Distributional vowel training may not be effective for Dutch adults

Wanrooij, K.; De Vos, J.F.; Boersma, P.

Publication date
2015

Document Version
Final published version

Published in
Proceedings of the 18th International Congress of Phonetic Sciences

DISTRIBUTIONAL VOWEL TRAINING MAY NOT BE EFFECTIVE FOR DUTCH ADULTS

Karin Wanrooij[1], Johanna F. de Vos[1,2] and Paul Boersma[1]

[1]University of Amsterdam, [2]Radboud University Nijmegen
karin.wanrooij@uva.nl, J.deVos@donders.ru.nl, paul.boersma@uva.nl

ABSTRACT

Distributional vowel training for adults has been reported as "effective" for Spanish and Bulgarian learners of Dutch vowels, in studies using a behavioural task. A recent study did not yield a similar clear learning effect for Dutch learners of the English vowel contrast /æ/~/ɛ/, as measured with event-related potentials (ERPs). The present study aimed to examine the possibility that the latter result was related to the method. As in the ERP study, we tested whether distributional training improved Dutch adult learners' perception of English /æ/~/ɛ/. However, we measured behaviour instead of ERPs, in a design identical to that used in the previous studies with Spanish learners. The results do not support an effect of distributional training and thus "replicate" the ERP study. We conclude that it remains unclear whether distributional vowel training is effective for Dutch adults.

Keywords: distributional training, category learning, speech perception, L2 acquisition.

1. INTRODUCTION

Distributional learning is learning from simple exposure to ambient distributions of stimuli. The mechanism supposedly plays a role in learning native speech sounds in infancy [12, 13, 18] and non-native speech sounds in adulthood [8, 10, 11]. Distributional learning has been used in the lab for training adult participants on difficult non-native speech sound contrasts. In such *distributional training* experiments the participants are exposed to manipulated speech sound distributions, after which the effect of the exposure on their speech perception is measured [4, 6, 17, 19, 20].

A typical distributional training experiment has an experimental group and a control group. The experimental group is exposed to a speech sound distribution that represents a contrast between *two* speech sound categories and is hence *bimodal*. The contrast to be trained is difficult for the participants (e.g., Spanish listeners), who tend to perceive tokens of the two non-native categories in the contrast (e.g., Dutch /ɑ/ and /aː/) as tokens of a single native category (e.g., Spanish /a/) [4, 17, 20]. The control group is usually exposed to a *unimodal* distribution reflecting the corresponding *single* native category [6, 10, 11, 19] or to non-speech [4, 17, 20]. If distributional learning occurs, the bimodally trained participants will be better at perceiving the non-native contrast after the training than the participants in the control group. Distributional training is a potentially interesting way of training due to its simplicity (an effect can be obtained without feedback or instruction) and brevity (exposure duration is limited to only a few minutes).

Previous studies report an effect of distributional vowel training in adults: one study observes such an effect in Bulgarian adult learners of the Dutch vowel contrasts /ɑ/~/aː/ and /ɪ/~/i/ [6], and three studies observe it in Spanish adult learners of the Dutch vowel contrast /ɑ/~/aː/ [4, 17, 20]. A recent study [19] examined whether distributional training can also be effective for *Dutch* adult learners of the Southern British English (SBE) vowel contrast /æ/~/ɛ/ (Dutch listeners tend to perceive both vowels in this contrast as Dutch /ɛ/, and thus find it difficult to perceive a difference between /æ/ and /ɛ/ [15, 21]). For this, a novel method was used to assess vowel perception after training, namely the measurement of the mismatch negativity (MMN), a brain response that can be computed from event-related potentials (ERPs), instead of the measurement of behaviour as used in the earlier experiments with Bulgarian and Spanish participants. The ERP study with the Dutch learners yielded a non-significant effect of distributional training.

To explore the possibility that the non-significance was related to the ERP method, we tested Dutch adults' capacity for distributional learning of the SBE contrast /æ/~/ɛ/, just as in [19], but in contrast to [19] we used the same behavioural method as in the previous studies on distributional vowel training in Spanish adults [4, 17, 20]. If, just as in [19], an effect of distributional training did not surface, the ERP method would not necessarily be inadequate for measuring distributional learning in adults. If an effect of distributional training emerged, the lack of clear positive evidence for it in [19] could be due to the ERP method (or to chance of course).

2. METHOD

The method was identical to that used in previous research on Spanish adults' distributional learning of the Dutch contrast /ɑ/~/aː/ ([4, 20], henceforth "the behavioural studies"), except for the use of a different contrast (SBE /æ/~/ɛ/) and of participants sampled from a different population (Dutch listeners), which were identical to the previous ERP study ([19], henceforth "the ERP study").

All participants performed a pre-test, a training and a post-test. During the *training* half of them (the experimental group) listened to training vowels representative of the SBE vowels /æ/ and /ɛ/, in a distributional training (section 2.2). The other half (the control group) listened to music. The *pre- and post-tests* served to assess participants' classification accuracy of vowel tokens representative of SBE /æ/ and /ɛ/ (section 2.3). It was expected that the listeners in the experimental group would improve more in classification accuracy than the control group, i.e., their improvement score (= post-test – pre-test accuracy score) would be higher.

2.1. Participants

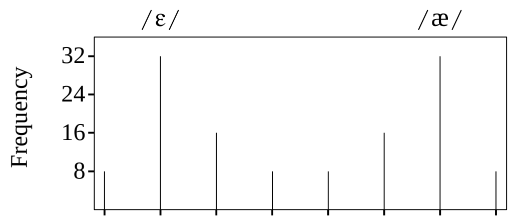
The 100 participants were native speakers of Dutch who had been raised monolingually. They were assigned semi-randomly (i.e., sex was controlled for, 18 men and 32 women in each group) to either the experimental group (mean age = 22.3 years; age range = 18–30 years) or the control group (mean age = 22.3 years; age range = 18–28 years).

2.2. Training

During training, the experimental group (the "Bimodal group") was presented with a vowel distribution representative of the SBE contrast /æ/~/ɛ/, while the control group (the "Music group") listened to instrumental classical music. As shown in Figure 1, the distribution presented to the experimental group contained eight acoustically different vowel tokens (eight vertical lines along the acoustic continuum on the x-axis), of which four represented /ɛ/ (token number 1 – 4) and four /æ/ (token number 5 – 8). Because the distribution represents *two* vowel categories, it is bimodal.

The eight vowel tokens were created with the Klatt synthesizer in Praat [2]. The manipulated acoustic properties were the first and second formants (F1 and F2). Figure 1 shows the F1 and F2 values (in ERB, Equivalent Rectangular Bandwidth) for each of the eight tokens.

Figure 1: Bimodal training distribution. (F1 and F2 in ERB).



Token:	1	2	3	4	5	6	7	8
F1:	9.9	10.4	10.8	11.3	11.7	12.1	12.6	13.0
F2:	20.7	20.4	20.1	19.8	19.5	19.2	18.9	18.6

The F1 and F2 values were calculated in the same way as in the behavioural studies [4, 20]. First, the mean F1 and F2 values of /ɛ/ and /æ/, and the standard deviation for the F1 and F2 values were determined on the basis of values reported in the literature [7, 19]. Then the edges of the F1 range were calculated by subtracting the standard deviation of F1 from the mean F1 value of /ɛ/ (token 1) and adding it to the mean F1 value of /æ/ (token 8). Similarly, the edges of the F2 range were computed by adding the standard deviation of F2 to the mean F2 value of /ɛ/ (token 1) and subtracting it from the mean F2 value of /æ/ (token 8). The F1 and F2 values of the intermediate tokens along the F1 and F2 ranges (i.e., stimuli 2 through 7) were calculated by linear extrapolation, where each step between consecutive tokens was roughly equal on the psychoacoustic ERB scale. The resulting step sizes (i.e., 0.4 ERB for F1 and 0.3 ERB for F2) were comparable to the step sizes in the behavioural studies [4, 20] (i.e., 0.4 ERB for F1 and F2).

Each token was filtered with eight additional formants (F3 = 2400 Hz, F4 = 3400 Hz, F5 = 4050 Hz, F6 through F10: previous formant plus 1000 Hz), and had a duration of 140 ms and a fundamental frequency (F0) that fell from 150 Hz to 100 Hz.

Each of the eight tokens was repeated a certain number of times (see the y-axis in Figure 1) so as to create a bimodal distribution. The bimodality is evident from the presence of *two* peaks (around tokens 2 and 7, which were presented most often). The total number of presentations during the training was 128: 32 times for stimuli 2 and 7, 16 times for stimuli 3 and 6, and 8 times for stimuli 4 and 5.

Token presentation was randomized per participant. The inter-stimulus interval (ISI) was 750 ms. Total training time was thus 1.9 minutes (=128 stimuli *[140 ms duration + 750 ms ISI]).

Before the training, participants in the experimental group were instructed to listen to the vowels carefully, because they would perform a second task (post-test) similar to the first one (pre-test) after the training. Participants in the control group were told that they would listen to classical music and could relax, after which they would perform another task similar to the first one.

2.3. Pre- and post-tests

During the pre- and post-tests, classification accuracy (in percent correct) of multiple tokens of [æ] and [ɛ] was assessed in a forced-choice XAB-task. Each task had 80 trials. In each of these trials, participants heard three stimuli: an X, A and B stimulus. After this, they had to indicate whether the first vowel (X) was more similar to the second vowel (A) or to the third vowel (B), by clicking on "1" (for A) or "2" (for B) on a computer screen.

The 80 *X stimuli* (40 for /æ/ and 40 for /ɛ/) in the tests were unique natural tokens of English /æ/ and /ɛ/ produced by six female and five male native speakers of SBE. Two productions of /æ/ and /ɛ/ each were provided by Daniel Williams [23]. Another two productions of /æ/ and /ɛ/ each were extracted from a subset of stimuli reported in [5]. The remaining stimuli were recorded in our lab.

Most tokens were extracted from a /h-V-d/ context (*head / had*) or a /f-V-f/ context (*fef / faf*). To add variation, some tokens were extracted from a /s-V-s/, /b-V-s/, /h-V-s/, /m-V-s/ or /t-V-s/-context. Table 1 lists the average F0, F1, F2 and duration of /æ/ and /ɛ/ for the female and male speakers separately.

Table 1: Mean F1, F2, F0 (in ERB) and duration (in ms) of the X stimuli in the XAB-test. Standard deviations between tokens are given between parentheses.

Acoustic property	Sex speaker	/æ/	/ɛ/
F1	**F**	14.96 (0.84)	12.38 (1.01)
	M	13.26 (0.45)	11.35 (0.57)
F2	**F**	19.18 (0.69)	20.97 (0.83)
	M	18.20 (0.73)	19.35 (0.69)
F0	**F**	5.34 (0.41)	5.51 (0.49)
	M	3.41 (0.39)	3.43 (0.39)
Duration	**F**	123.88 (26.50)	118.20 (23.00)
	M	113.57 (23.83)	97.40 (26.69)

The *response options A and B* were synthetic tokens created in Praat [2]. For [ɛ], F1 and F2 were 10.95 ERB and 20.04 ERB respectively. For [æ], F1 and F2 were 11.99 ERB and 19.32 ERB respectively. Both response options had eight additional formants with the same values as those for the training stimuli (section 2.2), a duration of 140 ms and an F0 that fell from 150 Hz to 100 Hz.

Trial order was randomized per participant and the presentation of the A and B stimuli was counterbalanced across trials. The ISI between the three stimuli (X, A and B) in each trial was 1.2 seconds. This relatively long ISI as well as the use of many different natural X tokens served to make participants *classify* rather than discriminate the X stimuli [16, 22].

Each test took circa 7 minutes. Test duration differed slightly per participant depending on how fast participants clicked on a response option (the next trial would only begin after clicking), and on whether the participant chose to take a short break (available after every 20 trials).

3. RESULTS

For each participant the percentage of correctly classified vowels was computed. Table 2 shows the mean pre-test and post-test accuracy percentages, and the improvement scores (= the post-test – pre-test accuracy percentage) for the Bimodal and Music groups separately.

Table 2: Pre- and post-test accuracy percentages, and the improvement score (= post-test – pre-test accuracy percentage), for the Bimodal and Music groups. Standard deviations between participants in each group are given between parentheses.

Group	**Bimodal**	**Music**
Pre-test	64.98 (12.03)	64.33 (9.97)
Post-test	68.55 (14.29)	71.50 (12.53)
Improvement (post – pre)	3.58 (7.51)	7.18 (7.68)

A repeated-measures ANOVA with the accuracy percentage as the dependent variable, Training Type (bimodal vs. music) as between-subject factor and Test (pre- vs. post-test) as within-subject factor showed a significant main effect of Test (mean difference = +5.38%, CI = +3.87 ~ +6.88%, $F[1,98] = 50.03$, $p < 0.001$): the accuracy percentage was higher after (70.03%, CI = +67.36 ~ +72.69%) than before (64.65%, CI = +62.46 ~ +66.84) the training phase. The main effect of Training Type was not significant (p = 0.62). Thus, the Bimodal group did not score significantly higher or lower than the Music group across the two tests. Crucially, the interaction between Training Type and Test was significant ($F[1,98] = 5.61$, $p = 0.02$). This indicates that the two groups did not improve equally. Table 2 illustrates, however, that the Bimodal group did not improve more than the Music group, as was expected, but *less*.

4. DISCUSSION

The current study did not yield a straightforward effect of distributional training in Dutch adult learners of the SBE vowel contrast /æ/~/ɛ/ when repeating a behavioural experiment that had shown such an effect in Spanish adult learners of Dutch /ɑ/~/aː/ [4, 17, 20]: contrary to expectation, the control group who listened to music improved *more* in their perception of this contrast than the bimodally trained participants. Since the control group did not receive distributional training, its larger improvement cannot be attributed to distributional learning. The reason why the control group improved more remains unclear. Notice, however, that the pattern of a control group outperforming the experimental group is not without precedent in adult distributional training experiments: Hayes-Harb [8] also obtained a better perception of the contrast in the bimodal distribution for a group that received no training than for the bimodally trained group.

The present result "replicates" the outcome of an ERP study [19] where a non-significant and thus also a non-transparent effect of distributional training in Dutch adult learners of /æ/~/ɛ/ was obtained with a different method, namely the measurement of ERPs. The purpose of the present paper was to explore the possibility that the ERP method somehow prevented the detection of a significant effect. The current outcome does not provide clear evidence that the ERP method is unsuitable for measuring distributional learning in adults.

One may wonder why the experiments in [19] and in the current study did not yield a clear distributional training effect for Dutch learners, while the experiments in [4], [6], [17] and [20] did for Spanish and Bulgarian learners. Unfortunately, the present study does not shed light on this issue. We can only conclude that it remains uncertain whether distributional vowel training is useful for Dutch adult learners.

5. REFERENCES

[1] Adank, P., Van Hout, R., Smits, R. 2004. An acoustic description of the vowels of Northern and Southern standard Dutch. *J. Acoust. Soc. Am*. 116, 1729–1738.

[2] Boersma, P., Weenink, D. 2012. Praat: Doing Phonetics by Computer. Available at: http://www.praat.org (accessed 2012).

[3] Bradlow, A.R. 1995. A comparative acoustic study of English and Spanish vowels. *J. Acoust. Soc. Am*. 97, 1916–1924.

[4] Escudero, P., Benders, T., Wanrooij, K. 2011. Enhanced bimodal distributions facilitate the

learning of second language vowels. *J. Acoust. Soc. Am*. 130, EL206–EL212.

[5] Escudero, P., Simon, E., Mitterer, H. 2012. The perception of English front vowels by North Holland and Flemish listeners: acoustic similarity predicts and explains cross-linguistic and L2 perception. *J. Phon.* 40, 280–288.

[6] Gulian, M., Escudero, P., Boersma, P. 2007. Supervision hampers distributional learning of vowel contrasts. *Proc. 16th ICPhS* Saarbrücken, 1893–1896.

[7] Hawkins, S., Midgley, J. 2005. Formant frequencies of RP monophthongs in four age groups of speakers. *J. Int. Phon. Assoc.* 35, 183–199.

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

[9] Klagstad, H.L. 1958. The phonemic system of colloquial Standard Bulgarian. *American Association of Teachers of Slavic and East European Languages*, 2, 42–54.

[10] Maye, J., Gerken, LA. 2000. Learning phonemes without minimal pairs, *BUCLD Proc.* 24, 522–533.

[11] Maye, J., Gerken, LA. 2001. Learning phonemes: how far can the input take us? *BUCLD Proc.* 25, 480–490.

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

[13] Maye, J., Werker, J.F., Gerken, LA. 2002. Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition* 82, B101–B111.

[14] Meunier, C., Frenk-Mestre, C., Lelekov-Boissard, T., Le Besnerais, M. 2003. Production and perception of vowels: does the density of the system play a role? *Proc. 15th ICPhS* Barcelona, 723–726.

[15] Schouten, M.E.H. 1975. *Native-language interference in the perception of second-language vowels: an investigation of certain aspects of the acquisition of a second language*. Doctoral dissertation, Utrecht University.

[16] Van Hessen, A.J., Schouten, M.E.H. 1999. Categorical perception as a function of stimulus quality. *Phonetica* 56, 56–72.

[17] Wanrooij, K., Boersma, P. 2013. Distributional training of speech sounds can be done with continuous distributions. *J. Acoust. Soc. Am*. 133, EL398–EL404.

[18] Wanrooij, K., Boersma, P., Van Zuijen, T.L. 2014a. Fast phonetic learning occurs already in 2-to-3-month old infants: an ERP study. *Front. Psychol. (Language Sciences)* 5, article 77, 1–12.

[19] Wanrooij, K., Boersma, P., Van Zuijen, T.L. 2014b. Distributional vowel training is less effective for adults than for infants. A study using the mismatch response. *PloS One* 9, 1–13.

[20] 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.

[21] Weber, A., Cutler, A. 2004. Lexical competition in non-native spoken-word recognition. *J. of Mem. Lang.* 50, 1–25.

[22] Werker, J.F., Logan, J.S. 1985. Cross-language evidence for three factors in speech perception. *Percept. Psychophys.* 37, 35–44.

[23] Williams, D., Escudero, P. 2014. A cross-dialectal acoustic comparison of Northern and Southern British English vowels. *J. Acoust. Soc. Am*. 136: 2751–2761.