DETECTING PHONETIC VARIATION VERSUS PHONEMIC DIFFERENCES

Daniel Williams[a,c], Paola Escudero[b,c]

[a]University of Potsdam, [b]Western Sydney University, [c]ARC Centre of Excellence for the Dynamics of Language
daniel.williams@uni-potsdam.de, paola.escudero@westernsydney.edu.au

ABSTRACT

Second-language (L2) speech perception research has largely focused on learning to detect phonetic variation which contributes to phonemic differences. However, phonetic variation may emerge for other reasons, e.g., different speakers. In native language acquisition, speech perception is flexible enough to permit speaker-related phonetic variation while at the same time preserving phonemic identity. In L2 speech learning, it remains unclear how speaker-related phonetic variation is learned or processed. To test how monolinguals and L2 learners handle speaker-related phonetic variation in detecting phonemic differences, the present study investigated the discrimination of novel words spoken in a familiar and unfamiliar accent. Compared to monolinguals, near-native L2 learners made slightly more errors overall and were quicker to attribute large phonetic differences to phonemic differences rather than speaker-related differences. This highlights the challenging task of learning naturally permissible variation in L2 phonemic categories.

Keywords: phonetic variation, phonemic contrasts, second-language learning, accents

1. INTRODUCTION

Research on second-language (L2) learning has tended to focus on phonetic learning in the context of distinguish phonemic contrasts in the L2 [5]. Understandably, this is an important goal for L2 learners, who strive to understand speakers of the L2 as well as be understood themselves. However, not all phonetic variation contributes directly to phonemic differences and it may arise for many other reasons, e.g., speakers of different regional accents.

L2 learners are indeed sensitive to speaker-related phonetic variation in the L2, at least as far as learning phonemic differences is concerned. Previous research has shown that L2 learners exhibit different learning trajectories depending on the L2 variety which they are exposed to. For example, very early on in the learning process, learners display different perceptual assimilation patterns, which are known to be predictive of the kinds of difficulties faced in later phonological learning [4]. Different learning strategies developed by L2 learners have been demonstrated in how new (non-native) contrasts are acquired. For example, Spanish learners of Scottish English focus more on the spectral differences between English /iː/ and /ɪ/ than Spanish learners of Southern British English who favour duration as a cue. This is in line with how the contrast is realised in the two varieties: in Scottish English vowel duration does not reliably distinguish the vowels, whereas it does so in Southern British English [3].

Although speakers of a language use the same phonemic categories, e.g., to make lexical distinctions, the phonetic realisations of these varies across speakers. Generally, listeners are well able to adapt to unfamiliar speakers of a familiar accent, including when differences are large, such as between genders [7]. During native language (L1) acquisition, individuals learn to permit a degree of phonetic variation whilst at the same time preserving phonemic identity [8]. However, adaption may sometimes be difficult when realisations of phonemic categories are substantially different due to an unfamiliar accent [7]. Little is currently known about the kinds of difficulties L2 learners may face in this context.

The present study investigated how monolinguals and near-native L2 listeners of Australian English (AusE) handle accent-related phonetic variation in detecting phonemic differences. To test this, we used novel words without lexical meaning in AusE, as our goal was to examine phonological processing without confounds of lexical clues. Pairs of novel words containing identical consonant frames but contrasting vowel categories were presented to listeners in order to exemplify phonemic differences in AusE. As a way of introducing accent-related phonetic variation, the novel word pairs contained stimuli spoken in a familiar accent (AusE) as well as in an unfamiliar English accent (Yorkshire accent).

We were interested in whether monolinguals show a greater capacity than L2 listeners to detect when phonetic variation is accent-related versus when it signals a phonemic difference. We expected that highly proficient L2 listeners would be able to detect phonemic differences, but less able to detect accent-related differences – unless they had developed L2 phonemic categories which are flexible enough to permit naturally occurring phonetic variation in the L2. Hence, we tested near-native L2 listeners and compared their performance to monolinguals.

2. METHOD

2.1. Listeners

20 AusE monolinguals and 9 L2 listeners participated and were aged 17-27. L2 listeners reported their proficiency to be near-native in the target variety of English and self-reported speaking a diverse range of other languages including Afrikaans, Arabic, Bulgarian, French, Korean and Spanish. This second group may be described as bilinguals under some definitions (for discussion, see [1]); for consistency, this group is referred to as L2 listeners to distinguish them from the group of monolingual listeners.

2.2. Stimuli

The stimuli were 14 novel /CVC/ words (which are non-existent in AusE) with the following format:

- /bVp/: GOAT, MOUTH, PALM, PRICE, THOUGHT;
- /dVk/: FACE, FLEECE, GOOSE, NURSE;
- /fVf/: FOOT, KIT, LOT, STRUT, TRAP;

where /V/ is one of the English vowels listed to the right of the colon. Each novel word was produced by two female AusE and two female Yorkshire speakers.

2.3. Procedure

Listeners were told they were going to hear new words spoken by different speakers and on each trial were asked whether the two speakers were saying the same or different words by pressing corresponding keyboard keys as quickly as possible. Every trial consisted of a pair of novel words being played – one in an AusE accent and one in the unfamiliar Yorkshire accent – with one second of silence separating the two presentations. Before the task started, listeners completed a familiarisation round and breaks were given during the task itself.

The novel word pairs were assembled as follows. Each novel word was paired with itself and all other words of the same consonantal frame, producing 40 novel word pairs (15 /bVp/ pairs, 10 /dVk/ pairs and 15 /fVf/ pairs). The 40 pairs belonged to two *Pair Types* (PT): 26 Different Pairs (DPs) and 14 Same Pairs (SPs). Within all pairs, one word was spoken by an AusE speaker and one by a Yorkshire speaker. The order of presentation of the speakers (*Accent Order*, AO) was counterbalanced across pairs, yielding 320 trials (i.e., 40 word pairs [26 DPs + 14 SPs] × 4 AusE-Yorkshire speaker combinations × 2 speaker orders).

A discriminant analysis was trained on log duration as well as F1 and F2 values corresponding to formant mean, slope and curvature from Elvin *et al.*'s [2] corpus of AusE vowels. For each phonemic category, a centroid in acoustic parameter space was generated with maximum separation from other categories. The stimuli from the present study were then tested on this model, which yielded predicted probabilities of each stimulus vowel falling into the AusE phonemic category originally intended by the speaker. Subsequently, for every unique stimulus pair, the difference between the two stimulus vowels' predicted probabilities was calculated (*Phonemic Distinctiveness Difference*, PDD). A PDD score of 0 (minimum possible, Min) indicates that both vowels in a stimulus pair are equally close to their respective AusE phonemic category centroids in acoustic parameter space. A PDD score of 1 (maximum possible, Max) indicates that the vowel in one stimulus is identical to its AusE phonemic category centroid, while the vowel in the other does not match its phonemic category centroid at all. It is expected that detecting phonemic sameness and difference in stimulus pairs displaying higher PDD scores will be more difficult, as listeners will need to decide how to handle the phonemically more ambiguous vowel – does it relate to a phonemic difference or an accent-related difference?

3. RESULTS

Trials with RTs < 50 ms or > 3,000 ms were removed (4.78% of all trials). Remaining trials were submitted to a mixed-effects bivariate generalised linear regression on the two dependent variables of RT and Accuracy in the program *R* [9] using the *MCMCglmm* package [6], which fits generalised linear regression models for Bayesian statistics. RTs were log-transformed and modelled with a normal distribution, while Accuracy (correct versus incorrect) was modelled with a binomial distribution. Priors were specified for residuals and random effects with the inverse-Wishart distribution with degree of belief parameters set to the lowest bounds. A MCMC chain moved through parameter space sampling the posterior distribution (for further details, see [6]); the first 20,000 were discarded and the next 100,000 were thinned, leaving 1000 samples. The model's fixed effects (and interactions) were PT, AO, PDD and *Language Background* (LB). All were centred on 0 so that the intercepts for RT and Accuracy represented means across the fixed effects which can be interpreted as main effects. For example, the two LB levels were coded as –0.5 for monolinguals and 0.5 for L2 listeners. Thus, the mean log RT or Accuracy of both groups is when LB is equal to 0 and the difference between groups is when LB is equal to 1. Random intercepts were included for listener, unique stimulus pair and trial number and by-listener slopes were included for effects and interactions repeated across listeners. The remainder of this section reports on the above fitted model.

Table 1: Posterior means for the fixed effects (Est.) with 95% credible intervals (C.I.) for Accuracy and RT expressed in odds ratios and ms, respectively. The direction of the differences represented by the main effects are indicated in square brackets. Significance is marked by * and indicates that C.I.s do not cross 1 for Accuracy or zero for RT.

Fixed effect	Accuracy		RT	
	Est.	C.I.	Est.	C.I.
Intercept	0.13*	0.09; 0.20	575*	494; 673
PT [SP → DP]	1.95*	1.26; 3.03	60*	3; 142
AO [Yor-Aus → Aus-Yor]	1.15	0.71; 1.82	40*	11; 75
PDD [Min → Max]	4.60*	2.36; 8.94	76*	34; 134
LB [Mono → L2]	1.72	0.88; 3.28	1	–122; 247
PT × AO	0.91	0.35; 2.26	16	–25; 77
PT × PDD	0.30	0.09; 1.11	–88*	–120; –30
AO × PDD	1.10	0.30; 4.27	–25	–75; 50
PT × LB	0.48*	0.33; 0.71	–63	–129; 52
AO × LB	1.39	0.99; 1.97	–5	–40; 45
PDD × LB	0.47*	0.28; 0.81	–6	–54; 61
PT × AO × PDD	0.83	0.06; 13.38	76	–47; 295
PT × AO × LB	0.63	0.30; 1.25	59	–25; 192
PT × PDD × LB	2.97*	1.07; 9.07	–27	–108; 112
AO × PDD × LB	0.72	0.26; 2.23	–1	–97; 153
PT × AO × PDD × LB	0.48	0.06; 3.67	339*	124; 565

Table 1 displays the model's fixed effects and significant effects are plotted in Figure 1 (Accuracy) and Figure 2 (RT). Turning first to Accuracy, listeners made few errors. For the main effect of PT, listeners made approximately twice as many errors on DPs than on SPs. Listeners were more than four times as likely to make an error when PDD was Max (= 1) than when it was Min (= 0). The PT × LB interaction (Figure 1, c) reveals that monolinguals made fewer errors on SPs than on DPs compared to L2 listeners. Similarly, the PDD × LB interaction (Figure 1, d) shows that, for monolinguals, PDD at Min resulted in

Figure 1: Posterior means of significant main effects and interactions for Accuracy. a) PT; b) PDD; c) PT × LB; d) PDD × LB; e) PT × PDD × LB with monolinguals (left) and L2 listeners (right). Error bars show 95% HPD intervals.

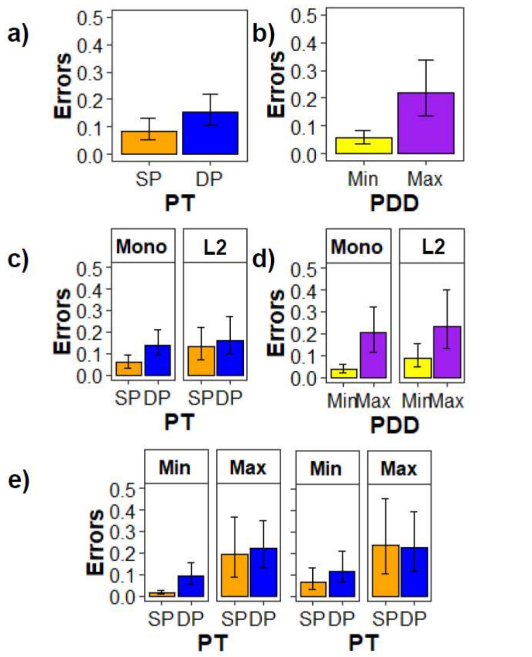
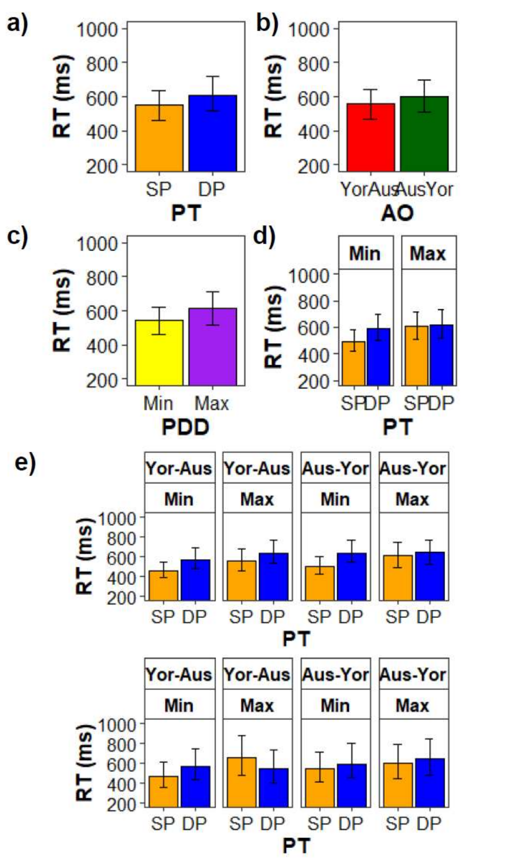


Figure 2: Posterior means of significant main effects and interactions for RT. a) PT; b) AO; c) PDD; d) PT × PDD; e) PT × AO × PDD × LB with monolinguals (upper) and L2 listeners (lower). Error bars show 95% HPD intervals.



fewer errors than at Max compared to L2 listeners. The PT × PDD × LB interaction (Figure 1, e) indicates that monolinguals made fewer errors than L2 listeners, but that the between-group differences varied across different PT × PDD combinations.

Turning to RT, for PT, responses to DPs were longer than to SPs. The main effect of AO indicates listeners took longer to respond when the first speaker was an AusE speaker. For PDD, listeners were slower to respond when PDD was Max. The PT × PDD interaction (Figure 2, d) shows that listeners were quicker to respond to SPs when PDD was Min, but were slower to respond when PDD was Max. Finally, the PT × AO × PDD × LB interaction (Figure 2, e) indicates that, aside from the effects or interactions already described, L2 listeners were on average 105 ms *slower* to respond to SPs than to DPs when PDD was Max and the first speaker was a Yorkshire speaker.

Lastly, the present bivariate analysis estimated a covariance matrix to control for simultaneous effects on Accuracy and RT. This reveals a relatively modest correlation at the level of trial, namely 0.19 (C.I.: 0.11; 0.27), suggesting that trials with a longer RT were more likely to result in an error.

4. DISCUSSION

The present study assessed how monolinguals and near-native L2 learners handle speaker-related phonetic variation in detecting phonemic differences.

Listeners found it easier to correctly attribute accent-related phonetic variation, as reflected in lower error rates and faster RTs, for SPs compared to DPs. As predicted, higher PDDs resulted in higher error rates and slower RTs, which indicates listeners found it more difficult to attribute accent-related phonetic variation when one of the speaker's realisations was phonemically more ambiguous than that of the other speaker. PDD also interacted with PT, as listeners' RTs for SPs increased when one of the two speakers' realisations was acoustically more divergent from its intended AusE phonemic category.

Interestingly, there is a directional asymmetry in the ordering of the speakers, as listeners were 40 ms slower to respond when the first speaker on a trial was an AusE speaker. This may be due to the familiar accent rendering greater phonemic incongruence between the two stimuli. When the first accent is familiar, listeners may strongly perceive the vowel as a single unambiguous phonemic category. Upon hearing the second stimulus in an unfamiliar accent, the phonemic "priming" of the first may therefore reinforce incongruence between the two stimuli. Conversely, when the first stimulus is in an unfamiliar accent, its vowel may be ambiguous and weakly perceived as several phonemic categories. When the second stimulus in a familiar accent is presented, listeners may not be as strongly "primed" on a specific category, leading to less perceived incongruence between stimuli and a faster response.

Despite the near-native proficiency of the L2 listeners, there were some clear cases of difficulty. They made more errors on SPs than monolinguals, indicating that they found it harder to identify accent-related differences in the realisations of the same phonemic category. Likewise, L2 listeners made more errors than monolinguals on SPs even when the two speakers' realisations of novel words were equally close to their intended AusE phonemic categories. Although listeners overall were faster to recognise SPs when the first speaker had an unfamiliar accent, L2 listeners showed the opposite trend when PDD was highest. This suggests a negative "priming" effect of the unfamiliar accent when it is especially phonetically divergent from the familiar accent's phonemic categories. It is possible that, upon hearing the first stimulus in the unfamiliar accent, L2 listeners *incorrectly* associated the vowel as an instance of another AusE phonemic category (i.e., different from the one intended by the speaker), which may have been due to inaccurate representations of some L2 AusE vowels. Ultimately, perceiving the first stimulus' vowel unambiguously (though erroneously) may have strengthened apparent phonemic incongruence with the second stimulus and increased processing time.

Finally, a larger and more controlled set of L2 listeners will elucidate these general findings by considering more explicitly potential influences of particular language backgrounds.

5. CONCLUSION

The present study examined the detection of speaker-related phonetic variation versus phonemic differences by monolinguals and near-native L2 listeners. In line with our expectations, L2 listeners performed largely as well as monolinguals at detecting phonemic differences (DPs). On the other hand, L2 listeners were less accurate than monolinguals at recognising the same phonemic category spoken across the familiar and unfamiliar accents (SPs). This suggests L2 listeners' phonemic categories may not "stretch" in the same ways as those of monolinguals to permit natural speaker- or accent-related phonetic variation, which is often not a focus of theoretical models of L2 learning (e.g., [5]). Additionally, representations of L2 phonemic categories may not always be entirely accurate, which may lead to attributing speaker-related phonetic differences to phonemic differences.

6. REFERENCES

[1] Butler, Y. G., Hakuta, K. 2004. Bilingualism and second language acquisition. In: Bhatia, T., Ritchie, W. (eds.), *Handbook of Bilingualism*. Malden, MA: Blackwell Publishing, 114–144.

[2] Elvin, J., Williams, D., Escudero, P. 2016. Dynamic acoustic properties of monophthongs and diphthongs in Western Sydney Australian English. *J. Acoust. Soc. Am.* 140, 576–581.

[3] Escudero, P, Boersma, P. 2003. Modelling the perceptual development of phonological contrasts with Optimality Theory and the Gradual Learning Algorithm. *U. Penn. Working Papers in Linguistics* 8, 71–85.

[4] Escudero, P., Chládková, K. 2010. Spanish listeners' perception of American and Southern British English vowels. *J. Acoust. Soc. Am* 128, EL254–EL259.

[5] Flege, J. E. 2007. Language contact in bilingualism: Phonetic system interactions. In: Cole, J., Hualde, J. I. (eds.), *Laboratory Phonology 9*. Berlin: Mouton de Gruyter, 353–380.

[6] Hadfield, J. D. 2010. MCMC methods for multi-response generalized mixed models: The MCMCglmm R Package. *J. Statistical Software* 33, 1–22.

[7] Kriengwatana, B., Terry, J., Chládková, K., Escudero, P. 2016. Speaker and Accent Variation Are Handled Differently: Evidence in Native and Non-Native Listeners. *PLOS ONE* 11, #e0156870.

[8] Mulak, K. E., Best, C. T., Tyler, M. D., Kitamura, C., Irwin, J. R. 2013. Development of phonological constancy: 18 months-olds, but not 15 months-old, identify words in a non-native regional accent. *Child Development* 84, 2064−2078.

[9] R Core Team. 2018. *R: A Language and Environment for Statistical Computing*. Version 3.5.1. Retrieved 2nd July 2018 from https://cran.ma.imperial.ac.uk/.