Poor phonetic perceivers are affected by cognitive load when resolving talker variability (L)

Mark Antoniou

MARCS Institute, University of Western Sydney, Locked Bag 1797, Penrith, New South Wales 2751, Australia

Patrick C. M. Wong^{a)}

Department of Linguistics and Modern Languages, The Chinese University of Hong Kong, Shatin, New Territories, Hong Kong Special Administrative Region, China

(Received 12 March 2015; revised 19 June 2015; accepted 19 June 2015; published online 5 August 2015)

Speech training paradigms aim to maximise learning outcomes by manipulating external factors such as talker variability. However, not all individuals may benefit from such manipulations because subject-external factors interact with subject-internal ones (e.g., aptitude) to determine speech perception and/or learning success. In a previous tone learning study, high-aptitude individuals benefitted from talker variability, whereas low-aptitude individuals were impaired. Because increases in cognitive load have been shown to hinder speech perception in mixed-talker conditions, it has been proposed that resolving talker variability requires cognitive resources. This proposal leads to the hypothesis that low-aptitude individuals do not use their cognitive resources as efficiently as those with high aptitude. Here, high- and low-aptitude subjects identified pitch contours spoken by multiple talkers under high and low cognitive load conditions established by a secondary task. While high-aptitude listeners outperformed low-aptitude listeners across load conditions, only low-aptitude listeners were impaired by increased cognitive load. The findings suggest that low-aptitude listeners either have fewer available cognitive resources or are poorer at allocating attention to the signal. Therefore, cognitive load is an important factor when considering individual differences in speech perception and training paradigms.

© 2015 Acoustical Society of America. [<http://dx.doi.org/10.1121/1.4923362>]

[TB]

Pages: 571–574

I. INTRODUCTION

Successful speech perception and word recognition depend on an interaction between subject-internal and -external factors. Speech and language training paradigms often attempt to maximise learning outcomes by manipulating external factors, such as providing feedback (McCandliss *et al.*, 2002), manipulating chunk size (Chen and Cowan, 2005), and introducing stimulus variability (Bradlow *et al.*, 1999). Stimulus variability exposes the learner to a variety of exemplars of the contrast to be learned and is thought to result in more robust categorical learning and better generalisation to novel stimuli (Logan *et al.*, 1991). In training paradigms, the inclusion of talker variability is thought to enhance learning because it is more effortful to process, and results in superior long-term retention of phonetic information (Barcroft and Sommers, 2005). However, although talker variability is thought to lead to better training outcomes, not all learners benefit because subject-external factors (e.g., talker variability) interact with subject-internal ones (e.g., pretraining aptitude) to determine learning success (Perrachione *et al.*, 2011). In perception, talker variability typically results in a processing cost, realised as a decrease in word identification accuracy and delayed response times (Mullennix *et al.*, 1989), and these effects are

exaggerated in nonnative listeners (Antoniou *et al.*, 2015). This evidence from the perception literature leads us to question what it is about talker variability that hinders learning in low-aptitude perceivers. We will propose that resolving talker variability requires the allocation of cognitive resources, and individual differences determine if talker variability will enhance or hinder performance.

Perrachione *et al.* (2011) investigated whether individual differences in perceptual abilities might modulate the effectiveness of multi-talker training in a pitch learning task. American English native speakers learned a vocabulary of 18 pseudowords comprised of six syllables produced by four talkers with one of three pitch contours (level, rising or falling). Although high variability is thought to improve learning, it was only advantageous for learners with strong pre-training pitch perception abilities. Importantly, learners with weaker abilities to perceive pitch did not benefit from high variability training, and were actually impaired relative to the low variability (i.e., single talker) condition. These findings demonstrate the importance of considering individual differences in pretraining aptitudes when evaluating the efficacy of speech training protocols.

A possible explanation for why high variability training might hinder low-aptitude individuals is that talker variability may incur additional processing costs compared to single talker speech (Martin *et al.*, 1989). According to the active control model (Magnuson and Nusbaum, 2007), computational mechanisms constantly monitor both indexical and

^{a)}Electronic mail: p.wong@cuhk.edu.hk

phonetic characteristics of speech. A change in talker triggers normalisation procedures, resulting in a processing cost due to the increased alternative interpretations that must be tested in order to recognise an utterance. From this active control perspective, resolving talker variability relies crucially on the availability of cognitive resources. Low-aptitude learners might be overwhelmed by high variability, hindering their ability to attend to relevant information in the contrasts being learned.

Support for the active control model comes from studies investigating speech perception under different levels of cognitive load. If talker normalisation depends on active control structures which in turn depend on the availability of cognitive resources then performance costs should be exacerbated under demanding conditions. In one such study, [Nusbaum and Morin \(1992\)](#) asked subjects to monitor words in blocked- vs mixed-talker presentations, and also store either one two-digit number (low load) or three two-digit numbers (high load) in memory. A performance decrement was observed for mixed-talker presentation under a high load, but no effect was observed for single-talker presentation, suggesting that resolving mixed-talker speech requires greater cognitive resources than single-talker speech.

We asked subjects to identify the pitch contours of vowels spoken by multiple talkers, but to do so under high and low cognitive load conditions to test the hypothesis that low-aptitude individuals have reduced availability of cognitive resources. Subjects were divided into high- and low-aptitude groups using a median split procedure on their pitch contour perception aptitude test scores. We then compared how they were affected by high vs low cognitive load conditions, established by employing a secondary task in which subjects made a same/different judgment concerning two letters of the alphabet. Crucially, some pairs or letters ended with the same phoneme and rhymed (e.g., C, V), whereas others were acoustically dissimilar and did not rhyme (e.g., F, J). Seminal studies on working memory have established that subjects can readily recall a sequence of acoustically dissimilar letters such as F, J, R, X, whereas they are likely to experience difficulty in retaining sequences of letters that are acoustically similar such as C, D, G, T ([Baddeley, 1966](#); [Conrad and Hull, 1964](#)). Based on these well-established findings, it was hypothesised that pairs of letters that rhymed would exert a higher cognitive load and require a greater allocation of cognitive resources to complete the judgment. If the low-aptitude group are overwhelmed by the cognitive demands of identifying pitch contours when there is talker variability, we would predict that the processing cost would be exacerbated under high cognitive load, whereas high-aptitude individuals will not show a performance decrement under the same cognitive load.

II. METHOD

A. Participants

Twenty-nine young adult native speakers of American English (10 males, 19 females) participated in this study ($M_{\text{age}} = 22.1$, $SD = 2.1$, range = 19–29). All were undergraduate students at Northwestern University. None of the subjects reported any history of audiologic or neurological

deficits. All passed a pure tone audiological screening at 25 dB hearing level (HL) at 500, 1000, 2000, and 4000 Hz. None possessed prior experience with a tone language.

B. Stimulus materials

Tone stimuli were created by superimposing one of three Mandarin-like pitch contours (level, rising, or falling) onto the English vowels /a, e, i, o, u/ using the pitch-synchronous overlap-add method in Praat. The vowels were recorded from four native speakers of American English (two males, two females). Letter stimuli were recorded from a new female native speaker of American English. The letters recorded were C, D, G, T, V (high cognitive load condition) and F, J, R, X (low cognitive load condition). All stimuli were normalised to 72 dB sound pressure level (SPL).

C. Procedure

1. Pitch contour perception test

Subjects first completed the pitch contour perception test in which identified the pitch contour of a vowel (/a, e, i, o, u/) by selecting the corresponding arrow ($\rightarrow$, $\nearrow$, $\searrow$). There were 180 randomised trials. On each trial, two arrows were presented onscreen as response options (counterbalanced) and subjects made their selection by pressing one of two buttons on a response box. They were instructed to respond as quickly as possible, and a response time limit was set to 3 s.

2. Cognitive load test

Subjects then completed the cognitive load test, which was comprised of 120 randomised trials. Each trial required subjects to make three judgments in the following order. First, a vowel was presented in isolation and subjects were asked to identify its pitch contour. Subjects then heard a letter of the alphabet (e.g., F) and were asked to keep it in memory for the time being. For the second judgment, another vowel was presented in isolation, and subjects identified its pitch contour. For the third and final judgment, a second letter of the alphabet was presented (e.g., R) and subjects had to indicate whether the two letters were the same or different (different in this example). Crucially, on half of the trials, the two letters were acoustically dissimilar and did not rhyme (e.g., F and R), whereas for the other half, their names ended with a common phoneme and rhymed (e.g., C and V), and this acoustic similarity effect is known to increase working memory demands ([Baddeley, 1966](#); [Conrad and Hull, 1964](#)). Subjects were permitted 1250 ms to indicate their response for each judgment. We calculated the accuracy of pitch contour identification and letter discrimination, as well as recognition times for correct trials.

III. RESULTS

A wide range of pitch contour perception aptitude scores were observed ($M = 82\%$, range = 49%–98%). A median split was performed on the basis of pitch contour perception aptitude scores, creating the low-aptitude ($n = 15$, $M = 69\%$, range = 49%–83%) and high-aptitude groups ($n = 14$, $M = 92\%$, range = 85%–98%). We examined if the

FIG. 1. Low-aptitude group show poorer pitch contour perception (A) and delayed response times (B) under high relative to low cognitive load conditions, whereas the high-aptitude group are unaffected. Errors bars depict standard error of the mean.

low-aptitude group was more disadvantaged in the cognitive load task than the high-aptitude group via a series of four $2 \times (2)$ ANOVAs with the between-subjects factor of group and the within-subjects factor of cognitive load (high vs low).

A. Pitch contour identification accuracy under cognitive load

Pitch contour identification accuracy under high and low cognitive load conditions is shown in Fig. 1(A). The high-aptitude group were more successful at identifying pitch contours than the low-aptitude group (86% vs 66%), $F(1, 27) = 31.5$, $p < 0.001$, $\eta_p^2 = 0.539$. There was also a main effect of cognitive load, $F(1, 27) = 10.5$, $p = 0.003$, $\eta_p^2 = 0.281$, indicating that overall accuracy was higher in the low load condition. A Group $\times$ Cognitive Load interaction suggested that the low-aptitude group were more affected by cognitive load, $F(1, 27) = 7.5$, $p = 0.011$, $\eta_p^2 = 0.218$. Simple effects tests confirmed that the low-aptitude group were poorer at identifying tones in the high cognitive load condition than the low load condition, $F(1, 27) = 18.5$, $p < 0.001$, whereas the high-aptitude group were unaffected by cognitive load, $F(1, 27) = 0.102$, $p = 0.752$.

B. Pitch contour identification recognition time under cognitive load

Recognition times for pitch contour identification are shown in Fig. 1(B). The low-aptitude group were slower to identify pitch contours than the high-aptitude group (960 vs 870 ms), $F(1, 27) = 7.2$, $p = 0.012$, $\eta_p^2 = 0.210$. There was no main effect of cognitive load, $F < 1$. A Group $\times$ Cognitive Load interaction suggested that the low-aptitude group were more affected by cognitive load, $F(1, 27) = 11.9$, $p = 0.002$, $\eta_p^2 = 0.305$. Simple effects tests confirmed that the low-aptitude group showed delayed recognition times in the high

relative to the low cognitive load condition, $F(1, 27) = 7.8$, $p = 0.010$. In contrast, the high-aptitude group did not show delayed recognition times under high cognitive load at the adjusted alpha level of 0.025, $F(1, 27) = 4.4$, $p = 0.047$.

C. Letter discrimination accuracy under cognitive load

Letter discrimination accuracy is shown in Fig. 2(A). There were no significant main effects of group, $F(1, 27) = 2.1$, $p = 0.16$, $\eta_p^2 = 0.071$, cognitive load, $F < 1$, or interaction, $F < 1$. Thus, group differences were not observed in terms of letter discrimination accuracy.

D. Letter discrimination recognition time under cognitive load

Letter discrimination recognition times are shown in Fig. 2(B). There was no main effect of group, $F(1, 27) = 1.2$, $p = 0.288$, $\eta_p^2 = 0.042$, but there was a significant main effect of cognitive load, $F(1, 27) = 4.7$, $p = 0.040$, $\eta_p^2 = 0.265$. A Group $\times$ Cognitive load interaction suggested that the low-aptitude group were more affected by cognitive load, $F(1, 27) = 9.8$, $p = 0.004$, $\eta_p^2 = 0.265$. Simple effects tests confirmed that the low-aptitude group were slower to identify letters in the high cognitive load condition than the low load condition, $F(1, 27) = 14.4$, $p = 0.001$, whereas the high-aptitude group were unaffected by cognitive load, $F(1, 27) < 1$, $p = 0.910$.

IV. DISCUSSION

The present findings provide converging evidence that low-aptitude individuals are impaired by cognitive load in pitch contour identification as well as letter discrimination. Low-aptitude subjects showed performance decrements in a high cognitive load condition for pitch contour identification accuracy, as well as slower identification of pitch contours

FIG. 2. (A) Letter discrimination accuracy of segments was unaffected by cognitive load, and the two groups did not statistically differ. (B) Low-aptitude group show delayed response times for letter discrimination under high relative to low cognitive load conditions, whereas the high-aptitude group are unaffected by cognitive load. Errors bars depict standard error of the mean.

and same/different judgment of letter pairs. High-aptitude individuals were unaffected by cognitive load.

Our findings lend support to the active control model of speech perception (Magnuson and Nusbaum, 2007). Under a high cognitive load, the model predicts that accuracy will be reduced and reaction times prolonged. Both predictions were supported in the present investigation. One possible reason for the performance decrement is due to the active control structures monitoring the incoming speech signal that activate when a change in talker is detected in order to achieve a stable mapping. This remapping requires the allocation of cognitive resources, and if the individual does not have sufficient resources in reserve, then task performance will deteriorate. In contrast, a low cognitive load condition frees additional resources for the process of speech recognition and therefore performance accuracy and response times improve. A second possible explanation is that a perceiver's ability to hone in on relevant auditory dimensions is modulated by how effectively they allocate auditory attention to the signal (Francis and Nusbaum, 2002). According to this attention-to-dimension perspective, low-aptitude individuals may be poorer at allocating attention to the signal. These two explanations are not necessarily mutually exclusive or incompatible. A third possible explanation is that low- and high-aptitude individuals have similar cognitive resources, but low-aptitude individuals have poorer pitch encoding, which places greater processing demands on other cognitive resources. However, this explanation seems unlikely because they were also poorer in the letter discrimination task, suggesting that a more general explanation than pitch encoding is required.

The results complement those of Perrachione *et al.* (2011) who found that successful learning of pitch contours depends on an interaction between individual differences in perceptual abilities and the structure of the training protocol. In that study, high variability training only benefitted individuals with high pretraining aptitudes, and actually hindered low-aptitude learners. However, it was not clear what was responsible for the learning decrement observed in low-aptitude individuals. The present results demonstrate that the learner by training interaction reported in Perrachione *et al.* (2011) may be attributed to reduced availability of cognitive resources in low-aptitude subjects in the mixed-talker training condition.

The findings have pedagogical implications for language training in educational settings. It is necessary to take into account the pretraining aptitudes of individuals when prescribing a training protocol. Low-aptitude individuals may benefit from low stimulus variability paradigms (e.g., single talker) because they may take advantage of processes that adapt to the consistent, predictable features of that talker's phonetics, facilitating recognition and improving accuracy relative to high stimulus variability paradigms. More generally, it may be beneficial for low-aptitude individuals to limit the cognitive load (i.e., dual-task) demands of training paradigms, phonetic details notwithstanding. Future work should test whether training outcomes can be maximised by tailoring training to suit individual learning profiles. Wong *et al.* (2012) recommend personalising learning

to meet the needs of individual learners. The present findings are consistent with their recommendations.

In conclusion, low-aptitude individuals may have insufficient cognitive resources in reserve to benefit from high stimulus variability paradigms. Mixed-talker paradigms require greater cognitive resources, and will only lead to enhanced learning outcomes if individuals possess sufficient cognitive resources to accommodate these cognitive demands. Measuring pretraining aptitudes is a simple and effective way of determining if individuals will be able to cope with these cognitive demands. For low-aptitude individuals, alternative training paradigms should be employed, both in research and educational settings.

ACKNOWLEDGMENTS

Supported by U.S. National Institutes of Health grants R01DC013315 and R01DC008333, National Science Foundation grant BCS-1125144 to P.C.M.W., and Australian Research Council Discovery Early Career Research Award DE150101053 to M.A.

- Antoniou, M., Wong, P. C. M., and Wang, S. (2015). "The effect of intensified language exposure on accommodating talker variability," *J. Speech Lang. Hear. Res.* **58**, 722–727.
- Baddeley, A. D. (1966). "Short-term memory for word sequences as a function of acoustic, semantic and formal similarity," *Q. J. Exp. Psychol.* **18**, 362–365.
- Barcroft, J., and Sommers, M. S. (2005). "Effects of acoustic variability on second language vocabulary learning," *Stud. Second Lang. Acquis.* **27**, 387–414.
- Bradlow, A. R., Nygaard, L. C., and Pisoni, D. B. (1999). "Effects of talker, rate, and amplitude variation on recognition memory for spoken words," *Percept. Psychophys.* **61**, 206–219.
- Chen, Z., and Cowan, N. (2005). "Chunk limits and length limits in immediate recall: A reconciliation," *J. Exp. Psychol. Learn. Mem. Cogn.* **31**, 1235–1249.
- Conrad, R., and Hull, A. J. (1964). "Information, acoustic confusion and memory span," *Br. J. Psychol.* **55**, 429–432.
- Francis, A. L., and Nusbaum, H. C. (2002). "Selective attention and the acquisition of new phonetic categories," *J. Exp. Psychol. Hum. Percept. Perform.* **28**, 349–366.
- Logan, J. S., Lively, S. E., and Pisoni, D. B. (1991). "Training Japanese listeners to identify English /r/ and /l/: A first report," *J. Acoust. Soc. Am.* **89**, 874–886.
- Magnuson, J. S., and Nusbaum, H. C. (2007). "Acoustic differences, listener expectations, and the perceptual accommodation of talker variability," *J. Exp. Psychol. Hum. Percept. Perform.* **33**, 391–409.
- Martin, C. S., Mullennix, J. W., Pisoni, D. B., and Summers, W. V. (1989). "Effects of talker variability on recall of spoken word lists," *J. Exp. Psychol. Learn. Mem. Cogn.* **15**, 676–684.
- McCandliss, B. D., Fiez, J. A., Protopapas, A., Conway, M., and McClelland, J. L. (2002). "Success and failure in teaching the [r]-[l] contrast to Japanese adults: Tests of a Hebbian model of plasticity and stabilization in spoken language perception," *Cogn. Affect. Behav. Neurosci.* **2**, 89–108.
- Mullennix, J. W., Pisoni, D. B., and Martin, C. S. (1989). "Some effects of talker variability on spoken word recognition," *J. Acoust. Soc. Am.* **85**, 365–378.
- Nusbaum, H. C., and Morin, T. M. (1992). "Paying attention to differences among talkers," in *Speech Perception Production And Linguistic Structure*, edited by Y. Tohkura, E. Vatikiotis-Bateson, and Y. Sagisaka (IOS Press, Tokyo), pp. 113–134.
- Perrachione, T. K., Lee, J., Ha, L. Y. Y., and Wong, P. C. M. (2011). "Learning a novel phonological contrast depends on interactions between individual differences and training paradigm design," *J. Acoust. Soc. Am.* **130**, 461–472.
- Wong, P. C. M., Morgan-Short, K., Ettlinger, M., and Zheng, J. (2012). "Linking neurogenetics and individual differences in language learning: The dopamine hypothesis," *Cortex* **48**, 1091–1102.