PERCEPTION, RECOGNITION, AND ENCODING OF CANTONESE SOUND CHANGE VARIANTS

Rachel Soo and Molly Babel

Department of Linguistics, The University of British Columbia, Canada
rachel.soo@ubc.ca, molly.babel@ubc.ca

ABSTRACT

In Cantonese, a sound change sees /n/ produced as [l] syllable-initially, creating [n]- and [l]-initial pronunciation variants for underlying /n/-initial words. We examine the perception, recognition, and encoding of these sound change variants through four online experiments utilizing Bayesian frameworks for equivalence interpretations. An immediate repetition priming paradigm with [l] targets (Exp. 1) demonstrates recognition equivalence between [n] and [l] forms, in spite of clear phonetic sensitivity to [n] and [l] in AX discrimination (Exp. 2a) and categorization tasks (Exp. 2b). A long distance repetition priming task (Exp. 3) suggests listeners encode both pronunciation variants by dually mapping them to a single lexical representation. Taken together, these findings contribute to our understanding of the [n]-[l] sound change and uniquely situates the study of phonetic variation, which has traditionally been studied through the lens of within-/cross-dialect pronunciation variants, in the context of diachronic sound change variants.

Keywords: Psycholinguistics, Speech perception, Social encoding, Cantonese, Sound change

1. INTRODUCTION

Listeners encounter phonetic variation in spoken language. Regular contextual exposure to pronunciation variants may provide listeners with the perceptual flexibility to recognize and encode multiple pronunciation variants [1, 2]. For example, [2] posit that Basque listeners are regularly exposed to Spanish-accented Basque, which allows for the dual-mapping of a "correct" laminar [t͡s̻] and "incorrect" palato-alveolar [t͡ʃ] pronunciation variant to a single Basque lexical item. Encoding strength may also vary as a function of variant typicality, or social weight, as canonical and socially prestigious dialect forms may be preferentially encoded [3, 4, 5, 6].

Outside of regional/foreign accent variation, pronunciation variants are also apparent in diachronic sound changes and this dual-mapping can also be realized for changes-in-progress. Much like within-/cross-dialect variation [1, 2, 3], particular pronunciation variants in a sound change may be specific to a demographic group [7]. Pronunciation variants may also be socially meaningful just as particular regional accents may be socially stigmatized or recognized as the standard [8]. This may be the case with a Cantonese sound change, where /n/ categorically becomes [l] syllable-initially [9, 10]. Words such as *nou5* 腦 "brain", historically pronounced with an initial /n/ (hereafter, historical variant) are pronounced with [l] (hereafter, innovative variant), making it homophonous with /l/-initial words such as *lou5* 老 "old". [n] pronunciations are considered "standard", associated with prestige and used in formal contexts [11], while [l] pronunciations are socially stigmatized. This sound change was initially observed in the first half of the 20th century [12], with early sociolinguistic work indicating that younger speakers were leading the sound change [11, 13, 14]. More recent phonetic work reveals that the sound change may be complete at the community-level [10, 15], with some hypercorrection to [n] pronunciations. We build on this work by examining the consequences of this sound change on the perception, recognition and encoding of [n] and [l].

2. GENERAL PROCEDURE

All experiments were delivered online in Gorilla [16]. Participants provided consent, completed a short listening task to establish that they were wearing adequate headphones [17], and listened to a short Cantonese story to familiarize them to the talker's voice. Then, they proceeded to the main task in each experiment. 39 Cantonese talkers completed Exp. 1; 16 were counterbalanced into Exp. 2a, and 23 into Exp. 2b. A separate group of 35 Cantonese talkers completed Exp. 3. All participants were recruited from student participant pools at Canadian and Hong Kong institutions, ranged in age from 18-35, acquired Cantonese before age 5, and spent > 10 years in a Cantonese-speaking family. This linguistic profile conspires to produce lifelong exposure to Cantonese, and given that the sound change has been around since the mid-20th century, regular exposure to [n] and [l] pronunciations as well. Participants completed the Bilingual Language Profile (BLP) [18], which quantifies relative dominance in English and Cantonese on a scale of +/- 218, respectively. Participants had a BLP dominance score below 50.

3. EXPERIMENT 1

Exp. 1 was an immediate repetition priming task. On each trial, participants were presented with two auditory items, a prime and a target, separated by a 500 ms inter-stimulus interval (ISI) and asked to make a lexical decision to the target. If the pronunciation affects word recognition, we predict that both [n]- and [l]-initial primes should facilitate the recognition of [l]-initial targets equivalently.

3.1. Materials

40 [l]-initial target words were used. Each [l]-initial target word (e.g., *lou5* 老 "old") was preceded by one of four different primes, representing different pair types. In identity pairs, the prime and target were identical words (though physically-distinct). In historical pairs, the prime was the historical and less frequent [n]-initial variant of the target (*nou5* 腦 "brain"). Critical pairs were matched with two control types. In rime controls, the prime and target shared rimes and differed in their initial consonants (*pou5* 抱 "embrace"). Unlike the historical trials, the prime did not share a consonant that is in a sound change with [l]. In unrelated controls, the prime and the target did not share any features (*caa4* 茶 "tea").

80 non-[l]-initial filler target words that were likewise preceded by different primes, representing different pair types, were used. This produced a 1:2 ratio of critical to filler items in the experiment, and a total of 120 real word targets. To balance the number of real word and nonword targets, 120 nonwords targets were created from gaps in Cantonese. Each of these prime-target pairs were counterbalanced across 4 lists. Each list contained a total of 240 trials, in which there were 120 real word targets (40 critical pairs and 80 filler pairs), and 120 nonword targets.

3.2. Analysis and results

Data wrangling for all experiments used the {tidyverse} package [19] in R [20]. For all

Table 1: Sample stimuli for Exp. 1 and 3.

Exp. 1	Pair	Prime	Target
Critical	Identity	*lou5* 老 "old"	*lou5* 老 "old"
	Historical	*nou5* 腦 "brain"	
Control	Rime	*pou5* 抱 "embrace"	
	Unrelated	*caa4* 茶 "tea"	
Exp. 3	Pair	Block 1	Block 2
Critical	Identity	*lou5* 老 "old"	*lou5* 老 "old"
	Historical	*naam4* 男 "male"	*laam4* 藍 "blue"
Control	Unmatched	–	*laai1* 拉 "pull"

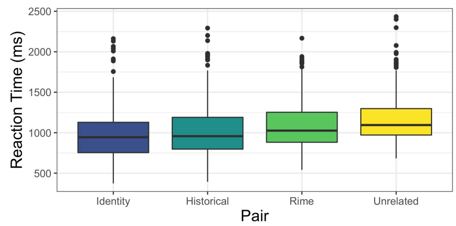

Figure 1: Reaction time (ms) to targets in critical and control pairs in Exp. 1.

experiments, reaction times < 200 ms or > 3000 ms were removed (1% of the responses), as well as reaction times < or > 2.5 standard deviations from each subject's mean (3% of the responses). Reaction times for correct responses to critical and control pairs are shown in Fig. 1.

The reaction times to correct responses were analyzed using a Bayesian mixed-effects model fitted in Stan using the {brms} package [21] in R [20]. The model included Pair (Historical, Identity, Rime, Unrelated) as a dummy-coded fixed effect. Historical pairs were used as the reference level to allow for the comparison of Historical pairs ([n]-initial primes) with Identity pairs ([l]-initial primes) and Rime pairs (primes with shared rimes). The model also included by-item random intercepts and by-subject random slopes for the effect of Pair. A lognormal response distribution was assumed for the reaction times. Regularizing, weakly informative priors were selected for all parameters. Hamiltonian Monte-Carlo sampling with four chains (each with 2000 iterations, 1000 warm-up) were used to draw samples from the posterior distribution. We report the mean of the posterior distribution, the 95% credible interval (CrI), the proportion of the CrI that falls within a region of practical equivalence (ROPE) of ±0.03, and the probability of direction (PD). The mean of the posterior distribution is a point estimate; the average value of the distribution. In general, if the 95% CrI does not encompasses 0, this is taken as evidence for a meaningful effect, while a 95% CrI that encompasses 0 is taken as evidence for a negligible effect. The ROPE specifies a region around 0 that is practically equivalent to 0 (in this case, ± 0.03) and thus, provides a measure of the probability that the credible effect values are non-negligible. A meaningful effect size should have a smaller proportion of 95% CrI within the ROPE, while negligible effects should have a larger proportion of the 95% CrI within the ROPE. Finally, the PD provides a measure of how probable the specific direction (positive or negative) of the effect is, with strong evidence for a non-null effect represented by a PD greater than 95% [22].

The model revealed weak evidence for a difference and strong evidence for equivalence between historical and identity pairs (Identity pairs: β =-0.03, CrI = [-0.07, 0.00], % ROPE = 38.60%, PD = 96.66%). While there is slightly greater priming for identity pairs, as indicated by the negative β value, the upper limit of the 95% CrI is 0, and 38.60% of the posterior distribution falls within the ROPE. In contrast, there is strong evidence for a difference and weak evidence for equivalence between historical and rime pairs, and between historical and unrelated pairs. For both comparisons, the 95% CrI does not encompass 0, 0% of the posterior distribution falls within the ROPE, and the PD is greater than 95% (Rime pairs: β =0.07, CrI = [0.03, 0.11], % ROPE = 0%, PD = 99.98%; Unrelated pairs β =0.14, CrI = [0.10, 0.18], % ROPE = 0%, PD = 100%).

3.3. Interim discussion

Exp. 1 showed that [n]-initial primes (in historical pairs) and [l]-initial primes (in identity pairs) equivalently facilitated the recognition of [l]-initial targets. This suggests that the sound change has rendered [n]- and [l]-initial items as recognition equivalents. There was also strong evidence for a difference between the historical and rime pairs, suggesting that the priming observed for [n]-initial primes in historical pairs to [l]-initial targets is unlikely to be a product of shared rimes between the prime and target. These data may be accounted for by (at least) two explanations. First, listeners may simply be poor at distinguishing [n] and [l] at a phonetic level. Alternatively, listeners may be able to perceive the difference between [n] and [l] but map the two pronunciation variants to a single lexical representation (e.g., *lou5* and *nou5* both map to *lou5* 老 "old") [2]. Exps. 2a and 2b were carried out to test the phonetic-level explanation.

4. EXPERIMENT 2A AND 2B

In Exp. 2a, AX discrimination task participants were presented with two items separated by a 500 ms ISI, and asked to make a same-different judgement. In Exp. 2b, categorization task participants were presented with an individual token and asked to categorize the initial consonant as "N" or "L".

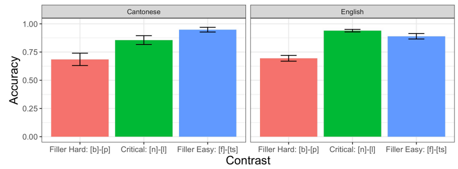

Figure 2: Discrimination accuracy for each nonword pair contrast in Exp. 2a for Cantonese and English listeners.

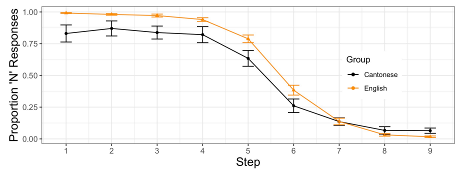

Figure 3: Proportion of 'N' responses in Exp. 2b for Cantonese and English listeners.

4.1. Materials

AX discrimination task items were three nonword pairs each representing different contrasts: *noe6-loe6* (critical), *foe6-tsoe6* (easy), and *boe6-poe6* (hard). The natural *noe6* and *loe6* tokens were used as end points from which nine-step continua were synthesized in TANDEM-STRAIGHT [23]. The individual tokens from this continuum were the items for the categorization task.

4.2. Analysis and results

Due to space, we do not detail the Bayesian analyses and instead provide figures visualizing the discrimination accuracy for pairs in Exp. 2a (see Fig. 2) and the proportion of 'N' responses in Exp. 2b (see Fig. 3). Responses from English control listeners are provided for comparison.

In the AX discrimination task, Cantonese talkers demonstrate high discrimination sensitivity to [n] and [l] in the critical *noe6-loe6* pair; this is on par with an acoustically clear contrast in *foe6-tsoe6*. Likewise, the sigmoidal response function in the categorization task demonstrates clear categorical perception of [n] and [l]. Together, these results demonstrate that listeners can distinguish [n] from [l] at a phonetic level. We turn now to testing the second, dual-mapping explanation in Exp. 3.

5. EXPERIMENT 3

This was a long distance repetition priming task. Items are presented across two blocks. On each trial, participants made a lexical decision to an individual auditory token. Block 1 contained the primes from

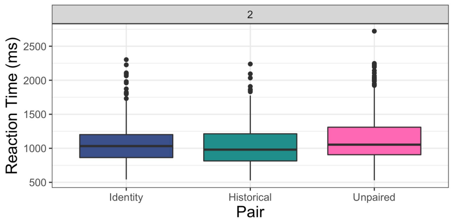

Figure 4: Reaction time (ms) to targets in critical and control pairs in block 2 of Exp. 3.

the pairs in Exp. 1 (e.g., *nou5* 腦 "brain") and block 2 had the corresponding targets (*lou5* 老 "old"). If [n]- and [l]-initial pronunciations are separately encoded variants, priming is not expected, as minimal pairs are known to not show priming across long distances. However, if both pronunciations are fully encoded as acceptable variants of a single lexical item, long distance priming is predicted.

5.1. Materials

Exp. 3 used a subset of stimuli from Exp. 1 (see Table 1). Block 1 contained the primes from 10 critical identity pairs (*lou5* 老 "old") and the primes from 10 critical historical pairs (*naam4* 男 "male"). Block 2 had the 10 corresponding targets of the critical identity pairs (*lou5* 老 "old") and the 10 corresponding targets of the critical historical pairs (*laam4* 藍 "blue"). 20 additional [l]-initial words with no prime in the preceding block were included as unmatched controls (*laai1* 拉 "pull"). Each block contained 160 real word fillers and 240 nonwords.

5.2. Analysis and results

Reaction time outlier removal was the same as in Exp. 1 (3% of the responses removed). These data are shown in Fig. 4. The reaction times to correct responses in block 2 were analyzed using a Bayesian mixed-effects model fitted in Stan using the {brms} package [21] in R [20]. The model included Pair (Historical, Identity, Unmatched), with Historical pairs as the reference level. The model also included by-item random intercepts and by-subject random slopes for the effect of Pair. A lognormal response distribution was assumed for the reaction times. Regularizing, weakly informative priors were selected for all parameters. Hamiltonian Monte-Carlo sampling with four chains (each with 2000 iterations, 1000 warm-up) were used to draw samples from the posterior distribution.

We observed weak evidence for a difference and strong evidence for equivalence between identity and historical pairs. While there was slightly more priming by [n]-initial primes in Historical pairs, indicated by the positive β estimate, the 95% CrI encompasses 0, and 23.27% of the posterior distribution falls within the ROPE (Identity pairs: β =0.06, CrI = [-0.02, 0.14], % ROPE = 23.27%, PD = 91.87%). However, there was strong evidence for a difference and weak evidence for equivalence between unmatched and historical pairs. The 95% CrI does not include 0, only 3.03% of the posterior distribution falls in the ROPE, and the PD is greater than 95% (Unmatched pairs: β =0.09, CrI = [0.01, 0.17], % ROPE = 3.03%, PD = 98.9%).

These results align with and extend those of Exp. 1; historical [n]-initial primes (in historical pairs) and innovative [l]-initial primes (in identity pairs) equivalently facilitate the recognition of [l]-initial targets not just at short time lags (Exp. 1), but also across a more extended interval of time (Exp. 3).

6. DISCUSSION AND CONCLUSION

In Exp. 1, historical [n]-initial primes and innovative [l]-initial primes equivalently facilitated the recognition of [l]-initial targets. Exps. 2a and 2b ruled out the possibility that these results were due to poor discriminability of [n] and [l] at a phonetic level. Listeners showed high discrimination sensitivity to [n] and [l] (on par with an acoustically clear contrast in Cantonese) in Exp. 2a and had crisp sigmoidal categorization functions, suggesting [n] and [l] are distinct perceptual categories. Having ruled out this phonetic explanation, Exp. 3 tested a lexical explanation, wherein both [n]- and [l]-initial pronunciations were dually mapped to a single lexical representation. We observed the same recognition equivalence between [n]- and [l]-initial primes even across a more extended interval of time in a long distance repetition priming task. As minimal pairs do not show priming in this paradigm, these results suggest that [n] and [l] are dually-mapped as acceptable pronunciation variants for a single lexical representation.

As mentioned, Cantonese talkers have metalinguistic awareness about these socially meaningful phonetic forms. As such, this dual-mapping may be a consequence of regular exposure to the pronunciation variants in the community [2]. Additionally, while the sound change has been described as complete [10], these findings also point toward enregistered, context-conditioned variation between [n] and [l] [24]. In this respect, a dual-mapping structure is an ideal representation, as it affords the ability to process both [n] and [l] pronunciation variants in the appropriate social contexts [10].

7. ACKNOWLEDGEMENTS

We thank Fion Fung and members of the Speech in Context Laboratory for technical assistance on this project. We also thank Ashley Farris-Trimble, Yao Yao, and Phil Monahan for assistance with participant recruitment. This work was funded by a SSHRC Insight Grant awarded to Molly Babel.

8. REFERENCES

[1] M. Sumner and A. G. Samuel, "The effect of experience on the perception and representation of dialect variants," *Journal of Memory and Language*, vol. 60, no. 4, pp. 487–501, 2009.

[2] A. G. Samuel and S. Larraza, "Does listening to non-native speech impair speech perception?" *Journal of Memory and Language*, vol. 81, pp. 51–71, 2015.

[3] M. Sumner and A. G. Samuel, "Perception and representation of regular variation: The case of final/t," *Journal of Memory and Language*, vol. 52, no. 3, pp. 322–338, 2005.

[4] M. Sumner and R. Kataoka, "Effects of phonetically-cued talker variation on semantic encoding," *The Journal of the Acoustical Society of America*, vol. 134, no. 6, pp. EL485–EL491, 2013.

[5] Z. Jones and C. G. Clopper, "Subphonemic variation and lexical processing: Social and stylistic factors," *Phonetica*, vol. 76, no. 2-3, pp. 163–178, 2019.

[6] M. Sumner, S. K. Kim, E. King, and K. B. McGowan, "The socially weighted encoding of spoken words: A dual-route approach to speech perception," *Frontiers in Psychology*, vol. 4, p. 1015, 2014.

[7] M. Rae, P. Warren *et al.*, "Goldilocks and the three beers: sound merger and word recognition in NZE," *New Zealand English Journal*, vol. 16, p. 33, 2002.

[8] M. Sumner, "A phonetic explanation of pronunciation variant effects," *The Journal of the Acoustical Society of America*, vol. 134, no. 1, pp. EL26–EL32, 2013.

[9] S.-l. Wong, "A Chinese syllabary pronounced according to the dialect of Cantonese," *Canton: Lingnam University*, 1941.

[10] L. S. Cheng, M. Babel, and Y. Yao, "Production and perception across three Hong Kong Cantonese consonant mergers: Community-and individual-level perspectives," *Laboratory Phonology*, vol. 13, no. 1, 2022.

[11] P. G. Pan, "Prestige forms and phonological variation in Hong Kong Cantonese speech." Master's Thesis, The University of Hong Kong, 1981.

[12] J. D. Ball, *Cantonese Made Easy: A book of simple sentences in the Cantonese dialect, with free and literal translations, and directions for the rendering of English grammatical forms in Chinese*. Kelly & Walsh, Limited, 1907.

[13] H. S.-W. Yeung, "Some aspects of phonological variations in the Cantonese spoken in Hong Kong," Master's Thesis, The University of Hong Kong, 1980.

[14] D. S. Bourgerie, "A quantitative study of sociolinguistic variation in Cantonese," Doctoral Dissertation, The Ohio State University, 1990.

[15] L. Cheng, "Mergers-in-Progress in Hong Kong and Vancouver Cantonese-English Bilinguals," Bachelor's Thesis, University of British Columbia, 2017.

[16] A. L. Anwyl-Irvine, J. Massonnié, A. Flitton, N. Kirkham, and J. K. Evershed, "Gorilla in our midst: An online behavioral experiment builder," *Behavior research methods*, vol. 52, no. 1, pp. 388–407, 2020.

[17] K. J. Woods, M. H. Siegel, J. Traer, and J. H. McDermott, "Headphone screening to facilitate web-based auditory experiments," *Attention, Perception, & Psychophysics*, vol. 79, no. 7, pp. 2064–2072, 2017.

[18] D. Birdsong, "Dominance and Age in Bilingualism," *Applied Linguistics*, vol. 35, no. 4, pp. 374–392, 2014, publisher: Oxford University Press. [Online]. Available: https://journals.scholarsportal.info/details/01426001/v35i0004/374_daaib.xml

[19] H. Wickham, M. Averick, J. Bryan, W. Chang, L. D. McGowan, R. François, G. Grolemund, A. Hayes, L. Henry, J. Hester, M. Kuhn, T. L. Pedersen, E. Miller, S. M. Bache, K. Müller, J. Ooms, D. Robinson, D. P. Seidel, V. Spinu, K. Takahashi, D. Vaughan, C. Wilke, K. Woo, and H. Yutani, "Welcome to the tidyverse," *Journal of Open Source Software*, vol. 4, no. 43, p. 1686, 2019.

[20] R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2021. [Online]. Available: https://www.R-project.org/

[21] P.-C. Bürkner, "brms: An R package for Bayesian multilevel models using Stan," *Journal of statistical software*, vol. 80, no. 1, pp. 1–28, 2017.

[22] B. Nicenboim and S. Vasishth, "Statistical methods for linguistic research: Foundational ideas—part ii," *Language and Linguistics Compass*, vol. 10, no. 11, pp. 591–613, 2016.

[23] H. Kawahara, M. Morise, T. Takahashi, R. Nisimura, T. Irino, and H. Banno, "Tandem-straight: A temporally stable power spectral representation for periodic signals and applications to interference-free spectrum, f0, and aperiodicity estimation," in *2008 IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE, 2008, pp. 3933–3936.

[24] A. Agha, "Voice, footing, enregisterment," *Journal of Linguistic Anthropology*, vol. 15, no. 1, pp. 38–59, 2005.