

Kalikow, D. W., & Swets, J. A. (1972). Experiments with computer-controlled displays in second-language learning. *IEEE Transactions in Audio and Electroacoustics, 20*, 23-28.

Kelly, L. G. (1969). *25 centuries of language teaching. Rowley*, MA: Newbury House.

Kim, H. (2004). The role of negative feedback in second language acquisition with a focus on recasts. *English Language & Literature Teaching, 9*(2), 151-171.

Krashen, S., & Terrell, T. (1983). *The natural approach: Language acquisition in the classroom*. Oxford: Pergamon Press.

Ladefoged, P. (1993). *A course in phonetics* (3rd ed.). New York: Harcourt Brace Jovanovich College Publishers.

Lambacher, K., Martens, W., & Kakeshi, K. (2002). The influence of identification training on identification and production of the American English mid and low vowels. *Language Processing*, *15*(2), 294-316.

Larsen-Freeman, D., & Long, M. (1991). *An introduction to second language acquisition research*. London: Longman.

Leather, J. (1983). Second-language pronunciation learning and teaching. *Language Teaching, 16*, 199-219.

Leon, P., & Martin, P. (1972). Applied linguistics and the teaching of intonation. *The Modern Language Journal, 56*(2), 139-144.

Long, M. H. (1996). The role of linguistic environment in second language acquisition. In W. C. Ritchie, & T. K. Bhatia (Eds.), *Handbook of second language acquisition* (pp. 413-468). San Diego: Academic Press.

Molholt, G. (1988). Computer-assisted instruction in pronunciation for Chinese speakers of American English. *TESOL Quarterly, 22*(1), 91-111.

Morley, J. (1991). The pronunciation component in teaching English to speakers of other languages. *TESOL Quarterly, 25*(3), 491-520.

Morley, J. (1994). *Pronunciation pedagogy and theory: New views, new dimensions*. Alexandria, VA: TESOL.

Munro, M., & Derwing, T. (1995). Processing time, accent, and comprehensibility in the perception of native and foreign-accented speech. *Language and Speech, 38*, 289-306.

Pennington, M. C., & Richards, J. C. (1986). Pronunciation revisited. *TESOL Quarterly, 20*(2), 207-225.

Prator, C. H. (1971). Phonetics vs. phonemics in the ESL classroom: When is allophonic accuracy important? *TESOL Quarterly, 5*(1), 61-72.

Seidlhofer, B. (2003). Controversies in Applied Lingusitics. Oxford uni. Press

Swain, M. (1985). Communicative competence: Some roles of comprehensible input and comprehensible output in its development. In S. M. Gass, & C. G. Madden (Eds.), *Input in second language acquisition* (pp. 235-253). Rowley, MA: Newbury House Publishers.

Swain, M., & Lapkin, S. (1995). Problems in output and the cognitive processes they generate: A step towards second language learning. *Applied Linguistics, 16*(2), 371-391.

Yavas, M. (2006). *Applied English phonology*. Oxford: Blackwell Publishing.

**Examples in: English**
**Applicable language: English**
**Applicable Levels; College**

Sujung Min
Dept of English Education
Kongju National University
182 Shinkwan-dong, Kongju
Chungnamdo, Korea
Tel: (041) 850-8192
Email: sujmin@kongju.ac.kr

Hubert H. Pak
Dept of English Education
Kongju National University
182 Shinkwan-dong, Kongju
Chungnamdo, Korea
Tel: (041) 850-8192
Email: ozasap@yahoo.co.kr

Received in April, 2007
Reviewed in May, 2007
Revised version received in June, 2007