Can Australian English listeners learn non-native vowels via distributional learning?

Jia Hoong Ong[1,2], Josephine Terry[1,2], Paola Escudero[1,2]

[1]The MARCS Institute for Brain, Behaviour and Development, Western Sydney University
[2]Australian Research Council Centre of Excellence for the Dynamics of Language

jia.ong@westernsydney.edu.au, j.terry@westernsydney.edu.au, paola.escudero@westernsydney.edu.au

Abstract

Inconsistent findings have been reported for distributional learning of vowels, possibly due to interference from learners' native phonological (L1) categories. Native Australian-English (AusE) listeners were exposed to unimodal and bimodal distributions of a continuum spanning Dutch /ɑ/-/a:/, which is perceived moderately well by AusE listeners. Despite sustaining learners' attention during the training phase (c.f. passive training), the distribution groups did not differ in their pre-post vowel discrimination, suggesting a lack of distributional learning. Our results imply that learners do not benefit from such rapid learning of contrasts that are perceived with high accuracy due to learners' L1 categories.

Index Terms: distributional learning, non-native speech perception, vowel perception, vowel discrimination

1. Introduction

Distributional learning refers to the process with which learners acquire knowledge by tracking frequency of occurrence of perceptual items [1]. This form of statistical learning is implicated in the acquisition of various phonological categories including consonants [2], [3], vowels [4], [5], and lexical tones [6]. For example, Hindi-learning infants will encounter speech sounds along a continuum of voice-onset time (VOT) ranging from -100ms to +20ms that can be modelled as two normal distributions with one peak at around -85ms and another at around +13ms [7]. Conversely, English-learning infants will encounter speech sounds along the same continuum that can be modelled as a normal distribution with a 0ms peak. This difference in the distributional peaks along that range of VOT—in particular, a bimodal or two-peak distribution for Hindi learners and a unimodal or single-peak distribution for the English learners—may explain the formation of two voicing categories by Hindi listeners (prevoiced and voiceless) and the formation of only one voicing category by English listeners (voiceless).

Distributional learning has been demonstrated in the laboratory by systematically manipulating the frequency of training tokens from a continuum spanning a contrast (e.g., /d/-/t/) presented to learners. Learners who were exposed to relatively more prototypical tokens of the contrast (i.e., those who were exposed to a bimodal distribution) were more likely to show greater discrimination of that contrast after training whereas those who were exposed to relatively more ambiguous tokens of that contrast (i.e., those exposed to a unimodal distribution) were less likely to show any improvement in discrimination after training [1], [3], [4], [8]. Research suggests that infant- and adult-learners are able to learn phonological categories distributionally and so distributional learning may play a crucial role in first-language (L1) and second-language (L2) acquisition [9]–[11]—though some argue that distributional learning may not be as effective for adults than for infants [12].

While distributional learning appears to be used by all language learners, its limits are still unclear. Limits may be related to individual and population differences or the learning environment in acquiring phonological categories via distributional learning. For instance, distributional learning appears to be influenced by whether an attentive or passive learning environment was used; the former appears to be more effective for distributional learning [6]. Concerning individual differences, distributional learning may not be equally effective for all learners as there may be individual differences in the perception of target speech sounds. [13]–[15]. This paper examines a possible factor that contributes to such perceptual differences; one that may arise from listeners' native language. It is widely accepted in non-native speech perception research that a particular non-native contrast may be perceived accurately or poorly by non-native listeners depending on their native phonological categories [10], [11], [16]–[18].

In this paper, we investigate whether native listeners of a particular language (Australian English, AusE) may be disadvantaged in learning a L2 contrast (Dutch /ɑ/ and /a:/) via distributional learning. Previous studies on distributional learning of Dutch /ɑ/-/a:/ have been successful when conducted with native Spanish listeners (i.e., improvement in discriminating that contrast by those in the bimodal condition but not those in the control condition in which participants listened to music; [4]) but not with native AusE listeners (i.e., both bimodal and unimodal conditions improve after training; [19]). These diverging results may be due to listeners' initial perception of the Dutch contrast. That is, Spanish listeners perceive the Dutch /ɑ/-/a:/ contrast as a single category, while AusE listeners perceive it as belonging predominantly to two categories [20]. This difference in categorisation across population leads to different degrees of discrimination accuracy: Spanish listeners will have poor discrimination performance while AusE listeners will have moderate to good discrimination performance (better than native Spanish listeners but not native-like as would be the case with native Dutch listeners). Indeed, a direct comparison of initial performance of this Dutch vowel contrast by native Spanish listeners (data from [8], [21]) and native AusE listeners (data from [19]) reveals that native AusE listeners have significantly higher accuracy than native Spanish listeners ($t(220)= 7.336, p< .001$).

We propose that distributional learning may be constrained by learners' initial perception of the target contrast: the mechanism may most benefit those that are poor perceivers of a contrast (i.e., those that perceive the contrast to be of a single category) but it may not be effective for listeners who are already relatively good perceivers of that contrast (i.e., those that perceive the contrast to belong to two separate categories generally). However, in order to conclude such, it is necessary to examine whether the lack of distributional learning of Dutch /ɑ/-/a:/ by AusE listeners is due to the learning environment. Thus, in this study, we presented the same number of training tokens as that presented to native Spanish listeners and to native AusE listeners in previous distributional learning studies [4], [19]. However, unlike those studies, we included an auditory vigilance task during the training phase to sustain learners' attention to training tokens, which has been shown to be more effective in eliciting distributional learning [6].

If distributional learning is constrained by how learners initially categorise the target contrast, then, we predict no distributional learning by native AusE listeners (i.e., both bimodal and unimodal distribution conditions improve in their discrimination of Dutch /ɑ/ and /a:/ after training). If this hypothesis is supported *despite* sustaining learners' attention to training tokens, then this can be taken as strong evidence of the role of initial perception of target contrast in distributional learning, particularly when the results of this experiment are considered with previous distributional learning studies with native Spanish listeners.

2. Method

2.1. Participants

Participants were 51 (28 females, 23 males) native AusE-listening undergraduates (Age range= 17-59, *M*= 22.45, *SD*= 6.44) from Western Sydney University. Forty-two participants spoke one or more languages other than English although none had prior exposure to Dutch. Participants were randomly allocated to one of two distribution conditions: Unimodal (*n*= 25) and Bimodal (*n*= 26). The average age of participants and the ratio of monolinguals to bilinguals were equivalent across distribution conditions (Age, $t(49)= 0.617$, $p= .540$; Monolinguals, $\chi^2(1)= 0.187$, $p= .666$). All provided their informed consent prior to participating and they were given course credit in return for participating.

2.2. Stimuli

2.2.1. *Test stimuli*

Test stimuli for this experiment consisted of naturalistic and synthetic tokens of Dutch vowels /ɑ/ and /a:/. The naturalistic tokens, which were selected from the corpus of Adank et al [22], were naturally produced Dutch vowels /ɑ/ and /a:/, taken from a /s-V-s/ syllable embedded in a carrier sentence. Sentences were produced by 10 male and 10 female speakers of Standard Northern Dutch, resulting in 20 tokens of /ɑ/ and 20 tokens of /a:/. The synthetic tokens were produced in Praat [23] and constituted good examples of /ɑ/ and /a:/ [21], [24]. They were 140ms in duration and the fundamental frequency (F0) fell between 150 and 100 Hz, representing a male voice.

2.2.2. *Training stimuli*

Training stimuli consisted of eight vowel training tokens that fall along the /ɑ/ to /a:/ continuum, produced in the same manner as the synthetic test stimuli. The training tokens differed from one another only in terms of F1, F2, and F3 spectral values such that from one end of the continuum to the other, the vowel morphs from an /ɑ/ (Token 1) to an /a:/ (Token 8).

2.3. Equipment

The experiment was presented using MATLAB 2012b [25] on an Acer TravelMate P653 laptop that used Windows 7 as the operating system. The auditory stimuli were presented at a comfortable volume using a pair of Sennheiser HD650 headphones connected to an Edirol USB Audio Capture UA-25EX audio interface.

2.4. Procedure

There were three phases to the experiment: pre-test discrimination, training, and post-test discrimination.

2.4.1. *Pre- and Post-test discrimination*

During pre-test and post-test, participants were required to complete a two-alternate forced choice discrimination task (XAB), in which participants determined whether the first sound (X) was similar to the second (A) or third sound (B). In each trial, the ISI was 1.2s. There were 84 trials in each test phase. The A and B tokens were always the synthetic test vowels in all trials, with the position of the synthetic /ɑ/ to /a:/ equally distributed across all trials as the second or third sound. The X tokens, on the other hand, were sampled from the naturally-produced test stimuli in 80 trials and in the remaining four trials, the X tokens were the synthetically-produced test-stimuli (these four trials constituted the 'catch trials'). The /ɑ/ to /a:/ vowels were X tokens an equal number of times in each test phase. Participants could only proceed to the next trial once a response had been collected via a mouse click of a button on the screen. Breaks were provided throughout each test phase.

2.4.2. *Training*

During training, participants were exposed to the training tokens. Participants in the two distribution conditions heard the same total number of training tokens (i.e., 128 tokens) but the conditions differed in the distribution of the training tokens. Specifically, participants in the Unimodal condition heard Tokens 4 and 5 (i.e., ambiguous /ɑ/ to /a:/ vowels) the most frequently whereas those in the Bimodal condition heard Tokens 2 and 7 (i.e., prototypical /ɑ/ to /a:/ vowels) the most frequently (see Table 1).

Table 1. *Frequency distribution of training tokens in Unimodal and Bimodal conditions*

Training Token	1	2	3	4	5	6	7	8
Unimodal	8	8	16	32	32	16	8	8
Bimodal	8	32	16	8	8	16	32	8

Following the training protocol of a previous distributional learning study [6] and as an extension of a previous distributional learning study involving the same vowel contrast [19], participants were required to completed an auditory vigilance task during training, which ensures attentive listening. Participants were instructed during the training phase that they were to listen to the sounds attentively and that some of the sounds would be beeps (sine wave tones). When they heard a beep, they were required to circle the sound number that corresponded to the beep on a paper response sheet. A total of 16 beeps were interspersed randomly within the 128 training tokens. The entire training phase took approximately 3mins to complete.

3. Results

We first determined whether the two distribution conditions differed in their Pre-test scores. The Pre-test scores of the two distribution conditions were not significantly different, either when the Pre-test scores were considered for each vowel (/ɑ/, $t(49)= 0.270$, $p= .788$; /a:/, $t(49)= 0.552$, $p= .584$) or as a whole ($t(49)= .496$, $p= .622$), suggesting that any difference at Post-test between the two distribution conditions may be attributed to the training itself.

A Mixed ANOVA with a between-subject factor Distribution Condition (Unimodal vs. Bimodal) and within-subject factors Session (Pre-test vs. Post-test) and Vowel (/ɑ/ vs. /a:/) revealed a main effect of Session ($F(1, 49)= 38.069$, $p< .001$, $\eta_p^2= 0.437$) and a significant Session by Vowel interaction ($F(1, 49)= 11.486$, $p= .001$, $\eta_p^2= 0.190$). Simple main-effect analysis revealed that, from Pre- to Post-test, participants showed greater improvement in discriminating /a:/ (Pre-test: $M= 0.661$, $SE= .023$; Post-test: $M= 0.764$, $SE= .023$) than /ɑ/ (Pre-test: $M= 0.693$, $SE= .018$; Post-test: $M= 0.732$, $SE= .020$). Importantly, the ANOVA revealed no significant Session by Distribution Condition interaction, which suggests that the two distribution conditions did not differ in their discrimination accuracy from Pre- to Post-test (see Figure 1).

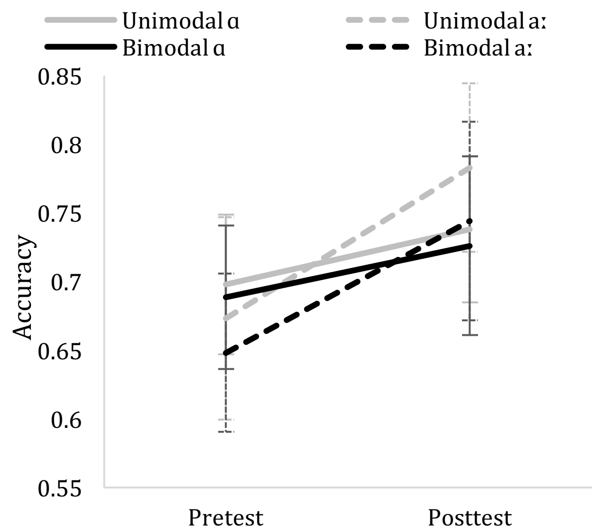

Figure 1: *Accuracy scores from Pre- to Post-test for Unimodal (gray) and Bimodal (black) distribution conditions by vowel (*ɑ *in solid and* a: *in dashed line)*

4. Discussion

Results of this study indicate that despite sustaining learners' attention to the training tokens, native AusE learners do *not* show distributional learning of Dutch /ɑ/-/a:/ vowels; both learners in the bimodal and unimodal distribution conditions improved in their discrimination of /ɑ/ and /a:/ after training. The improvement seen by both distribution conditions is likely due to the fact that participants performed the same test twice (practice effects). Unexpectedly, across both distribution conditions, learners showed greater pre-post improvement in discriminating /a:/ than /ɑ/. This larger improvement for a more peripheral vowel (Dutch /a:/) is reminiscent of the natural referent vowel framework [26], in which it is argued that listeners have an underlying perceptual bias for vowels in the peripheral of F1/F2 vowel space. Though it is unclear how distributional learning may contribute to such asymmetry, our finding on asymmetric improvement suggests a change in perception of each Dutch vowel by AusE listeners. Further work is needed to investigate this finding.

With respect to the lack of distributional learning seen among AusE listeners in this study, the results complement a similar previous study that used a passive training paradigm [19], suggesting that the learning environment (i.e., attentive or passive) does not impact their learning of Dutch /ɑ/-/a:/. It may be the case that on top of an attentive listening paradigm, the stimuli need to be exaggerated or enhanced for such learning to be successful [8]. Enhanced stimuli are argued to draw learners' attention to the acoustic differences of the target contrast [12] and so a larger gain in distributional learning may be observed. However, it may also be the case that since AusE listeners are already good perceivers of that contrast, exaggerating the stimuli will not yield any advantage. Work is currently underway in our laboratory to investigate whether enhanced bimodal distributions will lead to distributional learning in this population.

Native Spanish listeners, on the other hand, have been successful in demonstrating distributional learning of Dutch /ɑ/-/a:/ even when trained passively [4], [8]. Thus, we propose that this population difference in distributional learning of Dutch /ɑ/-/a:/ may be due to learners' L1 influence. Specifically, it appears that distributional learning is affected by how a L2 contrast is categorised in terms of learners' L1 phonological categories. When a L2 contrast is perceived as a single category—which leads to poor discrimination performance, as is the case with native Spanish listeners on the Dutch /ɑ/-/a:/ vowels—distributional learning appears to be effective. However, when the L2 contrast is already perceived as two different categories (although not to the extent of native listeners)—as is the case with native AusE listeners on the Dutch /ɑ/-/a:/ vowels—then learners do not appear to benefit from distributional learning.

Two caveats are worth mentioning. As demonstrated in the present study and in a previous study [19], native AusE listeners do not show distributional learning of Dutch /ɑ/-/a:/ vowels when the standard amount of training tokens as used in most distributional learning studies (i.e., 128 training tokens) is presented. It may be the case that listeners with relatively good perception of the target contrast may need longer training or more training tokens in order to learn the target contrast distributionally. Future studies can test this prediction by replicating the present experiment with double the amount of training [27]. Secondly, it could also be the case that our behavioural measure was not sensitive enough to register any learning by native AusE listeners after such rapid training. To investigate this possibility, future research using electroencephalography (EEG) could shed light on whether such rapid learning can occur preattentively using EEG measures such as mismatch negativity (MMN).

To further support our proposal, future studies could compare distributional learning of the Dutch /ɪ/-/i/ contrast by native AusE and native Spanish listeners. The categorisation pattern of the Dutch /ɪ/-/i/ contrast by native AusE and Spanish listeners is the reverse of the Dutch /ɑ/-/a:/: the Dutch /ɪ/-/i/ is perceived as a single category by AusE listeners but essentially as two categories by Spanish listeners [20]. Consequently, AusE listeners are reported to be poor perceivers of the Dutch /ɪ/-/i/ contrast whereas Spanish listeners are relatively good perceivers of the same Dutch contrast although not to the extent of native Dutch listeners [20]. Thus, if our proposal is true, that is, if distributional learning is influenced by how accurately learners' perceive the L2 contrast prior to distributional training, then we may expect AusE listeners, but not Spanish listeners, to show distributional learning of Dutch /ɪ/-/i/ contrast. Specifically, AusE listeners in the bimodal condition, but not those in the unimodal condition, should improve in their discrimination of Dutch /ɪ/-/i/ after training, whereas Spanish listeners should improve regardless of distribution condition due to practice effects from performing the test twice.

5. Conclusion

In conclusion, this study demonstrated that the lack of distributional learning of Dutch /ɑ/-/a:/ among AusE listeners is not due to the learning environment (i.e., attentive listening). When this study is considered with previous studies on distributional learning of Dutch /ɑ/-/a:/ by native AusE and native Spanish listeners, we propose that how L2 contrasts are categorised in learners' L1 can predict whether learners show distributional learning. Specifically, if learners perceive the L2 contrast as a single category (and therefore are poor perceivers of that contrast in the case of Spanish listeners), distributional learning appears to be effective. On the other hand, if learners perceive the L2 contrast as two separate categories (and thus are good perceivers of that contrast), they may not benefit from distributional learning—at least not with natural bimodal distributions with 128 training tokens. Further work is needed to examine this proposal more closely in order to broaden our understanding of the limits of distributional learning in acquiring L2 speech sounds.

6. Acknowledgements

We would like to thank Christina Quattropani for help with data collection. This research was supported by the Australian Research Council Centre of Excellence for the Dynamics of Language research grant (project number CE140100041) awarded to PE.

7. References

[1] J. Maye, J. F. Werker, and L. Gerken, "Infant sensitivity to distributional information can affect phonetic discrimination," *Cognition*, vol. 82, no. 3, pp. B101–B111, 2002.

[2] J. Maye, D. J. Weiss, and R. N. Aslin, "Statistical phonetic learning in infants: Facilitation and feature generalization," *Dev. Sci.*, vol. 11, no. 1, pp. 122–134, 2008.

[3] J. Maye and L. Gerken, "Learning phonemes without minimal pairs," in *Proceedings of the 24th Annual Boston University Conference on Language Development*, 2000, vol. 2, pp. 522–533.

[4] P. Escudero and D. Williams, "Distributional learning has immediate and long-lasting effects," *Cognition*, vol. 133, no. 2, pp. 408–13, Nov. 2014.

[5] K. Wanrooij, P. Boersma, and T. L. van Zuijen, "Fast phonetic learning occurs already in 2-to-3-month old infants: An ERP study," *Front. Psychol.*, vol. 5, no. FEB, 2014.

[6] J. H. Ong, D. Burnham, and P. Escudero, "Distributional learning of lexical tones: A comparison of attended vs. unattended listening," *PLoS One*, vol. 10, no. 7, p. e0133446, 2015.

[7] L. Lisker and A. S. Abramson, "A cross-language study of voicing in initial stops: Acoustical measurements," *Word*, vol. 20, no. 3, pp. 384–422, 1964.

[8] P. Escudero, T. Benders, and K. Wanrooij, "Enhanced bimodal distributions facilitate the learning of second language vowels.," *J. Acoust. Soc. Am.*, vol. 130, no. 4, pp. EL206–12, Oct. 2011.

[9] J. F. Werker, H. H. Yeung, and K. A. Yoshida, "How do infants become experts at native-speech perception?," *Curr. Dir. Psychol. Sci.*, vol. 21, no. 4, pp. 221–226, Jul. 2012.

[10] P. Escudero, "Linguistic perception and second language acquisition: Explaining the attainment of optimal phonological categorization," Utrecht University, 2005.

[11] J.-W. van Leussen and P. Escudero, "Learning to perceive and recognize a second language: The L2LP model revised," *Front. Psychol.*, vol. 6, pp. 1–12, 2015.

[12] K. Wanrooij, P. Boersma, and T. L. van Zuijen, "Distributional vowel training is less effective for adults than for infants. A study using the mismatch response," *PLoS One*, vol. 9, no. 10, 2014.

[13] R. Frost, B. C. Armstrong, N. Siegelman, and M. H. Christiansen, "Domain generality versus modality specificity: The paradox of statistical learning," *Trends Cogn. Sci.*, vol. 19, no. 3, pp. 117–125, 2015.

[14] A. Cristià, G. L. McGuire, A. Seidl, and A. L. Francis, "Effects of the distribution of acoustic cues on infants' perception of sibilants," *J. Phon.*, vol. 39, no. 3, pp. 388–402, 2011.

[15] K. Wanrooij, P. Escudero, and M. E. J. Raijmakers, "What do listeners learn from exposure to a vowel distribution? An analysis of listening strategies in distributional learning," *J. Phon.*, vol. 41, no. 5, pp. 307–319, Sep. 2013.

[16] C. T. Best and M. D. Tyler, "Nonnative and second-language speech perception," in *Language experience in second language speech learning: In honor of James Emil Flege*, M. J. Munro and O.-S. Bohn, Eds. Amsterdam: John Benjamins, 2007, pp. 13–34.

[17] C. T. Best, "A direct realist view of cross-language speech perception," in *Speech perception and linguistic experience: Issues in cross-language research*, W. Strange, Ed. Timonium, MD: York Press, 1995, pp. 171–204.

[18] J. E. Flege, "Second language speech learning: Theory, findings, and problems," in *Speech perception and linguistic experience: Issues in cross-language research*, W. Strange, Ed. Timonium, MD: York Press, 1995, pp. 233–277.

[19] J. Terry, J. H. Ong, and P. Escudero, "Passive distributional learning of non-native vowel contrasts does not work for all listeners," in *Proceedings of the 18th International Congress of Phonetic Sciences (ICPhS)*, 2015.

[20] S. Alispahic, P. Escudero, and K. E. Mulak, "More vowels are not always better: Australian English and Peruvian Spanish learners ' comparable perception of Dutch vowels," in *Proceedings of the 39th annual Boston University Conference on Language Development*, 2015, pp. 40–51.

[21] P. Escudero and K. Wanrooij, "The effect of L1 orthography on non-native vowel perception," *Lang. Speech*, vol. 53, no. 3, pp. 343–365, 2010.

[22] P. Adank, R. van Hout, and R. Smits, "An acoustic description of the vowels of Northern and Southern Standard Dutch," *J. Acoust. Soc. Am.*, vol. 116, no. 3, pp. 1729–1738, 2004.

[23] P. Boersma and D. Weenink, "Praat: Doing phonetics by computer." 2013.

[24] L. C. W. Pols, H. R. C. Tromp, and R. Plomp, "Frequency analysis of Dutch vowels from 50 male speakers," *J. Acoust. Soc. Am.*, vol. 53, pp. 1093–1101, 1973.

[25] "MATLAB Release 2012b." Natick, Massachusetts, United States, 2012.

[26] L. Polka and O. S. Bohn, "Natural Referent Vowel (NRV) framework: An emerging view of early phonetic development," *J. Phon.*, vol. 39, no. 4, pp. 467–478, 2011.

[27] K. A. Yoshida, F. Pons, J. Maye, and J. F. Werker, "Distributional phonetic learning at 10 months of age," *Infancy*, vol. 15, no. 4, pp. 420–433, 2010.