Tailoring phonetic learning to the needs of individuals on the basis of language aptitude

Mark Antoniou and Melissa Blair

The MARCS Institute for Brain, Behaviour and Development, Western Sydney University
m.antoniou@westernsydney.edu.au, 17647700@student.westernsydney.edu.au

Abstract

Many learners have difficulty discerning non-native speech sounds. The present study sought to tailor training to the needs of learners to improve learning. Participants learned five artificial languages, containing word pairs that differed on a single, critical phonetic feature. Previous research established that language aptitude predicted learning for these artificial languages. Therefore, we divided subjects into high and low aptitude groups and modified training in several ways. High aptitude subjects benefitted from numerous training methods, whereas low aptitude subjects only benefitted from increased exposure. These results are useful for identifying which training types benefit high and low aptitude learners.

Index Terms: phonetic learning, language aptitude, tailoring, individual differences

1. Introduction

Many learners commonly experience difficulty when learning the sounds of a non-native language. It is well documented that non-native talkers speak with a foreign accent. Studies on cross-language speech perception have demonstrated that listeners also hear with an 'accent' [1], meaning that non-native sounds are perceived in relation to native categories, a process termed assimilation [2]. Theoretical models have been developed to account for these language-specific effects on speech perception [3]. Difficulties in phonetic perception may lead to subsequent difficulties in language acquisition, such as vocabulary, comprehension, and literacy. Thus, by improving phonetic learning, it may be possible to ultimately improve language learning more generally. Although it is widely accepted that learners differ from one another, the majority of language teaching materials have assumed that all individuals learn in the same way. This consequently results in a mismatch in variation between learners in classrooms, and the homogeneity implied by the one-size-fits-all approach of most language course books. The aim of the present study was to examine the interaction between individual differences and training paradigm design.

For language learning to be successful, it is necessary to strike a balance between a complex combination of internal and external factors [4]. Internal factors refer to characteristics of the individual learner, such as cognitive abilities and motivation. External factors, on the other hand, refer to characteristics of the learning situation, including the language to be learned, and the characteristics of the teaching method, such as whether corrective feedback is provided. Recent work shows that the interaction between subject-internal and -external factors determines how successful learning will be. For example, talker variability (i.e., exposing learners to the speech of numerous talkers) is thought to result in more robust learning outcomes, better generalisation to novel stimuli, and greater long-term retention of phonetic information [5]. However, not all learners benefit from talker variability. Perrachione et al. [6] asked American-English native speakers to learn 18 pseudo-words. Four talkers produced six pseudo-words, each comprised of a single syllable with one of three pitch contours: level, rising or falling. It was found that high talker variability only benefitted learners with strong pre-training abilities. Learners with weak pre-training abilities were actually impaired relative to the single-talker condition. The implication of these findings is that it is crucial to take both individual differences and training methods (internal and external factors) into consideration when prescribing speech training.

Although foreign language learning can be challenging, research has identified several factors that are considered to be advantageous for non-native learners. First, it has been proposed that closely related languages may be easier to learn than others. This may be due to knowledge transfer from a learner's native language [7], or cross-language phonetic similarity [3]. Second, it has been suggested that certain language features might be easier to acquire for all learners regardless of past language experience [8]. Third, individual differences in cognitive abilities, such as working memory [9] and language aptitude [10] may contribute to language learning, and explain the variability that is typically observed between learners. The central idea in foreign language aptitude research is that the underlying reason why learners vary considerably in foreign language learning is because some possess a specific talent for learning languages. The concept of language aptitude is controversial, and raises questions concerning whether it is an innate ability or responsive to training, and whether it predicts language learning success and thus might be used to tailor training [11].

The present study is the latest in a series examining the interaction between subject-internal and -external factors in the learning of non-native phonetic contrasts.

Antoniou et al. [12] compared English monolinguals to Mandarin-English and Korean-English bilinguals in their ability to learn novel words from three artificial languages containing phonetic contrasts similar to those that occur in English, Mandarin, and Korean. Results indicated that all groups learned the Mandarin-like language most successfully, followed by the English-like, whereas the Korean-like language was the most difficult to learn. For the difficult Korean-like language, Korean-English bilinguals showed an advantage, suggesting that similarity to the native language may aid the learning of 'difficult' non-native phonetic contrasts.

Antoniou and Low [13] extended this work by presenting five artificial languages to English monolinguals and also measured the learners' pre-training cognitive profiles (language aptitude and working memory). Results revealed that, like in Antoniou et al. [12], the Mandarin-like language was the easiest to learn, followed by the English-like, whilst the Korean-like, Vietnamese-like, and Arabic-like languages were all learned equally poorly. In terms of individual differences, language aptitude reliably predicted successful learning in four of the five languages, and may therefore be a useful variable for separating good from poor learners when prescribing training.

The aim of the present study was to tailor training proactively to maximise learning outcomes based on the needs of individual learners. Based on the findings of [13], subjects were separated into high and low aptitude groups by administering a standardised assessment of language aptitude prior to the commencement of training. Training was modified in three ways (one for each of the three difficult to learn languages from [13]). The first training modification doubled the number of exposure trials to provide learners with more opportunity to learn the critical phonetic contrast without increasing other demands. Second, minimal pair words were chunked, that is, they were presented in couplets with the aim of highlighting the critical phonetic contrast that differentiates the words. Third, during the exposure phase, subjects were quizzed and corrective feedback was provided. These different training types were expected to benefit the high and low aptitude groups in different ways. Specifically, we hypothesized that:

1. High aptitude learners will outperform low aptitude learners.
2. High aptitude learners will benefit from multiple training types, and will show the largest benefit from corrective feedback.
3. Low aptitude learners may need more exposure to learn nonnative phonetic contrasts.
4. Low aptitude learners may be overwhelmed by feedback, especially for hard-to-learn languages.

2. Method

2.1. Participants

Twenty-eight Australian English native speakers (*M* age = 22.1 years; *SD* = 5.3) took part in the study. All were undergraduate Psychology students at Western Sydney University. None reported any history of neurological deficits. All passed an air conduction audiogram at 25 dB HL at 500, 1000, 2000, and 4000 Hz.

2.2. Language aptitude assessment

Prior to training, subjects' language aptitude was assessed using the LLAMA standardised tests [14]. Specifically, LLAMA subtest B was used to measure vocabulary learning ability. In the LLAMA B test, 20 pictures are presented on-screen, and subjects are given two minutes to click on each picture and learn to associate the visually presented novel word that corresponds with each picture. An audible *ding* is presented as feedback for a correct response, and a *beep* is presented for incorrect responses. LLAMA B scores range from 0-100 in increments of 5, with higher scores indicating greater language aptitude. Subjects were divided into two groups based on their LLAMA B scores, creating high (>50) and low aptitude (≤50) groups (see Table 1).

Table 1. *High and low aptitude group sample sizes, mean ages, and LLAMA B scores.*

Group	*n*	Age	LLAMA B (*SD*)
High aptitude	14	23.7	68.9 (9.0)
Low aptitude	14	20.5	35.0 (13.4)

2.3. Stimulus materials

Each language was comprised of four pairs of consonant-vowel monosyllables, and ended with vowels /e/, /i/, /o/, and /u/. The word pairs differed on a single, critical phonetic feature that was non-native, but similar to that of a phonetic feature which occurs in a natural language (see Table 2). The English-like artificial language contained a bilabial fricative voicing contrast /ɸ/-/β/, phonetically similar to that of the English labiodental fricative voicing contrast /f/-/v/, but differing in place of articulation (bilabial as opposed to labiodental). The English-like stimuli were produced by a phonetically trained male English speaker. The Mandarin-like language contained a dental-retroflex stop contrast /t/-/ʈ/, phonetically similar to the dental-retroflex sibilant contrast found in Mandarin /s/-/ʂ/, but realised with a different manner. As the retroflex versus non-retroflex stop contrast is native to Gujarati, the stimuli were produced by a female Gujarati native speaker. The Korean-like artificial language contained a voiceless fricative lenition contrast /θ/-/θ˙/, phonetically similar to the Korean voiceless stop lenition contrast /t/-/t*/, but differing in manner. The Korean-like stimuli were produced by a phonetically trained female native speaker of Korean. The Arabic-like language differentiated words using a voiceless velar-uvular ejective contrast /k'/-/q'/, similar to the native Arabic voiceless velar-uvular plosive contrast /k/-/q/, but produced with a different airstream mechanism. The Arabic-like stimuli were produced by a male native speaker of Quechua, a language containing velar and uvular ejectives. The Vietnamese-like language differentiated words using velar voiced plosive and implosive stops /g/-/ɠ/, similar to the bilabial voiced plosive and implosive stop contrast that occurs in Vietnamese /b/-/ɓ/, but with a different place of articulation. The Vietnamese-like stimuli were produced by a female native speaker of Sindhi, a language that contrasts velar voiced plosive and implosive stops.

The stimulus recordings were conducted inside a sound attenuated booth using a Shure SM58 cardioid microphone attached to a boom stand. The duration of the vowels were normalised to 350 ms for the Korean-like, Vietnamese-like, Arabic-like and Mandarin-like languages, and to 300 ms for the English-like language. The same vowel was spliced into both constants within a minimal pair to ensure that subjects were distinguishing the words based solely on the critical consonant distinction.

2.4. Procedure

The artificial language learning experiment was presented using Sennheiser HD 280 Pro headphones connected to a HP Pro Book 650 laptop running E-Prime software. Stimulus output level was calibrated to 72 dB SPL. The presentation order of the five artificial languages was counter-balanced across subjects.

A passive exposure training paradigm was adapted from past research [12], [13]. All training methods began with an exposure phase, which paired a picture and sound onscreen.

Table 2. *Natural and artificial language contrasts and their frequency of occurrence in the world's languages.*

Artificial language	Natural language contrast	Occurrence in world's languages (%)	Artificial language contrast
English-like	/f/-/v/ Labiodental **fricative voicing**	50.8	/ɸ/-/β/ Bilabial **fricative voicing**
Mandarin-like	/s/-/ʂ/ **Dental-retroflex** sibilants	20.2	/t/-/ʈ/ **Dental-retroflex** stops
Arabic-like	/k/ - /q/ **Voiceless velar-uvular** plosives	20.0	/k'/-/q'/ **Voiceless velar-uvular** ejectives
Vietnamese-like	/b/-/ɓ/ Bilabial **voiced plosive-voiced implosive stops**	10.0	/g/-/ɠ/ Velar **voiced plosive-voiced implosive stops**
Korean-like	/t/-/t*/ Voiceless alveolar stop **lenition**	1.8	/θ/-/θ'/ Voiceless dental fricative **lenition**

The exposure phase consisted of eight words (four pairs) and 12 repetitions, giving 96 trials in total. The number of exposure trials were doubled to 192 in the double exposure training type (used for the Arabic-like language). Words were presented in random order at a rate of every 3.5 seconds with no requirement for response to proceed to the next word. In the chunking training type (used for the Vietnamese-like and Mandarin-like languages), minimal pair words were presented in couplets (e.g., /gu/ followed by /ɠu/). In the feedback training type (used for the Korean-like and English-like languages), following exposure to each couplet, subjects were quizzed by having both pictures presented onscreen and then selecting the correct picture following playback of each of the two words, and corrective feedback was provided.

The exposure phase was immediately followed by the test phase, which was identical for all three training types. In the test phase, subjects were presented with eight repetitions of each word, giving 64 test trials in total per language. Each word was auditorily presented and all eight corresponding pictures were displayed on the screen. Participants were required to match the heard word to the correct picture by pressing the corresponding number (1-8) on a keyboard. The test was self-paced and no corrective feedback was provided.

3. Results

Word identification scores for each of the five artificial languages are depicted in Figure 1. Initial inspection of Figure 1 suggests that both groups found the Mandarin-like language easiest to learn. The high aptitude group appears to have a general learning advantage over the low aptitude group, with the exception of the Vietnamese-like language, which both groups learned poorly.

To determine how language aptitude relates to training paradigm design, we conducted a 2 (Group) × (5 Languages) factorial ANOVA. A main effect of group, $F(1, 26) = 16.4$, $p < .001$, $\eta_p^2 = .386$, revealed that the high aptitude group exhibited an overall learning advantage over the low aptitude group. Further, there was a main effect of language, $F(4, 104) = 39.5$, $p < .001$, $\eta_p^2 = .603$, as well as a Group × Language interaction, $F(4, 104) = 3.1$, $p = .02$, $\eta_p^2 = .106$. To analyse the interaction, a series of posthoc *t*-tests were conducted. These posthoc analyses confirmed that the high aptitude group outperformed the low aptitude group in learning of the Mandarin-like language, $t(26) = 2.82$, $p = .009$, the Arabic-like language, $t(26) = 3.75$, $p = .001$, and the English-like language, $t(26) = 3.07$, $p = .005$. The high and low aptitude groups did not differ in their learning of the Vietnamese-like, $t(26) = 0.53$, $p = .604$, or Korean-like languages, $t(26) = 2.36$, $p = .026$ (note that Bonferroni correction requires an adjusted alpha level of .05 / 5 = .01).

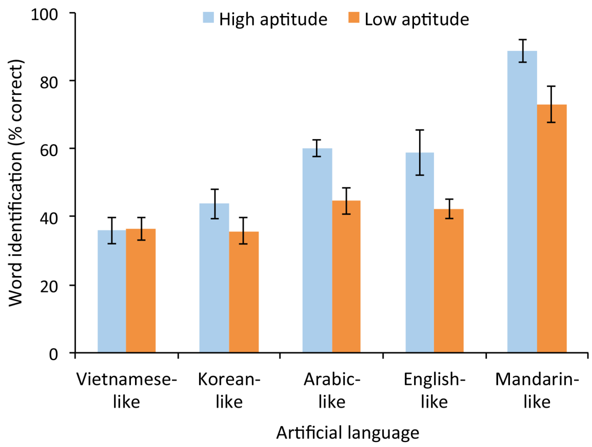

Figure 1. *High and low aptitude groups' mean word identification scores for each artificial language. Error bars depict standard error of the mean.*

By comparing the present results to those of Antoniou and Low, we may evaluate the effectiveness of the three training types relative to the baseline passive exposure condition used in their study (i.e., 96 exposure trials, no chunking, no feedback). Recall that in that study, the Vietnamese-like, Korean-like, and Arabic-like languages were learned equally poorly (~40%). Also bear in mind that Antoniou and Low did not separate good from poor learners. Our double exposure training improved learning of the Arabic-like language for both the high and low aptitude groups (grand average = 52.4%). The low aptitude group benefitted from chunking for the easy-to-learn Mandarin-like language. However, low aptitude learners did not benefit from chunking or feedback for the other languages. The high aptitude group benefitted from all three training types (double exposure for the Arabic-like language, chunking for the Mandarin-like, and feedback for the Korean-like).

4. Discussion

The present study investigated the interaction between language aptitude and training paradigm design in phonetic learning. As predicted, high aptitude learners showed a general learning advantage over those with low aptitude. High aptitude learners also benefitted from a variety of training types. In contrast, low aptitude learners benefitted from doubled exposure for the difficult-to-learn Arabic-like language. They did not benefit from either chunking or feedback for the difficult-to-learn languages, and learned the Korean-like and Vietnamese-like languages equally poorly.

These findings are consistent with a cognitive resource limitation view of speech perception and phonetic learning. Such an account is offered by the active control model [15], according to which closed-loop processing routines called active control structures monitor incoming speech stimuli in a context sensitive way. Crucially, this active control draws from a finite pool of cognitive resources and thus when the task facing the learner becomes more difficult, more cognitive resources are required, and perceptual performance will decline when these resources are depleted. Support for the active control model comes from studies investigating speech perception under different levels of cognitive load [16]. If speech processing depends on active control structures which in turn depend on the availability of cognitive resources then performance costs should be exacerbated under cognitively demanding conditions. Therefore, it stands to reason that low-aptitude learners might be overwhelmed by a phonetic learning task in which they are faced with the challenge of learning a difficult non-native phonetic contrast. This resource limitation hinders their ability to attend to the relevant information in the contrasts being learned.

High aptitude learners benefitted from several training types. First, they benefitted from chunking of minimal pair stimuli when learning the Mandarin-like language. Second, they benefitted from corrective feedback for the English-like language. Third, they benefitted from increased passive exposure to the Arabic-like language. These results suggest that for all of the above languages, high aptitude learners possess sufficient cognitive resources in reserve to take advantage of the additional information presented in each training paradigm.

In contrast, low aptitude learners do not have sufficient cognitive resources in reserve and thus are unable to attend to the additional information presented during training (e.g., chunking or feedback). They did, however, benefit from increased exposure to the Arabic-like language, probably because this training method does not increase processing demands.

The findings are consistent with [12], [13]. As in those studies, English listeners found the Mandarin-like language easiest to learn, whereas the Vietnamese-like and Korean-like languages were the most difficult to learn. However, our training modifications resulted in improved learning outcomes relative to these previous studies. This improvement was most pronounced for the Arabic-like language. The results demonstrate that cognitive abilities interact with language learning outcomes. Specifically, they support the assertion that language aptitude is a useful variable for tailoring training [11]. Thus, we strongly advocate training individuals differently when it comes to phonetic learning.

5. Conclusions

We divided learners into high and low language aptitude groups and exposed them to several different training methods. High aptitude subjects benefitted from numerous training methods, whereas low aptitude subjects only benefitted from increased exposure. In sum, the present findings demonstrate that when it comes to language learning, one size does not fit all. Our results have implications for speech training paradigms.

6. References

[1] M. Antoniou, M. D. Tyler, and C. T. Best, "Two ways to listen: Do L2-dominant bilinguals perceive stop voicing according to language mode?," *J Phon*, vol. 40, no. 4, pp. 582–594, Jul. 2012.

[2] M. Antoniou, C. T. Best, and M. D. Tyler, "Focusing the lens of language experience: Perception of Ma'di stops by Greek and English bilinguals and monolinguals," *J Acoust Soc Am*, vol. 133, no. 4, pp. 2397–2411, Apr. 2013.

[3] C. T. Best and M. D. Tyler, "Nonnative and second-language speech perception: Commonalities and complementarities," in *Language experience in second language speech learning: In honor of James Emil Flege*, O.-S. Bohn and M. J. Munro, Eds. Amsterdam: John Benjamins, 2007, pp. 13–34.

[4] M. Antoniou, M. Ettlinger, and P. C. M. Wong, "Complexity, training paradigm design, and the contribution of memory subsystems to grammar learning," *PLoS ONE*, vol. 11, no. 7, p. e0158812, Jul. 2016.

[5] J. S. Logan, S. E. Lively, and D. B. Pisoni, "Training Japanese listeners to identify English /r/ and /l/: A first report," *J Acoust Soc Am*, vol. 89, no. 2, pp. 874–886, Feb. 1991.

[6] T. K. Perrachione, J. Lee, L. Y. Y. Ha, and P. C. M. Wong, "Learning a novel phonological contrast depends on interactions between individual differences and training paradigm design," *J Acoust Soc Am*, vol. 130, no. 1, pp. 461–472, 2011.

[7] R. C. Major, "Transfer in second language phonology: A review," in *Phonology and second language acquisition*, vol. 36, J. G. Hansen Edwards and M. L. Zampini, Eds. Amsterdam: John Benjamins, 2008, pp. 63–94.

[8] F. R. Eckman, "Typological markedness and second language phonology," in *Phonology and second language acquisition*, vol. 36, J. G. Hansen Edwards and M. L. Zampini, Eds. Amsterdam: John Benjamins, 2008, pp. 95–115.

[9] A. Baddeley, "Working memory and language: An overview," *J Comm Dis*, vol. 36, no. 3, pp. 189–208, May 2003.

[10] J. B. Carroll, "Foreign language proficiency levels attained by language majors near graduation from college," *Foreign Lang Ann*, vol. 1, no. 2, pp. 131–151, Dec. 1967.

[11] Z. Dörnyei and P. Skehan, "Individual differences in second language learning," in *The handbook of second language acquisition*, C. J. Doughty and M. H. Long, Eds. 2003, pp. 589–630.

[12] M. Antoniou, E. Liang, M. Ettlinger, and P. C. M. Wong, "The bilingual advantage in phonetic learning," *Biling Lang Cogn*, vol. 18, no. 4, pp. 683–695, Oct. 2015.

[13] M. Antoniou and T. Low, "Individual differences in language learning," presented at the CoEDL Fest, Western Sydney University, 2016.

[14] P. Meara, "Llama, Language Aptitude Tests: The manual," *Swansea: Lognostics*, 2005.

[15] J. S. Magnuson and H. C. Nusbaum, "Acoustic differences, listener expectations, and the perceptual accommodation of talker variability," *J Exp Psychol Hum Percept Perform*, vol. 33, no. 2, pp. 391–409, 2007.

[16] M. Antoniou and P. C. M. Wong, "Poor phonetic perceivers are affected by cognitive load when resolving talker variability," *J Acoust Soc Am*, vol. 138, no. 2, pp. 571–574, Aug. 2015.