LANGUAGE EXPOSURE BENEFIT TO TALKER LEARNING IN AN UNFAMILIAR LANGUAGE

Adriel John Orena[1], Rachel M. Theodore[2], Linda Polka[1]

[1]McGill University, [2]University of Connecticut
adriel.orena@mail.mcgill.ca

ABSTRACT

Recent research shows that exposure to an unfamiliar language is sufficient in improving talker learning. Here, we further investigated the nature of this effect by exploring individual differences and methodological issues. Two groups of English-monolingual adults were recruited: one with regular exposure to French (Montréal), and the other without (Storrs). Both groups learned the voices of English talkers faster than French talkers; however, in contrast to previous findings, no group differences were found in talker learning, which may be due to task differences (4-AFC vs. 2-AFC task). However, stable patterns in individual differences emerged: performance in identifying English talkers was correlated with performance in identifying French talkers for Storrs residents, but not for Montréal residents. These findings suggest that a language-general "talent" contributes to talker learning only in encounters with highly novel languages, and that exposure to the surface acoustic-phonetic properties of an unfamiliar language drives the *language exposure benefit*.

Keywords: Speech Perception, Talker Learning, Talker Identification, Language Exposure

1. INTRODUCTION

Over two decades ago, Goggin, Thompson, Strube and Simental [3] showed that listeners are better at learning the voices of talkers who speak a familiar language than talkers who speak an unfamiliar language. This *language familiarity effect* is robust and has been observed in many cross-language studies with different language pairs [5, 9]. Subsequent studies have extended this effect to a gradient of unfamiliar talkers: listeners are better at learning the voices of talkers who speak like them than talkers with different regional accents [8], talkers with foreign accents [10], and talkers who speak the listener's second language [5]. Collectively, these studies suggest a relationship between talker-learning performance and familiarity with a dialect or language: the more familiar you are with a language, the better you are at recognizing talkers in that language.

There is strong evidence that phonological knowledge mediates talker learning [3, 7]. However, infants also show a language familiarity effect for talker discrimination [4], suggesting that language exposure alone can promote talker learning. This hypothesis was recently tested by [6], who confirmed a role for phonetic sensitivity in the gradient effects of the language familiarity benefit. In [6], groups of monolingual-English adults were trained to learn the voices of talkers speaking either a familiar language (i.e., English) or an unfamiliar language (i.e., French). The results showed that listeners who have had regular exposure to French via their living situation (Montréal, Québec residents) outperformed listeners who have had no French exposure (Storrs, Connecticut residents) in learning the voices of French talkers. Despite not having any proficiency in French, monolingual-English Montréal residents already gained some benefit in talker learning. Thus, these findings extend the gradient effects of language experience on talker learning to include perceptual learning that precedes word comprehension.

The current study investigates the nature of this *language exposure benefit* to talker learning by exploring individual differences in talker learning. First, we aimed to replicate the language exposure benefit found in [6] using an easier training paradigm than previous studies (see Methods section for justification). Second, we asked whether individual differences contribute to talker learning, and if so, how they might present themselves differently according to our language groups. For example, it is possible that some listeners may be better at recognizing talkers solely based on language-general acoustic properties of the voice. Thus, we asked whether such a 'talent' exists for identifying talkers speaking a highly novel language, and whether it continues to play a large role after having systematic exposure to that language. Finally, we conducted exploratory analysis to determine what kinds of language exposure (e.g., print media? TV/film?) are important for predicting talker learning abilities in individuals with exposure to an unfamiliar language.

2. METHODS

To assess talker learning, we used a talker learning design similar to that of [6] with one critical change in the training phase. In [6], a 4-alternative forced choice (4-AFC) task was used to train and test participants on the voice-face criterion. Before advancing to the test phase, participants were required to achieve criterion in the training phase, which was defined as 85% accuracy in a single block of 60 trials, or the completion of 8 training blocks. All participants achieved an 85% accuracy level for English talkers before completing eight training blocks; however, for the French talkers, 3 Montréal and 11 Storrs participants completed the maximum 8 training blocks rather than achieving the 85% accuracy level. Results from the training phase show clear language exposure benefits for Montréal residents in the training phase; however, the interpretability of the test phase were less clear due to differences in the amount of expertise participants achieved with the talkers during the training phase. Thus, for this study, we used an easier training paradigm (2-AFC) so that participants would achieve more similar amounts of expertise in the *training phase* for both languages. The *test phase* in the current study kept the 4-AFC task used in [6].

2.1. Participants

We recruited two groups of adult participants from cities with different language environments. One group consisted of 16 monolingual-English adults from Montréal, Québec (Age range = 19 – 30), and the other consisted of 16 monolingual-English adults from Stoors, Connecticut (Age range = 19 – 31). Montréal is a bilingual city, where English and French co-exist in many aspects of life. In contrast, Storrs is an overwhelmingly predominant English-monolingual environment. Comparing these two types of groups allowed us to control for proficiency in their native language (English) while manipulating exposure to an unfamiliar language (French) – i.e., Montréal residents having regular exposure to French, and Storrs residents having no regular exposure to French. We used language questionnaires to verify that their exposure to French was as predicted by their residence.

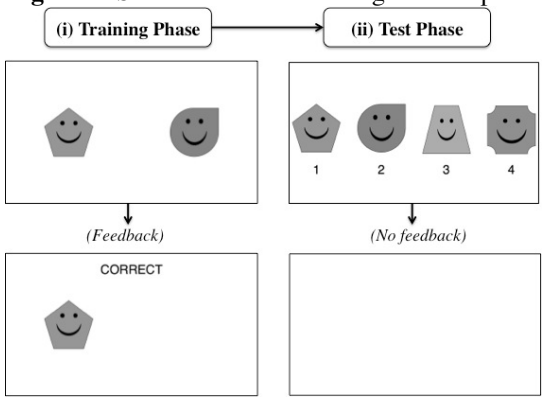
2.2. Stimuli

We used the same stimuli that were used in [6]. Four female native speakers of each language (English and French) were recorded reciting 10 sentences (5 *trained* and 5 *novel*). The speakers were depicted as cartoon avatars in the experiment. In total, we had 8 speakers (4 in each language), and 40 sentences (4 speakers x 10 sentences). Stimuli testing from [6] confirms that the stimuli were equally "easy" for native speakers of each language.

2.3. Procedure

All participants were tested on their ability to learn and identify talkers in both languages; each language was tested separately, with order of which language went first counter-balanced within each group. For each language, each participant completed a training phase followed by a test phase (Figure 1). Participants were tested in labs in their respective cities, and precautions were taken to ensure that testing procedures were identical across both labs.

Figure 1. Schematic of the training and test phase



(i) Training phase: The training phase was a 2-AFC task, which provided participants with an opportunity to learn the voice-avatar pairings. On each trial, auditory stimuli were simultaneously presented with the picture display (two avatars, one on each side of the screen). Participants were instructed to identity the voice by pressing one of two labelled buttons. Feedback was provided on each trial, with the word "Correct" or "Incorrect" appearing on the screen with the correct avatar. Each training block consisted of 60 randomized trials (4 talkers X 5 sentences X 3 repetitions). The training block was repeated until the participant met learning criterion, defined as 85% correct talker identification within a single block (52/60 trials).

(ii) Test phase: In the test phase, participants were tested on their knowledge of the voice-identity pairings through a 4-AFC. In this phase, all four avatars appeared on the screen, while one sentence is played from one of the speakers. Participants were then instructed to identity the voice by pressing one of four buttons in the button box. The test phase consisted of 120 randomized trials (4 talkers X 10 sentences (5 trained and 5 novel) X 3 repetitions). No feedback was provided in this phase.

3. RESULTS

3.1. Learning Rate

First, we assessed learning rate by measuring the number of blocks required to reach learning criterion for each language. An ANOVA with the within-subjects factor of stimulus language (English vs. French) and the between-subjects of location (Montréal vs. Storrs) showed a main effect of stimulus language [$F(1,30) = 26.69$, $p < .001$, $\eta_p^2 = 0.47$], with fewer training blocks required to reach criterion for English [$M = 1.22$, $SE = 0.07$] compared to French [$M = 3.53$, $SE = 0.42$]. However, there was no main effect of location [$F(1,30) = 0.02$, $p > .250$, $\eta_p^2 = 0.001$], nor an interaction between stimulus language and location [$F(1,30) = 0.08$, $p > .250$, $\eta_p^2 = 0.003$]. Thus, in contrast to [6], we were unable to find a language exposure benefit for unfamiliar languages in learning rate.

3.2 Test Phase Accuracy

In the test phase, we assessed participants' ability to retain memory of the voices they just learned, and generalize their learning of voices to novel sentences. An ANOVA with the within-subjects factors of stimulus language and sentence type, and the between-subjects factor of location showed a main effect of stimulus language [$F(1,30) = 85.29$, $p < .001$, $\eta_p^2 = 0.74$], with higher accuracy performance for English talkers [$M = 87.60\%$, $SE = 0.85$] compared to French talkers [$M = 72.66\%$, $SE = 1.92$].

Figure 2: Mean % correct for trained and novel sentences, separated by English and French talkers (collapsed across groups). Error bars indicate standard error of the mean.

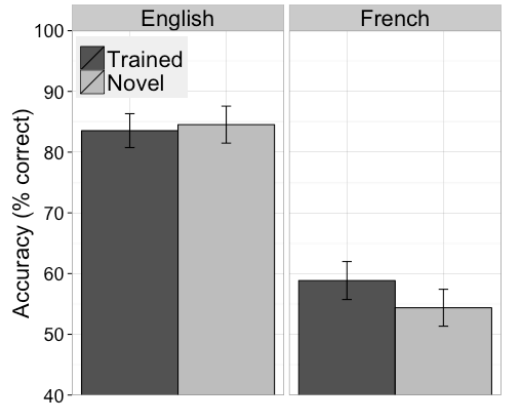
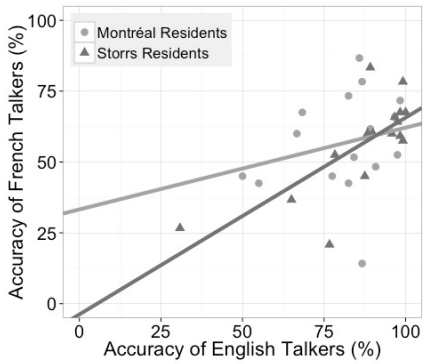


Figure 2 shows the mean percent correct for the trained and novel sentences, separated by English and French talkers. There was a main effect of sentence type [$F(1,30) = 4.56$, $p = .041$, $\eta_p^2 = 0.13$], and an interaction between stimulus language and sentence type [$F(1,30) = 11.44$, $p = .002$, $\eta_p^2 = 0.28$]. Post-hoc tests showed that participants did not differ in their performance on trained and novel English sentences [$t(31) = -.99$, $p > .250$, $d = -0.06$], but instead differed in their performance in French [$t(31) = 3.54$, $p < .001$, $d = 0.26$], with participants having higher accuracy for trained compared to novel sentences in French. These results replicate the findings in [6] that show that listeners have a more stable representation of voices for talkers speaking their native language than talkers speaking an unfamiliar language. However, no other main effects nor interactions were significant (all *p*s > .33), including those with location as a between-subjects factor.

3.3 Individual Differences

First, we asked whether there was an underlying talent for learning and identifying voices, regardless of language. If so, there should be a correlation between participants' talker identification accuracy for English and French talkers. Indeed, we find that participants' percentage accuracy for the English voices and the French voices in the test phase are correlated [$r = .49$, $n = 32$, $p = .004$]. When separated by location, we found that the correlation held only for the Storrs group [$r = .73$, $p < .001$], but not for the Montréal group [$r = .23$, $p = .38$]. These findings suggest that a language-general "talent" for talker-learning can be found for unfamiliar languages, but sensitivity to the surface phonetic properties of an unfamiliar language drives the language exposure benefit.

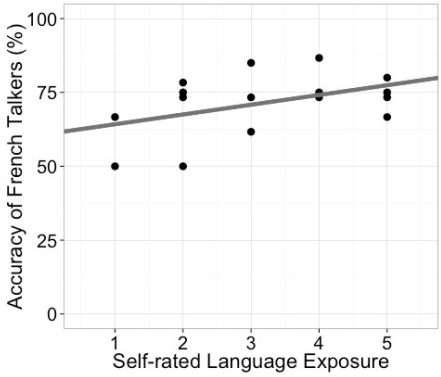
Figure 3: Relationship between accuracy of identifying English talkers and French talkers in the test phase. Light circle points represent Montréal group, while dark triangle points represent Storrs group.



Since a language-general mechanism could not explain the variability in French talker-learning performance for Montréal residents, we conducted

exploratory analysis to assess whether individual differences in language exposure could predict French talker-learning performance. We examined the relationship between participants' self-rated language exposure in French and their talker-learning performance for French talkers. Participants were asked to rate their exposure to French from two activities (watching TV/films, reading print media) on a 1-5 likert scale. Figure 4 shows a marginal correlation between participants self-rated exposure to TV/film and their percent correct on the first French training block of the training phase [r(16)= .45, p .08]. Note, however that this correlation does not hold for their percent correct on the test phase [r(16) = .25, p = .36]. There was no correlation between participants' self-rated exposure to print media and their performance on French in either the training or test phase (both *p*s > .250).

Figure 4: Relationship between exposure to French TV/film and accuracy of identifying French talkers in the training phase. Each point represents a Montréal resident. For the self-rated measure, 1 = "Never", and 5 = "Daily".



4. DISCUSSION

These results support previous findings that show the *language familiarity effect* [3]: participants were faster at learning the voices of talkers who speak their native language (English) than talkers who speak an unfamiliar language (French). However, in contrast to [6], we did not find differences in talker learning of a foreign language between groups with and without exposure to that language. A potential explanation for our results is that the present training paradigm was not an ideal context for showing talker-learning differences between groups that differ only in language exposure. In [6], participants were trained to recognize voices in a 4-AFC task, while the present study used a 2-AFC task. Previous research show that learning rates for 2-AFC tasks are faster than 4-AFC [2]; thus, due to the easier nature of the task, the representation of the learning may not be as strong [1]. It may be case that varying the training paradigm in laboratory analogs of talker identification can induce different types of talker learning. Thus, the combined data from [6] and the present study suggests that phonological sensitivity to an unfamiliar language is sufficient in boosting talker-learning skills, but that this effect is observable only under challenging learning situations. In other words, the language exposure benefit arises only when the talker learning system is "pushed". Indeed, comparisons across data from [6] and the present study show significantly higher accuracy for French talkers in the test phase in [6] compared to the present study [t(69) = 3.23, p <.002], confirming that the 4-AFC task implemented in the training phase in [6] had a stronger impact on talker learning compared to the 2-AFC task implemented here.

Though group level patterns did not emerge in terms of number of training blocks to hit criterion or accuracy during test, the influence of location emerged when considering more fine-grained measures, including individual differences within groups. For Storrs residents, talker learning performance for English talkers was correlated with that for French talkers; however, this correlation is absent for Montréal residents. This suggests that when surface phonetic properties of a language are highly novel, listeners rely on their own individual talker-learning talent to detect language-general sound properties of a talker. Consistent with [2], our results show that once listeners are sensitized to the phonetic properties of the unfamiliar language, their talker-learning abilities are more strongly influenced by their representation of these non-native sounds, rather than language-general talent.

Finally, we explored individual differences in language exposure. Potentially due to limited power in our sample size, we did not find significant correlations between self-rated language exposure measures and talker learning abilities. Our results do, however, suggest a marginal relationship between participants' exposure to French audio-visual media (such as TV/film) and their French talker-learning abilities. Nevertheless, difficulty in finding significant correlations for self-reported measures suggests that more precise measures of language exposure may be necessary to find concrete individual differences in talker learning. For example, future research could manipulate exposure levels directly or make use of speech perception tests to examine participants' phonetic sensitivity to the target foreign language. These exploratory analyses, combined with our behavioural data, suggest that this is a promising area for further investigation.

5. REFERENCES

[1] Ahissar, M., Hochstein, S. (1997). Task difficulty and the specificity of perceptual learning. *Nature*, 387(6631), 401-406.

[2] Bregman, M. R., Creel, S. C. (2014). Gradient language dominance affects talker learning. *Cognition*, 130(1), 85-95.

[3] Goggin, J. P., Thompson, C. P., Strube, G., Simental, L. R. (1991). The role of language familiarity in voice identification. *Memory & Cognition*, 19(5), 448-458.

[4] Johnson, E. K., Westrek, E., Nazzi, T., Cutler, A. (2011). Infant ability to tell voices apart rests on language experience. *Developmental Science*, 14(5), 1002-1011.

[5] Köster, O., Hess, M. M., Schiller, N. O., Kunzel, H. J. (1998). The correlation between auditory speech sensitivity and speaker recognition ability. *Forensic Linguistics*, 5, 22-32.

[6] Orena, A.J., Theodore, R.M., Polka, L. Exposure is all it takes to improve talker identification in a foreign language. Submitted.

[7] Perrachione, T. K., Del Tufo, S. N., Gabrieli, J. D. E. (2011). Human voice recognition depends on language ability. *Science*, 333(6042), 595.

[8] Vanags, T., Carroll, M., Perfect, T. J. (2005). Verbal overshadowing: a sound theory in voice recognition? *Applied Cognitive Psychology*, 19, 1127-1144

[9] Winters, S. J., Levi, S. V., Pisoni, D. B. (2008). Identification and discrimination of bilingual talkers across languages. *Journal of the Acoustical Society of America,* 123(6), 4524-4538.

[10] Xu, R. B., Zuo, D., Mok, P. (2012). Speaker Identity of Cantonese-English Bilingual Speakers. *14th Australasian International Conference on Speech Science and Technology*, 14, 61-64