Does training on a phonemic contrast absent in the listener's dialect influence word recognition?

Sophie Dufour[a] **and Noël Nguyen**
Laboratoire Parole et Langage, CNRS, Aix-Marseille University, Aix-en-Provence 13604, France
sophie.dufour@univ-provence.fr; noel.nguyen@lpl-aix.fr

Ulrich Hans Frauenfelder
Laboratoire de Psycholinguistique Expérimentale, University of Geneva, Geneva, Switzerland and Fondation Universitaire à Distance, Sierre, Switzerland
ulrich.frauenfelder@unige.ch

Abstract: Southern French listeners were trained on the word final Standard French /e/-/ɛ/ contrast that does not exist in their dialect. They learned to associate minimal pairs of new words with visual shapes. Although final training session performance was relatively high, the learning did not transfer to a lexical decision task with phonological priming. Thus successful training on a phonemic contrast did not guarantee the efficient use of this contrast in spoken word recognition tasks. These findings are discussed in light of abstractionist and exemplarist models.

PACS numbers: 43.71.−k, 43.71.Es [DS]
Date Received: February 24, 2010 **Date Accepted:** April 6, 2010

1. Introduction

Previous research has shown that listeners experience difficulties in discriminating non-native contrasts, particularly when the two phonemes overlap perceptually with a single native phonemic category (see perceptual assimilation model, Best *et al.*, 1988). Perhaps the most typical example is the difficulty that adult Japanese have in identifying the English /r/-/l/ contrast. This phenomenon has been attributed to the fact that Japanese listeners only have a single liquid phoneme /l/ to which both /l/ and /r/ are assimilated (Goto, 1971). This difficulty has proven to be resistant to long-term exposure, as it still occurs in listeners despite years of experience in the foreign language (Pallier *et al.*, 2001; Takagi and Mann, 1995).

Despite the robustness of the difficulty, improvements in non-native contrasts identification have been found in the laboratory thanks to controlled training procedures. For example, Bradlow *et al.* (1997) showed that the forced-choice identification of /r/ and /l/ by Japanese listeners significantly improved after several weeks of intensive training using the productions of five American English speakers. The improved performance level generalized to novel stimuli produced by new speakers, and was maintained 3 months after completion of the perceptual training procedure (Bradlow *et al.*, 1999). Although their /r/-/l/ identification accuracy increased by about 16% (Bradlow *et al.*, 1997), the Japanese trainees still performed significantly below a control group of native listeners.

The studies on the training of non-native contrasts generally examined how this training transferred to new sets of stimuli by using a closely matched testing task. Both training and testing tasks generally involved distinguishing between members of minimal pairs. However, few studies have examined the transfer of training to tasks which reflect word recognition and in particular lexical activation, and in which the participants' attention is shifted away from the target phonemic contrast.

[a]Author to whom correspondence should be addressed.

In the present study, we examined the impact on word recognition of training on a non-native contrast. Unlike studies described above, the particular contrast we studied belonged to a non-native regional variety of the listener's native language. Recent studies have shown that perceptual difficulties arise in the native language for contrasts that do not occur in the listener's dialect (e.g., Conrey *et al.*, 2005; Dufour *et al.*, 2007). In a recent study (Dufour *et al.*, 2007), we examined the word-final /e/-/ɛ/ contrast perception that exists in Standard but not Southern French, which only has the close-mid /e/ vowel in this position. We observed that Southern French listeners treated auditory words like [epe] (Standard French form for *épée*, 'sword') and [epɛ] (Standard French form for *épais*, 'thick') as homophones in a lexical decision task. This finding suggests that the words *épée* and *épais* are associated with a single phonological representation, namely, /epe/ in Southern French, and that at an early stage of phonemic categorization both [e] and [ɛ] are assimilated to the same phoneme /e/.

Here, we tested whether we could improve the ability of Southern French listeners to discriminate between the Standard French /e/ and /ɛ/ phonemes, by means of a training procedure in which listeners learned minimal pairs of new words based on the /e/-/ɛ/ contrast. Two fundamental questions are raised: Can Southern French listeners be trained to learn the /e/-/ɛ/ contrast in word final position? If so, can they use this newly learned contrast in the recognition of words they already know?

The experiment involved three phases: a pre-test, training and post-test phase. During the pre-test, participants performed a primed lexical decision task with both minimal and identical pairs. In Dufour *et al.*, 2007, Southern French listeners exhibited shorter reaction times (RTs) to both Standard French forms /epe/ and /epɛ/ when either the word /epe/ or the word /epɛ/ was presented first. This phase allowed us to replicate the minimal-pair priming effect found in Dufour *et al.*, 2007 and to confirm that the participants tested did not possess the vowel contrast. During the second phase, participants learned minimal pairs of novel words based on the /e/-/ɛ/ contrast by associating these words with visual shapes (see Magnuson *et al.*, 2003 for the same experimental design). During the post-test, the same procedure as the pre-test was used to assess changes in the perception of the /e/-/ɛ/ contrast during spoken word recognition. To examine the persistence of the training, post-tests were administered on three occasions: immediately after the training, 1 day after, and 1 week after. We reasoned that listeners would exhibit no or a reduced priming effect for the minimal pairs of known /e/-/ɛ/ words in the post-test relative to the pre-test, if the training led them to differentiate the /e/ and /ɛ/ phonemic categories during word recognition. This is because there should no longer be an exact match between how the word's final vowel is categorized and the representation associated with the target word's form in the listener's memory. Contrary to earlier training studies, the task used to assess change in target /e/-/ɛ/ contrast perception (pre- and post-test) was different from that used in the training procedure. Using a different task allowed us to assess the generalization of the training to another task and guarantee that any improvement in /e/-/ɛ/ contrast perception between the pre- and post-test was not due to mere habituation to the task.

2. Method

2.1 Participants

Twenty-four Southern French native listeners randomly divided into four groups of six participated in the experiment. They came from the University of Provence and reported no hearing or speech disorders.

2.2 Materials

The stimuli used in the pre- and post-test phases were taken from Dufour *et al.*, 2007. They included 32 bisyllabic /e/-/ɛ/ minimal word pairs. For the purpose of the lexical decision task, the material also contained 32 bisyllabic /e/-/ɛ/ minimal non-word pairs.

The word and non-word pairs were split into two sets with 16 word and 16 non-word pairs in each set. One set was used in the pre-test and the other in the post-test. The presentation order of each set was counterbalanced across participants so that each minimal pair was heard

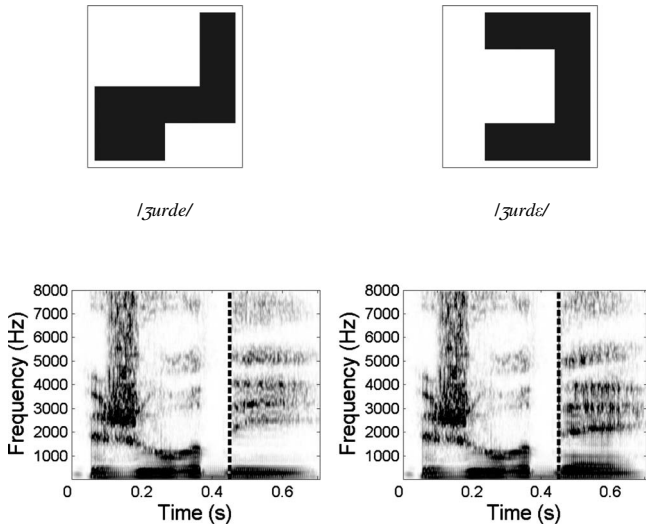

Fig. 1. Examples of shapes with their assigned names and corresponding spectrograms. On each spectrogram, the vertical dotted line indicates the location of the cross-splicing point.

in the pre-test for half of the participants and in the post-test for the other half. The two sets were matched as closely as possible on variables such as lexical frequency and uniqueness point (for words) as well as number of phonemes and overall duration (for both words and nonwords), which are known to affect lexical decision times. In both the pre-test and post-tests, four counterbalanced lists were created from the corresponding set, so that each member of a minimal pair was either repeated or followed by the other member of the minimal pair. Finally, 66 words and 66 non-words were also included as fillers in the experimental lists. The prime preceded the target (i.e., the same item or the other member of the minimal pair) by 8 to 11 items. The items forming a minimal pair appeared in the same positions across the four lists.

For the training phase, 12 bisyllabic novel words (6 /e/-/ɛ/ minimal pairs) were created. Following Magnuson *et al.*'s (2003) procedure, 12 visual patterns were also generated and were randomly associated pairwise with the novel words (see Fig. 1).

2.3 Acoustic stimuli

The six novel minimal pairs were recorded by a female native speaker of Standard French for whom the /e/-/ɛ/ contrast is preserved. By means of cross-splicing, both members of each pair were made acoustically identical up to the onset of the final vowel. For three pairs, the first part of the /e/ word up to the onset of the final vowel was cross-spliced with the final vowel in the corresponding original /ɛ/ word. For the three other pairs, the first part of the /ɛ/ word was cross-spliced with the final vowel in the corresponding original /e/ word. Thus, participants could only rely on the final vowels to distinguish the members of the novel pairs (see Fig. 1).

2.4 Procedure

Pre- and post-tests measured the priming effect for the /e/-/ɛ/ minimal and repetition pairs by having participants make lexical decisions as quickly and accurately as possible, giving the "word" response with their dominant hand. RTs were measured from the onset of the test item. An interval of 2500 ms elapsed between the participant's response and the presentation of the next stimulus. If participants failed to respond within 1800 ms from stimulus onset, no response was recorded and the next item was presented. Each group of participants was presented with one of the four stimulus lists. All participants began the session with six practice trials.

Training consisted of six blocks: five with feed-back on the correct response followed by one without. This final block allowed us to assess participants' performance after training.

Table 1. Percent correct word-to-shape associations during training.

Block	Overall	Novel words with final /e/	Novel words with final /ɛ/
1	38	33	43
2	55	50	60
3	68	63	73
4	75	72	78
5	80	77	83
6 (without feedback)	82	80	84

The structure of the training was as follows. First, a fixation point appeared for 1000 ms on the screen. Next, four shapes were presented on the screen, and then participants heard 1 of the 12 novel words. They were instructed to click on the shape that they thought corresponded to the word. In the first five blocks, after each response, the three distractor shapes disappeared, leaving only the correct referent, and the name of the shape was presented again. In the last block, all four shapes disappeared following the participant's response, and the next trial started.

Each training block consisted of 60 trials. Within each block, each of the 12 novel words appeared as targets five times. Of the three distractor shapes in each trial, one was the shape associated with the word forming the other member of the minimal pair. The other two were selected randomly from the remaining ten shapes, so that each shape appeared the same number of times per block. The shapes were positioned randomly with respect to each other in each trial.

3. Results

Accuracy in the training across the six blocks is shown in Table 1. At the end of the training, participants reached 80% correct responses for the novel final-/e/words and 84% correct responses for those with a final /ɛ/. This performance is relatively high and reflects the ability of Southern French speakers to perceive the /e/-/ɛ/ contrast in word final position in this particular training task.

Mean correct RTs for the lexical decision task are shown in Fig. 2. ANOVAs by participants ($F1$) and items ($F2$) were performed with pair (same, minimal) and presentation (first, second) as variables.

3.1 Pre-test

The main effect of the pair was not significant [Fs <1]. The main effect of presentation was significant [$F1(1,23)=35.21$, $p<0.0001$; $F2(1,31)=44.06$, $p<0.0001$]. RTs were shorter for the second presentation than the first. The pair × presentation interaction was not significant [Fs <1], indicating that the magnitude of the priming effect did not vary as a function of pair type. This finding replicates our earlier observation that Southern French speakers treat the two members of /e/-/ɛ/ minimal pairs as homophones (Dufour *et al.*, 2007).

3.2 Post-test

Only the main effect of presentation was significant [$F1(1,23)=45.67$, $p<0.0001$; $F2(1,31)=37.28$, $p<0.0001$]. RTs were shorter for the second presentation than the first. Crucially, a combined analysis of the pre- and post-test results showed no interaction between the session (pre- vs post-test) and presentation (first vs second) for the minimal pairs [Fs <1]. Thus training had no impact on the magnitude of the facilitation effect observed in /e/-/ɛ/ minimal pairs.

3.3 1 day post-test

As for the first post-test session, only the effect of presentation was significant [$F1(1,23)=20.85$, $p<0.001$; $F2(1,31)=34.97$, $p<0.0001$]. There was no interaction between session (pre- vs 1-day post-test) and presentation (first vs second) for the minimal pairs [Fs <1].

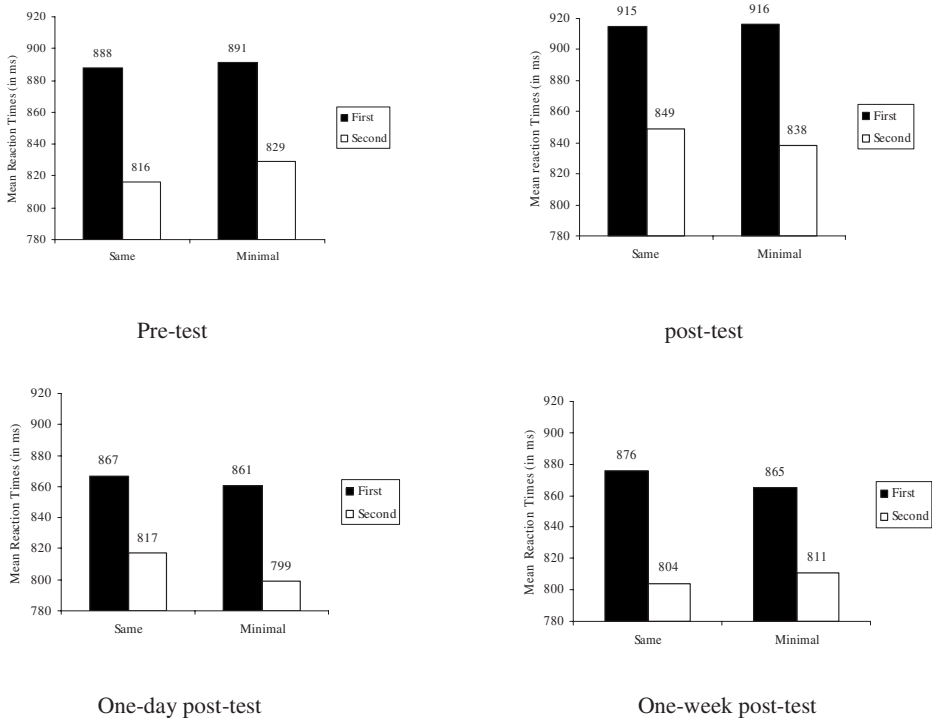

Fig. 2. Mean reaction times (in ms) for the first (black bars) and second (white bars) presentation as a function of pair type and for each session.

3.4 1 week post-test

Again, only the effect of presentation was significant [$F1(1,23)=50.67$, $p<0.0001$; $F2(1,31)=49.72, p<0.0001$]. There was also no interaction between session (pre- vs 1week post-test) and presentation (first vs second) for the minimal pairs [Fs <1].

4. Discussion

This study has shown that Southern French listeners, for whom the /e/-/ɛ/ contrast does not occur in word-final position, are able to learn minimal pairs of new words based on this contrast. This finding reflects these listeners' ability to use the /e/-/ɛ/ contrast when they associate novel words with the corresponding visual shapes. However, our study also showed that participants did not use the knowledge of the contrast they acquired during training in the subsequent recognition of words they already knew. After training, they still treated words such as /epe/ and /epɛ/ as homophones, in a lexical decision task. Hence, despite training on the /e/-/ɛ/ contrast in novel words, Southern French listeners did not use this contrast to contact differentially their already existing lexical representations.

Our findings feed the debate on the general format of the lexical representations. Some theories assume that word forms are stored in the lexicon only as abstract phonological representations (e.g., McClelland and Elman, 1986). According to others, all exemplars of words are stored in memory with their acoustic details (Goldinger, 1998). Our priming results are incompatible with a simple episodic model of word recognition. According to this view, the facilitatory priming effect should have been stronger in the case of a repetition of the same word with a perfect acoustic match between the first and the second presentation than for the minimal pairs. However, even though the acoustic realization of the second presentation differed from that of the first presentation in /e/-/ɛ/ minimal pairs, the minimal-pair priming effect was of the

same magnitude as the identical-pair priming effect. Although there is evidence suggesting that idiosyncratic properties of words are retained in memory and affect speech processing (e.g., Nygaard, 2005 for a review), our results suggest that abstract representations also exist and mediate spoken word recognition.

As in other studies (e.g., Bradlow *et al.*, 1997, 1999), we showed that training to discriminate between minimal pairs is possible. This finding is consistent with both an abstractionist view in which listeners construct new phonemic categories for /e/ and /ɛ/ during training, and an exemplar view in which listeners store memory traces for each novel word. Interestingly, we showed that the training had no impact on the recognition of already known words. In an exemplar view, this may be attributed to the fact that the memory traces associated with the novel words are both too specific and insufficiently established to influence the recognition of words already stored in the lexicon. In an abstractionist view, the absence of impact of the training may be due to the fact that in the training listeners focus their attention on the sublexical phonemic level as opposed to the lexical level in the lexical decision task.

To conclude, successful training on a phonemic contrast does not guarantee efficient use of this contrast in spoken word recognition. Discrimination tasks generally used to assess changes in the perception of non-native contrasts tend to overestimate the listeners' processing abilities with these contrasts. As pointed out by Dupoux *et al.* (2008), it is crucial to test these abilities through a wide assortment of tasks ranging from phonemic discrimination tasks to tasks like lexical decision known to measure lexical activation.

Acknowledgments

This research was supported by Grant No. ANR-08-BLAN-0276-01 from the Agence Nationale de la Recherche (France) and by the Marie Curie Research Training Network, Sound to Sense (S2S).

References and links

Best, C. T., McRoberts, G. W., and Sithole, N. M. (**1988**). "Examination of perceptual reorganization for non-native speech contrasts: Zulu click discrimination by English-speaking adults and infants," J. Exp. Psychol. Hum. Percept. Perform. **14**, 345–360.

Bradlow, A. R., Akahane-Yamada, R., Pisoni, D. B., and Tohkura, Y. (**1999**). "Training Japanese listeners to identify English /r/ and /l/: Long-term retention of learning in speech perception and production," Percept. Psychophys. **61**, 977–985.

Bradlow, A. R., Pisoni, D. B., 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**, 2299–2310.

Conrey, B., Potts, G., and Niedzielski, N. (**2005**). "Effects of dialect on merger perception: ERP and behavioral correlates," Brain Lang **95**, 435–449.

Dufour, S., Nguyen, N., and Frauenfelder, U. H. (**2007**). "The perception of phonemic contrasts in a non-native dialect," J. Acoust. Soc. Am. **121**, EL131–EL136.

Dupoux, E., Sebastian-Galles, N., Navarrete, E., and Peperkamp, S. (**2008**). "Persistent stress 'deafness': The case of French learners of Spanish," Cognition **106**, 682–706.

Goldinger, S. D. (**1998**). "Echoes of echoes? An episodic theory of lexical access," Psychol. Rev. **105**, 251–279.

Goto, H. (**1971**). "Auditory perception by normal Japanese adults of the sounds 'r' and 'l'," Neuropsychologia **9**, 317–323.

Magnuson, J. S., Tanenhaus, M. K., Aslin, R. N., and Dahan, D. (**2003**). "The time course of spoken word recognition and learning: Studies with artificial lexicons," J. Exp. Psychol. Gen. **132**, 202–227.

McClelland, J. L., and Elman, J. L. (**1986**). "The TRACE model of speech perception," Cognit. Psychol. **18**, 1–86.

Nygaard, L. C. (**2005**). "Linguistic and paralinguistic factors in speech perception," in *Handbook of Speech Perception*, edited by D. B. Pisoni and R. E. Remez (Blackwell, Oxford, England).

Pallier, C., Colomé, A., and Sebastián-Gallés, N. (**2001**). "The influence of native-language phonology on lexical access: Exemplar-based versus abstract lexical entries," Psychol. Sci. **12**, 445–449.

Takagi, N., and Mann, V. (**1995**). "The limits of extended naturalistic exposure on the perceptual mastery of English /r/ and /l/ by adult Japanese learners of English," Appl. Psycholinguist. **16**, 379–405.