ResearchGate

See discussions, stats, and author profiles for this publication at:
https://www.researchgate.net/publication/223501271

Computer-based training for learning English vowel contrasts

Article *in* System · December 2004

DOI: 10.1016/j.system.2004.09.011

CITATIONS
54

READS
196

2 authors:

Xinchun Wang
California State University, Fr…
18 PUBLICATIONS **101** CITATIONS
SEE PROFILE

Murray J. Munro
Simon Fraser University
82 PUBLICATIONS **4,123** CITATIONS
SEE PROFILE

All content following this page was uploaded by Xinchun Wang on 03 December 2015.

The user has requested enhancement of the downloaded file.

Available online at www.sciencedirect.com
SCIENCE DIRECT®
System 32 (2004) 539–552
SYSTEM
www.elsevier.com/locate/system

Computer-based training for learning English vowel contrasts

Xinchun Wang [a,*], Murray J. Munro [b]

[a] *Department of Linguistics, California State University, Fresno, 5245 North Backer Avenue, M/S PB 92, Fresno, CA 93740, USA*
[b] *Department of Linguistics, Simon Fraser University, 8888 University Drive, Burnaby, British Columbia, Canada V5A 1S6*

Abstract

Computer-based training can be effective in improving second language learners' perceptions and productions of segmental speech contrasts. However, because most previous studies have addressed specific theoretical problems in speech learning, an impact on pedagogy has hardly been felt. Research participants are commonly subjected to rigid research paradigms with strict schedules, so that it is unclear whether research results can be applied in pedagogical contexts. Moreover, a limited range of speech sounds have been examined, with a disproportionate number of studies considering only consonants. In this investigation, we applied training techniques from previous work in a pedagogically oriented approach in which participants had some control over lesson content and worked at a self-determined pace. Sixteen native Mandarin and Cantonese speakers received training on three English vowel contrasts over two months. The training stimuli consisted of synthetic and natural utterances, presented in a graded fashion. In comparison with an untrained control group, trainees showed improved perceptual performance, transferred their knowledge to new contexts, and maintained their improvement three months after training. These findings support the feasibility of computer-based, learner-centered programs for second language pronunciation instruction.
© 2004 Elsevier Ltd. All rights reserved.

Keywords: Pronunciation; Computer assisted language learning; Speech perception

* Corresponding author. Tel.: +1 559 278 2300; fax: +1 559 278 7299.
E-mail address: xinw@csufresno.edu (X. Wang).

0346-251X/$ - see front matter © 2004 Elsevier Ltd. All rights reserved.
doi:10.1016/j.system.2004.09.011

1. Introduction

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. However, if CALL is to reach its full potential in L2 pronunciation instruction, it must make use of the knowledge about L2 speech learning that has been derived from empirical research. At present, there is a significant gap between some of the key research findings of laboratory studies from the past two decades and techniques that have actually been put into practice. In this paper, we illustrate how this disparity may be overcome by applying some of the techniques used in previous laboratory-style training studies. In particular, we show that computer-based perceptual training in English vowels, using synthetic and natural speech, can be effectively implemented with students who are given control over the amount of practice they receive.

2. Review of literature

The trend toward CALL in pronunciation instruction follows many years of laboratory work on L2 phonetic learning. Some studies have evaluated the impact of perceptual training using discrimination and identification tasks in which learners hear stimuli, respond to them, and receive simple "right-wrong" feedback (e.g., Strange and Dittmann, 1984; Jamieson and Morosan, 1986). In discrimination tasks, for example, listeners hear pairs of words in sequence, such as /right/-/light/ or /light/-/light/ and are asked to say whether the two words are the same or different. In identification tasks, they may hear a single word, such as /right/, and select the letter representing the first sound from a choice of two ("r" or "l"). Other studies have used visual feedback (Leather, 1990) or a combination of auditory and visual stimuli (Hardison, 2003) to illustrate linguistic phenomena such as pitch. Still other work has used feedback and evaluation from automatic speech recognition (Dalby and Kewley-Port, 1999; Hincks, 2003) to assess learners' productions. In general, such studies follow well-developed procedures that have yielded important insights into L2 speech phenomena (Logan and Pruitt, 1995). However, most of this work has yet to be implemented by teachers and learners in CALL settings.

Several reasons for the disparity between research and practice can be identified. First, the phonetic focus of training studies has been limited. Many researchers have examined consonant acquisition, most notably of the English /r/-/l/ distinction by Japanese speakers (Bradlow et al., 1997; Lively et al., 1993; Logan et al., 1991; Strange and Dittmann, 1984). While learners' notorious problems with this pair make it an appropriate focus for research, this concern has eclipsed other issues in pronunciation, including other consonantal distinctions, vowels in general, and prosody. More recently, however, some researchers have directed their attention to a wider range of concerns, including tone and intonation (Hardison, 2004; Wang et al., 1999).

A second reason is that the sophisticated software (e.g., speech synthesizers, such as Klatt (1980)) and digitally manipulated stimuli used in laboratory studies create the impression that the development and use of pedagogical tools are beyond the capabilities of those without expertise in software design and digital signal processing. In fact, such studies are usually published in journals that are not commonly accessed by applied linguists and almost never by L2 teachers (e.g., Journal of the Acoustical Society of America, Speech Communication). This is problematic, because the results of research may not actually reach an important audience, namely those who create and use pronunciation software.

A final reason is that laboratory studies are usually theoretically rather than pedagogically motivated (see Logan and Pruitt, 1995 for an extensive review). They have investigated the nature of perceptual representations (Bohn, 1995), the relationship between perception and production (Sheldon and Strange, 1982; Rochet, 1995), and the benefits of different types of tasks and feedback (Logan and Pruitt, 1995). The strict laboratory procedures used, as well as the theoretical constructs tested, make such studies appear irrelevant to pedagogy. For instance, training in the studies cited above has been carried out according to fixed schedules, with all trainees receiving the same amount of training, irrespective of individual differences. In actual teaching settings, of course, individual differences must be accommodated. Moreover, performance in research studies is often evaluated after a few hours or days, a time frame that is pedagogically unrealistic. While these constraints are appropriate for laboratory research, they do not permit exploration of the benefits of a student-centred, longer-term approach.

2.1. Some key research findings

In the development of CALL materials for pronunciation instruction, a number of valuable insights can be derived from previous work. Here we describe three particularly important findings that will be put into practice in the present study.

2.1.1. 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 (Jamieson and Morosan, 1986, 1989; Logan and Pruitt, 1995). 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.

2.1.2. The perceptual fading technique

In everyday situations the speech that we hear varies in clarity. Noting this, Jamieson and Morosan (1986) 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.

2.1.3. 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 (Logan et al., 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 (see e.g., Wang et al., 1999).

3. Research purpose and questions

3.1. Rationale for the study

Here we test the effectiveness of computer-based training on three English vowel contrasts – /i/-/ɪ/, as in *beat* vs. *bit*, /u/-/ʊ/, as in *Luke* vs. *look*, and /ɛ/-/æ/, as in *bet* vs. *bat*. The participants are speakers of Mandarin and Cantonese, who are likely to conflate these pairs because they do not contrast phonemically in either language (Hung, 2000; Rogers, 1997; Wang, 1997). For at least the /i/-/ɪ/ contrast, many learners tend to rely inappropriately on length as a means of distinguishing the two sounds. Bohn (1995), for instance, describes how Mandarin speakers tend to label long high front vowels as /i/ and short ones as /ɪ/ and 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/-/ɪ/ contrast as "long" and "short" categories (Wang and Munro, 1999). Although North American English /i/-/ɪ/ do tend to differ in length, it is the vowel *quality*, determined by articulatory characteristics such as tongue position that serves as the primary basis for the distinction (see Hillenbrand and Clark, 2000). 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.

Very few studies have examined the effect of computer-based training on the acquisition of L2 vowels. Akahane-Yamada et al. (1998) observed improvement after providing such training to 10 Japanese speakers on three English vowels. Wang and Munro (1999) trained native Mandarin speakers to identify synthetic English vowel contrasts (e.g., /i/ vs. /ɪ/), but did not test performance on natural speech. To establish the overall effectiveness of computer-based training for vowels, more work needs to be carried out on a wider range of vowel contrasts and learners.

3.2. 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 ESL 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 Mandarin and Cantonese 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?

4. Participants

The participants were advanced ESL speakers, all degree-seeking students at a major English-speaking Canadian university. They were assigned to one of two groups: 16 (13 Mandarin, three Cantonese) to a trainee group, and five (four Mandarin, one Cantonese) to a control group. They ranged in age from 18 to 42 years ($M = 28$) and had lived in Canada between 6 months and 5 years ($M = 2$). Most of the Mandarin speakers had been born and raised in Mainland China, but two were from Taiwan. The Cantonese speakers had been born and raised in Hong Kong. All participants had normal hearing as determined through a pure tone hearing test (250–4000 Hz at 20 dB).

5. Design and materials

5.1. Overview of the study

Table 1 summarizes the design of the study. Both the trainees and controls completed a pretest, a posttest, and a generalization test. The trainees also received perceptual training and a retention test 3 months later. The identical pretest and posttest allowed us to observe any improvement in performance in the trainees, the generalization test evaluated the participants' performance on new, untrained test items, and the retention test determined the persistence of training effects.

Table 1
Design of the study

Phase of study	Tasks	Participants
Pretest	Identification tasks (no feedback)	Trainees, controls Synthetic and natural speech
Training (2 months)	Identification tasks (with feedback) 1. Synthetic Speech: low variability 2. Natural Speech: 4 speakers 3. Synthetic Speech: increased variability	Trainees
Posttest	Same as pretest	Trainees, controls
Generalization test	Same format as pretest, but with new speakers and new stimuli	Trainees, controls
Retention test (3 months later)	Same as pretest	Trainees

5.2. Materials

Testing and training materials were recordings of words containing the three vowel contrasts /i/-/ɪ/, /u/-/ʊ/, and /ɛ/-/æ/. We used custom-designed synthetic speech to train the participants to ignore duration differences between vowel pairs and to pay attention to vowel quality instead. We then provided further training with more variable synthetic speech and with natural speech to expand the range of test items and voices.

5.2.1. Synthetic speech stimuli

The synthetic stimuli were created using a Klatt (1980) synthesizer. Tokens of *heed*, *hid*, *who'd*, *hood*, *head*, and *had* were designed with two training goals in mind. One was to draw the trainee's attention away from vowel length and toward vowel quality. The stimuli therefore varied in duration so that the listeners would learn that duration alone was not sufficient to distinguish the categories. For instance, *heed* was synthesized with six different vowel durations, ranging from 125 to 250 ms in equal increments. This progression, from long to short can be heard in Audio File 1, while Audio File 2 contains a parallel set for *hid*. Native English speakers hear the former items as *heed* and the latter as *hid*, regardless of duration, as confirmed by three native listeners. However, ESL learners with non-native perceptual representations for these vowels might pay inappropriate attention to length, such that they would identify short *heed* as *hid* or long *hid* as *heed*. Analogous problems might also occur with the other pairs. Our stimuli, however, allowed them to hear both short and long *heed* followed by feedback indicating that the correct identification in both cases was *heed*. This might help them realize that quality rather than length should be the focus of their attention.

A second goal in the stimulus design was to incorporate stimulus variability using the fading approach described earlier. For this reason, we created synthetic variants of the original stimuli described above, with four different versions of each stimulus varying in pitch. Audio File 3 illustrates this variation for *heed*. We added one fur-

ther kind of variability to the stimuli in the form of vowels that were slightly less acoustically distinct (in terms of acoustic phonetic detail) than the original ones. These were introduced later in the training process. (For full acoustic details, see Wang, 2002). One full set of the six words is presented in Audio File 4.

5.2.2. Natural speech stimuli

We recorded six Canadian English speakers (three male, three female) producing 71 word pairs containing the target vowels. Most were CVC, VC, CCVC, and CVCC monosyllables with various onsets and codas. However, a few disyllabic pairs (e.g. *beaten*/*bitten*, and *kettle*/*cattle*) were also included for /i/-/ɪ/ and /ɛ/-/æ/, but not for /u/-/ʊ/, because no appropriate pairs exist. Instead, some additional common words (e.g., *boot*, *book*) were included. Two recordings of each word were made by each speaker and saved as audio files. A sampling of eight natural tokens is provided in Audio File 5.

6. Procedure

6.1. Pretest

All testing was carried out individually, using custom-designed software on a Macintosh computer. The pretest was an identification task in which the listeners heard words through headphones (e.g,, *heed*, *hid*, etc.) and identified them by pressing labeled buttons (*heed*, *hid*, etc.) on a computer screen. The trainee and control groups were tested first on the synthetic stimuli and then on the natural stimuli. For the former, 12 synthetic items were used for each contrast. For the latter, the participants identified words from one male and one female speaker, leaving the rest of the speakers' recordings for the training and the generalization test. For natural /i/-/ɪ/ and /ɛ/-/æ/, 3 minimal pairs were included, with 24 tokens per pair (6 words × 2 speakers × 2 productions). For the /u/-/ʊ/ distinction, one minimal pair was included because of the small number available. The natural tokens for /u/-/ʊ/, therefore, were 16 (2 words × 2 speakers × 2 productions × 2 repetitions). Scoring on the pretest and all subsequent tests was in terms of %-correct identifications.

6.2. Perceptual training

The training tasks were identical to those in the pretest, except that immediate feedback was provided. The user interface was very simple. After the trainee responded, feedback in the form of the words "Correct" or "Sorry, that should be___." flashed on the screen. The next item was then played.

A self-paced, trainee-centered procedure was adopted to accommodate different rates of progress, and to keep the unpaid participants motivated and committed. They were invited to work on the tasks over two months by visiting the lab 2–3 times per week. Each session lasted 50–60 min and consisted of blocks of 24 items each, which took 5–7 min to complete. Although the first author helped initiate the

sessions and gave suggestions for the pace of training, the trainees took responsibility for committing to a schedule and for recycling the training blocks as desired. The number of blocks the participants chose to complete varied ($M = 69$, with a range of 36–129, including repetitions). At the end of each block, the program presented the %-correct score. The trainee could then replay the block or move on.

The training progressed from low variability to high variability. Trainees began with the distinct synthetic tokens and then moved on to natural tokens from four speakers. Stimuli representing only one vowel contrast were presented in each block. However, for the natural speech, multiple talkers were presented during the same block to enhance speaker variability. ''Fading'' blocks that introduced the new pitch patterns and the less distinct synthetic tokens were gradually added late in the training.

6.3. Posttest, test of generalization, and retention test

After two months, the trainees and controls took the posttest, which was identical to the pretest, as well as a generalization test with new tokens produced by the four familiar speakers (heard during the training phase) and tokens by two new speakers. Because of different numbers of stimuli available for each contrast, there were 64 stimuli for /i/-/ɪ/, 34 for /u/-/ʊ/, and 52 for/ɛ/-/æ/. Three months after the posttest, the trainees returned for an identical retention test.

7. Analyses

The only measures used in the study were %-correct identification scores. To answer our first research question, whether computer-based training would improve trainees' perception of L2 vowel contrasts, we compared the participants' scores on the natural speech stimuli in the pretest and posttest. If the trainee group were to improve while the control group did not, we could conclude that the improvement was due to training.

To answer research question 2, whether training would help learners ignore vowel duration differences and attend to vowel quality, we examined the participants' identifications of the long and short synthetic stimuli before and after training. In particular, we determined whether trainees identified synthetic words more correctly in the posttest, regardless of their duration. Such a finding would result in an affirmative answer to the research question.

The results of the test of generalization were used to answer research question 3, whether trainees would generalize new knowledge about vowels to speech and speakers not heard before. If the trainee and control groups did not differ on the pretest, it is likely that higher scores on the test of generalization by the trainee group would be due to the training.

To answer research question 4, whether the learners' new knowledge about vowel categories would be retained after training, we compared the results of the posttest

with those of the retention test. A significant decline in scores would suggest that the trainees had not retained all their newly acquired knowledge.

8. Results

8.1. Pretest–posttest comparison

Fig. 1 shows the mean %-correct identification scores for both tests by vowel contrast. The trainees and controls performed similarly on the pretest, but only the trainees showed higher scores on all three vowel contrasts in the posttest. Overall, in terms of %-correct scores, the trainees showed a 14-point improvement on the /i/-/ɪ/ pair, a 32-point improvement on the /u/-/ʊ/ pair, and a 16-point improvement on the /ɛ/-/æ/ pair. The control group showed a 3-point increase on the /i/-/ɪ/ pair but a 3- to 4-point decrease on the other two pairs.

A mixed-design ANOVA on Groups (trainee and control), with Test (pre and post) and Vowel (three contrasts) as within-groups factors revealed significant effects of Group [$F(1,19) = 9.775$, $p = .0056$], Test [$F(1,19) = 9.462$, $p = .0062$], and Vowel [$F(2,19) = 14.642$, $p < .0001$]. The effect of Vowel was evidently due to better performance on /i/-/ɪ/ than on the other contrasts. However, there was also a significant Group × Test interaction [$F(1,1) = 11.325$, $p = .0032$]. Tests of simple main effects (Howell, 1992) on this interaction indicated a significant difference between groups at posttest [$F(1,35) = 20.335$, $p < .001$] but not at pretest [$F(1,35) = .293$, $p = .592$]. A significant effect of Test occurred in the trainees [$F(1,19) = 14.522$, $p < .001$], but not in the controls [$F(1,19) = .009$, $p = .925$]. In other words, the trainees' perceptual scores on all three contrasts improved significantly in the posttest, while the controls did not improve on any contrasts. The Vowel × Test [$F(2,1) = .466$, $p = .6307$], Group × Vowel [$F(2,1) = .2659$, $p = .0831$], and Group × Vowel × Test

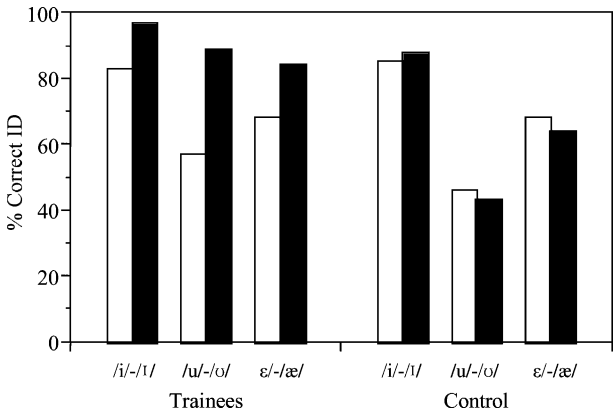

Fig. 1. Mean scores for the trainees and controls on the natural speech in the pretest (light bars) and the posttest (dark bars) for the three contrasts.

[$F(2,38) = .937$, $p = .4007$] interactions were not significant. These results support an affirmative answer to research question 1.

8.2. Vowel duration vs. vowel quality

Fig. 2 illustrates the effect of duration differences on the trainee and control groups' identifications of *heed* and *hid* in the pretest and posttest. As noted earlier, native speakers of English show virtually no effect of duration differences in their perceptions of these stimuli. For instance, *heed* will be correctly identified whether the vowel is long or short. However, in the pretest (dark squares), the trainees tended to identify *heed* less and less correctly as the duration of the stimuli decreased from 250 to 125 ms (as shown on the horizontal axis). In the posttest, however, (light circles) *heed* was correctly identified at high levels of accuracy whether the stimulus was long or short. This result suggests that the trainees did indeed learn to focus their attention away from duration and toward vowel quality. Research question 2 is therefore answered in the affirmative.

8.3. Test of generalization

The mean %-identification scores on the generalization test for the trainees and controls appear in Fig. 3. Here the controls scored lower on all three contrasts. A mixed-design ANOVA yielded significant effects of Group [$F(1,19) = 44.803$, $p < .0001$] and Vowel [$F(2, 19) = 10.951$, $p = .0002$], and a significant Group × Vowel interaction [$F(2, 38) = 9.523$, $p = .0004$]. Tests of simple main effects (Howell, 1992) revealed significant between-group differences on all vowel contrasts ($p < .05$), with the trainees outperforming the controls on all contrasts in the generalization test. This outcome suggests an affirmative answer to research question 3. As before, both

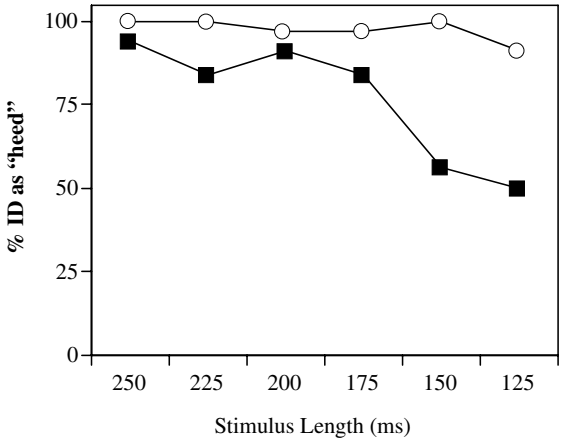

Fig. 2. The effect of stimulus length on identification of synthetic *heed* by the trainees in the pretest (dark squares) and posttest (open circles).

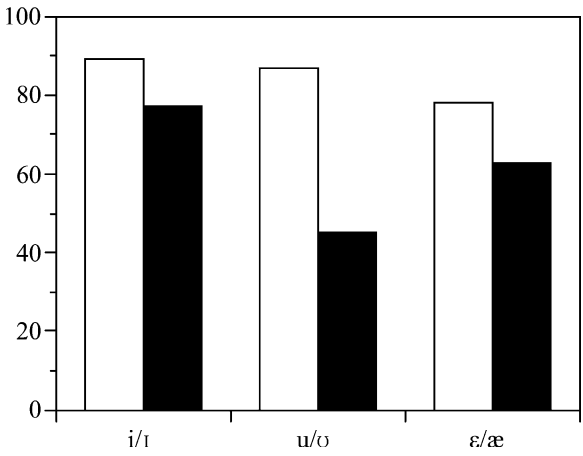

Fig. 3. Mean scores for the trainees (light bars) and controls (dark bars) on the test of generalization for the three contrasts.

groups identified the /i/-/ɪ/ pair more accurately than the other pairs, but the control group appears to have performed somewhat worse on the /u/-/ʊ/ contrast.

8.4. Three-month retention test

The mean %-identification scores on the natural tokens in the retention test were 95%, 83%, and 79% for /i/-/ɪ/, /u/-/ʊ/, and /ɛ/-/æ/ respectively. These had dropped slightly from the posttest scores of 97%, 89%, and 84%, but were still substantially higher than the corresponding pretest scores of 83%, 57%, and 68%.

A mixed-design ANOVA with Test (pre, post, and retention) and Vowel as factors was performed. The effect of Test [$F(2,15) = 22.329$, $p < .0001$], and Vowel [$F(2,15) = 6.705$, $p = .0039$] were significant, but the Vowel × Test interaction [$F(4,30) = 1.937$, $p = .1558$] was not. Post hoc (Tukey) tests revealed that the differences were significant between pretest and posttest, and between pretest and retention test, but not between posttest and retention test. In other words, there was no significant decline in performance on any of the pairs after three months. This evidence supports an affirmative response to our fourth research question. Also, the trainees continued to perform better on /i/-/ɪ/ than on /u/-/ʊ/ and /ɛ/-/æ/.

9. Discussion

The results obtained in this study support positive responses to all four of the research questions posed. A key 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. Overall, correct identifications of

/i/-/I/ tended to be highest, even in the control group, perhaps because this distinction was inherently easier for these learners than the other contrasts.

While our approach to training was based on findings of previous research, it differed from earlier work in that the trainees determined the quantity and schedule of training. This resulted in considerable variability across trainees in the amount of training received. The observed improvement, in a study in which rigid schedules were not enforced, suggests that the techniques used in laboratory training can be successfully applied in settings in which learners participate according to their own preferences. In this case, the unpaid participants willingly scheduled and attended sessions, and even reported that they enjoyed the activity. While we cannot conclude that other learners would feel as motivated as this group, we suspect that their positive response resulted from the novelty of the training and their feelings of success as their performance improved.

This study also provides evidence that identification training with feedback can improve ESL speakers' performance on English vowel contrasts, a research focus that so far has received little attention. With respect to the secondary issues raised in our rationale, we first found that, for the /i/-/I/ contrast, the trainees shifted their attention away from durational differences and focused on the more relevant quality difference between the two members of the pair. Second, the trainees were able to generalize their new knowledge to unique test items and new voices on which they did not actually receive training. Finally, the improvement persisted for at least three months after the training.

One issue that is beyond the scope of this paper is the potential transfer of perceptual training to production. According to Flege's (1995) Speech Learning Model, over time, learners' L2 speech productions come to correspond to perceptual representations. While correct pronunciation of speech categories need not *always* be based on accurate perception (see Sheldon and Strange, 1982), some work indicates that computer-based perceptual training can lead to automatic, immediate improvement in segmental and prosodic production, even when no production training is provided (Bradlow et al., 1997; Rochet, 1995; Wang et al., 2003). At the very least, such work indicates that the use of perceptual training has an important place in the teaching of L2 pronunciation, whether alone or combined with production-specific training.

9.1. Implications for CALL

Vowels are difficult to teach because their articulatory properties cannot always be clearly described and because vowel articulation is not easy to observe without special instrumentation. This is especially true of English "tense-lax" pairs, such as /i/-/I/. Consequently, vowels may be excellent candidates for perception-based training in CALL. Moreover, two of the pairs successfully trained here (/i/-/I/ and /ɛ/-/æ/) have a relatively high functional load and are therefore likely to be important in speech intelligibility. With the growing awareness of the importance of enhanced intelligibility as opposed to "accent reduction" (Derwing et al., 1998), these vowel contrasts would likely rank higher in importance than many other segmental distinctions in a carefully designed pronunciation curriculum.

Here we make no claim to having developed a software training package that is suitable for use by ESL teachers. Rather, we have attempted to show that the design of such a package is a feasible goal and that it has the potential to be effective in CALL. The completion of such a project requires collaboration between pedagogical specialists and those with the technical expertise to develop appropriate speech stimuli and an advanced user interface. We know of no pronunciation software – commercial or other – that makes systematic use of synthetic speech, fading, or stimulus variability to teach segmental contrasts. As, there is now evidence that these techniques can benefit learners, we believe that the development of such materials would be an important contribution to pronunciation pedagogy.

Acknowledgements

We thank Cliff Burgess and the participants in this study for their invaluable contributions. We also appreciate many helpful comments from Greta Gorsuch and Bryan Smith. Portions of this work were supported by a grant to the second author from the Social Sciences and Humanities Research Council of Canada.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.system.2004.09.011.

References

Akahane-Yamada, R., Strange, W., Downs-Pruitt, J., Masuda, Y., 1998. Modification of L2 vowel production by perception training as evaluated by acoustic analysis and native speakers. Journal of the Acoustical Society of America 103, 3089.

Bohn, O.-S., 1995. Cross-language speech perception in adults: first language transfer doesn't tell it all. In: Strange, W. (Ed.), Speech Perception and Linguistic Experience. York Press, Timonium, MD, pp. 279–304.

Bradlow, A.R., Pisoni, D.B., Akahana-Yamada, R., Tohkura, Y., 1997. Training Japanese listeners to identify English /r/ and /l/: IV. Some effects of perceptual learning on speech production. Journal of the Acoustical Society of America 101, 2299–2310.

Dalby, J., Kewley-Port, D., 1999. Explicit pronunciation training using automatic speech recognition technology. CALICO Journal 16, 425–445.

Derwing, T.M., Munro, M.J., Wiebe, G., 1998. Evidence in favour of a broad framework for pronunciation instruction. Language Learning 48, 393–410.

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

Hardison, D., 2003. Acquisition of second-language speech: effects of visual cues, context, and talker variability. Applied Psycholinguistics 24, 495–522.

Hardison, D., 2004. Generalization of computer-assisted prosody training: quantitative and qualitative findings. Language Learning and Technology 8, 34–52.

Hillenbrand, J.M., Clark, M.J., 2000. Some effects of duration on vowel recognition. Journal of the Acoustical Society of America 108, 3013–3022.

Hincks, R., 2003. Speech technologies for pronunciation feedback and evaluation. ReCALL 15, 3–20.
Howell, D.C., 1992. Statistical Methods for Psychology, third ed. PWS-Kent, Boston.
Hung, T.N., 2000. Towards a Phonology of Hong Kong English. World Englishes 19, 337–356.
Jamieson, D., Morosan, D., 1986. Training non-native speech contrasts in adults: acquisition of the English /θ/ and /ð/ contrast by francophones. Perception and Psychophysics 40, 205–215.
Jamieson, D., Morosan, D., 1989. Training new, nonnative speech contrasts: a comparison of the prototype and perceptual fading techniques. Canadian Journal of Psychology 43, 88–96.
Klatt, D., 1980. Software for a cascade/parallel formant synthesizer. Journal of the Acoustical Society of America 67, 971–995.
Leather, J., 1990. Perceptual and productive learning of Chinese lexical tone by Dutch and English speakers. In: Leather, J., James, A. (Eds.), New Sounds 90: Proceedings of the Amsterdam Symposium on the Acquisition of Second Language Speech. University of Amsterdam, Amsterdam, pp. 72–89.
Lively, S.E., Logan, J.S., Pisoni, D.B., 1993. Training Japanese listeners to identify English /r/ and /l/: II. The role of phonetic environment and talker variability in learning new perceptual categories. Journal of the Acoustical Society of America 94, 1242–1255.
Logan, J.S., Lively, S.E., Pisoni, D.B., 1991. Training Japanese listeners to identify English /r/ and /l/: a first report. Journal of the Acoustical Society of America 89, 874–886.
Logan, J.S., Pruitt, J.S., 1995. Methodological issues in training listeners to perceive non-native phonemes. In: Strange, W. (Ed.), Speech Perception and Linguistic Experience. York Press, Timonium, MD, pp. 351–377.
Rochet, B., 1995. Perception and production of second-language speech sounds by adults. In: Strange, W. (Ed.), Speech Perception and Linguistic Experience. York Press, Timonium, MD, pp. 379–410.
Rogers, C.L., 1997. Intelligibility of Chinese-accented English. Unpublished Ph.D. dissertation, Indiana University.
Sheldon, A., Strange, W., 1982. The acquisition of /r/ and /l/ by Japanese learners of English: evidence that speech production can precede perception. Applied Psycholinguistics 3, 243–261.
Strange, W., Dittmann, S., 1984. Effects of discrimination training on the perception of /r-l/ by Japanese adults learning English. Perception and Psychophysics 36, 131–145.
Wang, X., 1997. The acquisition of English vowels by Mandarin ESL learners: a study of production and perception. Unpublished MA Thesis, Department of Linguistics, Simon Fraser University.
Wang, X., 2002. Training Mandarin and Cantonese speakers to identify English vowel contrasts: long-term retention and effects on production. Unpublished Ph.D. dissertation, Department of Linguistics, Simon Fraser University.
Wang, X., Munro, M.J., 1999. The perception of English tense-lax vowel pairs by native Mandarin speakers: the effect of training on attention to temporal and spectral cues. In: Proceedings of the 14th International Congress of Phonetic Sciences, vol. 3, pp. 125–128.
Wang, Y., Jongman, A., Sereno, J., 2003. Acoustic and perceptual evaluation of Mandarin tone productions before and after training. Journal of the Acoustical Society of America 113, 1033–1043.
Wang, Y., Spence, M., Jongman, A., Sereno, J., 1999. Training American listeners to perceive Mandarin tones. Journal of the Acoustical Society of America 106, 3649–3658.

View publication stats