L2-Mandarin regional accent variability facilitates Mandarin-naïve English listeners' learning of Mandarin tone-words

Yanping Li[1], Michael D. Tyler[2], Denis Burnham[1], Catherine T. Best[1]

[1]The MARCS Institute, Western Sydney University, Australia [2]Independent researcher
yanping.li@westernsydney.edu.au, tylerspeechscience@gmail.com
denis.burnham@westernsydney.edu.au, c.best@westernsydney.edu.au

ABSTRACT

English learners have difficulties using tones in Mandarin words [1]. As high phonetic variability facilitates second language (L2) word learning [2], we investigated how L2-Mandarin accent variability in minimal-tone-contrast word training affects Mandarin-naïve English participants' tone-word learning. 48 English learners completed 6 training sessions on 16 Mandarin pseudowords spoken by 12 Beijing talkers or four talkers each from Beijing, Yantai, and Guangzhou. Participants took word identification tests after each session, and generalisation tests to novel talkers with familiar and unfamiliar accents after session 6. Growth curve analysis on word identification accuracy and response times across training sessions revealed no significant differences between training conditions. However, linear mixed effects modelling of the generalization results showed that while both groups succeeded on novel talkers with familiar and unfamiliar accents, the multiple accent group identified words significantly faster than the Beijing accent group, indicating that L2-Mandarin accent variability facilitates English learners' tone-word learning.

Keywords: High variability word training; L2-Mandarin regional accents; tone-word identification, talker and accent generalization

1. INTRODUCTION

Lexical tones are used as phonological contrasts at the segmental level in tone languages [3] such as Mandarin Chinese. There are four Mandarin lexical tones, i.e., T1 with a high-level contour, T2 high-rising, T3 low-dipping, and T4 high-falling [4]. Imposing the four tones on the identical consonant-vowel [CV] syllable /*ma*/ yields four different words in Mandarin Chinese: level = *mother*, rising = *hemp*, dipping = *horse*, and falling = *curse*. Non-tone language listeners (e.g., English and French) are able to rely on their native intonation patterns to discriminate dissimilar tones, e.g., dipping vs. level, dipping vs. falling, and rising vs. falling [5, 6] and to improve categorization between similar tones, e.g., level vs. rising, level vs. falling, and rising vs. dipping [7]–[9]. However, perceiving Mandarin tone contours as intonation patterns does not help English listeners to overcome persistent difficulties with lexical tones in learning and using words, especially in larger utterances in communicative contexts [1, 10, 11].

Word training is a possible solution to establish L2 phonological representations and connect them to lexical meanings [e.g., 12]. Phonological constancy, which maintains word recognition across lexically irrelevant talker and accent variation [13], appears important to such training. Talker variability facilitates word learning in non-tone languages, presumably by providing greater variability in phonetic realizations of words, thereby yielding more robust L2 word representations. For example, native (L1) English learners trained on Spanish words with high talker variability had better accuracy and response time for identifying the newly learned words than those trained with low talker variability [2]. High talker variability word training has also been applied to non-tone language learners during tone language acquisition, to direct their attention to lexical tones [10, 14, 15]. For example, naïve English participants exposed to high talker variability in training on minimal-tone-contrast Mandarin words performed better on word identification than those exposed to low talker variability [14].

Studies with high talker variability have rarely if ever examined the effects of *accent* variability on minimal-tone-contrast word training; as talkers have nearly always been selected from the same region, usually an accepted standard variety of the target language. The present study examines accent variability effects on Mandarin tone word learning. Standard Mandarin (Mandarin, henceforth) developed historically from the dialect spoken in Beijing, China. Talker(s) from regions outside of Beijing acquire Mandarin as an L2, given that their L1 Chinese dialects are mutually unintelligible languages [16]. Thus, there is variability in L2-accented Mandarin tones, triggered by similarities and dissimilarities between their native dialect tone systems and Mandarin. For example, Yantai, Guangzhou, and Shanghai dialects' tone systems differ from each other and from Mandarin, resulting in L2-Mandarin tone pronunciations that deviate from native Mandarin in the slopes for rising (T2) and falling (T4) tones, and in the depth of the dip in dipping (T3) tone [17] (see also [18] for acoustic details on L2-accented Mandarin tones).

This study adapted the high talker variability training paradigm in [10] to manipulate accent variability during naïve English participants' learning of minimal-tone-contrast Mandarin words, by using stimuli produced by multiple talkers from either (i) Beijing (single accent condition: talker-only variability) or (ii) Beijing, Yantai and Guangzhou (multiple accent condition: talker and accent variability). Learners in both training conditions should be able to learn words and to generalize recognition of the newly learned words to novel talkers, given the high talker variability in both training conditions. In addition, the same words spoken in an unfamiliar accent (Shanghai) should be more difficult for both conditions to identify than those with the familiar (trained) Beijing accent, resulting in slower response times to the unfamiliar accent talker than to the familiar Beijing accent talker. However, we expected the multiple accent training condition learners to generalize words to novel talkers in both accents faster and/or more accurately than those in the single accent condition.

2. EXPERIMENT

2.1. Method

2.1.1. Participants

Mandarin-naïve Australian English participants (n = 48) were recruited online (see [19] for details), and randomly assigned to the single (n = 24, M_{age} = 24.5 years, SD = 5.8, 14 females) vs. multiple accent (n = 24, M_{age} = 25.5 years, SD = 5.1, 15 females) training conditions.

2.1.2. Stimuli

There were 16 Mandarin pseudowords based on four CV syllables (/ba/, /di/, /du/, /gu/). Imposing the four tones on each syllable yielded 16 Mandarin real words. But artificial meanings were assigned for this study, using 16 frequent English words from [20], which were represented with 16 grey-scaled drawings selected from the Multilingual Picture database [21].

Training stimuli were produced by female talkers, either 12 from Beijing (single accent condition) or four each from Beijing, Yantai, and Guangzhou (multiple accent condition). Both training groups thus heard 12 talkers (i.e., high talker variability) in total. There were four tokens for each word, resulting in 768 stimuli (12 talkers × 16 words × 4 tokens) in each training condition. Generalization stimuli (n = 128, 2 talkers × 16 words × 4 tokens) were the newly trained words but produced by a novel female talker (19.0 years old) with a familiar Beijing accent, and by a novel female talker (24.0 years old) with an unfamiliar Shanghai accent.

2.1.3. Procedure

This study was run remotely in a quiet room with E-Prime Go. The experimenter conducted it via ZOOM meetings with the participants to ensure data quality. Participants in both training conditions completed six training sessions on the 16 words in six consecutive days with a picture-to-word paradigm [10], followed immediately by two generalization tests.

There were four blocked talkers of the same accent in each training session (45 min), yielding three sets of 12 talkers, which were counterbalanced across participants; those three sessions were repeated a second time for a total of six sessions. Correspondingly, the four talkers each of the Beijing, Yantai, and Guangzhou Mandarin accents in the multiple accent condition were blocked by session. Accent order of the sessions was counterbalanced across participants.

There were 256 randomized trials (4 talkers × 4 syllables × 4 tones × 4 tokens) in each session with 16 words blocked by talker (n = 4) to optimise word learning [22]. For the same reason, talker blocks were further subdivided by syllable (i.e., by minimal-tone-contrast word set), yielding 16 randomized trials per word set (4 words × 4 tokens). Talker orders and syllable orders within talker were randomized across participants. Immediately after training on words for a given syllable by talker, participants completed a quiz with feedback on the four words in that syllable set, to optimize the learning on the just-trained words [23]. This word training + quiz cycle was repeated for the other three syllables produced by the same talker for all 16 target words by that talker. Participants then learned the 16 words again produced by each of the other three talkers with the same cycle.

On each word training trial, a drawing was displayed, and the paired word was played out; participants were instructed to remember as many audio-visual pairs as possible. After training on the four minimal-tone-contrast words of a syllable set by a given talker, they completed the corresponding 16-trial quiz. In each quiz trial, they heard one target word and had to choose the correct item in the display of 4 drawings. Correct answers were followed by a green tick on the screen, incorrect answers with a red cross mark. If the participant gave no response within 5.5 s, a reminder appeared asking them to respond more quickly. Following wrong or no answers, the correct drawing and audio word were presented as feedback.

To track learning progress across the six sessions, participants completed word identification tests at the end of each session. Each post-session test had 64 randomized trials (16 words × 4 trained talkers × 1 token; token per talker per word was counterbalanced across participants), and the procedure was identical to the quizzes except that no feedback was given and

there was no time-out, i.e., participants could take as long as needed to respond. Generalization tests to novel talkers with a familiar Beijing and an unfamiliar Shanghai accent were run, in that order, with the same procedure, after session 6. They included 128 randomized trials (2 tests × 16 words × 4 tokens), blocked by talker, and took ~ 7 min total to complete.

2.2. Results

2.2.1. Identification tests across training sessions

Indexed by their identification tests in each training session, English learners in both training conditions improved more than 30% from the first (M_{single} = 44.54%, SD = 22.50; $M_{multiple}$ = 55.33%, SD = 23.33) to the sixth session (M_{single} = 85.75%, SD = 13.52; $M_{multiple}$ = 86.96%, SD = 9.36), with greater improvement during the first three sessions than the last three, in which they asymptoted at 75-80% accuracy.

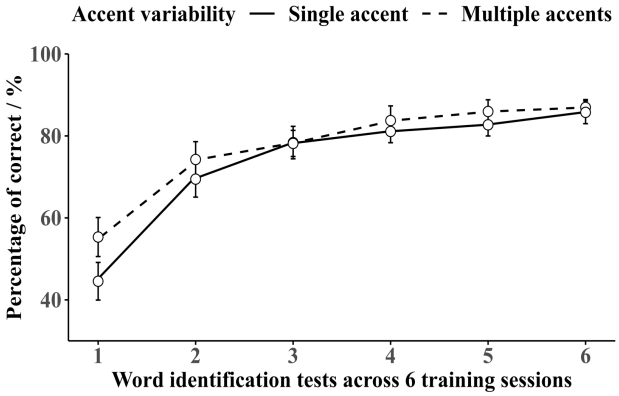

Figure 1: Mean percent correct word identifications across training session tests in single and multiple accent conditions. Error bars = ± 1 standard error of the mean.

Response times (RTs) for correct responses are shown in Figure 2. Values beyond ± 2.5 standard deviations from the mean for each participant were excluded – 449 occurrences, 3.22% of trials across participants, as recommended by [24]. The final RT data (n = 13475) were log transformed for data visualization and statistical analyses. Higher log values indicate slower responses. RTs in both training conditions decreased more sharply in the first three training sessions than in the last three (asymptote ~7.8).

Both accuracy and log-transformed RTs across the six training sessions were subjected to growth curve analysis [25], which captures changes in longitudinal data. Given that the curves in Figures 1 and 2 each had a single inflection, growth curve data were modelled using the *lmer* function from package *lmerTest* in *R* version 4.2.1 [26] including up to second-order orthogonal polynomials: intercept (mean) and linear (+/- slope) and quadratic terms (degree of inflection, i. e., the depth of peak or valley). Two mixed-effects models (for accuracy and RTs) were built with training condition as a fixed effect, and participants as a random effect. While learners in the multiple accent training condition appear somewhat more accurate and faster than those in the single accent condition (see Figures 1 & 2), neither groups' accuracy nor RT differed significantly in mean, linear, or quadratic trends, suggesting that talker variability during training is already optimally effective, with neither a deficit nor further benefit due to accent variability.

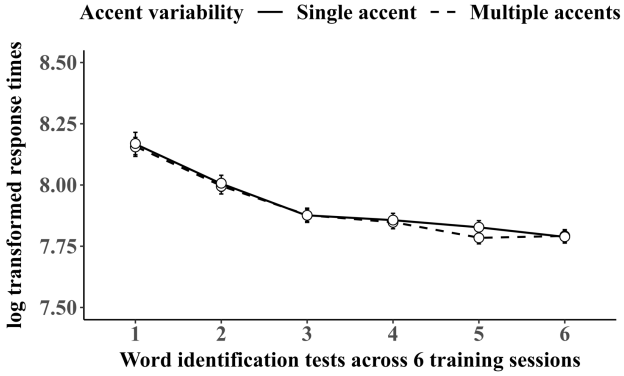

Figure 2: Log transformed RTs for correctly identified words across training sessions in single and multiple accent conditions. Error bars = ± 1 standard error of the mean.

2.2.2. Generalization tests

Both training groups also retained accuracy in generalization tests to novel talkers with familiar Beijing (M_{single} = 85.13%, SD = 13.57; $M_{multiple}$ = 91.17%, SD = 7.21) and unfamiliar Shanghai accents (M_{single} = 79.58%, SD = 14.60; $M_{multiple}$ = 86.04%, SD = 10.33). However, the multiple accent group slightly outperformed the single accent group on the generalization tests (see Figure 3).

One linear mixed-effects model was built on each participant's accuracy (%) with training conditions and generalization tests as fixed effects, and participants as random effects, using the *lmer* function from package *lme4*. We used the Kenward-Roger degrees of freedom approximation to calculate p values for the fixed-effects factors and the *Anova* function from package *car* to calculate F. Pairwise comparisons were conducted with *lsmeans* in *R* when necessary.

The main effect of generalization tests was significant, $F(1, 47)$ = 21.87, $p < .001$, and that for training conditions was marginally significant, $F(1, 46)$ = 3.81, p = 0.05. No interactions were significant; pairwise comparisons showed that participants in both training conditions correctly identified more words with the familiar Beijing accent than those with the unfamiliar Shanghai accent, i.e., Single: Estimate = 5.54, SE = 1.63, $t(46)$ = 3.40, p = 0.007; Multiple: Estimate = 5.13, SE = 1.63, $t(46)$ = 3.15, p = 0.01.

Figure 4 displays log transformed RTs for correct

identifications in the two generalization tests by participants from the single (familiar: M = 7.75, SD = 0.46, vs. unfamiliar: M = 7.75, SD = 0.47) and multiple accent (familiar: M = 7.70, SD = 0.44, vs. unfamiliar: M = 7.70, SD = 0.47) conditions.

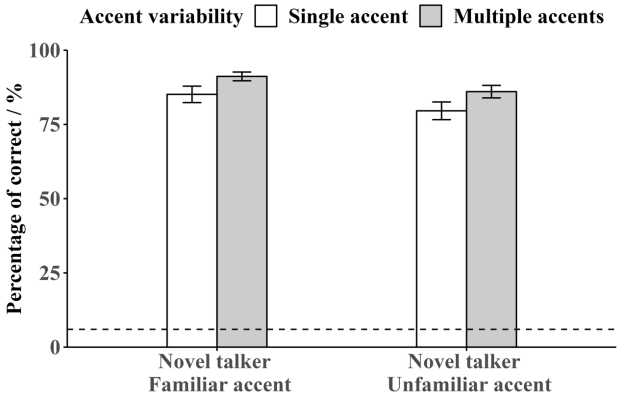

Figure 3: Mean percent correct responses in word generalization tests to novel talkers of familiar Beijing or unfamiliar Shanghai accents. Error bars = ± 1 standard error of the mean. Dashed line = chance level (1/16 × 100 = 6%).

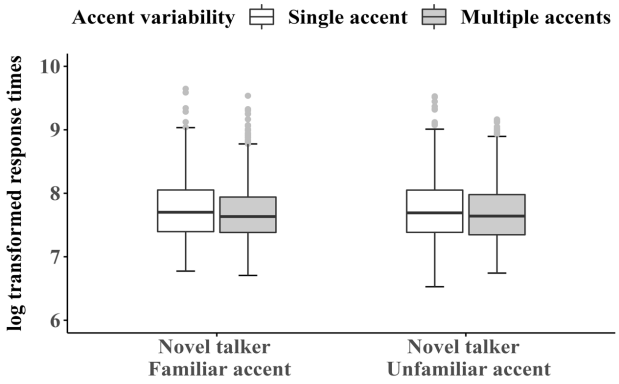

Figure 4: Log transformed RTs for correctly identified words in generalization tests for familiar and unfamiliar accents. Error bars = 95% confidence interval, top and bottom of the box are 25% and 75% percentiles; line inside box is 50% percentile (median). Grey points are outliers.

Another linear mixed-effects model was built for RTs with training conditions and generalization tests as fixed-effects, and participants and target words as random effects. Calculation of p and F values was the same as in the accuracy model. While there was no significant main effect of generalization tests, the main effect of training conditions, $F(1, 5046) = 14.29$, $p < .001$, and the interaction of training conditions × generalization tests, $F(3, 5042) = 4.82$, $p = 0.002$, were both significant. To tease the interaction apart, pairwise comparisons were conducted to assess differences in RTs between the two training conditions for each generalization test. While there were no significant differences between the two generalization tests in each training condition, the participants in the multiple accent condition identified the words produced by novel talkers with Beijing, Estimate = -0.05, SE = 0.02, $t(5042)$ = -2.73, p = 0.03, and Shanghai accents, Estimate = -0.05, SE = 0.02, $t(5043)$ = -2.63, p = 0.04, significantly faster than those in the single accent condition.

3. DISCUSSION

This study investigated effects of L2-Mandarin accent variability on naïve English participants' learning of Mandarin minimal-tone-contrast pseudowords. The single and multiple accent condition learners were comparable in identifying tone-words during training. However, participants in the multiple accent condition identified words in the two generalization tests more quickly than those in the single accent training condition, indicating that accent variability during training on minimal-tone-contrast words established more robust phonological representations of the four Mandarin tones, which facilitated their generalization of the learned words to novel talkers and a novel accent.

For each training condition, word identification accuracy in the generalization tests was significantly lower for the novel talker with the unfamiliar Shanghai accent than for the novel talker with the familiar Beijing accent, suggesting that both groups experienced difficulties with the unfamiliar accent. That inference would thus predict that RTs for words with the unfamiliar accent should be greater than for a novel talker with a familiar accent. However, RTs did not differ in either training condition for identification of words produced by novel talkers with familiar versus unfamiliar accents. Given that identifying words spoken by a novel talker with an unfamiliar accent was less accurate than for words with a familiar accent, it is possible that the learners spent more time identifying words spoken by the novel talker with the familiar accent, thereby achieving higher accuracy at a cost of RTs for that accent. This suggests that achieving phonological constancy across talker variability, and perhaps especially across accent variability, may require additional cognitive effort, i.e., pose additional cognitive load, in tone-word identification (e.g., [27]).

4. ACKNOWLEDGEMENT

Supported by a Western Sydney University (WSU) – China Scholarship Council (CSC) joint scholarship, Candidature Research Funds from the MARCS Institute, WSU, and partly by the Australian Linguistic Society Research Grants, which were all awarded to the first author. We are grateful for online recruitment assistance of Careers at Australian universities, technical support from Johnson Chen and Chris Wang at MARCS, and above all to the participants.

5. REFERENCES

[1] E. Pelzl, E. F. Lau, T. Guo, and R. DeKeyser, "Even in the best-case scenario L2 learners have persistent difficulty perceiving and utilizing tones in Mandarin: Findings from behavioural and event-related potentials experiments," *Stud. Second Lang. Acquis.*, vol. 43, no. 2, pp. 268–296, 2021, doi: 10.1017/S027226312000039X.

[2] J. Barcroft and M. S. Sommers, "Effects of acoustic variability on second language vocabulary learning," *Stud. Second Lang. Acquis.*, vol. 27, no. 03, 2005, doi: 10.1017/S0272263105050175.

[3] M. Yip, *Tone*. Cambridge: Cambridge University Press, 2002.

[4] Chao Y.-R., "A system of tone letters," *Maître Phon.*, vol. 45, pp. 24–27, 1930.

[5] C. K. So and C. T. Best, "Categorizing Mandarin tones into listeners' native prosodic categories: The role of phonetic properties," *Poznan Stud. Contemp. Linguist.*, vol. 47, no. 1, pp. 133–145, 2011, doi: 10.2478/psicl-2011-0011.

[6] P. A. Hallé, Y.-C. Chang, and C. T. Best, "Identification and discrimination of Mandarin Chinese tones by Mandarin Chinese vs. French listeners," *J. Phon.*, vol. 32, pp. 395–421, 2004, doi: 10.1016/S0095-4470(03)00016-0.

[7] S. Wiener, K. Ito, and S. R. Speer, "Effects of multitalker input and instructional method on the dimension-based statistical learning of syllable-tone combinatons," *Stud. Second Lang. Acquis.*, vol. 43, no. 1, pp. 155–180, 2021, doi: 10.1017/S0272263120000418.

[8] Y. Wang, M. M. Spence, A. Jongman, and J. A. Sereno, "Training American listeners to perceive Mandarin tones," *J. Acoust. Soc. Am.*, vol. 106, no. 6, pp. 3649–3658, 1999, doi: 10.1121/1.428217.

[9] G. Shen and K. Froud, "Categorical perception of lexical tones by English learners of Mandarin Chinese," *J. Acoust. Soc. Am.*, vol. 140, no. 6, pp. 4396–4403, 2016, doi: 10.1121/1.4971765.

[10] P. C. M. Wong and T. K. Perrachione, "Learning pitch patterns in lexical identification by native English-speaking adults," *Appl. Psycholinguist.*, vol. 28, no. 4, pp. 565–585, 2007, doi: 10.1017/S0142716407070312.

[11] Jie Xi, Hongkai Xu, Ying Zhu, Linjun Zhang, Hua Shu, and Yang Zhang, "Categorical perception of Chinese lexical tones by late second language learners with high proficiency: Behavioral and electrophysiological measures," *J. Speech Lang. Hear. Res.*, vol. 64, no. 12, pp. 4695–4704, 2021, doi: 10.1044/2021_JSLHR-20-00210.

[12] D. J. Bolger, M. Balass, E. Landen, and C. A. Perfetti, "Context variation and definitions in learning the meanings of words: An instance-based learning approach," *Discourse Process.*, vol. 45, no. 2, pp. 122–159, 2008, doi: 10.1080/01638530701792826.

[13] C. T. Best, "Devil or angel in the details? : Perceiving phonetic variation as information about phonological structure," in *Phonetics-Phonology Interface: Representations and Methodologies*, J. Romero and M. Riera, Eds., 2015, pp. 3–31.

[14] M. Sadakata and J. M. McQueen, "Individual aptitude in Mandarin lexical tone perception predicts effectiveness of high-variability training," *Front. Psychol.*, vol. 5, no. NOV, pp. 1–15, 2014, doi: 10.3389/fpsyg.2014.01318.

[15] H. Dong, M. Clayards, H. Brown, and E. Wonnacott, "The effects of high versus low talker variability and individual aptitude on phonetic training of Mandarin lexical tones," *PeerJ*, vol. 7, p. e7191, 2019, doi: 10.7717/peerj.7191.

[16] C. N. Li and S. A. Thompson, *Mandarin Chinese: A Functional Reference Grammar*. University of California Press, 1989.

[17] Y. Li, C. T. Best, M. D. Tyler, and D. Burnham, "Regionally accented Mandarin lexical tones," *J. Acoust. Soc. Am.*, vol. 148, no. 4, pp. 2474–2475, 2020, doi: 10.1121/1.5146856.

[18] Y. Li, C. T. Best, M. D. Tyler, and D. Burnham, "Tone variations in regionally accented Mandarin," in *Interspeech 2020*, ISCA, 2020, pp. 4158–4162. doi: 10.21437/Interspeech.2020-1235.

[19] Y. Li, C. T. Best, M. D. Tyler, and D. Burnham, "L2-Mandarin regional accent variability during lexical tone word training facilitates naive English listeners' tone categorization and discrimination," in *Proceedings of SST 2022*, R. Billington, Ed., 2022, pp. 156–160.

[20] W. J. B. van Heuven, P. Mandera, E. Keuleers, and M. Brysbaert, "Subtlex-UK: A new and improved word frequency database for British English," *Q. J. Exp. Psychol.*, vol. 67, no. 6, pp. 1176–1190, 2014, doi: 10.1080/17470218.2013.850521.

[21] J. A. Duñabeitia *et al.*, "The multilingual picture database," *Sci. Data*, vol. 9, no. 1, Art. no. 1, 2022, doi: 10.1038/s41597-022-01552-7.

[22] 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.*, pp. 461–472, 2011, doi: 10.1121/1.3593366.

[23] M. A. Pyc and K. A. Rawson, "Why testing improves memory: mediator effectiveness hypothesis," *Science*, vol. 330, no. 6002, p. 335, 2010, doi: 10.1126/science.1191465.

[24] R. K. Chan and J. H. Leung, "Why are lexical tones difficult to learn?: Insights from the incidental learning of tone-segment connections," *Stud. Second Lang. Acquis.*, vol. 42, no. 1, pp. 33–59, 2020, doi: 10.1017/S0272263119000482.

[25] D. Mirman, *Growth Curve Analysis and Visualization Using R*. in The R Series. Boca Raton, Florida: CRC Press, 2014.

[26] R Core Team, "R: The R Project for Statistical Computing." 2022. [Online]. Available: https://www.r-project.org/

[27] T. Bent and R. Frush Holt, "The influence of talker and foreign-accent variability on spoken word identification," *J. Acoust. Soc. Am.*, vol. 133, no. 3, pp. 1677–1686, 2013, doi: 10.1121/1.4776212.