LASTING STRESS 'DEAFNESS' AFTER AUDITORY TRAINING: FRENCH LISTENERS REVISITED

Sharon Peperkamp and Jason Brazeal

Laboratoire de Sciences Cognitives et Psycholinguistique (CNRS, ENS-PSL, EHESS)
sharon.peperkamp@ens.psl.eu, mail@jasonbrazeal.com

ABSTRACT

Using a sequence recall task, we trained native speakers of French–a language without contrastive stress–to perceive a stress contrast. Contrary to a previous training study with French listeners, but in line with the persistent stress 'deafness' effect observed in advanced French learners of Spanish–a language with contrastive stress–we found no effect of training. We discuss these results in terms of differences in processing mode tapped by different perception tasks.

Keywords: speech perception, stress 'deafness', auditory training, French.

1. INTRODUCTION

A large body of research has shown that the perception of even the most difficult non-native sound contrasts can be improved by means of auditory training. Much of this successful training uses the high-variability phonetic training paradigm [1], focusing on a variety of segmental [1-7] or tonal [8-11] contrasts. Here, we examine another dimension, i.e. word stress.

Native speakers of French, a language without contrastive stress, have difficulty perceiving stress contrasts [12,13]. Previous research found that eight 30-minute training sessions improve stress perception in French listeners as assessed in an odd-one out task, but only in the absence of a large amount of phonetic variability [15]. (The results were the same for implicit training with a shape-word matching task and explicit training with several different tasks.) This moderate result may seem surprising, given that stress is signalled by substantial acoustic cues, i.e. duration, pitch, and intensity. Yet, research with L2 speakers has shown that the so-called stress 'deafness' effect is extremely robust: French speakers who are advanced learners of Spanish, a language with contrastive stress, have as much difficulty with perceiving stress as French monolinguals [14]. Thus, from this perspective, what is surprising is not so much that the training was only successful with little phonetic variability but that it was successful at all.

It is of course possible that focused training with explicit feedback is more effective than years of L2 learning. In the present study we examine this issue, and train French listeners to perceive stress by means of the same task as used with the advanced bilinguals in [14], i.e. sequence recall [13]. In this task, participants must recall sequences of two auditorily presented non-words that differ either in the position of stress (test condition) or in a phoneme (control condition), with the various tokens of the non-words being phonetically different. The combination of a memory load and phonetic variability ensures that this task taps a phonological processing level. The control condition concerns a native segmental contrast and provides an individual baseline with regard to phonological short-term memory.

Here, we use the sequence recall task in a pretest-posttest design, with six intervening training sessions.

2. METHOD

2.1. Materials

We used eight minimal pairs of CVCV non-words, one involving a phoneme contrast, i.e. /fíku/-/fítu/ (for the pre- and posttest), and seven involving a stress contrast (one for a pre- and posttest, and six for the training sessions), for instance /páku/-/pakú/. Each pair was instantiated by eight tokens, four per non-word item, recorded either by a single speaker or by two speakers (see Table 1); in the latter case, each speaker contributed two tokens per non-word item. Overall, there were three different speakers.

A resynthesis algorithm in the audio editor Adobe Audition was used to introduce global pitch variation into all stimuli except those used in three of the training sessions (see Table 1). For items that were recorded by a single speaker, the percentages of pitch modification were 94, 98, 102, and 106. For items that were recorded by two speakers, the percentages were 97 and 103, each applied to one token of each speaker.

The mean duration of the stimuli with the phoneme contrast was 454ms (/fiku/: 450; /fitu/: 459). The mean durations of the different stimuli sets with the stress contrast are shown in Table 2. The same table shows that in four sets, stressed vowels are longer, louder, and have a higher F0 peak than unstressed vowels. In the remaining three sets, one or two of these measures differ between stressed and unstressed vowels.

Table 1. Materials.

Session	Contrast(s)	ISI	Pitch change	Speaker(s)
Pre- and posttest	/fíku/ - /fítu/ /númi/ - /numí/	80ms	yes	French male
Training				
Session 1	/mípa/ - /mipá/	240ms	no	Dutch female
Session 2	/páku/ - /pakú/	240ms	yes	Dutch female
Session 3	/mítu/ - /mitú/	160ms	no	English male
Session 4	/támi/ - /tamí/	160ms	yes	English male
Session 5	/núta/ - /nutá/	80ms	no	Dutch female + English male
Session 6	/kánu/ - /kanú/	80ms	yes	Dutch female + English male

Table 2. Mean durations of stress-initial and stress-final stimuli, and mean differences between stressed and unstressed vowels in terms of duration, F0, and energy.

Session	Total duration		Difference between stressed and unstressed vowels		
	Stress-initial	Stress-final	Duration (ms)	Max F0 (Hz)	Energy (dB)
Pre- and posttest	452	406	38.9 **	72.8 ***	5.1 ***
Training					
Session 1	504	473	17.6 **	51.7 ***	4.3 ***
Session 2	445	427	10.5	54.8 ***	5.1 ***
Session 3	451	523	48.3 **	25.9 ***	4.6 ***
Session 4	455	508	60.6 ***	9.9 **	5.7 ***
Session 5					
Speaker 1	503	435	-5.4	65.9 ***	5.4 ***
Speaker 2	511	505	14.3 ~	13.4 ***	4.8 ***
Mean	*507*	*505*	*4.5*	*39.7*	*5.1* ***
Session 6					
Speaker 1	416	394	10.2	61.8 **	5.7 ***
Speaker 2	411	426	39.8 **	15.4 *	4.9 ***
Mean	*414*	*426*	*25.0* *	*38.6*	*5.3* ***

2.2. Procedure

The pre- and posttest were identical. They consisted of two parts, for the phoneme and the stress contrast, respectively. In each part, participants were first asked to press the number keys 1and 2, upon which they heard all the tokens of the first and second item, respectively. Subsequently, they could continue listening to the various tokens of the two items by pressing the associated keys; pressing each one of these keys resulted in the playing of one token of the corresponding item. Next, it was verified that they had learned the distinction between the two items as well as the correct association between the items and the number keys. That is, they heard a token of one of the items and had to press the associated key, 1 or 2. A message on the screen informed participants whether their response was correct or incorrect. After having given seven correct responses in a row, participants turned to the test phase, during which they were presented with 5 blocks of 16 sequences constituted by repetitions of the two items. The sequence length increased at each block, varying from two in the first block (e.g., /pakú páku/) to six in the last block (e.g. /pakú pakú páku páku pakú páku/). Participants had to reproduce each sequence by typing the associated keys in the correct order. The order of the 16 sequences in each block was randomized, and each item was instantiated randomly by one of the four recorded tokens. Each trial consisted of a sequence with an ISI of 80ms, followed by the word 'OK'. Participants could not begin typing their response until they had heard this word, and the short ISI prevented them from mentally translating the non-words into the associated numbers while they listened to the sequence. A 1500ms pause separated each response from the next trial.

The pre- and posttest lasted around 30 minutes. They were separated by between 9 and 17 days (mean controls: 12.0; mean trainees: 11.5; $t<1$). During this period, the trainees but not the controls went through six training sessions, in which they were trained on novel pairs of non-words differing only in the position of stress. The training started the day after the pretest and ended between one and six days before

the posttest (mean: 1.7). Trainees could not complete more than one session per day, and no more than four days elapsed between sessions, except for one person who had a seven-day lapse between two training sessions. Each session lasted around 30 minutes.

At the beginning of the first training session the objective of the study was explained. That is, participants were told that in some languages words can differ solely by their stress pattern; that since this type of contrast does not exist in their language, French listeners find it hard to perceive; and that the goal of the study was to train their ears to perceive it better. Each of the following training sessions began with a reminder that the two items they would be trained on differ only in the position of stress and that stressed syllables are longer and pronounced with a higher and louder voice than unstressed ones.

The procedure for the training sessions differed from that of the pre- and posttest in two aspects. First, there was no phoneme contrast. Second, participants received feedback during the test blocks: If the answer was correct, positive feedback was displayed on the screen. An error message was displayed otherwise, together with a message asking the participant to re-listen to the sequence and re-enter a response; after 2500ms, the same sequence as before was played. Once more, if the answer was correct, positive feedback was provided. If the second answer was still incorrect, the computer screen displayed the error message as well as the correct transcription and an invitation to listen to the sequence once more; after 2500ms, the sequence was played one last time with the correct answer still displayed on the screen.

Differences across the training sessions with regard to the ISI, the presence vs. absence of global pitch variation, and the number of speakers (Table 1) were loosely meant to gradually increase the level of difficulty.

2.3. Participants

Twenty native speakers of French were tested in Paris. Half of them, two men and eight women aged between 20 and 30 (mean: 24), participated in the pre- and posttest only. The other half, three men and seven women aged between 20 and 31 (mean: 23), also participated in the training sessions. None had ever lived abroad, but all had studied one or more foreign languages—including languages with non-predictable stress—at school.[1]

3. RESULTS AND DISCUSSION

3.1. Pre- and posttest

Responses in the test phase that were a 100% correct transcription of the input sequence were coded as correct; all other responses were coded as incorrect. Table 3 shows the mean error rates.

Table 3. Mean error rates in the pre- and posttest. Standard errors are shown in parentheses.

		Pretest	Posttest
Trainees	Phoneme	27.8 (4.2)	22.1 (5.6)
	Stress	55.9 (7.1)	48.6 (7.4)
Controls	Phoneme	25.6 (3.5)	21.5 (3.3)
	Stress	53.1 (5.9)	50.5 (6.0)

Using the *lme4* package [16] in the R environment [17], we analysed these data in a logistic regression model with contrast-coded fixed factors Group (trained vs. control), Test (pre- vs. posttest), Contrast (phoneme vs. stress), and all the interactions, and a random intercept for Participant (with one or more added slopes the model would not converge). Statistical significance was assessed by means of model comparison using a likelihood ratio test. The results showed effects of Contrast (β=0.66, SE=0.03, z=23.2, χ^2=575, p<.0001) and Test (β=-0.13, SE=0.03, |z|=4.4, χ^2=19.8, p<.0001), but no effect of Group (β=-0.001, SE=0.15, |z|<1) and no interactions (Group × Test: β=0.04, SE=0.03, z=1.31, χ^2=1.72, p>.1; Group × Contrast: β=0.004, SE=0.03, z<1; Group × Test × Contrast: β=-0.02, SE=0.03, |z|<1).

Thus, for all participants performance was better on the phoneme than on the stress contrast, and better in the post- than in the pretest. Even though the interpretation relies on a null result (i.e. the absence of a triple interaction), this pattern of results strongly suggests that training did not have an effect. Rather, compared to the pretest, both trainees and controls showed a small, contrast-independent, increase in performance in the posttest, presumably due to having been familiarized with the task.

3.2. Training sessions

Trials with a 100% correct transcription of the input sequence on either the first or the second try were coded as correct, all others as incorrect. Table 4 shows the mean error rates. (When only the first try is taken into account, the error rates are between 11.5 and 19.0 percentage points higher (mean: 15.8).)

Recall that variations in the training sessions were loosely meant to gradually increase the level of difficulty. It is easy to see that this was not the case.

Table 4. Mean error rates in the six training sessions. Standard errors are shown in parentheses.

1	2	3	4	5	6
15.12 (11.3)	39.12 (15.4)	7.50 (8.3)	12.12 (10.3)	33.00 (14.9)	36.40 (15.2)

As to the factors that may influence task difficulty, it is impossible to draw conclusions because of collinearity and the fact that each session contained unique items. Yet, comparing sessions 1, 3, and 5 to sessions 2, 4, and 6, respectively, we note that the added global pitch variation in the latter always yielded higher error rates, as expected given [13]. Manipulation of ISI, however, did not seem to have a linear effect, contrary to what we expected. In particular, while compared to the ISI of 80ms (sessions 5-6), lower error rates were obtained with the ISI of 160ms (sessions 3-4), presumably because it leaves participants the time to mentally translate the non-words online into the associated numbers, there was no additional drop in error rate for the ISI of 240ms (sessions 1-2). Rather, the 240ms ISI yielded higher error rates than the 160ms ISI.

We speculate that the 240ms ISI may be hard because the associated numbers have to be kept in short-term memory for a longer time (see [18,19] for a similar effect in AX discrimination). According to this interpretation, the difference in performance with a 240ms compared to a 160ms ISI should increase as the sequence length—and hence the total trial duration—increases. And indeed, from shortest to longest sequence length the differences in mean error rates are 4.4, 11.7, 16.6, 28.1, and 26.3 percentage points, showing a robust linear increase ($R^2 = 0.89$, $p < 0.04$).

4. GENERAL DISCUSSION

Using the sequence recall task of [13], we failed to improve stress perception in French listeners. This finding meshes well with the persisting stress 'deafness' effect seen in advanced French L2 learners [14], but it contrasts with a previous study reporting successful training with an odd-one out task [15].

What might account for the difference between the two training studies? The amounts of training were comparable, i.e. six 30-minute sessions for a total of 480 trials consisting of 1920 tokens in our study vs. eight 30-minute sessions for a total of 1728 single-token trials in [15]. Furthermore, in studies on segmental and tonal contrasts that provide analyses of the training sessions, the largest improvement is often reported for the earlier sessions [2,3,6,9,20]. Thus, it seems unlikely that making the training last longer would allow us to observe an effect.

Rather, we argue that our trainees failed to improve their stress perception for other reasons. It is well-known that differences in both stimulus complexity and task demands influence non-native sound perception [18,19,21]. For instance, the automatic selective perception model [23] distinguishes a phonetic and a phonological processing mode. The latter is called upon when rapid processing is required, and performance in this mode is worse than in the former. As to stimuli complexity, our test stimuli had a large amount of phonetic variability (four single-speaker tokens per item, with additional global pitch variation), while successful training as assessed in an odd-one out task was obtained only when there was little variability (each item recorded once by two female speakers) [15]. As to task demands, the sequence recall task requires participants to encode the auditory input in their short-term memory buffer in order to recall the sequence. With phonetically varied stimuli this encoding cannot use a low-level acoustic or phonetic representation but must be phonological in nature. As French listeners do not represent stress phonologically, they fail to improve in the sequence recall task. By contrast, the odd-one out task of [15] is less constraining, especially in the version with very limited phonetic variability.

We also note that a short ISI prevents participants from rehearsing the stimuli subvocally. Findings that successful training can generalize from perception to production [22-24] suggests that standard training tasks may trigger subvocal rehearsal. A long ISI of 500ms was used in the odd-one out task of [15], leaving ample time for this strategy; as there were only three items in each trial, this benefit of a long ISI may not have been counteracted by the drawback of the increased memory load that we saw in the sequence recall task.

Overall, then, successful training appears to rely on a low-level response strategy and/or activation of the perception-production loop in pre- and posttest [25,26], both of which are largely ruled out in the sequence recall task. Yet even successful training can fail to withstand a stricter test: in [15], the exact same training did not improve stress perception when the amount of phonetic variability in pre- and posttest was increased. We note that possibly, improved perception of segmental and tonal contrasts after training in the high-variability phonetic training paradigm [1] may similarly be observed only in the relatively unconstraining identification and discrimination tasks that are typically used in this paradigm. Thus, future research could investigate the extent to which the benefits of this widely used training paradigm resist a cognitively more demanding perception task. This is especially relevant when considering the question of the extent to which training aids real life L2 processing, where the cognitive load is particularly high.

To conclude, we have provided evidence that French listeners' inability to accurately represent and process stress at a phonological level resists not only L2 learning but also explicit auditory training.

7. REFERENCES

[1] Logan, J, Lively, S, Pisoni, D. 1991. Training Japanese listeners to identify English /r/and/l/: A first report. *J. Acoust. Soc. Am.* 89, 874–886.

[2] Flege, J. 1995. Two procedures for training a novel second language phonetic contrast. *Appl. Psycholinguist.* 16, 425–442.

[3] Pruitt, J., Jenkins, J., Strange, W. 2006. Training the perception of Hindi dental and retroflex stops by native speakers of American English and Japanese. *J. Acoust. Soc. Am*. 119, 1684-1696.

[4] Shinohara, Y., Iverson, P. 2018. High variability identification and discrimination training for Japanese speakers learning English /r/–/l/. *J. Phonetics*, 66, 242–251.

[5] Kingston, J. 2003. Learning foreign vowels. *Lang. Speech* 46, 295-349.

[6] Nishi, K., Kewley-Port, D. 2007. Training Japanese listeners to perceive American English Vowels. Influence of training sets. *J. Speech Lang. Hear. R*. 50, 1496-1509.

[7] Iverson, P., Evans, B. 2009. Learning English vowels with different first-language vowel systems II: Auditory training for native Spanish and German speakers. *J. Acoust. Soc. Am*. 126, 866-877.

[8] Sadakata, M., McQueen, J. 2014. Individual aptitude in Mandarin lexical tone perception predicts effectiveness of high-variability training. *Front. Psychol.* 5, 1318.

[9] Wang, Y., Spence, M., Jongman, A., Sereno, J. 1999. Training American listeners to perceive Mandarin tones. *J. Acoust. Soc. Am*. 106, 3649–3658.

[10] Wang, Y., Jongman, A., Sereno, J. 2003. Acoustic and perceptual evaluation of Mandarin tone productions before and after perceptual training. *J. Acoust. Soc. Am*. 113, 1033–1043.

[11] Kaan, E., Barkley, C., Bao, M., Wayland, R. 2008. Thai lexical tone perception in native speakers of Thai, English and Mandarin Chinese: An event-related potentials training study. *BMC Neurosci.* 9, 53.

[12] Dupoux, E., Pallier, C., Sebastián-Gallés, N., Mehler, J. (1997) A destressing 'deafness' in French? *J. Mem. Lang.* 36, 406-421.

[13] Dupoux, E., Peperkamp, S., Sebastián-Gallés, N. (2001) A robust method to study stress 'deafness'. *J. Acoust. Soc. Am*., 110, 1606-1618.

[14] Dupoux, E., Sebastián-Gallés, N, Navarrete, E., Peperkamp, S. 2008. Persistent stress 'deafness': the case of French learners of Spanish. *Cognition* 106, 682-706.

[15] Schwab, S., Dellwo, V. 2021. Explicit versus non-explicit prosodic training in the learning of Spanish L2 stress contrasts by French listeners. *J. Second Lang. Stud*. DOI:10.1075/jsls.21017.sch

[16] Bates, D., Maechler, M., Bolker, B., Walker, S. 2015. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software* 671, 1-48.

[17] R Core Team 2021. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

[18] Werker, J., Tees, R. 1984. Phonemic and phonetic factors in adult cross-language speech perception. *J. Acoust. Soc. Am*. 75, 1866–1878.

[19] Werker, J., Logan, J. 1985. Cross-language evidence for three factors in speech perception. *Percept. Psychophysics* 37, 35–44.

[20] Lively, S., Pisoni, D., Yamada, R., Tohkura, Y., Yamada, T. 1994. Training Japanese listeners to identify English R and L. III. Long-term retention of new phonetic categories. *J. Acoust. Soc. Am*. 89, 874–886.

[21] Strange, W. 2011. Automatic selective perception ASP of first and second language speech. A working model. *J. Phonetics* 39, 456-466.

[22] Rochet, B. 1995. Perception and production of second-language speech sounds by adults. In: W. Strange (ed) *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*. York Press, 379-410.

[23] Bradlow, A., Pisoni, D., Akahane-Yamada, R., 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.

[24] Lambacher, S., Martens, W., Kakehi, K., Marasinghe, C., Molholt, G. 2005. The effects of identification training on the identification and production of American English vowels by native speakers of Japanese. *Appl. Psycholinguist*, 262, 227-247.

[25] Baddeley, A., Gathercole, S., Papagno, C. 1998. The phonological loop as a language learning device. *Psych. Review* 105, 158–173.

[26] Jacquemot, C., Scott, S. 2006. What is the relationship between phonological short-term memory and speech processing? *Trends Cogn. Sci*. 10, 480–486.

[1] One additional participant was withdrawn from the experiment, because for the stress contrast in the pretest they had not passed the success criterion of the verification phase after 200 trials. The data from a second additional participant were excluded, because in the posttest they made twice as many complete reversals (for instance, 1211 instead of 2122) than correct responses on the stress contrast, suggesting a potential confusion in the association between the items and the response keys.