# Training novel phonemic contrasts: a comparison of identification and oddity discrimination training

Conference or Workshop Item

# Training Novel Phonemic Contrasts: A Comparison of Identification and Oddity Discrimination Training

*Zöe Handley[a], MikeSharples[a], Dave Moore[b]*

[a]Learning Sciences Research Institute, University of Nottingham, UK
[b]Medical Research Council (MRC) Institute of Hearing Research (IHR), Nottingham, UK

zoe.handley@nottingham.ac.uk, mike.sharples@nottingham.ac.uk, dave@ihr.mrc.ac.uk

## Abstract

High Variability Pronunciation Training (HVPT) is a highly successful alternative to ASR-based pronunciation training. It has been demonstrated that HVPT is effective in teaching the perception of non-native phonemic contrasts, and that this skill generalizes to the perception of unfamiliar words and talkers, transfers to pronunciation, and is retained long-term. HVPT is, however, not efficient and hence not motivating for the learner. In this study, we therefore compare HVPT with an alternative, namely oddity discrimination training. This comparison, in which Mandarin-Chinese speakers were trained to pronounce the English /r/-/l/ phonemic contrast, provides preliminary evidence to support the use of discrimination tasks in addition to identification tasks to add variety to HVPT.

## 1. Introduction

High Variability Pronunciation Training (HVPT), a technique in which learners are presented minimal pairs containing non-native phonemes in a forced-choice identification (ID) task with immediate feedback (high variability comes from the use of stimuli which vary in terms of phonetic context and talker), is an attractive alternative to ASR-based pronunciation training for two main reasons. First, HVPT has been shown to improve pronunciation without recourse to repetition or articulatory training which can be stressful for the learner [1, 2]. Second, unlike ASR-based training, HVPT provides reliable feedback to the learner.

Moreover, HVPT has made significant progress towards achieving the goals of Computer-Assisted Pronunciation Training (CAPT) (see [3]). HVPT has been demonstrated to be effective in teaching the perception and pronunciation of non-native consonants and vowels [2]. Moreover, it has been demonstrated that this skill generalizes to the perception and pronunciation of unfamiliar words and talkers, and is retained long-term [1].

HVPT is, however, not very efficient. Training typically consists of fifteen one-hour sessions over three weeks. And, only one phonetic contrast is trained! The training is therefore boring and not very motivating for the learner. Alternatives to and variations on HVPT therefore merit consideration.

*Discrimination training* is an alternative to ID training, which focuses on the differences between speech sounds. It has the potential to enhance the learners' experience of phonetic training by increasing the variety of tasks proposed to the learner. Discrimination training has, however, been neglected on the basis that it promotes fine-grained discrimination of speech sounds, which is inconsistent with what we know about speech perception; native speakers have difficulty discriminating between sounds which belong to the same phonemic class [4]. We, however, believe that discrimination training merits further investigation. Firstly, we believe that there is a flaw in Strange and Dittman's study. Strange and Dittman's training was based on the discrimination of pairs of adjacent stimuli on a synthetic /r/-/l/ continuum. Yet, as previous studies had demonstrated, native speakers perceive stimuli on the /r/-/l/ continuum categorically [5]. That is, they are able to discriminate between stimuli at either ends of the continuum which belong to different phonemic categories, but not able to discriminate between adjacent stimuli which belong to the same phonemic category. Secondly, there are variations of discrimination training, which promote classification in addition to discrimination [3]. An example is oddity training. In oddity training, learners are presented three stimuli, two from one class of speech sounds and one from the other and asked to identify which stimulus was the "odd one out", i.e. different.

The main purpose of this study is to determine whether oddity training improves learners' perception and pronunciation of non-native phonemic contrasts. A secondary goal of this study is to compare the effectiveness of oddity discrimination training with that of ID training. These questions are explored within the context of training Mandarin-Chinese speakers to perceive and pronounce the English /r/-/l/ phonemic contrast [6], a contrast which they do not have in their first language in word-final positions [7].

## 2. Method

### 2.1. Design

The study had a between subjects design with a pre-test-post-test format. The dependent variables were: (1) ID accuracy, the number of words correctly classified by the learners in the perceptual tests, and (2) pronunciation accuracy, the number of learner pronunciations correctly classified by the evaluators. The independent variable was the mode of stimulus presentation (ID or oddity). Two groups of six Mandarin speaking Chinese learners of English received training. The first group received ID training and the second oddity training.

### 2.2. Participants

Three groups of participants were recruited from among the staff and students at the University of Nottingham: (1) ten native speakers of British English (henceforth *native speakers*) were recruited to provide baseline data; (2) twelve Mandarin speaking Chinese learners of English were recruited to provide learner data; and, (3) ten native speakers of British English (henceforth *evaluators*) were recruited to evaluate the learners' pronunciations. All of the participants reported no history of any speech or hearing disorder and were compensated at a rate of £5 per hour.

The *native speakers* comprised one male and nine females. Their ages ranged from 18 to 33 (average 23.10 years).

The *learners* comprised six males and six females. Their ages ranged from 22 to 32 (average 25.42 years). All learners spoke Pǔtōnghuà (Standard Chinese/Mandarin). Nine spoke a regional dialect in addition to Pǔtōnghuà. Their age of onset of learning English ranged from 10 to 15 (average 12.58); their length of learning English ranged from 6 to 21 years (average 12.50 years); their age of arrival ranged from 17 to 27 (average 22.83 years); and, their length of residence ranged from 1 to 6 years (average 2.54 years).

The *evaluators* comprised four males and six females. Their ages ranged from 24 to 42 (average 28.40 years).

### 2.3. Apparatus

All sessions were run individually in a quiet laboratory on a PC equipped with a headset, namely Sennheiser eH150. The perceptual test was presented using E-Prime and the pronunciation test was presented via custom-made Web pages using Internet Explorer. The learners' pronunciation data was collected using Audacity.

### 2.4. Materials

The materials consisted of a language background questionnaire and a corpus of minimal pairs which contrast /r/ and /l/. The corpus of minimal pairs which was used in both the perceptual and pronunciation tests, and the training was based on [1].[1] It comprised 100 words, ten minimal pairs which contrast /r/-/l/ in each of the following five phonetic contexts: (1) initial singleton (IS; e.g. *rock* vs. *lock*), (2) initial cluster (IC; e.g. *pray* vs. *play*), (3) intervocalic (IV; e.g. *marrow* vs. *mallow*), (4) final cluster (FC; e.g. *cord* vs. *called*), and (5) final singleton (FS; e.g. *war* vs. *wall*). This corpus was recorded by eleven native speakers of Southern British English, five male and six female in a sound-attenuated room using a Marrantz PMD 660 equipped with an ES961 Uniplate microphone. The training set comprised a sub-set of the corpus, namely five minimal pairs which contrast /r/-/l/ per phonetic context, recorded by a sub-set of the talkers, namely four male and four female talkers. The perceptual test set comprised the full corpus pronounced by two of the trained-on talkers, one male and one female, and two new talkers, one male and one female. The pronunciation test set comprised the full corpus pronounced by a new female talker. This set-up allowed us to test for generalization from (1) trained-on words pronounced by trained-on talkers (TWTT) to (2) new words pronounced by trained-on talkers (NWTT), (3) trained-on words pronounced by new talkers (TWNT), and (4) new words pronounced by new talkers (NWNT) in the perceptual test, and, for generalization from (1) trained-on words (TW) to (2) new words (NW) in the pronunciation test.

### 2.5. Procedure

#### *2.5.1. Pre- and post-tests*

The post-test was identical to the pre-test. Both consisted of two parts, a pronunciation and a perceptual test, which were run in the same session. The pronunciation test was run first.

*Pronunciation test*: The test took around 20 minutes to complete. The stimuli were presented twice in random order. During the first cycle through the stimuli, on each trial, one member of a minimal pair was presented orthographically on screen accompanied by an auditory pronunciation model presented over headphones, and the learners' task was to repeat after the model. The pronunciation models were presented by a female speaker. During the second cycle, the auditory pronunciation model was not provided and the learners' task was to read the word on the screen. Each cycle of 100 trials was presented in five blocks of 20 trials. Before beginning the test in earnest, on each cycle the learners were presented 10 practice trials which were not scored.

*Perceptual test:* The test took around 30 minutes to complete. Each of the 50 stimuli were presented four times, once by each of four different speakers, two male and two female. The stimuli were blocked by speaker and presented in blocks of 50 trials. On each trial, one member of a minimal pair was presented auditorily over headphones. The letters "R" and "L" then appeared on the screen. The learners' task was to identify whether the word presented contained an "R" or an "L" by clicking the corresponding letter on the screen using the mouse. Learners had a maximum of 10000ms to respond and there was an inter-trial interval of 2000ms. Before beginning the test in earnest, the learners were presented ten practice trials twice, once by a male speaker and once by a female speaker.

#### *2.5.2. Training*

Both experimental groups received four consecutive days of training. Each day, they received two sessions of training, one presented by a male talker and one presented by a female talker. During each training session, each of the 100 training stimuli were presented three times. In ID training, learners are presented one stimulus per trial, whereas in oddity training, learners are presented three stimuli per trial, two instances of one member of a minimal pair, and one instance of the other member of the minimal pair. In order to balance exposure to the stimuli across experimental conditions, learners who received ID training were presented 300 trials per speaker and learners who received oddity training were presented 100 trials per speaker, two trials per minimal pair with the stimulus containing /r/ being the odd one out on one trial and the stimulus containing /l/ being the odd one out on the other trial.

*ID training:* The procedure for ID training was the same as the procedure for the ID test, except that learners received feedback during ID training. Feedback consisted of a chime for a correct response and a buzz followed by repetition of the stimulus for an incorrect response. In addition, the correct response was highlighted.

*Oddity training:* On each trial, one instance of one member of a minimal pair and two instances of the other member of the minimal pair were presented over headphones in random order with an inter-stimulus interval of 500ms. The numbers "1", "2", and "3" then appeared on the screen. The learners' task was to identify which stimulus was the odd one out, "1", "2," or "3" by clicking the corresponding number on the screen using the mouse. Feedback consisted of a chime for a correct response and a buzz followed by repetition of the stimuli for an incorrect response. In addition, the correct response was highlighted.

[1]Minor modifications were made to account for differences between British and American English.

*2.5.3. Evaluations*

The pronunciation evaluations took four one-hour sessions to complete. During each session, the evaluators evaluated three learners. The procedure was the same as the procedure for the perceptual test, but with learner productions as stimuli and native speakers as participants.

## 3. Results

Due to space considerations, in this paper, we focus exclusively on the effects of the training on ID accuracy, i.e. perception.

### 3.1. Baseline ID scores

The mean score on the perceptual test was calculated across each experimental group, learners and native speakers. These results are compared in *Figure 1*.

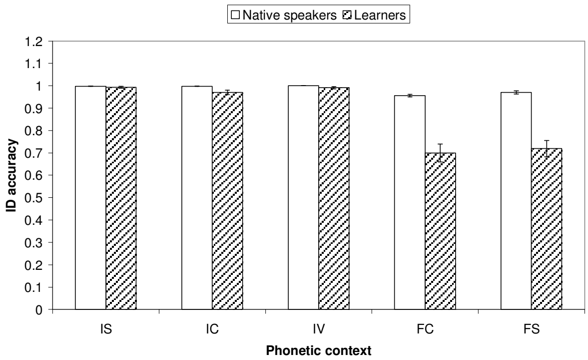


*Figure 1:* Comparison of native and learner ID accuracy scores for the five phonetic contexts: IS (initial singleton), IC (initial cluster), IV (Intervocalic), FC (final cluster) and FS (final singleton). The error bars in this figure and those that follow represent one standard deviation from the mean.

*Figure 1* shows that native speakers score near ceiling across all phonetic contexts. Learners, on the other hand, while they also score near ceiling for IS, IC, and IV, score significantly lower than native speakers for FC and FS. This suggests that there is an effect of phonetic context for learners and an effect of listener for FC and FS.

The learners' perception data were submitted to a one-way ANOVA with phonetic context (IS, IC, IV, FC, or FS) as a within-participants factor. This analysis showed a statistically significant main effect for phonetic context [$F(4, 44) = 54.763$, $p < 0.001$, partial $\eta^2 = 0.833$]. These statistics show that a large percentage of the variation in performance on the perception test for learners is attributable to phonetic context (83%). Regarding differences between learners' and native speakers' performance on the perceptual test, the learners' scores were compared with native speakers' scores for FC and FS separately using an independent t-test. These analyses revealed significant differences between learners' and native speakers' scores for both FC [$t(20) = 5.801$, $p < 0.001$] and FS [$t(20) = 6.086$, $p < 0.001$].

### 3.2. Training

Given the findings presented in section 3.1 regarding Mandarin-Chinese learners' perception and pronunciation of English /r/-/l/, the analysis that follows will focus on their performance on words which contrast /r/ and /l/ in FC and FS positions. In order to determine whether the two experimental groups were matched, the pre-test perceptual data for FC and FS were subjected to separate one-way ANOVA with training as a between-participants factor. The effect of training was not found to be significant.

Mean pre- and post-test scores for TWTT are compared across ID and oddity training in *Figure* , with data for FC presented on the left and data for FS on the right. This figure shows clear improvements from pre- to post-test for FS for both ID and oddity training. For FC, there are also improvements from pre- to post-test, however, to a lesser extent. The data for FC and FS were subjected to separate two-way ANOVA with time (pre- vs. post-test) as a within-participants factor and training (ID vs. oddity) as a between-participants factor. These analyses revealed a significant main effect for time for FS [$F(1, 10) = 6.875$, $p = 0.026$, partial $\eta^2 = 0.407$], but not for FC. The effect of training was not found to be significant for either FC or FS.

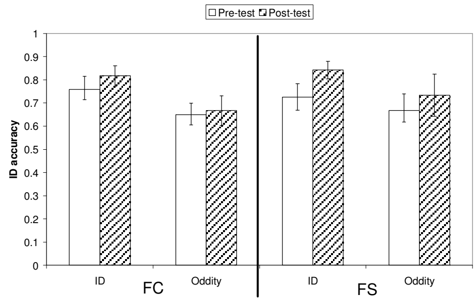


*Figure 2:* Pre- and post-test scores for TWTT compared across ID and oddity training for FC and FS.

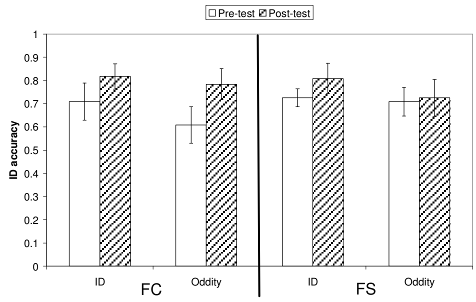


*Figure 3:* Pre- and post-test scores for NWTT compared across ID and oddity training for FC and FS.

Regarding generalization to untrained words, *Figure 3* presents pre- and post-test scores for NWTT. It also shows improvements for both ID and oddity training. However, in contrast with the results for TWTT, this figure shows greater improvement from pre- to post-test for FC than for FS, and that the improvement is marginal for FS for learners who received oddity training. The data were analyzed in the same way as the TWTT data. These analyses revealed a significant main effect for time for FC [$F(1, 10) = 14.523$, $p = 0.003$, partial $\eta^2 = 0.592$], but not FS. The effect of training was not found to be significant for either FC or FS.

*Figure 4* presents the results for generalization to untrained talkers. It shows improvements from pre- to post-test for FC for both ID and oddity training, but no improvements for FS for either ID or oddity training. The data were analyzed as above. Neither the effect of time, nor the effect of training was found to be significant for either FC or FS.

*Figure 5* presents the results for generalization to untrained words and untrained talkers. It shows improvements from pre- to post-test for both FC and FS for both ID and oddity training. The data were again analyzed as above. These analyses revealed a significant main effect for time for FC [$F(1, 10) = 4.076$, $p = 0.071$, partial $\eta2 = 0.290$] and FS [$F(1, 10) = 11.211$, $p = 0.007$, partial $\eta2 = 0.529$]. Again, the effect of training was not found to be significant for either FC or FS.

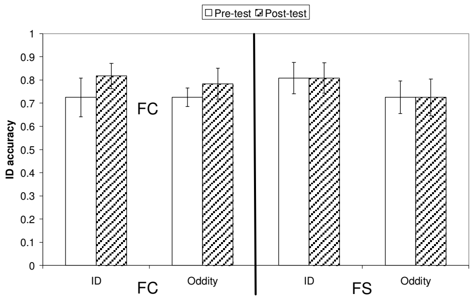


*Figure 4:* Pre- and post-test scores for TWNT compared across ID and oddity training for FC and FS.

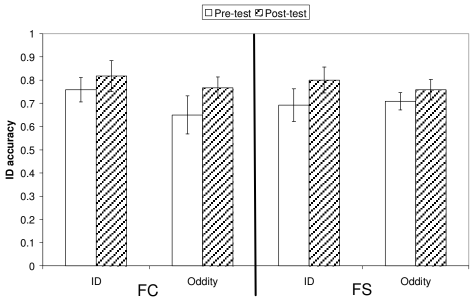


*Figure 5:* Pre- and post-test scores for NWNT compared across ID and oddity training for FC and FS.

## 4. Discussion

In summary, the results provide some evidence to support the use of discrimination training as a method of perceptual training for learners of foreign languages; both discrimination and ID training were found to improve Mandarin-Chinese speakers' perception of the English /r/-/l/ contrast in this study. We should, however, be cautious in making generalizations from this study: 1) improvement from pre- to post-test was not observed for all parts of the test, and 2) due to the absence of a control group which did not receive training, we cannot say whether improvements were simply due to re-administration of the same test. A possible explanation for the absence of improvements from pre- to post-test for both phonetic contexts for all parts of the test might be that the contexts in which the Mandarin-Chinese learners have difficulty perceiving and pronouncing the English /r/-/l/ phonemic contrast were more restricted than we originally thought – a more detailed analysis of Mandarin-Chinese learners' perception and pronunciation of the English /r/-/l/ phonemic contrast has revealed that they only have difficulty with the contrast in a very restricted set of phonetic contexts, namely when they appear in FC and FS positions and are preceded by the vowel /ɔː/ [7]; the differences from pre- to post-test may therefore be very small and due to the limited sample size could not be detected. Regarding re-administration of the same test, given that on some parts of the test improvements were observed for learners who received ID training, but not for those who received oddity training and vice versa, ID and oddity training can be considered to be controls for one another.

The results of this study also suggest that there are no differences between oddity and ID training. One possible explanation, as mentioned in the introduction, is that oddity training, like ID training, promotes classification in addition to discrimination [2]. Another is that exposure is more important than the task; as said, exposure was balanced across the two types of training. We must, however, again exercise caution in interpreting these results for the reasons stated above. Whichever explanation is correct, the implications for CAPT are the same: being as effective as ID training, discrimination can be used in combination with ID training to make HVPT more interesting for the learner.

## 5. Conclusions

In conclusion, the results of this study provide some preliminary evidence to support the use of discrimination tasks in addition to ID tasks in perceptual training. There were, however, limitations to the study. First, the learners' difficulty with the selected phonemic contrast was restricted to a small set of phonetic contexts. Second, there was no control group. Third, the sample was very small. Before drawing any firm conclusions, further research with a larger sample, a control group and a phonemic contrast which presents learners difficulty across a wider range of phonetic contexts is recommended. Regarding exposure, future studies should compare the two types of training with simple exposure.

## 6. Acknowledgements


This study was funded by the Humanities Research Centre (HRC), University of Nottingham.


## 7. References


[1] Bradlow, A, Pisoni, D, Akahane-Yamada, R, and Tohkura, Y (1997). Training Japanese listeners to identify English /r/ and /l/: IV. Some effects of perceptual learning on speech production. *J. Acoust. Soc. Am.* 101(4): 2299-2310

[2] Lambacher, S., Martens, W., Kakehi, K., Marasinghe, C. and Molholt, G. (2005). The effects of identification training on the identification and production of English vowels by native speakers of Japanese. *Applied Psycholinguistics*. 26: 227-247

[3] Logan, J. S., & Pruitt, J. S. (1995). Methodological issues in training listeners to perceive non-native phonemes. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 351-377). Baltimore: York Press.

[4] Strange, W., & Dittmann, S. (1984). Effects of discrimination training on the perception of /r-l/ by Japanese adults learning English. *Perception & Psychophysics*, 36(2), 131-145.

[5] Miyawaki, K., Strange, W., Verbugge, R., Liberman, A., Jenkins, J., and Fujimura, O. (1975). An effect of linguistic experience: The discrimination of [r] and [l] by native speakers of Japanese and English. *Perception & Psychophysics.* 18 (5): 331-340

[6] Chen, S and Fon, J (2007). The effects of phonetic distance, learning context and learner proficiency on L2 perception of English liquids. In Procs. *ICPhS XVI*. Saarbrüken, 6-10 August 2007.

[7] Lin, Y-H (2007). *The sounds of Chinese*. Cambridge: Cambridge University Press.

[8] Handley, Z, Sharples, M, and Moore, D (Submitted) Mandarin-Chinese learners' perception and pronunciation of English /r/ and /l/: The Role of Phonetic Context. *Interspeech 2009.*