Can we predict who will benefit from computer-based phonetic training?

Valerie Hazan, Yoon Hyun Kim

Speech, Hearing and Phonetic Sciences, UCL, UK.

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

Abstract

This study investigated whether specific auditory or cognitive skills were linked to initial sensitivity to a novel phonetic contrast (Korean lenis-aspirated contrast) or to the degree of learning following computer-based phonetic training. Correlations between auditory or cognitive skills and phonetic perception were generally fairly weak, with measures of frequency acuity and attentional switching most often correlated with phonetic ability. The ability to sort stimuli according to a particular acoustic cue was also correlated with performance on the syllable and word identification tests. However, rate of learning was not correlated with any of the auditory or cognitive skills tested.

Index Terms: computer-based phonetic training, L2 perception, individual variability.

1. Introduction

The development of computer-based language training software for second-language (L2) learning has been a central topic of interest since the very first workshop on speech technology in language learning (STiLL) in 1998. Computer-based training can be applied at different levels of language learning, from the 'analytic' training of specific phonetic categories that are difficult for the L2 learner because they do not occur in the speaker's native language [e.g., 1] to total immersion programmes that place the learner within a virtual environment in which they converse with a virtual native speaker [e.g., 2]. One common finding of studies on the effectiveness of such training is that individual learners vary drastically in the extent to which they benefit from such language training. In fact, this issue of individual variability in training was discussed at the very first STiLL workshop by Eskenazi, who suggested that individual learners might use quite different learning strategies (e.g. global vs analytic approach) and that it would be beneficial to match training approaches to this learning style. Apart from the issue of learning strategy, it is also possible that successful learning, at least at the phonetic level, might be linked to more general cognitive skills such as short-term memory [3], attention or the ability to associate two unrelated events [4]. Successful learning may also be related to auditory skills such as the ability to discriminate small changes in frequency in a signal [5]. Such skills might act at two levels: first, they may give some learners an increased initial sensitivity to a novel L2 phonetic contrast; second, they may facilitate the learning process itself and ultimate attainment, so that the effectiveness of phonetic training sessions is enhanced. Having a better understanding of which skills might help promote better learning is potentially very useful in that it might be possible to use this information to tailor the training that is presented to the learner.

The study involved training English participants on a novel phonetic contrast (the Korean lenis-aspirated voicing contrast) to which they had not previously been exposed. In addition to the four sessions of training, they were presented with a broad range of tests evaluating their auditory, linguistic and cognitive skills. We then examined whether any of these skills correlated either with initial performance on this novel phonetic contrast, ultimate attainment or rate of learning during training.

2. Method

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 years) and 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

The aim of this task was to see to what extent the participants could learn the novel contrast as a result of a short period of intensive auditory training. The computer-based training followed the high-variability phonetic training approach [1]. Participants were trained to identify the Korean alveolar lenis, /t/ and aspirated, /t^h/ stop, a contrast primarily marked by changes in VOT (with VOT values for both the lenis and aspirated falling within the VOT range for the English voiceless stop) and in initial F0 [e.g., 6]. In this computer-based training task, participants heard a syllable and had to decide whether the initial phoneme they heard was 't' or 'th' by by clicking on the appropriate button. 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 Korean speakers (3 females, 3 males). 36 syllables were each repeated four times per training session; there were four training sessions in total.

2.3. Pre/post training tests

2.3.1. Word and syllable identification tests

Identical tests were presented pre- and post training to look at the impact of the training on the identification of syllables or words containing the lenis-aspirated voicing contrast. In these tests, participants did the same task as in the training but received no feedback. In the syllable ID test, CV syllables with the lenis and aspirated stops were combined with the vowels /i/, /a/, or /u/. The Korean speakers recorded for the pre/post tests were different to those in the training sessions. Test materials consisted of 6 CV stimuli recorded by four new speakers, each presented 5 times (total: 120 trials). In the Word ID test, Korean words with /t/ or / t^h/ in initial position were used to test the generalization of the training. There were 64 words, each presented twice (total: 128 trials).

2.3.2. Analytic cue weighting tests

Two analytic tests were presented to assess which acoustic cues to the novel contrast the participants were sensitive to both pre- and post-training. For the cue-weighting test, 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 (total: 112 stimuli). In the test, participants had to judge whether the stimuli started with /t/ or / t^h/; in the analysis of results, we can see whether identification was affected by the absence of one of the cues. A second analytic task involved 'attentional filtering' [7]. In this task, participants had 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 set described above, were used in the two orthogonal tasks with VOT of either 20 ms or 80 ms and F0 of 180 or 230 mel.

2.4. Native language, auditory and cognitive tasks

2.4.1. Native language tests

Two tests were carried out to evaluate whether the ability to learn a novel phonetic contrast was related to native language skills. An adaptive identification test using a 'pea'-'bee' synthetic speech continuum was used to test labeling consistency for an English voicing contrast (see [8] for details). Our hypothesis was that participants who were less categorical in their labeling might more easily adapt their phonetic categories when learning the Korean voicing contrast. A second test assessed the perception of L1 sentences in noise. Materials were derived from English speech in noise (SPIN) sentences [9] in which the final word varies in predictability from the context. The test design included 15 keywords 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'). The prediction here is that participants showing good speech discrimination in noise might again show more plasticity in their L2 perception.

2.4.2. Auditory acuity tests

These tests assessed the ability to make fine discriminations between simple stimuli varying in frequency. It is predicted that participants with good auditory acuity might be more sensitive to novel acoustic cues in an new contrast. An adaptive procedure was used to determine difference limens for nonspeech tokens varying in frequency in the F2 range (F2 test) and fundamental frequency range (F0 test). In this discrimination test, participants were asked to choose the odd one out among three nonspeech sounds.

2.4.3. Cognitive tests

The aim of the cognitive tests is to assess whether good selective attention, short-term memory or ability to associate unrelated events helps in learning a novel phonetic contrast. To test attention, three subtests of the Test of Everyday Attention (TEA, [10]) were presented. These included tests of selective attention, of attentional switching and of sustained attention. To assess working memory, a backward digit span task was presented. 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. The third task was a test of paired association learning, which link two unrelated items and is related to processes of memory encoding and retrieval [11]. 20 meaningless syllables (CVC) were paired with a number (1 or 2). A test is then presented where a syllable was given as a cue on the screen and participants had to say which number was paired with it.

2.5. Test procedure

Each participant was tested individually over 10 sessions lasting between 30 and 50 mins, 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, after a brief familiarization with the new phonetic labels, and the attentional filtering task; at Session 2, the acoustic-cue weighting test; at Sessions 3 to 6, their training sessions; at Sessions 7 and 8, the post-tests (same as pre-tests); at Session 9, the paired-association task, the working memory span task and the speech perception in noise task; at Session 10, the attention and auditory acuity tasks.

3. Results

3.1.1. How do participants perform on first exposure to a novel phonetic contrast?

Table 1: *Correct identification scores (%) for the Syllable and Word identification tests pre- and post-training.*

% correct	*Mean*	*s.d.*	*Min.*	*Max.*
Syllable ID pre	55.5	11.9	36	72
Syllable ID post	66.5	9.8	49	85
Word ID pre	51.5	8.2	38	71
Word ID post	58.8	9.0	44	81

Pre-training tests included 'global' tests, which investigated the identification of the novel phonetic contrast within syllables or words, and 'analytic' tests (the cue-weighting and attentional filtering tasks described in 2.3.2), which assessed which acoustic cues the participant was attending to. In terms of the global tests, there was wide individual variability in the identification of the novel contrast at first exposure (after a brief initial familiarisation), with many participants performing at chance level but some individuals scoring around 70% correct (See Table 1).

We examined whether pre-training performance on the global tests was correlated with any of the auditory or cognitive skills. Although correlations with a significance of $p<0.05$ are shown in the tables, we will highlight correlations with a more stringent criterion of a $p<0.01$ significance level. Performance on the word and syllable ID tests was correlated, showing that both tasks were tapping into the same ability, although scores were typically higher in the easier syllable test. In terms of relation with cognitive or auditory skills (see Table 2), the only significant correlation at 0.01 level was between pre-test Syllable ID score and F2 discrimination accuracy.

Table 2: *Correlations between the Syllable (SID) and Word ID (WID) tests pre- and post-training and auditory/cognitive tests for which significant correlations occured. Correlations in bold are significant (* p<0.05, ** p<0.01).*

Test	SID pre	SID post	WID pre	WID post
F2 limen	**0.547****	0.270	0.296	**0.397***
WM	0.157	**0.397***	0.005	0.150
TEA switch	0.023	0.272	**-0.413***	0.048
VOT acc pre	0.366	**0.412***	**0.414***	**0.693****
VOT acc post	0.247	**0.444***	0.261	**0.531****

Next, we classified participants as 'good' or 'poor performers' on global tests pre-training, based on performance being above or below 0.5 SD of the mean on the Syllable ID pretest. This was to see whether the two groups varied on any of the auditory or cognitive tests. Independent-samples t-test suggested that the groups varied only in terms of their performance on the F2 frequency acuity test, as also suggested by the above-mentioned correlation. The two groups also varied significantly in terms of the percent change on the syllable ID score pre/post training (See Fig. 1 and 2): participants who showed initial sensitivity to the contrast improved much less (mean of 2.6 percentage points) than participants initially performing at chance level (mean of 57.6 percentage points).

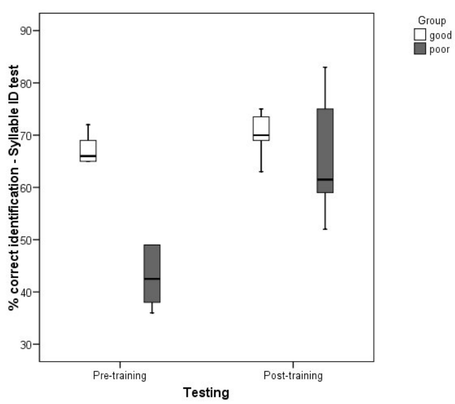

Figure 1*: Box plot showing correct identification for the Syllable ID task pre/post training for two groups of participants classified in terms of their performance on the Syllable ID pre-training test.*

As regards performance on the analytic tests, initial sorting ability on the basis of VOT (VOT acc pre) was correlated with Word ID post-training, so an initial sensitivity to VOT as a cue promoted better ultimate attainment on this task.

3.1.2. Are good analytic skills linked to early sensitivity to a novel phonetic contrast?

We classified participants in terms of their performance on the analytic cue-weighting test, i.e. whether they were able to consistently categorize using VOT, F0 or a combination of both cues. 12 out of the 25 participants could be classified as 'early categorizers' in that they showed a consistent use of one or both cues pre-training. Independent-samples t-test showed that the early-categorizers showed a better performance (77.2% vs 56.5%) in the sorting task (in which they had to sort stimuli on the basis of their VOT) than the non-categorizers ($t=2.35$; $p< 0.05$), denoting a general 'analytic' skill. Interestingly, though, the two groups did not differ significantly in their performance on the Syllable or Word ID pre-tests. Good performance on the analytic tests involving acoustic cue categorization therefore gave little advantage when identifying natural syllables or words containing the novel sounds prior to training, although there was a trend for initial sensitivity to VOT to lead to higher word ID scores post-training (i.e. better generalization of the learning).

3.1.3. What factors appear to impact on ultimate attainment (performance in post-tests)?

The post-test scores for the Syllable and Word ID tests are shown in Table 1. The next question is whether ultimate attainment is related to any of the auditory, cognitive or L1 language skills. The evidence for this was limited: the only significant correlations were between post-test Word ID and accuracy of VOT sorting both pre- and post-training; no other correlations at the 0.01 level were obtained with performance on the cognitive or auditory tests (see Table 2).

3.1.4. Is rate of learning related to any auditory, cognitive or native language skills?

Table 3: *Relative change (in percentage points) pre- and post- test, and between the 1st and 4th training session.*

% change	Mean	s.d.	Min.	Max.
syllable ID	11.0	13.6	-16	42
word ID	7.4	9.1	-7	28
training	8.7	5.6	-1.3	17.4

The rate of learning varied widely across individuals (see Table 3 and Fig. 2). First let's examine learning in the global tasks. The relative change for the Syllable ID test was correlated with the relative change for the Word ID test (with typically greater learning for the easier Syllable test). This suggests that learning generalized over these two global tasks.

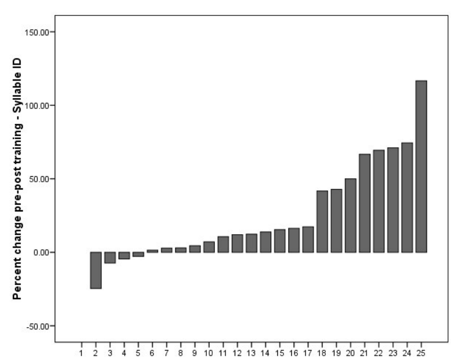

Figure 2: *Degree of learning shown by the 25 participants for the Syllable ID task, expressed as relative change (percentage points).*

Surprisingly, though, relative change in the Word and Syllable ID tasks was not significantly correlated with the rate of change between the first and last training tasks, nor with the change in the accuracy in VOT sorting pre/post training. However, percent change in the Word ID test was correlated with performance in the first training session (r=.522; p=0.009): the better the score at the initial training session, the more learning took place for the more difficult Word ID test, which again suggests better generalization of the learning. The only correlation obtained between rate of learning and auditory/cognitive skills were between the percent change in word ID and performance on the attentional switching task (TEA swi), although this correlation was fairly weak.

Table 4: *Correlations between the scores at the first and last training sessions (TR first, last) and relative change in the Word ID test (WID) and the auditory/cognitive tests for which correlations occurred (in bold, * p<0.05, ** p<0.01).*

Test	TR first	TR last	%change WID
PA	**0.473***	**0.468***	-0.105
TEA swi	**0.464***	**0.446***	**0.448***
F0 rt pre	**-0.487***	**-0.437***	-0.099

4. Discussion

This study examined whether initial sensitivity to a novel phonetic contrast, ultimate attainment post-training or rate of learning in phonetic training were correlated with native language, auditory or cognitive skills as tested in a large test battery. As expected, significant individual variability was found. Prior to training, and with only minimal exposure to this novel Korean voicing contrast, some English participants were able to identify the novel sounds in natural syllables and words with up to 70% accuracy, even though most participants were performing at chance. These 'good initial performers' were mostly distinguished from the 'poor initial performers' in terms of having better frequency acuity. Participants were also classified in terms of their analytic skills (i.e., ability to tune into the key acoustic cues). There was not a clear dichotomy between participants showing an initial global or analytic skill: 6 out of the 11 'good initial performers' were also 'early-categorizers', but were 3 out of the 10 'poor initial performers'. Good analytic skills did not appear to necessarily guarantee that these participants could tune into the appropriate acoustic cues in the natural speech tasks.

The degree of learning following phonetic training also varied drastically across individual participants. It is noteworthy that the 'good initial performers' improved much less in the training sessions than the 'poor initial performers' even though they were far from showing a ceiling effect. Initial sensitivity to a novel contrast seems therefore not to be a predictor of ultimate attainment as a result of learning. Rate of learning was not strongly correlated with the auditory or cognitive skills that we tested, with only a weak correlation with a measure of attentional switching. Performance on the phonetic training sessions themselves was weakly correlated with attentional switching and paired-associative learning. This is not unexpected as phonetic training involves learning the link between a new label and a sound. A number of the cognitive, language or auditory measures did not correlate significantly with performance on the phonetic training task. These include: measures of selective and sustained attention, working memory, native-language tasks and F0 frequency limen. The tasks used may not have been sensitive enough to measure the skills evaluated. Also, previous studies had found correlations with phonological short-term memory [e.g., 3], but we did not include that task in our battery. Further investigations are required before definite conclusions can be made.

5. Conclusions

Given that L2 phonetic training requires a significant investment of time on the part of the learner, it would be useful to be able to predict who might benefit from such training on the basis of a initial evaluation tasks. This study provides some initial pointers. First, initial sensitivity to a novel phonetic contrast is not predictive of ultimate attainment, so that a selection of candidates for phonetic training based on pre-test performance would not be recommended. Second, good analytic acoustic-phonetic skills are not necessarily predictive of performance on natural words, although early sensitivity to a key acoustic cue to the contrast seemed to lead to better generalization of the learning to the more difficult stimuli. Third, paired-associative and attention switching skills seem to be correlated with performance in phonetic training sessions. As these skills have been shown to be trainable, they might be practiced as precursors to phonetic training.

6. Acknowledgements

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

References

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

[2] Johnson, W.L. and Wu, S., "Assessing aptitude for learning with a serious game for foreign language and culture", 9th Int. conference on Intelligent Tutoring Systems (ITS 2008),

[3] Aliaga-Garcia, C., Mora, J.C., Cerviño-Povedano, E. "Phonological short-term memory and L2 speech learning in adulthood", Proc. New Sounds 2010, 19-24.

[4] Goldstone, R. "Perceptual learning", Annual Reviews in Psychology, 49, 585-612, 1998

[5] Wong, P., Perrachione, T. "Learning pitch patterns in lexical identification by native English-speaking adults", App. Psycholing., 28, 565-585, 2007.

[6] Kang, K., Guion, S. "Phonological systems in bilinguals: Age of learning effects on the stop consonant systems of Korean-English bilinguals", J. Acoust. Soc. Am., 119, 1672-1683, 2006.

[7] Garner, W. "The processing of information and structure", Erlbaum Potomac, MD, 1974.

[8] Hazan, V. et al. "Speech perception abilities of adults with dyslexia: is there any evidence for a true deficit?", JSLHR, 52, 1510-1529, 2009

[9] Bradlow, A. R. and Alexander, J. A. "Semantic-contextual and acoustic-phonetic enhancements for English sentence-in-noise recognition by native and non-native listeners", J. Acoust. Soc. Am, 121, 2339-2349, 2007.

[10] Robertson, I. H., Ward, T., Ridgeway, V Nimmo-Smith, I. "The Test of Every day Attention". Bury St Edmunds: Thames Valley Test Company. 1994.

[11] Buckner, R., Wheeler, M. "The cognitive neuroscience of remembering", Nature reviews Neuroscience 2, 624-634, 2001.