

Individual variability in the perceptual learning of L2 speech sounds and its cognitive correlates

Yoon Hyun Kim and Valerie Hazan

Speech, Hearing and Phonetic Sciences, UCL, London, UK

toithaka@gmail.com, v.hazan@ucl.ac.uk

ABSTRACT

This study explored which cognitive processes are related to individual variability in the learning of novel phonemic contrasts in a second language. 25 English participants were trained to perceive a Korean stop voicing contrast which is novel for English speakers. They were also presented with a large battery of tests which investigated different aspects of their perceptual and cognitive abilities, as well as pre- and post-training tests of their ability to discriminate this novel consonant contrast. The battery included: adaptive psychoacoustic tasks to determine frequency limens, a paired-association task looking at the ability to memorise the pairing of two items, a backward digit span task measuring working memory span, a sentence perception in noise task that quantifies the effect of top-down information as well as signal detection ability, a sorting task investigating the attentional filtering of the key acoustic features. The general measures that were the most often correlated with the ability to learn the novel phonetic contrast were measures of attentional switching (i.e. the ability to reallocate attention), the ability to sort stimuli according to a particular dimension, which is also somewhat linked to allocation of attention, frequency acuity and the ability to associate two unrelated events.

Keywords: Perceptual training, L2 perception, individual variability.

1. INTRODUCTION

Many second-language learners have problems in perceiving sound distinctions in the foreign language that do not occur in their native language (e.g., Iverson et al., 2003; Miyawaki et al., 1975). Studies have now shown conclusively that even short periods of intensive phonetic training can improve the perception of novel consonant or vowel contrasts (e.g. Logan et al., 1991; Iverson & Evans, 2009). Perceptual training can transfer to improvements in the pronunciation of the novel sounds (e.g., Hazan et al., 2005) and the effects of perceptual training can be retained over a long time span (e.g. Bradlow et al. 1999). Studies that have provided data for individual learners have noted that both the ability to discriminate novel contrasts before training and the effect of perceptual training itself can vary quite dramatically across learners. For example Bradlow et al. (1999) found that pre-test English /r/-/l/ identification by Japanese speakers ranged from 52% to 86% for /r/ and 56% to 99% for /l/, and that individual gains after training ranged from 6% to 25% for /r/ identification and from -0.57% to +17% for /l/ identification.

Many potential factors could explain this individual variability. First, a number of factors relating to L2 experience have been shown to affect the learning of the phonetic aspects of a second language. These include age of L2 learning (e.g., Flege et al., 1995), duration of L2 exposure (e.g., Jia et al., 2006), degree of ongoing use of L1 (e.g., Flege et al., 1997). However, even when these factors are carefully controlled, individual variability in the effects of training remain. It has been suggested that individual variability in L2 training could be related to speech processing abilities in the L1 or to more general auditory abilities such as frequency and temporal discrimination (Wong and Perrachione, 2007). Finally, it has been suggested that perceptual learning may be related to more general cognitive abilities such as short-term memory, attention or the learning of associations between two unrelated items (e.g., Goldstone, 1998).

To our knowledge, no study of L2 phonetic training has included a broad battery of cognitive, psychoacoustic and phonetic tests that could help explore the correlates of individual variability in training. In our study, the phonetic contrast chosen for training was the Korean contrast between lenis and aspirated stops. Korean has a three-way stop voicing contrast; the contrast between the lenis and aspirated Korean stop

is difficult for English listeners as they are both assimilated to the English voiceless category (Kang & Guion, 2006). For Korean native speakers, F0 at vowel onset and VOT are important cues to this contrast (e.g., Kim, 2007). Our study included a set of training sessions aimed at improving English listeners' ability to hear this novel distinction as well as a battery of pre/post tests. Our aim was to investigate whether the degree of learning and/or ultimate attainment in the perception of this novel contrast could be related to L1 speech processing abilities, auditory acuity or cognitive abilities.

2. METHODOLOGY

2.1. Participants

Participants were 25 British English speakers (18 females, 7 males), all students at UCL in London. They were aged between 18 and 29 years (median: 21) and were screened for normal hearing thresholds. Although some of the participants spoke other languages, none had studied Korean or another language with lenis/aspirated stop contrasts.

2.2. Training task and materials

The computer-based training followed the high-variability phonetic training approach (Logan et al., 1991). Participants were trained to identify the Korean alveolar lenis, /t/ and aspirated, /t^h/ stop. Participants heard a token and had to decide which phoneme they heard by clicking on a button labeled 't' or 'th'. If they made an error, both the sound and the correct label were repeated. Each phoneme was presented in a CV syllable with the vowels, /a/, /i/, or /u/, produced by six speakers (3 females, 3 males). 36 syllables were each repeated four times per training session; there were four training sessions in total.

2.3. Test Battery

2.3.1. *Measuring individual variability in the perceptual learning of L2 speech sounds*

2.3.1.1. *Phoneme identification within syllables*

Pre- and post-training tests were carried out to look at the impact of the training. In these tests, participants did the same task as in the training but received no feedback. The test stimuli were CV syllables with the lenis and aspirated stops combined with the vowels /i/, /a/, or /u/. The speakers recorded for the pre/post tests were different to those in the training sessions. Test materials consisted of 6 CV stimuli recorded by 2 men and 2 women, each presented 5 times (total: 120 trials).

2.3.1.2. *Phoneme identification within word*

Korean words with /t/ or /t^h/ in initial position were used to test the generalization of the training. There were 64 words (2 phonemes * 8 words * 4 speakers), each presented twice (total: 128 trials). The speakers were different from those of the syllable test and training sessions.

2.3.1.3. *Acoustic-cue weighting*

This test examined which acoustic cues English participants attended to before and after training. A natural syllable, /ta/, was manipulated with Praat to synthesize stimuli with a range of VOT (20 to 80 ms in 10 ms steps) and F0 (170 to 320 mels in 10 mel steps) values. Thus there were 112 stimuli (7 * 16). Participants had to judge whether the stimuli were /t/ or /t^h/.

2.3.2. *Measuring individual variability in auditory acuity, L1 speech perception and cognitive tasks*

2.3.2.1. *Frequency discrimination*

An adaptive procedure was used to determine difference limens for single-formant tokens varying in frequency in the F2 range (F2 test) and fundamental frequency range (F0 test). Participants were asked to choose the odd one out among three nonspeech sounds.

2.3.2.2. Attentional filtering of VOT and F0

In this sorting task (Garner, 1974), participants were asked to sort four stimuli into two categories by either F0 or VOT. In correlated-dimension sorting tasks, they classified stimuli into two groups that differed in both dimensions. In orthogonal-dimension sorting tasks, participants sorted stimuli according to one dimension while ignoring differences in the other. The same stimuli, selected from the stimulus set of the acoustic-cue weighting task, were used in the two orthogonal tasks. Their VOT was 20 ms or 80 ms and F0 was 180 mel or 230 mel.

2.3.2.3. Categorical perception of L1 voicing distinction

An adaptive identification test using a synthetic speech continuum was used to test the consistency of labeling (steepness of the identification function) for an English voicing contrast between the syllables 'pea' and 'bee' (see Hazan et al., 2009 for details).

2.3.2.4. L1 Speech perception in noise

This test assessed participants' speech perceptual abilities in their native language. Materials were derived from the speech in noise (SPIN) sentences of Bradlow and Alexander (2007) in which the final word varies in predictability from the context. As in that study, the test design included keywords (15) presented in 'right context (RC)' sentences (e.g., 'the meat from a pig is called pork') and 'neutral context (NC)' sentences (e.g., 'he talked about the pork'); in addition, a further 15 keywords were presented in neutral context and 'wrong context (WC)' (e.g., 'the meat from a pig is called dinner').

2.3.2.5. Attention

A number of subtests of the Test of Everyday Attention (TEA, Robertson et al, 1994) were presented. These included: 'elevator counting with distraction' (a measure of selective attention), 'elevator counting with reversal' (a measure of attentional switching) and 'lottery' (a measure of sustained attention).

2.3.2.6. Working memory

A backward digit span task was used to assess complex working memory (Gathercole, 1999). After remembering numbers in order, participants had to reproduce the numbers in inverse order and press corresponding number keys on a keyboard. Number string length varied from 3 to 10.

2.3.2.7. Associative learning

Paired association learning which link two items is related to processes of memory encoding and retrieval (Buckner & Wheeler, 2001). 20 meaningless syllables (CVC) were paired with a number (1 or 2). The cued-recall test was repeated 4 times. A syllable was given as a cue on the screen, and participants had to say which number was paired with it.

2.4. Test procedure

Each participant was tested individually over 10 sessions, lasting between 30 and 50 minutes, carried out on different days. Testing and training took place in a sound-treated booth, with sounds presented via high-quality headphones at a comfortable listening level fixed across participants. At Session 1, participants carried out the syllable and word identification pre-tests and the attentional filtering task. At Session 2, they performed the acoustic-cue weighting test. At Sessions 3 to 6, they carried out their training sessions. At Sessions 7 and 8, they carried out the post-tests (same as pre-tests). At Session 9, they carried out the paired-association task, the working memory span task and the speech perception in noise task. At Session 10, participants were tested on the attention tasks and the F0/F2 difference limen tests.

3. RESULTS

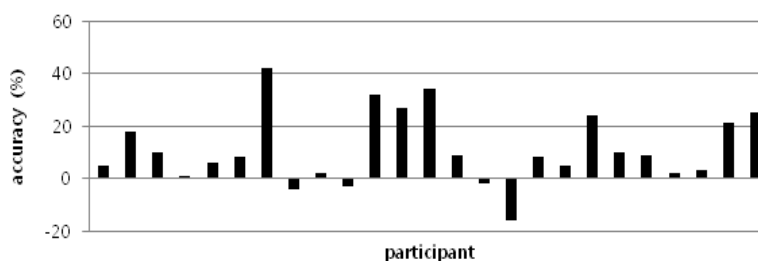
3.1. Pre and post-test results for Korean stop voicing contrast

The mean identification accuracy for the lenis-aspirated Korean voicing contrast in the pretest was just above chance level for both words and syllables. Significant individual variability was present in all measures as can be seen in Table 1. The mean percentage difference between pre- and post test was 11% for syllables and 7.4% for words. The degree of change between pre- and post-test varied across individual learners from -16% to +42% for the syllable test (see Fig. 1) and -7% to +28% for the word test.

Table 1. Accuracy of L2 phoneme identification in training sessions and tests

1.	Mean (%)	s.d.	Minimum	Maximum
syllable identification (pre-test)	55.5	11.9	36	72
syllable identification (post-test)	66.5	9.7	49	85
word identification (pre-test)	51.5	8.2	38	71
word identification (post-test)	58.8	9.0	44	81
training (first session)	66.1	8.6	50	79.9
training (last session)	74.8	11.2	52.1	96.5
training (change across four sessions)	8.7	5.6	-1.3	17.4

Figure 1: Change in syllable identification accuracy between the pre- and post-test for individual participants.



Next, we examined the rate of change between the first and final training sessions (Table 1). The degree of variability was such that, despite minimal previous exposure to this contrast, some participants, after four short training sessions were at ceiling (97%) while others were performing near chance.

Correlations analyses were carried out on the z-scores of pre/post data for the syllable and word tests and training performance. Syllable identification accuracy in the pre-test was only significantly correlated with word accuracy in the pre-test ($r=0.566$, $p=.003$), while post-test syllable accuracy was correlated with scores in the first ($r=.671$) and last ($r=.547$) training sessions. Pre-test scores of word identification were correlated with word identification post-test scores ($r=.443$). The performance at first training was highly correlated with the achievement of the last training session ($r=.873$), and the difference in score between the first and last training sessions was significantly correlated with accuracy in the final training session ($r=.654$).

3.2. Correlations with cognitive, auditory acuity and speech processing tasks

The next phase of the analysis involved looking at which of the additional measures collected correlated with measures related either to initial ability, to the rate of learning of the novel contrast, or to ultimate attainment. Pearson's correlations were applied to the normalised data (see Table 2).

First, let's examine what measures correlated with the initial sensitivity to the novel contrast. Syllable pre-test scores were significantly correlated with frequency acuity in the F2 region. Word-level pre-test scores were correlated with accuracy of sorting by the VOT dimension (attentional filtering test) and with attention switching ability. Performance at the initial and final training sessions was correlated with

performance on the paired association task, with attentional switching ability and with the response time of sorting by the F0 dimension (attentional filtering test). Ultimate attainment (post-test scores) was again correlated with accuracy of sorting by the VOT dimension (both word and syllable tests), with frequency acuity in the F2 region (word test only) and with working memory (syllable test only). There were no significant correlations between performance measures for the novel contrast and measures reflecting speech processing ability in the L1, F0 limen and measures of sustained and selective attention.

In summary, the general measures that were the most often correlated with the ability to learn the novel phonetic contrast were measures of attentional switching (i.e. the ability to reallocate attention), the ability to sort stimuli according to a particular dimension, which is also somewhat linked to allocation of attention, frequency acuity and the ability to associate two unrelated events.

Table 2. Correlation between measures related to the L2 phonetic contrast and measures of speech, auditory or cognitive processing abilities.

	SID_pre	SID_post	WID_pre	WID_post	TR_first	TR_last	TR_diff
F2_limen	0.547**	0.270	0.296	0.397*	0.246	0.143	-0.094
F0_limen	-0.037	0.096	-0.045	0.230	0.204	0.244	0.173
PB_slop	-0.291	0.156	-0.197	-0.084	-0.024	-0.175	-0.316
VOT_acc	0.366	0.412*	0.414*	0.693**	0.273	0.250	0.079
F0_acc	0.196	0.386	-0.081	0.219	0.260	0.235	0.068
VOT_rt	-0.047	0.065	-0.079	-0.183	0.020	0.065	0.099
F0_rt	-0.097	-0.285	0.068	-0.030	-0.487*	-0.437*	-0.122
Cind	0.223	-0.192	0.092	0.108	-0.316	-0.273	-0.058
Cdep	-0.245	-0.311	-0.031	-0.237	-0.265	-0.364	-0.320
PA_last	0.225	0.090	0.108	0.006	0.473*	0.468*	0.206
PA_diff	0.019	0.105	-0.063	0.035	0.352	0.302	0.060
WM	0.157	0.397*	0.005	0.150	0.355	0.246	-0.058
TEA_sel	0.025	-0.074	-0.137	0.150	0.203	0.243	0.173
TEA_swi	0.023	0.272	-0.413*	0.048	0.464*	0.446*	0.176
TEA_sus	-0.018	0.167	0.176	0.141	0.164	0.082	-0.089

1. * p<.05, ** p<.01

2. SID_pre (syllable identification pre-test), SID_post (syllable identification post-test), WID_pre (word identification pre-test), WID_post (word identification post-test), TR_first (training first session), TR_last (training last session), TR_diff (difference between the first and the last training session), PA_last (paired association last session), PA_diff (difference between the first and the second session of paired association task), WM (working memory span), Cind (context-independent rate of SPIN perception), Cdep (context-dependent rate of SPIN perception), F2_limen (F2 discrimination limen), F0_limen (F0 discrimination limen), PB_slop (slope of categorical perception for L1 voicing contrast), VOT_acc (accuracy of sorting by VOT dimension), F0_acc (accuracy of sorting by F0 dimension), VOT_rt (response time of sorting by VOT dimension), F0_rt (response time of sorting by F0 dimension), TEA_sel (selective attention test), TEA_swi (attentional switching test), TEA_sus (sustained attention test)

4. DISCUSSION

The results of our study confirmed our expectations that English listeners with no experience of Korean would vary significantly in their ability to both initially perceive and learn a novel phonetic contrast which is difficult as it involves two Korean phonemes that both assimilate to a single English category. At the pre-test, some English listeners identified the lenis-aspirated contrast with a high degree of accuracy while many were at chance; after four short training sessions, a few listeners were at ceiling while others had not learned anything about the contrast. In order to identify the Korean voicing contrast, it is necessary to use acoustic cue information from VOT and F0 (Francis and Nusbaum, 2002; Kim, 2007; Kang and Guion, 2006).

Indeed, some learning measures were correlated with VOT sorting accuracy or F0 sorting time. Successful acquisition of L2 speech sounds might depend on whether L2 learners can adjust their sensitivity to small acoustic differences through attention. Indeed, we found correlations between accuracy and frequency limens as well as with measures of attentional switching. Finally, the learning of a new contrast is dependent on being able to associate a novel sound with a new category. There again, learners who were successful at a paired-association task tended to be more successful at learning the novel contrast. It should be noted that the correlations between cognitive or frequency acuity measures and the learning of a novel contrast, although significant, are not necessarily very strong. It is therefore clear that no single dimension can predict successful learning, and that the picture is complex given the multiplicity of factors that contribute to the learning of the sounds of a new language. This study though provides a further contribution to the understanding of which factors can contribute to successful language learning at the phonetic level.

5. REFERENCES

- Bradlow, A., Akahane-Yamada, R., Pisoni, D., Tohkura, Y. 1999. Training Japanese listeners to identify English /r/ and /l/: Long-term retention of learning in perception and production. *Perception & Psychophysics*, 61, 977-985.
- Bradlow, A. R. and Alexander, J. A. 2007. Semantic-contextual and acoustic-phonetic enhancements for English sentence-in-noise recognition by native and non-native listeners. *Journal of the Acoustical Society of America*, 121(4), 2339-2349.
- Buckner, R., Wheeler, M. 2001. The cognitive neuroscience of remembering. *Nature reviews. Neuroscience* 2(9), 624-634.
- Flege, J., Frieda, E., Nozawa, T. 1997. Amount of native-language (L1) use affects the pronunciation of an L2. *Journal of Phonetics*, 25(2), 169-186.
- Flege, J., Munro, M., MacKay, I. 1995. Effects of age of second-language learning on the production of English consonants. *Speech Communication*, 16(1), 1-26.
- Flege, J., Takagi, N., Mann, V. 1996. Lexical familiarity and English-language experience affect Japanese adults' perception of /i/ and /l/. *Journal of the Acoustical Society of America*, 99(2), 1161-1173.
- Garner, W. 1974. *The processing of information and structure*: Erlbaum Potomac, MD.
- Gathercole, S. 1999. Cognitive approaches to the development of short-term memory. *Trends in Cognitive Sciences*, 3(11), 410-419.
- Goldstone, R. 1998. Perceptual learning. *Annual Reviews in Psychology*, 49(1), 585-612.
- Hazan, V., Messaoud-Galusi, S., Rosen, S., Nouwens, S., Shakespeare, B. 2009. Speech perception abilities of adults with dyslexia: is there any evidence for a true deficit? *Journal of Speech, Language, and Hearing Research*, 52(6), 1510-1529.
- Hazan, V., Sennema, A., Iba, M., Faulkner, A. 2005. Effect of audiovisual perceptual training on the perception and production of consonants by Japanese learners of English. *Speech Communication*, 47(3), 360-378.
- Iverson, P., Evans, B. (2009). Learning English vowels with different first-language vowel systems II: Auditory training for native Spanish and German speakers. *The Journal of the Acoustical Society of America*, 126, 866-879.
- Iverson, P., Kuhl, P., Akahane-Yamada, R., Diesch, E., Tohkura, Y., Kettermann, A., Siebert, C. 2003. A perceptual interference account of acquisition difficulties for non-native phonemes. *Cognition*, 87(1), 47-57.
- Jia, G., Strange, W., Wu, Y., Collado, J., Guan, Q. 2006. Perception and production of English vowels by Mandarin speakers: Age-related differences vary with amount of L2 exposure. *Journal of the Acoustical Society of America*, 119-1118.
- Kang, K., Guion, S. 2006. Phonological systems in bilinguals: Age of learning effects on the stop consonant systems of Korean-English bilinguals. *Journal of the Acoustical Society of America*, 119, 1672-1683.
- Kim, Y. H. 2007. The Effect of Selective Attention on Discrimination between Lenis /t/ and Aspirated /th/ of Korean Alveolar Stops as shown by the Korean and the Japanese. Unpublished doctoral dissertation, Seoul National University, Seoul, Korea.
- Logan, J., Lively, S., Pisoni, D. 1991. Training Japanese listeners to identify English /r/ and /l/: A first report. *Journal of the acoustical society of America*, 89(2), 874-886.
- Miyawaki, K., Strange, W., Verbrugge, R., Liberman, A., Jenkins, J., Fujimura, O. 1975. An effect of linguistic experience: The discrimination of [r] and [l] by native speakers of Japanese and English. *Perception and Psychophysics*, 18(5), 331-340.
- Robertson, I. H., Ward, T., Ridgeway, V Nimmo-Smith, I. 1994. *The Test of Every day Attention*. BurySt Edmunds: Thames Valley Test Company.
- Schmidt, A. M. 1997. Cross-language consonant identification: English and Korean. In O.-S. Bohn M. J. Munro (Eds.), *Language experience in second language speech learning: In honor of James Emil Flege* (pp. 185-200). Philadelphia: John Benjamins Publishing Company.
- Wong, P., Perrachione, T. 2007. Learning pitch patterns in lexical identification by native English-speaking adults. *Applied Psycholinguistics*, 28(04), 565-585.

6. ACKNOWLEDGEMENTS

This work was supported by the Korea Research Foundation Grant funded by the Korea Government [KRF-2008-356-H00004] and UCL Department of Speech, Hearing and Phonetic Sciences