Effects of perceptual training on Cantonese productions of English plosives

Ana N. Rinzler[1]

[1] Rutgers University, the State University of New Jersey

Author Note

Cognitive Psychology

The Center for Cognitive Science

I would like to thank Dr. Terry Kit-fong Au from the University of Hong Kong for permitting us to conduct acoustic analyses on production data from her research.

Correspondence concerning this article should be addressed to Ana N. Rinzler, 1498 State Rte. 28 West Hurley, NY 12491. E-mail: anb136@psych.rutgers.edu

Effects of perceptual training on Cantonese productions of English plosives

Abstract

The Critical Period Hypothesis (Lenneberg, 1976) posits that biological predispositions (i.e. the loss of neural plasticity with age) is the primary reason for discrepancies in language acquisition between younger and older learners. Alternatively, it is possible that environmental factors, such as the quality and quantity of input can influence a speaker's ability to acquire a second language. This research examines the latter hypothesis posited by Terry Au (ms) by conducting acoustic analysis via PRAAT on production data collected from Cantonese speaking adults who were trained on notoriously difficult English contrasts. Specifically, we measured the vowel duration of English words with voiced /b, d, g/ and voiceless plosives /p, t, k/ in coda position of phonological minimal pairs (i.e., bag and back). A general linear mixed effects model revealed no effect of training; such that trained adult Cantonese speakers did not differ significantly from untrained adult Cantonese speakers. However, there was a significant difference between the vowel durations of voiced and voiceless plosives, and an interaction between the training factor (i.e. whether participants received training or not), and the voicing factor (whether the plosive was voiced or voiceless). This suggests a interdependence between these factors, but further analysis is needed to understand how these factors interact.

Introduction

The Critical Period Hypothesis (CPH) posited by Lenneberg (1967), states that language acquisition increases in difficulty with age due to a loss of neural plasticity. This hypothesis has much empirical support, both for first and second language (L2) acquisition. Additionally, there are observational factors that demonstrate a potential discrepancy in acquisition proficiency. For instance, children of immigrant families often acquire more native-like sounding productions, which is equated to greater L2 proficiency. A recent hypothesis posted by Au (ms) suggests that biological factors are not the only influential factor for L2 acquisition. Au (ms) posits that the quality and quantity of L2 input can influence a speaker's ability to acquire that language. For instance, children often receive more input than adults due to greater cultural immersion (e.g. engaging with peers and native English speaking teachers at school); while adults, often have less of these interactive experiences. Consequently, younger speakers often develop a greater proficiency in the language than adult speakers.

To investigate this latter hypothesis, Au (ms) conducted a 4-6 week intensive training program that focused on improving Cantonese speakers ability to produce and perceive notoriously difficult English contrasts. Specifically, she trained Cantonese speakers on phonological minimal word pairs with English voiced and voiceless plosives (i.e. /b d g/ and /p t k) in onset, medial and coda position of the words. Au (ms) measured comprehension through a forced-choice task in which participants heard a word, and had to choose the word they thought they heard from two options. They were auditorily presented with one of two phonological minimal pair words and their comprehension was measured by the percent they got correct. In addition, Au (ms) compared the participants pre-performance (i.e. before training) to their post-performance (i.e. after training); as well as the comprehension performance of trained participants to untrained participants. The participants in her studied ranged from 9 yrs. to 22 yrs. in age, and were divided into trained and wait-list control conditions. Au's (ms) ANCOVA analysis revealed that training improved the comprehension ability of her adult group (18 - 22 yrs.) under certain conditions, even more so than the younger age group. As for how training influenced the production abilities of these speakers, Au (ms) had three native English speakers listen to post-productions of trained speakers and judge whether they thought the sound was native or non-native. Again, she found that adults out-performed children and adolescents.

This study aims to further examine the production data from Au's (ms) study through acoustic analysis of the post-productions of adults who were trained. Specifically, PRAAT was used to measure vowel duration of minimal pairs with voiced and voiceless plosives in coda position. Vowel duration is a robust acoustic cue to English production of plosives in coda position. Specifically, vowel duration is longer before English words with voiced plosives in coda position, than voiceless plosives in coda position. We hypothesize that training will affect the vowel duration such that trained participants will have longer vowel durations for voiced plosives than untrained participants. This finding would suggest that Cantonese speakers may utilize the cue of vowel duration in a similar manner to native English speakers. Further, this suggests that receiving good quality and a greater quantity of native input can influence an L2 speaker's ability to produce words with a more native-like quality than predicted by the Critical Period Hypothesis.

Methods

In this study, acoustic analyses were conducted on Cantonese speakers' productions of English phonological minimal word pairs with voiced (i.e. /b d g/) and voiceless (i.e. /p t k/) plosives in coda position. The production of the word "got" was excluded from these analyses as it was the only word that did not have a minimal pair. For all other productions, PRAAT was used to measure the duration of the vowel. Measuring vowel duration was motivated by evidence that vowel length is an acoustic cue that English speakers use when distinguishing between the following plosive as voiced or voiceless. For instance, the duration of the vowel preceding a voiced plosive is typically longer than the duration of a vowel preceding a voiceless plosive (House and Fairbanks, 1953; Peterson and Lehiste, 1960; House, 1961; Umeda, 1975; Klatt, 1976). Please note that the production data analyzed in this study was collected and generously provided by Dr. Terry Kit-fong Au, from the University of Hong Kong.

Participants

There were a total of 36 undergraduate students from the University of Hong Kong. 18 of the participants were in the training group (33% men), and 18 of the participants were in a wait-list control group (28% men).

Material

The following analyses are based on productions of phonological minimal word pairs with voiced and voiceless plosives in coda position. The vowel duration from the following words with a voiced coda were analyzed:/bæd, bæg, kæb, kʌb , dɔg , fæd, fid, pɪg, tæb/. The following words with a voiceless coda were analyzed: / bæt, bæk, kæp, kʌp, dɑk, fæt, fit, pɪk, tæb/. Only post-training productions were examined in this study. For the wait-list control participants, these productions represent the second time that participants produced the minimal pairs. In other words, they did not receive training in between the first time and second time that they produced these words. Conversely, for trained participants, these productions represent the second time that they produced these minimal pairs after receiving training.

Procedure

Participants in Au's (ms) study completed a 4 - 6 week training program compromised of comprehending and producing English phonological minimal word pairs. The purpose of training was to improve Cantonese speaker's ability to perceive and produce notoriously difficult English contrasts. The data analyzed in this study are words in which the contrast occurs in coda position. The productions were then sent to our lab for acoustic analyses.

The software PRAAT was used to conduct acoustic analyses. Textgrids were created from the .wav sound files in order to mark the beginning and end of the vowel boundary. Utilizing Sennheiser HD 555 headphones, the beginning of the vowel was marked using the *wav* method and the end of the vowel was marked using the *F2* method. All boundaries were marked at the zero-crossing line. Measurements at present, were only taken by one researcher. Thus, future cross-validation through concordance rates is required. PRAAT scripting was then used to export vowel duration measurements.

Data analysis

All analyses were conducted using R (R Core Team, 2012) and *lme4* (Bates, Maechler & Bolker, 2012). See *Footnotes* for complete list of R packages used. Data from the production task were analyzed using a general linear mixed-effects model. The criterion variable was *vowel duration* which was convereted to milliseconds and normalized for speaker. There were two predictors which were fixed factors: (1) training *trained/untrained* and voicing (2) *voiced/voiceless*. Both factors were categorical and were sum coded. For the training variable, *trained* (i.e. participants who were trained) were assigned a 1, and *untrained* (i.e. participants who were not trained) were assigned a 0; while *voiced* (i.e. words with voiced plosives in coda position) were assigned a 1 and *voiceless* were assigned a 0 (i.e. words with voiceless plosives in coda position). Two new columns in the data frame were generated to represent the sum variables of the training and the voicing conditions. The variable participant was treated as a random effect as each participant had multiple productions. In other words, each participant produced all of the 36 voiced (18) and voiceless (18) words. Visual inspection of the Q-Q plots and plots of residuals against fitted values revealed no violations of normality or homoscedasticity. Lastly, statistical significance of voicing and training, and the voicing by training interaction was accessed using hierarchical portioning of variance via nested model comparisons, with the voicing variable entered first into all models. P-values were obtained using likelihood ratio tests comparing all models against the null model, and the alpha level was set a $p < .05$.

Results

The interaction model explained the most variance (marginal R^2 of .052; conditional R^2 of .305) and thus was used to interpret the data. The interaction model explained more variance than just voicing or training alone. The model with only voicing had a marginal R^2 of .045, conditional R^2 of .295; and the model with only training had a marginal R^2 of .003, and a conditional R^2 of .253. It is clear that training factor contributed the least to overall variance explained. The interaction model, however revealed that there was a main effect of *voicing* [χ^2 (1)=44.664, p <.001], but no main effect of *training* [χ^2 (1)=0.4302, p =0.5119]. There was however, an interaction between *voicing x training* [χ^2 (4)=40.342, p <.05]. The intercept mean was 177.98 (ms). The effect of *voicing* (t = 6.478, $p < .001$) on this intercept was 11.65 +/- resulting in a range of *voiced* plosives having a mean vowel duration of approximately 189.63 (ms) +/- 1.8 (se); while the mean vowel duration for *voiceless* plosives was approximately 166.34 (ms) +/- 1.8 (se). However, this effect was modulated by the interaction between *voicing* and *training* - although this effect just reached significance (t = 1.978, p = 0.048). When the interaction between *voicing* and *training* was considered, the vowel duration mean increased *voiced* plosive duration to 193.19 (ms) +/- 4.805 (se) and *voiceless* plosive duration to 162.78 (ms) +/- 4.805 (se). Visual inspection of the data in *Figure 1.* below suggests that vowel durations were longer for trained participants, however this can not be concluded from sum coding directly. Multiple comparisons of the means are required.

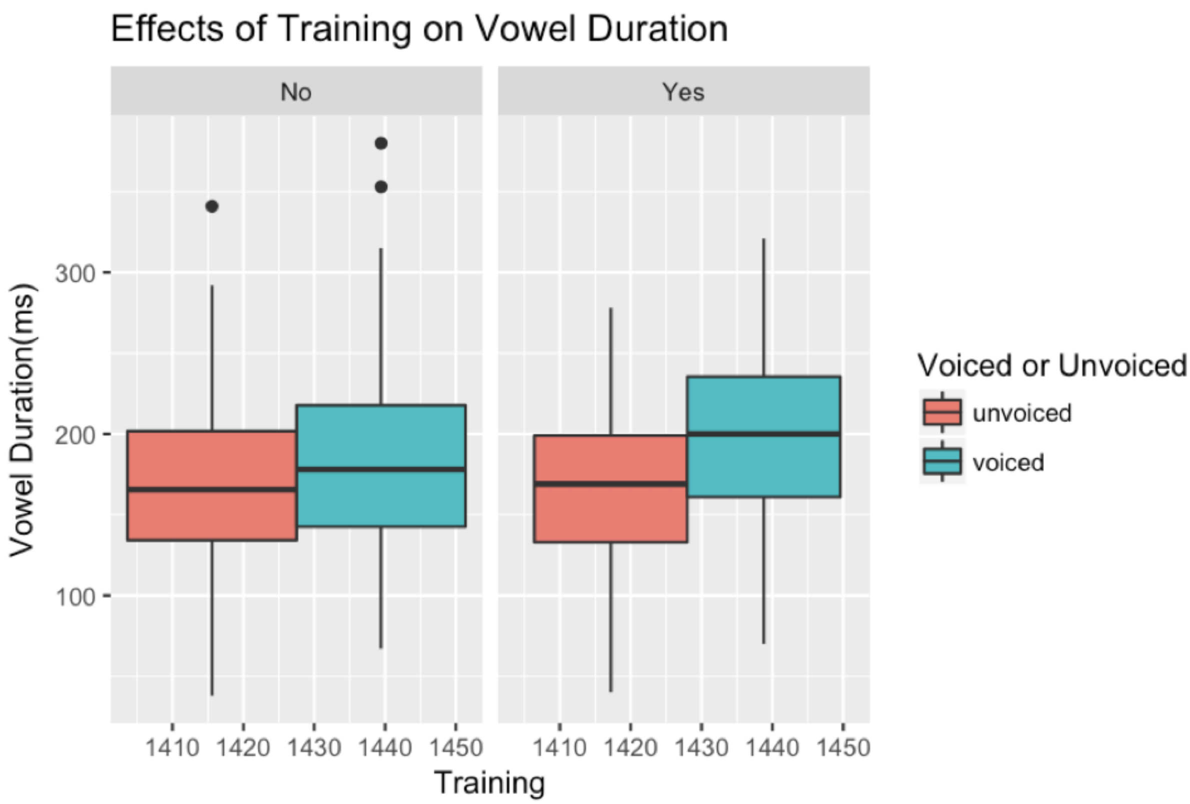

Figure 1. The above boxplot illustrates the differences in vowel duration for trained and untrained participants for words with voiced and voiceless plosives in coda position.

Discussion

Overall, the results suggest that participants produced words with voiced plosives longer than words with voiceless plosives, and that this effect was modulated by training. This suggests that vowel duration is an acoustic cue that Cantonese speakers use when making the distinction in English between words with voiced and voiceless plosives in coda position. However, there was no main effect of training, so trained participants did not differ significantly from untrained participants in their productions of vowel duration.

Interestingly, there was an interaction between training and voicing. This suggests that the factor *training* may influence vowel duration; at present it is unclear as to how this factor plays a role. Graphical representations of the data suggest that trained and untrained participants had about equal vowel lengths for voiceless plosives, and that trained participants had longer vowel durations for voiced plosives, than untrained participants. This is in the direction that I would predict, as trained participants received greater exposure to instances of English productions.

The question then becomes: what is the nature of this exposure that participants received? In other words, did the native English productions contain vowels that had canonical and predictable vowel durations, or was there extensive variation in the input they received? It would not be surprising if participants received varying vowel durations for two main reasons: (1) not all of the vowels in the minimal pairs were identical (i.e. dɔg and dɑk), and furthermore some vowels are typically always longer in duration than others (i.e. fæd), and (2) the participants received input from four different native English speakers. There is a debate as to whether presenting listeners with input from multiple speakers has a positive or negative effect on learnability. Regardless, it seems likely that greater perceptual variability may obscure vowel duration as a cue, and consequently diminish the probability of this utilizing this cue.

In future research, I plan to record monolingual English speakers producing the minimal pairs from Au's (ms) study to measure their vowel duration. I then plan to compare these durations to the Cantonese speakers' productions in order to get a better idea of what native-like vowel duration is. Additionally, I aim to explore other potential acoustic cues that may be relevant for Cantonese speakers when making these contrasts. For instance, I plan to measure aspiration duration, mean aspiration intensity, F2 at the end of the vowel, as well as closure duration. Aspiration is a phonological feature of Cantonese and thus I predict it will play a role. Further, Cantonese is a tonal language, and therefore F2 at the end of the vowel (a predictor of the following plosive in English) may also be of interest.

References

Aust, F., & Barth, M. (2018). *papaja: Create APA manuscripts with R Markdown*. Retrieved from https://github.com/crsh/papaja

Bartoń, K. (2018). *MuMIn: Multi-model inference*. Retrieved from https://CRAN.R-project.org/package=MuMIn

Bates, D., & Maechler, M. (2018). *Matrix: Sparse and dense matrix classes and methods*. Retrieved from https://CRAN.R-project.org/package=Matrix

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. doi:10.18637/jss.v067.i01

Henry, L., & Wickham, H. (2017). *Purrr: Functional programming tools*. Retrieved from https://CRAN.R-project.org/package=purrr

Højsgaard, S., & Halekoh, U. (2018). *DoBy: Groupwise statistics, lsmeans, linear contrasts, utilities*. Retrieved from https://CRAN.R-project.org/package=doBy

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13), 1–26. doi:10.18637/jss.v082.i13

Murphy, L. (2015). *Likelihood: Methods for maximum likelihood estimation*. Retrieved from https://CRAN.R-project.org/package=likelihood

Müller, K. (2018). *Bindrcpp: An 'rcpp' interface to active bindings*. Retrieved from https://CRAN.R-project.org/package=bindrcpp

Müller, K., & Wickham, H. (2018). *Tibble: Simple data frames*. Retrieved from https://CRAN.R-project.org/package=tibble

Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., & R Core Team. (2018). *nlme: Linear and nonlinear mixed effects models*. Retrieved from https://CRAN.R-project.org/package=nlme

R Core Team. (2017). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

Robinson, D. (2018). *Broom: Convert statistical analysis objects into tidy data frames.* Retrieved from https://CRAN.R-project.org/package=broom

Tang, Y., Horikoshi, M., & Li, W. (2016). Ggfortify: Unified interface to visualize statistical result of popular r packages. *The R Journal*, *8*(2). Retrieved from https://journal.r-project.org/

Wickham, H. (2009). *Ggplot2: Elegant graphics for data analysis.* Springer-Verlag New York. Retrieved from http://ggplot2.org

Wickham, H. (2017). *Tidyverse: Easily install and load the 'tidyverse'.* Retrieved from https://CRAN.R-project.org/package=tidyverse

Wickham, H. (2018a). *Forcats: Tools for working with categorical variables (factors).* Retrieved from https://CRAN.R-project.org/package=forcats

Wickham, H. (2018b). *Stringr: Simple, consistent wrappers for common string operations.* Retrieved from https://CRAN.R-project.org/package=stringr

Wickham, H., & Henry, L. (2018). *Tidyr: Easily tidy data with 'spread()' and 'gather()' functions.* Retrieved from https://CRAN.R-project.org/package=tidyr

Wickham, H., Francois, R., Henry, L., & Müller, K. (2017). *Dplyr: A grammar of data manipulation.* Retrieved from https://CRAN.R-project.org/package=dplyr

Wickham, H., Hester, J., & Francois, R. (2017). *Readr: Read rectangular text data.* Retrieved from https://CRAN.R-project.org/package=readr

Xie, Y. (n.d.). *Xaringan: Presentation ninja.* Retrieved from https://github.com/yihui/xaringan

Zhu, H. (2018). *KableExtra: Construct complex table with 'kable' and pipe syntax.* Retrieved from https://CRAN.R-project.org/package=kableExtra

Footnotes

The following lists the specific packages used in R:

R (Version 3.4.3; R Core Team, 2017) and the R-packages *bindrcpp* (Version 0.2.2; Müller, 2018), *broom* (Version 0.4.4; Robinson, 2018), *doBy* (Version 4.6.1; Højsgaard & Halekoh, 2018), *dplyr* (Version 0.7.4; Wickham, Francois, Henry, & Müller, 2017), *forcats* (Version 0.3.0; Wickham, 2018a), *ggfortify* (Version 0.4.4; Tang, Horikoshi, & Li, 2016), *ggplot2* (Version 2.2.1; Wickham, 2009), *kableExtra* (Version 0.8.0; Zhu, 2018), *likelihood* (Version 1.7; Murphy, 2015), *lme4* (Version 1.1.17; Bates, Mächler, Bolker, & Walker, 2015), *lmerTest* (Version 3.0.1; Kuznetsova, Brockhoff, & Christensen, 2017), *Matrix* (Version 1.2.14; Bates & Maechler, 2018), *MuMIn* (Version 1.40.4; Bartoń, 2018), *nlme* (Version 3.1.137; Pinheiro, Bates, DebRoy, Sarkar, & R Core Team, 2018), *papaja* (Version 0.1.0.9709; Aust & Barth, 2018), *purrr* (Version 0.2.4; Henry & Wickham, 2017), *readr* (Version 1.1.1; Wickham, Hester, & Francois, 2017), *stringr* (Version 1.3.0; Wickham, 2018b), *tibble* (Version 1.4.2; Müller & Wickham, 2018), *tidyr* (Version 0.8.0; Wickham & Henry, 2018), *tidyverse* (Version 1.2.1; Wickham, 2017), and *xaringan* (Version 0.6.4; Xie, n.d.)