SEPTEMBER 09 2011

Enhanced bimodal distributions facilitate the learning of second language vowels

Paola Escudero; Titia Benders; Karin Wanrooij

J. Acoust. Soc. Am. 130, EL206–EL212 (2011)
https://doi.org/10.1121/1.3629144

Escudero *et al.*: JASA Express Letters [DOI: 10.1121/1.3629144] Published Online 09 September 2011

Enhanced bimodal distributions facilitate the learning of second language vowels

Paola Escudero
MARCS Auditory Laboratories, Building 1, University of Western Sydney, Bullecourt Avenue, Milperra, New South Wales 2214, Australia
paola.escudero@uws.edu.au

Titia Benders and Karin Wanrooij
Amsterdam Center for Language and Communication, University of Amsterdam, Spuistraat 210, 1012 VT Amsterdam, The Netherlands
titia.benders@uva.nl, karin.wanrooij@uva.nl

This study addresses the questions of whether listening to a bimodal distribution of vowels improves adult learners' categorization of a difficult L2 vowel contrast and whether enhancing the acoustic differences between the vowels in the distribution yields better categorization performance. Spanish learners of Dutch were trained on a natural bimodal or an enhanced bimodal distribution of the Dutch vowels /ɑ/ and /aː/, with the average productions of the vowels or more extreme values as the endpoints respectively. Categorization improved for learners who listened to the enhanced distribution, which suggests that adults profit from input with properties similar to infant-directed speech.

PACS numbers: 43.71.Ft, 43.71.Es, 43.71.An [James Hillenbrand]
Date Received: May 25, 2011 **Date Accepted:** July 01, 2011

1. Introduction

Adult second language (L2) learners are known to experience prolonged difficulty in identifying certain L2 phoneme contrasts. A well-known example of a difficult L2 contrast is English /ɹ/ - /l/ for Japanese listeners (Aoyama *et al.*, 2004), who are unable to perceive the third formant differences between the two sounds (Miyawaki *et al.*, 1973; Iverson *et al.*, 2003). The Dutch vowels /ɑ/ and /aː/ are also notoriously difficult for Spanish listeners even after extensive exposure to the Dutch language (Escudero and Wanrooij, 2010). This difficulty seems to be due to the fact that Spanish listeners perceive both Dutch vowels as the single Spanish phoneme /a/. Escudero *et al.* (2009) show that Spanish learners of Dutch are unable to use the spectral difference between /ɑ/ and /aː/ in a native-like manner and tend to resort to the duration difference between the vowels to classify them correctly.

Previous research has shown that adults can improve their discrimination of a difficult non-native contrast through *distributional learning,* i.e., through listening to a bimodal distribution of speech sounds along an acoustic continuum that encompasses the two categories (Maye and Gerken, 2000, 2001; Gulian *et al.*, 2007). These bimodal distributions have two peaks of frequently occurring values, with each peak representing one of the categories. Listeners thus learn to discriminate categories by being exposed to frequent tokens of the sounds near the two ends of the acoustic continuum.

The first objective of the present study was to replicate the finding that distributional learning aids non-native sound perception. This seems pertinent because to date, only a handful of studies have reported positive effects (Maye and Gerken, 2000, 2001; Gulian *et al.*, 2007; Hayes-Harb, 2007). Whereas these previous studies trained listeners who were exposed to a contrast for the first time, we examine whether distributional learning is helpful for learners who live in the L2-speaking country and have prolonged difficulty with the target contrast. Additionally, we were interested in the effect of training on sound categorization rather than discrimination, and therefore

used a highly variable corpus of natural tokens to test the learners' vowel categorization before and after training.

In the present study, Spanish learners of Dutch were exposed to a bimodal distribution based on the spectral difference between the vowels /ɑ/ and /aː/. The duration difference between the vowels was not included in the training because categories that vary in multiple dimensions are more difficult to learn than those that vary in a single dimension (Goudbeek *et al.*, 2009). In addition, it has been shown that learners' perception of difficult sound contrasts improves through extensive training that involves the selective attention to the most relevant acoustic dimension (Iverson *et al.*, 2005; Kondaurova and Francis, 2010). Recall that Spanish learners of Dutch tend to rely on vowel duration rather than spectral properties to classify Dutch /ɑ/ and /aː/ (Escudero *et al.*, 2009). Therefore, we examined whether a two-minute exposure to a bimodal distribution of their least favored acoustic dimension for this Dutch contrast, namely, spectral quality, would help them to classify these vowels more accurately.

The second objective of the present study was to establish whether an *enhanced* bimodal distribution would lead to a better classification performance than a *natural* bimodal distribution. As shown in Fig. 1, the two distributions differ in the values of their tokens. That is, the endpoints of the vowel continuum in the natural bimodal distribution (dashed line in the figure) have values that are based on typical productions of /ɑ/ and /aː/ in the Dutch language. In contrast, the endpoints of the enhanced bimodal distribution (solid line in the figure) have values that increase the acoustic difference between the two vowels.

The enhancement of a bimodal distribution was inspired by the properties of *infant-directed speech* and *foreigner-directed speech*. It has been shown that caregivers increase the differences between sound categories when addressing infants (e.g., Kuhl *et al.*, 1997; Burnham *et al.*, 2002) and that these enhanced properties may facilitate their speech discrimination (Liu *et al.*, 2003). Similarly, native speakers increase the differences between speech sound categories when talking to foreigners (Scarborough *et al.*, 2007). We thus predicted that adult learners would benefit more from an enhanced than from a natural bimodal distribution. Although the enhancement of speech properties has led to success within other training methods such as *perceptual fading* (Jamieson and Morosan, 1986; Iverson *et al.*, 2005; Kondaurova and Francis, 2010), sessions usually involve explicit feedback and are always substantially longer than the two-minute training of the present study.

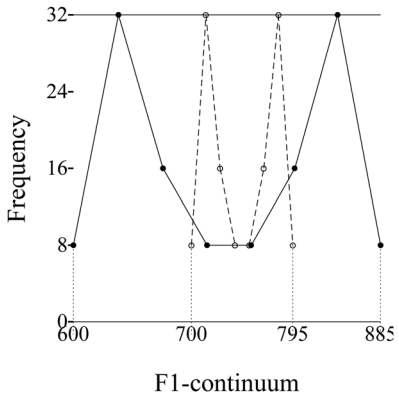

Fig. 1. Bimodal (dashed) and enhanced bimodal (solid) distributions. The F1-values on the *x* axis are displayed along the Erb scale. The endpoints of the continua are given in Hz.

2. Methodology

2.1 Participants

A total of 159 Spanish-speaking learners of Dutch from Spain and various Latin American countries participated in the study. Their ages ranged between 24 and 63 yr. They had lived in the Netherlands between 2 weeks and 20 yr at the time of testing. There was a large variability in their level of proficiency in the Dutch language, ranging from basic to advanced. This level was measured in self-reports and in a general listening comprehension test (Dialang by Alderson and Huhta, 2005), which was administered at the end of the session to those listeners who reported to have some knowledge of the Dutch language. Listeners were randomly assigned to three different groups, namely, enhanced, bimodal and music, of 53 participants each. The average age, length of residence in the Netherlands, and proficiency in the Dutch language was comparable across groups.

2.2 Stimuli and procedure

Listeners performed a perception task (pre-test) which was followed by a short training phase and a second presentation of the same perception task (post-test). The pre- and post-tests were administered to all listeners. The three groups of listeners differed in the stimuli they heard during the training phase: the two experimental groups were presented with bimodal or enhanced vowel distributions, while the control group listened to classical music.

The pre- and post-tests were identical two-alternative forced choice categorization tasks in an *XAB* format. In each trial, listeners heard three sounds and were asked to decide whether the first sound (*X*) was more like the second (*A*) or the third (*B*). The 40 *X* stimuli were naturally produced tokens of Dutch /ɑ/ and /aː/, which had been extracted from a /s-V-s/ context produced in a carrier sentence. They were a subset of the corpus of Adank *et al.* (2004) and were produced by 10 male and 10 female speakers of Standard Northern Dutch. Table 1 shows the average first formant (F1), second formant (F2), measured at the midpoint of the vowel, and duration values of /ɑ/ and /aː/ for the female and male speakers separately.

The table also shows the acoustic properties of the two auditory responses, *A* and *B*, which were synthetic tokens created in the Praat program (Boersma and Weenink, 2011). These tokens were generated from a source signal of 140 ms. with a fundamental frequency (F0) that fell from 150 to 100 Hz. This source signal was then filtered with 10 formants. The values of F1 and F2 were based on those reported in Pols *et al.* (1973) for natural productions of Dutch /ɑ/ and /aː/. Eight formants were added for more natural sounding stimuli. Given that the response tokens had the same F0 and duration, listeners could only use spectral properties to classify the *X* stimuli. Escudero and Wanrooij (2010) found, in an identical task with the same stimuli, that 20 Dutch natives accurately classified the *X* tokens 88% of the time, which indicates that the *X*'s are good examples of the Dutch vowels (though with considerable

Table 1. Average duration (in milliseconds), F1 and F2 values (in Hz) of the *X*-stimuli (females and males) and the auditory responses.

		/aː/	/ɑ/	/aː/	/ɑ/
		X-stimuli		Auditory response options A and B	
Duration	Females	216	93	140	140
	Males	204	94		
F1	Females	923	719	770	687
	Males	652	584		
F2	Females	1 552	1 239	1 303	1 104
	Males	1 424	1 156		

variability) and that the *A* and *B* tokens are good auditory labels for them (despite the absence of duration differences).

There were 20 trials where the *X* stimulus was the vowel /ɑ/ and 20 where it was /aː/. The presentation of the *A* and *B* stimuli was counterbalanced across trials and trial order was randomized per participant. The inter-stimulus interval (ISI) between the three sounds in each *XAB*-trial was 1.2 s. Listeners were encouraged to respond as quickly as possible and to guess if unsure. They were told that the next trial would only appear after their response.

The bimodal and enhanced training conditions contained eight synthetic vowel tokens, which were generated using the same procedure as for the *A* and *B* stimuli. These eight tokens were presented to listeners according to the frequency distributions shown in Table 2 and Fig. 1. The shape of the distributions was identical to that used in previous studies with adult listeners (e.g., Maye and Gerken, 2000, 2001; Gulian *et al.*, 2007; Hayes-Harb, 2007), with the near-endpoint tokens 2 and 7 being presented four times as often as the center tokens 4 and 5. The training contained 128 tokens with an ISI of 750 ms., for a total of two minutes.

Table 2 also shows the F1 and F2 values of each of the eight tokens for the two training conditions. The endpoints of both the bimodal and enhanced continua (tokens 1 and 8) were based on the values measured for male Dutch speakers by Pols *et al.* (1973). The endpoints of the bimodal continuum have values that are comparable to the mean productions of the vowels /ɑ/ and /aː/, respectively, while the endpoints of the enhanced continuum were chosen to exaggerate the natural difference between the vowels in the bimodal continuum. For the enhanced endpoint of /ɑ/, we *subtracted* the standard deviations of F1 and F2 from the vowel's average F1 and F2 values, while for the enhanced endpoint of /aː/, we *added* the standard deviations to its average F1 and F2 values. The steps between the eight tokens for both continua were approximately equal on the psychoacoustic Erb scale (bimodal: F1: 0.1 Erb, F2: 0.2 Erb; enhanced: F1: 0.4 Erb, F2: 0.4 Erb). All tokens in the training continua had a duration of 140 ms. and an F0 that fell from 150 to 100 Hz. Although the most frequent tokens in the bimodal continuum (tokens 2 and 7) differ little in their F1 and F2 values (in Hz), these values and their acoustic distance closely resemble those of the *A* and *B* tokens, which Dutch natives could easily classify as Dutch /ɑ/ and /aː/.

During training, participants in the experimental groups were instructed to listen to the vowels carefully because they would perform another vowel classification test afterward. The participants in the control group were instructed to relax while listening to classical music and were also told that they would perform a second classification task afterward.

3. Results

The average number of correct responses in the pre- and post-tests are given in Table 3. An analysis of variance (ANOVA) on the correct pre-test responses with group (music, bimodal, and enhanced) as the independent variable revealed no significant difference ($F(1, 157) = 0.22, p = 0.64$). This finding allowed us to compare the effect of

Table 2. Frequency of presentation and acoustic properties (F1 and F2 in Hz) of the eight tokens, in the bimodal and enhanced training distributions.

Token number		1	2	3	4	5	6	7	8
Token frequency		8	32	16	8	8	16	32	8
Bimodal	*F1*	700	713	726	740	753	767	781	795
	F2	1 115	1 144	1 174	1 204	1 235	1 266	1 298	1 330
Enhanced	*F1*	600	637	675	714	755	797	840	885
	F2	1 000	1 055	1 112	1 171	1 233	1 296	1 362	1 430

Table 3. Average number of correct responses (out of 40) for the pre- and post-tests and their average difference in the three conditions. Standard deviations are between parentheses.

	Music	Bimodal	Enhanced
Pre-test	22.72 (4.55)	23.49 (4.61)	22.28 (5.02)
Post-test	22.66 (5.69)	23.81 (4.82)	24.70 (5.12)
Difference	−0.06 (4.87)	0.32 (4.39)	2.42 (4.75)

training between the three groups. Three one-sample *t*-tests confirmed that the subjects in the three groups scored significantly above chance (music: $t(52) = 4.35$, $p < 0.01$, $t(52) = 4.35$, $p < 0.01$; bimodal: $t(52) = 5.51$, $p < 0.01$; enhanced: $t(52) = 3.31, p < 0.01$).

The table also shows the *difference* between the number of correct responses in the post- and pre-tests, which is a measure of improvement after training. An ANOVA on this difference score with group as the independent variable yielded a significant main effect of group ($F(2, 156) = 4.30, p < 0.05$), which indicates that the three groups did not have equal improvement after training. *Post hoc* comparisons using Tukey's HSD revealed significant differences between the music and enhanced groups (difference = 2.47, $p < 0.05$) and an almost significant difference between the bimodal and enhanced groups (difference = 2.09, $p = 0.058$), but no significant difference between the music and bimodal groups (difference = 0.38, $p = 0.91$). One-sample *t*-tests comparing the difference between post- and pre-tests in each of the three groups to 0 only yielded significance for the enhanced training group ($t(52) = 3.70$, $p < 0.01$; bimodal: $t(52) = 0.53$, $p = 0.60$; music: $t(52) = -0.08$, $p = 0.93$).

4. Discussion

The findings of the present study confirm that distributional learning results in an improvement of L2 sound perception. Moreover, they show that a bimodal distribution with enhanced differences, i.e., with larger distances between endpoints, is beneficial. From these results, we can conclude that adult learners' classification of a difficult L2 vowel contrast is facilitated by enhanced distributional training.

The finding that enhanced distributional training is more beneficial than natural bimodal training relates to the literature on foreigner-directed speech. The current study suggests that the way native speakers talk to foreigners (Scarborough *et al.*, 2007) helps second language learners to learn sound contrasts more efficiently. Recall that infant-directed speech is also characterized by the increased contrast of acoustic properties, which is hypothesized to aid speech sound discrimination in infants (Liu *et al.*, 2003). Therefore, it would be interesting to compare the acoustic enhancement in infant- and foreigner-directed speech and relate the outcome to the difference in native and non-native perception.

Kuhl *et al.* (1997) suggest that the enhanced properties of infant-directed speech allow for more variation within each sound category and that this enhancement therefore helps infants to generalize their learning to tokens produced by new speakers. Our enhanced distributional training may have also helped the L2 learners to more accurately categorize the highly variable *X* tokens, which were produced by many different female and male speakers. Another possible reason for the success of the enhanced training may lie in the fact that its extreme values facilitated the learners' selective attention to the relevant cue, i.e., the spectral difference between Dutch /ɑ/ and /aː/, and thus lead to higher categorization accuracy. This possibility is supported by previous L2 training studies (Iverson *et al.*, 2005; Kondaurova and Francis, 2010) which used enhanced input as well as explicit feedback to promote selective attention.

In sum, we have shown that enhanced distributional training yields an improvement of 6% (2 to 3 tokens out of 40) in two minutes, while other training

methods, which may lead to higher levels of improvement, demand much longer sessions (e.g., Tremblay *et al.*, 1998; Iverson *et al.*, 2005; Wade and Holt, 2005; Zhang *et al.*, 2009; Kondaurova and Francis, 2011). Importantly, distributional learning does not require any kind of feedback, which is commonly used in other training methods. Thus, the rapid efficacy and relative simplicity of our enhanced distributional training make it a potentially powerful tool for the perceptual training of difficult non-native contrasts.

Acknowledgments

This research was supported by Grant No. 275.75.005 from the Netherlands Organization for Scientific Research (NWO) awarded to the first author. Research assistants were also supported by NWO Grant No. 016.024.018 awarded to Paul Boersma. We would like to thank Paul Boersma and Kateřina Chládková for their comments on earlier versions.

References and links

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

Alderson, J. C., and Huhta, A. (**2005**). "The development of a suite of computer-based diagnostic tests based on the Common European Framework," Lang. Test. **22**, 301–320.

Aoyama, K., Flege, J. E., Guion, S. G., Akahane-Yamada, R., and Yamada, T. (**2004**). "Perceived Phonetic dissimilarity and L2 speech learning: The case of Japanese /ɾ/ and English /l/ and /ɹ/," J. Phonetics **32**, 233–250.

Boersma, P., and Weenink, D. (**2011**). "Praat: Doing phonetics by computer," http://www.praat.org (Last viewed 10/12/2007).

Burnham, D., Kitamura, C., and Vollmer-Conna, U. (**2002**). "What's new, pussycat? On talking to babies and Animals," Science **296**(5572), 1435–1435.

Escudero, P., and Wanrooij, K., (**2010**). "The effect of L1 orthography on non-native and L2 vowel perception," Lang. Speech **53**, 343–365.

Escudero, P., Benders, T., and Lipski, S. (**2009**). "Native, non-native and L2 perceptual cue weighting for Dutch vowels: the case of Dutch, German and Spanish listeners," J. Phonetics **37**, 452–465.

Gulian, M., Escudero, P., and Boersma, P. (**2007**). "Supervision hampers distributional learning of vowel contrasts," *Proceedings of the 15th International Congress of Phonetic Sciences*, pp. 893–1896.

Goudbeek, M., Swingley, D., and Smits, R. (**2009**). "Supervised and unsupervised learning of multidimensional acoustic categories," J. Exp. Psychol. **35**, 1913–1933.

Hayes-Harb, R. (**2007**). "Lexical and statistical evidence in the acquisition of second language phonemes," Second Lang. Res. **23**, 1–31.

Iverson, P., Kuhl, P. K., Akahane-Yamada, R., Diesch, E., Tohkura, Y., Kettermann, A., and Siebert, C. (**2003**). "A perceptual interference account of acquisition difficulties for non-native phonemes," Cognition **87**, B47–B57.

Iverson, P., Hazan, V., and Bannister, K. (**2005**). "Phonetic training with acoustic cue manipulation: A comparison of methods for teaching English /r/-/l/ to Japanese adults," J. Acoust. Soc. Am. **118**, 3267–3278.

Jamieson, D. G., and Morosan, D. E. (**1986**). "Training non-native speech contrasts in adults: Acquisition of the English /ð/ - /θ/ contrast by francophones," Perception Psychophys. **40**(4), 205–215.

Kondaurova, M., and Francis, A. (**2010**). "The role of selective attention in the acquisition of English tense and lax vowels by native Spanish listeners: Comparison of three training methods," J. Phonetics **38**, 569–587.

Maye, J., and Gerken, L. (**2000**). "Learning phonemes without minimal pairs," in *BUCLD 24 Proceedings*, edited by C. Howell, S. A. Fish, and T. Keith-Lucas (Cascadilla, Somerville, MA), pp. 522–533.

Kuhl, P., Andruski, J., Chistovich, I., Chistovich, L., Kozhevnikova, E., Ryskina, V., Stolyarova, E., Sundberg, U., and Lacerda, F. (**1997**). "Cross-language analysis of phonetic units in language addressed to infants," Science **227**(5326), 684–686.

Liu, H-M., Kuhl, P., and Tsao, F-M. (**2003**). "An association between mothers' speech clarity and infants' speech discrimination skills," Develop. Sci. **6**(3), 1–10.

Maye, J., and Gerken, L. (**2001**). "Learning phonemes: how far can the input take us?," in *BUCLD 25 Proceedings*, edited by A. H.-J. Do, L. Dominquez, and A. Johansen (Cascadilla, Somerville, MA), pp. 480–490.

Maye, J. C., Werker, J. F., and Gerken, L. A. (**2002**). "Infant sensitivity to distributional information can affect phonetic discrimination," Cognition **82**, B101–B111.

Miyawaki, K., Strange, W., Verbrugge, R. R., Liberman, A. M., Jenkins, J. J., and Fujimura, O. (**1975**). "An effect of linguistic experience: The discrimination of [r] and [l] by native speakers of Japanese and English," Perception Psychophys. **18**, 331–340.
Pols, L. C. W., Tromp, H. R. C., and Plomp, R. (**1973**). "Frequency analysis of Dutch vowels from 50 male speakers," J. Acoust. Soc. Am. **53**, 1093–1101.
Scarborough, R., Brenier, J., Zhao, Y., Hall-Lew, L., and Dmitrieva, O. (**2007**). "An acoustic study of real and imagined foreigner-directed speech," *Proceedings of the 15th International Congress of Phonetic Sciences*, pp. 2165–2168.
Tremblay, K., Kraus, N., and McGee, T. (**1998**). "The time course of auditory perceptual learning: Neurophysiological changes during speech-sound training," NeuroReport **9**, 3557–3560.
Wade, T., and Holt, L. (**2005**). "Incidental categorization of spectrally complex non-invariant auditory stimuli in a computer game task," J. Acoust. Soc. Am. **118**, 2618–2633.
Zhang, Y., Kuhl, P., Imada, T., Iverson, P., Pruitt, J., Stevens, E., Kawakatsu, M., Tohkura, Y., and Nemoto, I. (**2009**). "Neural signatures of phonetic learning in adulthood: A magnetoencephalography study," NeuroImage **46**, 226–240.