PASSIVE DISTRIBUTIONAL LEARNING OF NON-NATIVE VOWEL CONTRASTS DOES NOT WORK FOR ALL LISTENERS

Josephine Terry, Jia Hoong Ong, Paola Escudero

The MARCS Institute, University of Western Sydney
J.Terry@uws.edu.au, jia.ong@uws.edu.au, paola.escudero@uws.edu.au

ABSTRACT

Distributional learning studies investigating the acquisition of the Dutch contrast /ɑ/-/aː/ by non-native Dutch learners have reported mixed results. The present study extends the literature by examining whether (i) *naïve* listeners are able to extract the distribution structure of a sequence of /ɑ/-/aː/ tokens drawn from a continuum,; and (ii) differential effects exist between naturalistic vs. exaggerated distributions. Australian-English listeners were randomly assigned to a flat, unimodal, bimodal or enhanced distribution training condition. Their performance was assessed using a categorisation task before and after training.

Our findings showed that while categorisation accuracy was higher at post-test vs. pre-test (perhaps due to task learning), naïve learners did not show the predicted distributional learning effects: the bimodal and enhanced groups did not outperform the flat and unimodal groups. The results could be attributed to individual differences in the ability to sustain attention throughout the training phase, which may be necessary for highly variable speech sounds such as vowels.

Keywords: distributional learning; phonetic categories; vowel; attention; statistical learning

1. INTRODUCTION

Learners are able to use statistical information in a highly structured environment to acquire knowledge specific to that environment, including various aspects of language. For instance, learners track the frequency of speech sounds in their linguistic environment in order to acquire and adapt phonetic categories [13]. Namely, they form categories around sounds that occur with the highest frequency. As an example, native Japanese infants infer from their linguistic environment that along a particular acoustic dimension, speech sounds tend to cluster around a single peak (unimodal), forming the Japanese /r/, whereas native English infants infer along that same acoustic dimension that there are two peaks (bimodal) in their linguistic environment, the English /r/ and /l/. This type of acquisition is termed distributional learning and it has been proposed as a mechanism underpinning perceptual attunement [24]: the phenomenon that by their first birthday, infants become accustomed to speech sounds that are in their linguistic environment [e.g., 11].

Empirically, it has been shown that infants are able to acquire phonetic categories base on statistical, or distributional properties [15]. For example, 8-month-old English infants were exposed to one of two distributions of a non-native Hindi contrast (e.g., /da/-/ta/): unimodal or bimodal. Following habituation, the bimodal infants, but not the unimodal infants, successfully discriminated that Hindi contrast. Moreover, the bimodal infants were able to generalise their output to a similar voicing contrast at another place of articulation (e.g., /g/-/k/). Thus, it appears that infants are able to utilise frequency cue to discover phonetic categories in their linguistic environment.

This form of acquisition extends to learners who have established native language phonetic categories. Indeed, adult second-language (L2) learners are able to acquire non-native contrasts via distributional learning [10, 13, 14, 18]. However, the mechanism is constrained for certain kinds of speech sounds (e.g. vowels produced with relatively high variability) [6, 8]. For example, Escudero et al. [6] found that Spanish-speaking learners of Dutch were only able to acquire a difficult Dutch vocalic contrast (/ɑ/-/aː/) with exposure to a bimodal distribution with artificially exaggerated acoustic cues: an *enhanced* distribution. On the other hand, Escudero and Williams [8] found that the same population group were in fact able to learn that contrast following exposure to both naturalistic and enhanced bimodal distributions and they were able to retain this improvement for at least 12 months. Thus, research using naturalistic bimodal distributions to elicit learning of non-native vowel contrasts in adult populations has reported mixed results.

In response to these mixed findings, it has been proposed that infants and adults respond to and utilise distributional structures in different ways [22]. Specifically, it is suggested that the distributional learning effect seen in adult L2

learners is weaker than that observed with infants because adults are less reliant on stimulus-driven learning and they have existing higher-order linguistic representations. Wanrooij et al. [22] suggest that the enhanced distribution drives learning in adults (i.e. in [6]) not because of the distribution structure per se, but rather because of the participants' *awareness* of the relevant acoustic cues. This awareness is in fact driven by the enhanced differences between the end points of the acoustic continuum. While this explanation is possible, it does not account for the observed learning effects elicited via exposure to a naturalistic bimodal distribution [8].

To add to our understanding of distributional learning as a potential L2 learning mechanism, we replicate the general procedure of previous studies in comparing distributional learning of the same Dutch /ɑ/-/aː/ vocalic contrasts [6, 8]. However, rather than examining learning in a population with some (albeit minimal) experience with Dutch we employ listeners who have had no prior experience with the Dutch language at all.

Specifically, the present study seeks to directly compare whether Australian English adults are able to learn the Dutch contrast using three different naturalistic distributions (flat, unimodal and bimodal) and an enhanced distribution. If naïve adult learners can infer the number of categories based on the distribution structure, then it is predicted that those trained on a bimodal distribution (naturalistic or enhanced) will outperform those trained on a flat or unimodal distribution on discriminating the target contrast. Furthermore, if the effect of an enhanced distribution is additive, then it is expected that those trained on an enhanced distribution will outperform those trained on a bimodal distribution.

2. METHOD

2.1. Participants

One-hundred and eleven psychology students (92 females, 19 males) from the University of Western Sydney participated in exchange for course credit. The mean age of participants was 21.77 years ($SD = 4.82$, range = 18 – 41 years). While all participants were Australian-English speakers, 63 spoke one or more languages other than English. None had prior exposure to Dutch. Participants were randomly allocated to one of the four distribution conditions: Flat ($N = 28$), Unimodal ($N = 28$), Bimodal ($N = 28$), and Enhanced ($N = 27$). The average age of participants, and the ratio of monolinguals to bilinguals was equivalent across groups.

2.2. Stimuli and procedure

Following a previously reported procedure [6, 8, 21], participants performed an initial categorisation task (pre-test), followed by a short listening phase (training) and then a repeat of the initial categorisation task (post-test).

2.1.1. Categorisation pre- and post-test

The pre- and post-tests were two-alternate forced choice categorisation tasks (*XAB*). In each trial, the participants listened to three vowels (1.2 sec ISI) and judged whether the first vowel (*X*) was more like the second (*A*) or third vowel (*B*). They were instructed to respond as quickly and accurately as possible, and to guess if unsure. The *X* tokens were selected from the corpus of Adank et al. [2]. They were naturally produced Dutch vowels /ɑ/ and /aː/, 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.

The *A* and *B* stimuli were two vowel tokens synthetically produced in Praat [3] and constituted good examples of /ɑ/ and /aː/ [7, 19]. They were 140 ms in duration and the fundamental frequency (F0) fell between 150 and 100 Hz, representing a male voice. In both the pre- and post-tests, participants responded to 84 randomly presented trials. The /ɑ/ and /aː/ syllables were X tokens an equal number of times (i.e. 40 trials each) and for the remaining 4 trials, the X tokens were the synthetic A and B stimuli (i.e. "catch-trials"). The A and B stimuli were also equally distributed across all trials. Each trial was initiated with a key-press and breaks were provided.

2.1.1. Training

Between the pre- and post-tests, participants were exposed to a two minute presentation of eight repeating synthetic vowels (produced in the same manner as the *A* and *B* vowels). The eight vowels fell along the /ɑ/ to /aː/ continuum and differed from each another only in terms of F1, F2, and F3 spectral values. In the flat, unimodal, and bimodal distributions, the end-point values (i.e., tokens 1 & 8) were similar to the average natural production values of /ɑ/ and /aː/, respectively. However, in the enhanced distribution, the value of token 1 was one standard deviation lower in F1 and F2 than those of the average production of /ɑ/. The value of token 8 was one standard deviation higher in F1 and F2 than those of the average production of /aː/. The

frequency with which particular vowels were presented differed according to the four distributions (flat, unimodal, bimodal, enhanced). The distribution frequencies are shown in Table 1. In total, listeners heard 128 tokens (750 ms ISI).

Table 1: Frequency distribution of vowels in the training phase.

| Token number	1	2	3	4	5	6	7	8
Flat	16	16	16	16	16	16	16	16
Unimodal	8	8	16	32	32	16	8	8
Bimodal	8	32	16	8	8	16	32	8
Enhanced	8	32	16	8	8	16	32	8

Prior to the training phase, participants were instructed to listen carefully to the vowels as they would undertake a second vowel classification task afterwards (post-test).

3. RESULTS

At Pre-test, accuracy ranged from 41.25% to 92.5% correct, and mean accuracy ($M = 69.53\%$, $SD = 12.67\%$) was significantly above chance (50%), $p = .00$. To ensure that accuracy was equivalent across conditions in the pre-tests, an ANOVA compared % correct categorisation. We found no differences between the conditions, $F(3, 107) = .35$, $p = .79$, partial $\eta^2 = .01$. To examine the effects of exposure to the four distributions on categorisation accuracy, a 2 (pre-test, post-test) x 4 (flat, unimodal, bimodal, enhanced) ANOVA was conducted. There was a significant main effect of test, with accuracy higher after exposure to a vowel distribution, $F(1, 107) = 35.66$, $p = .00$, partial $\eta^2 = .25$. However, there was no main effect of the type of distribution, $F(3, 107) =.16$, $p = .92$, partial $\eta^2 = .01$, and no test*distribution interaction, $F(3, 107) = 1.37$, $p = .26$, partial $\eta^2 = .04$. Categorisation accuracy (% correct) at pre- and post-training and across the four distribution conditions is presented in Figure 1.

A difference score for each participant was calculated by subtracting Pre-test % correct from Post-test % correct. A one-way ANOVA on these difference scores revealed no effect of distribution condition (flat, unimodal, bimodal, enhanced), $F(3, 107) = 1.37$, $p = .26$. Planned comparisons were conducted in order to examine specific effects. Firstly, the mean difference scores from the *flat+unimodal* conditions were compared to the *bimodal+enhanced* conditions. A t-test revealed no effect, $t(109) = .45$, $p = .65$. Secondly, mean difference scores in the *enhanced* and *bimodal* conditions were compared. Again, a t-test revealed no difference, $t(53) = .35$, $p = .73$.

Figure 1: Accuracy (% correct) of vowel categorisation at pre-training and post-training as a function of training distribution type. Error bars are *SEM*.
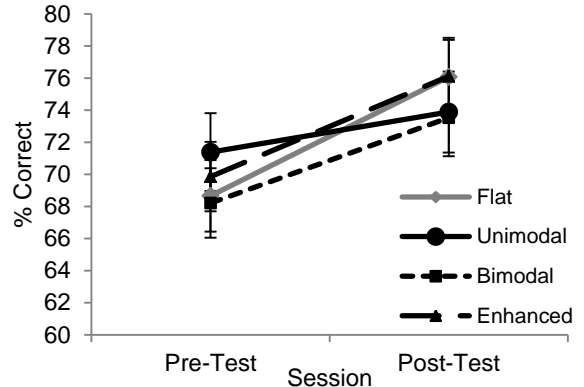
4. DISCUSSION

The present experiment examined whether naïve adult learners were able to acquire a difficult Dutch /ɑ/-/a:/ contrast via distributional learning. While we did find that categorisation performance was higher at the post-test compared to the pre-test, (perhaps due to task learning), there was no effect of the specific distributional properties. The bimodal groups (naturalistic or enhanced) did not outperform the unimodal and flat groups. Neither did the enhanced group outperform the naturalistic bimodal group.

Previous studies have reported mixed results with the same Dutch contrast tested on Spanish learners of Dutch. Escudero and Williams [8] found that exposure to bimodal and enhanced distributions elicited improved categorisation performance. However, Escudero et al. [6] found improvement only after exposure to an enhanced distribution. Adding to the already conflicting findings, our naïve learners showed no differential distributional learning effects, not even in the enhanced condition. If it is that case that awareness of the exaggerated acoustic cues is required to elicit learning in adults [22], then our naïve adult learners may not have become aware of these cues. In order words, the widening of the acoustic cues that are used to distinguish the contrast in the enhanced condition did not elicit awareness.

This is contrary to Wanrooij et al. [22] who argue that the exaggerated cues should elicit awareness. However, it is possible that our results are due to individual differences in attending to the: i) relevant acoustic cue [23]; and ii) entire training phase in general [17]. Concerning the first possibility, it may be the case that the participants in this study, more than half of which speak a language other than

English, attend to the formant values that defined the distribution in different ways. While additional analyses found no performance differences between monolingual and multilingual participants, the cues to which listeners attend as a function of prior language exposure requires further examination. For instance, it is known that Australian English speakers can better discriminate the Dutch contrast /ɪ-ʏ/ more accurately than Spanish listeners [1].

The second possibility for the unexpected set of results relates to a more general type of attention during acquisition. Despite being told to listen to the training tokens carefully, not all participants may have listened attentively. Previous research has shown that when passive listening and attentive listening are compared directly, the distributional learning effect is more readily observed in the attentive listening condition [17]. Indeed, attentive listening is important in naturalistic language acquisition. Specifically, infants acquiring their native phonetic categories do so with infant-directed speech (IDS) as their input. One of the defining characteristics of IDS is the use of exaggerated prosody, which serves to capture the infants' attention to the speech sounds [4]. This is in contrast with adult-directed speech, to which infants do not appear to attend as much [4]. The role of attention is also argued to be important in second-language acquisition; for learners to acquire any linguistic knowledge, the learner must first *notice* the L2-input, which can only happen if the learner attends to the input [20]. Thus, it appears that general attention to the linguistic input is important in language learning.

In saying that, there have been reports that consonants can be learned passively [e.g., 10, 18]. We speculate that this difference is due to the variability of the speech sounds. Consonants tend to be produced with less variability and therefore, have narrower peaks in their distribution, whereas vowel production tends to be more variable [8, 12]. Therefore, attention to the speech sounds may be necessary when the to-be-learned speech sounds are highly variable in nature. Indeed, listeners are more/better able to utilise the acoustic cue to discriminate between speech sounds from a less noisy (low variability) distribution than from a highly variable distribution [16]. Future studies should, therefore, replicate the present experiment with the additional of a training task that sustains the participants' attention throughout the training phase, similar to that of a previous study [17]. This manipulation would allow an examination of the role of attention in learning vocalic categories.

5. CONCLUSION

The present research aimed to examine whether naïve adult listeners could acquire a Dutch vocalic contrast (/ɑ/-/aː/) via distributional learning, and if so, whether there were differential effects of a naturalistic distribution vs. an enhanced distribution. We did not find any support for distributional learning of that contrast, adding to the already inconsistent findings reported in the literature. However, we speculate that our findings may be due to individual differences in sustaining attention throughout the training phase, which may be necessary in acquiring highly-variable speech sounds (e.g. vowels). Future research in this area is needed in order for us to fully understand the mechanism that underpins learning the phonemes of first and subsequent languages.

6. REFERENCES

[1] Alispahic, S., Escudero, P., Mulak, K. 2014. Difficulty in discriminating non-native vowels: Are Dutch vowels easier for Australian English than Spanish listeners?. *Interspeech*, Singapore.

[2] Adank, P., Van Hout, R., Smits, R. 2004. An acoustic description of the vowels of Northern and Southern standard Dutch. *JASA,* 116, 1729-1738.

[3] Boersma, P., Weenink, D. 2011. "Praat: Doing phonetics by computer. http://www.praat.org.

[4] Burnham, D., Kitamura, C., Vollmer-Conna, U. 2002. What's new, pussycat? On talking to babies and animals. *Science (80).* 296, 1435.

[5] Cooper, R. P., Aslin, R. N. 1990. Preference for infant-directed speech in the first month after birth. *Child Dev.* 61, 1584–1595.

[6] Escudero, P., Benders, T., Wanrooij, K. 2011. Enhanced bimodal distributions facilitate the learning of second language vowels. *JASA.* 130, EL206–12.

[7] Escudero, P., Wanrooij, K. 2010. The effect of L1 orthography on on-native and L2 vowel perception. *Lang. Speech* 53, 343-365.

[8] Escudero, P., Williams, D. 2014. Distributional learning has immediate and long-lasting effects. *Cognition* 133, 408–13.

[9] Fry, D. B., Abramson, A. S., Eimas, P. D., Liberman, A. M. 1962. The identification and discrimination of synthetic vowels. *Lang. Speech* 5, 171–189.

[10] Hayes-Harb, R. 2007. Lexical and statistical evidence in the acquisition of second language phonemes. *Second Lang. Res.* 23, 65–94.

[11] Kuhl, P. K. 2004. Early language acquisition: Cracking the speech code. *Nat. Rev. Neurosci.* 5, 831–843.

[12] Liberman, A. M., Harris, K. S., Howard, S., Griffith, B. C. 1957. The discrimination of speech sounds within and across phoneme boundaries. *J. Exp. Psychol.* 54, 358–368.

[13] Maye, J., Gerken, L. 2000. Learning phonemes without minimal pairs. *Proc. 24th BUCLD*, Boston, 522–533.

[14] Maye, J., Gerken, L. 2001. Learning phonemes: How far can the input take us. *Proc 25th BUCLD, Boston,* 480–490.

[15] Maye, J., Weiss, D. J., Aslin, R. N. 2008. Statistical phonetic learning in infants: Facilitation and feature generalization. *Dev. Sci.* 11, 122–134.

[16] Nixon, J., van Rij, J., Mok, P. P. K., Baayen, H., Chen, Y. 2014. Eye movements reveal Cantonese listeners use statistical information to assess category membership of acoustic cues. in *Proc. 15th ASSTA*, 246.

[17] Ong, J. H., Burnham, D., Escudero, P. Distributional learning of lexical tones: A comparison of attended vs. unattended listening. Submitted manuscript.

[18] Perfors, A., Dunbar, D. 2010. Phonetic training makes word learning easier. *Proc. 32nd CogSci,* Portland, 1613–1618.

[19] Pols, L. C. W., Tromp, H. R. C., Plomp, R. 1973. Frequency analysis of Dutch vowels from 50 male speakers, *JASA* 53, 1093–1101.

[20] Schmidt, R. W. 1990. The role of consciousness in second language learning. *Appl. Linguist.* 11, 129–158.

[21] Wanrooij,K., Boersma, P. 2013. Distributional training of speech sounds can be done with continuous distributions. *JASA,* 133 (5), EL398-EL394.

[22] Wanrooij, K., Boersma, P., Zuijen, T. L. Van. 2014. Distributional vowel training is less effective for adults than for infants. A study using the mismatch response. *PLoS One* 9.

[23] Wanrooij, K., Escudero, P., Raijmakers, M. E. J. 2013. What do listeners learn from exposure to a vowel distribution? An analysis of listening strategies in distributional learning. *J. Phon.* 41, 307–319.

[24] Werker, J. F., Yeung, H. H., Yoshida, K. A. 2012. How do infants become experts at native-speech perception? *Curr. Dir. Psychol. Sci.* 21, 221–226.