Journal of Phonetics 41 (2013) 307–319

Contents lists available at SciVerse ScienceDirect

Journal of Phonetics

journal homepage: www.elsevier.com/locate/phonetics

What do listeners learn from exposure to a vowel distribution? An analysis of listening strategies in distributional learning

Karin Wanrooij [a,*], Paola Escudero [b], Maartje E.J. Raijmakers [c]

[a] Amsterdam Center for Language and Communication, University of Amsterdam, Spuistraat 210, 1012 VT Amsterdam, The Netherlands
[b] MARCS Institute, University of Western Sydney, Locked Bag 1797, Penrith South DC, NSW 2751, Australia
[c] Department of Developmental Psychology, University of Amsterdam, Weesperplein 4, 1018 XA Amsterdam, The Netherlands

ARTICLE INFO

Article history:
Received 13 March 2012
Received in revised form
24 March 2013
Accepted 29 March 2013
Available online 20 June 2013

ABSTRACT

This study first confirms the previous finding that Spanish learners improve their perception of a difficult Dutch vowel contrast through listening to a frequency distribution of the vowels involved in the contrast, a technique also known as *distributional training*. Secondly, it is demonstrated that learners' initial use of acoustic cues influences their performance after distributional training. To that end, types of unique *listening strategies*, i.e., specific ways of using acoustic cues in vowel perception, are identified using *latent class regression models*. The results before training show a split between "low performers", who did not use the two most important cues to the Dutch vowel contrast, namely the first and second vowel formants, and "high performers", who did. Distributional training diversified the strategies and influenced the two types of listeners differently. Crucially, not only did it bootstrap the use of cues present in the training stimuli but also the use of an untrained cue, namely vowel duration. We discuss the implications of our findings for the general field of distributional learning, and compare our listening strategies to the developmental stages that have been proposed for the acquisition of second-language vowels in Spanish learners.

1. Introduction

Learning speech sounds on the basis of frequency distributions is commonly known as 'distributional learning' (Gulian, Escudero, & Boersma, 2007; Maye & Gerken, 2000, 2001; Maye, Weiss, & Aslin, 2008). Distributional learning is considered to be the main mechanism that underlies the acquisition of speech sounds in the first year of life, when infants' sensitivity to native speech sound contrasts (which occur frequently in the infant's environment) increases (e.g., Cheour et al., 1998), while that to non-native speech sound contrasts (which occur infrequently) declines (e.g., Werker & Tees, 1984/2002). Since infants' vocabularies are non-existent or small in the first months of life, another way of learning speech sounds, namely from noticing the difference in meaning between words whose forms differ in only one speech sound, cannot play a dominant role yet (Maye, Werker, & Gerken, 2002; Stager & Werker, 1997). The probable existence of distributional learning as a real mechanism for learning speech sounds has been supported by computer simulations (Guenther & Gjaja, 1996; Lacerda, 1995) and also by observations in the lab, not only for infants (Cristià, McGuire, Seidl, & Francis, 2011; Maye et al., 2002, 2008; Yoshida, Pons, Maye, & Werker, 2010), but also for adults (Escudero, Benders, & Wanrooij, 2011; Gulian et al., 2007; Hayes-Harb, 2007; Maye & Gerken, 2000, 2001).

In distributional learning experiments in the lab, listeners hear a randomly presented series of stimuli that vary in steps along a continuous dimension. Crucially, each stimulus is presented with a certain frequency, such that some stimuli appear more often than others. In this way listeners hear a *distribution* of speech sounds. Two groups of listeners usually participate; one presented with a bimodal and another with a unimodal distribution of speech sounds (e.g., Gulian et al., 2007; Hayes-Harb, 2007; Maye & Gerken, 2000, 2001; Maye et al., 2008). In the former distribution, stimuli with properties near the two endpoints of the acoustic continuum are presented most often, while in the latter, stimuli with properties near the middle of the acoustic continuum are most frequent. After the training phase, both groups of listeners are tested on their ability to discriminate the same two stimuli, which had occurred equally often in both trained distributions. If there is an effect of distributional learning, better discrimination is expected after exposure to bimodal than unimodal distributions. This is because exposure to a bimodal distribution induces the perception of the two test stimuli as exemplars of two different speech sound categories, while listening to a unimodal distribution leads to hearing the same two test stimuli as exemplars of a single speech sound category (e.g., Gulian et al., 2007; Hayes-Harb, 2007; Maye & Gerken, 2000, 2001; Maye et al., 2008).

* Corresponding author. Tel.: +31 20 5253857.
E-mail address: karin.wanrooij@uva.nl (K. Wanrooij).

0095-4470/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.wocn.2013.03.005

Although the original distributional learning studies (Maye & Gerken, 2000, 2001; Maye et al., 2002) aimed at demonstrating that this mechanism underlies the learning of phonetic categories, recent studies have exploited the technique to train difficult non-native speech sound contrasts. Gulian et al. (2007) exposed native Bulgarian speakers to bimodal distributions of the Dutch vowel contrasts /ɑ/–/aː/ and /ɪ/–/i/, which these listeners tend to perceive as the single Bulgarian vowels /a/ and /i/ respectively. After a training phase of only 5 min per vowel contrast, listeners exposed to a bimodal distribution classified the vowels in each contrast more accurately than those exposed to a unimodal distribution. More recently, Escudero et al. (2011) presented Spanish-speaking learners of Dutch with bimodal distributions of Dutch /ɑ/–/aː/. In natural speech, these Dutch vowels differ both in their spectral (/aː/ has higher first and second formants) and durational (/aː/ is longer) properties (Adank, Van Hout, & Smits, 2004; Pols, Tromp, & Plomp, 1973). When classifying the vowels, Spanish learners of Dutch tend to rely on the durational differences, while Dutch natives use spectral differences primarily (Escudero, Benders, & Lipski, 2009; Giezen, Escudero, & Baker, 2010). To direct Spanish listeners' attention to the dimension that is most important to native Dutch listeners, Escudero et al.'s (2011) training vowels differed in spectral properties only. Further, rather than comparing the effect of bimodal and unimodal training, the authors presented listeners with either a natural bimodal (hence 'bimodal') or an enhanced bimodal (hence 'enhanced') distribution. In the former distribution, the endpoint stimuli had average values for the first and second formants (Pols et al., 1973; Section 2.2.2 of the present manuscript), while the stimuli in the latter had an enlarged spectral difference, i.e., the endpoint tokens had exaggerated properties similar to those of infant-directed (Burnham, Kitamura, & Vollmer-Conna, 2002; Kuhl et al., 1997; Sundberg, 2001; Sundberg & Lacerda, 1999) and foreigner-directed speech (Uther, Knoll, & Burnham, 2007). In this way, the acoustic difference between training stimuli in the enhanced distribution was more pronounced and presumably easier to perceive than the difference between training stimuli in the bimodal distribution (Section 2.2.2 and Table 3). As expected from previous studies that suggest facilitation of speech discrimination with enhanced differences between stimuli (e.g., Kuhl et al., 1997; Liu, Kuhl, & Tsao, 2003), the results showed that vowel classification accuracy (as measured in pre- and post-tests; Section 2.2.1) increased after enhanced training, and that this improvement was larger than in the control condition. (Improvement after bimodal training was not larger than in the control condition). The authors concluded that difficult non-native contrasts can be trained effectively with a distributional learning paradigm, which requires only a few minutes of stimuli exposure and no feedback.

In the present study we first aimed to show again Escudero et al.'s (2011) distributional training results in adult second-language (L2) learners (Section 2.2). To this end, we exposed two new groups of Spanish learners of Dutch to the same bimodal and enhanced distributions of the Dutch vowel contrast /ɑ/ – /aː/. Their classification performance of multiple natural realizations of the two vowels was evaluated in pre- and post-tests, which were identical to those used in Escudero et al. The second and primary aim of the present study was to probe the causes of the increase in vowel classification accuracy after enhanced training, found in Escudero et al. (2011). Specifically, we examined whether distributional training could promote the use of the main acoustic cues for distinguishing the Dutch vowels, i.e., their first and second formants, and whether an enhanced distribution is more effective in this respect than a bimodal distribution, for which the difference in formant values between the training vowels is smaller. To investigate listeners' use of acoustic cues, we employed a statistical technique called *latent class regression analysis* (Huang & Bandeen-Roche, 2004; Sections 1.2 and 2.3 of the present manuscript). With this technique one can identify classes of listeners, with each class representing a subgroup with a unique *listening strategy*, i.e., a specific way of using acoustic cues. This approach thus allowed us to examine the relationship between initial listening strategies, improvement after training, and exposure to bimodal versus enhanced distributions.

1.1. Theoretical background and definition of listening strategies

Recall that, as mentioned in Section 1, we use the term *listening strategy* to refer to a specific use of acoustic cues in the perception of speech sound contrasts (also known as *acoustic cue-weighting*). Accordingly, we do *not* address general learning strategies (as in e.g., Oxford, Nyikos, & Ehrman, 1988), or individual differences in L2 speech sound perception that may result from a variety of other factors such as the length of residence in an L2 country (e.g., Flege , Bohn, & Jang, 1997) or the type of task presented to the listeners (Díaz, Mitterer, Broersma, & Sebastián-Gallés, 2012).

Extensive research has demonstrated cross-linguistic differences in acoustic cue-weighting (e.g., Bohn & Flege, 1990; Escudero & Boersma, 2004; Escudero et al., 2009; Iverson, Hazan, & Bannister, 2005; Iverson et al., 2003; Morrison, 2008, 2009). These studies show that when discriminating speech sounds, native and non-native listeners may favor different acoustic cues. For instance, the well-known observation that Japanese adults have trouble perceiving English /r/ and /l/ as two different speech sounds (e.g., Goto, 1971; Iverson et al., 2003; Miyawaki et al., 1975; Yamada, 1995) has been attributed to the Japanese focus on the irrelevant second formant rather than the relevant third formant, which is used by English natives (Iverson et al., 2003). Similarly and as mentioned above, Dutch natives favor spectral cues when distinguishing between Dutch /ɑ/ and /aː/, while Spanish learners of Dutch tend to resort to duration (Escudero et al., 2009).

In addition to reporting group differences, previous research reveals substantial individual differences in the use of acoustic cues (e.g., Chandrasekaran, Sampath, & Wong, 2010; Escudero & Boersma, 2004; Escudero et al., 2009; Morrison, 2008, 2009). For instance, Escudero et al. (2009) report that over a third of their Spanish learners of Dutch relied more on spectral cues than on duration when categorizing Dutch /ɑ/ and /aː/. Accordingly, it is likely that, in the current study, not all Spanish learners of Dutch will solely focus on duration before training.

Individual differences in cue-weighting are not commonly addressed in theories and models on L2 speech perception, which tend to focus on general group differences. That is, well-known theoretical accounts of non-native speech perception explain the *general* difficulty that Spanish listeners have with discriminating and classifying certain L2 vowels. For instance, for Dutch /ɑ/ and /aː/, both Flege's Speech Learning Model (SLM; Flege, 1995, 2002, 2003; Flege & MacKay, 2004) and Best's Perceptual Assimilation Model (PAM; Best, 1994) posit that the difficulty arises from the similarity of both Dutch vowels to a single Spanish vowel, namely /a/. Mayr and Escudero (2010) present an extensive review of these and other explanations for listeners' difficulties in perceiving non-native speech sounds.

In the current study, where we expect to find differences in the perceptual patterns of Spanish learners of Dutch vowels, we will compare our results to Escudero's Second Language Linguistic Perception (L2LP) model (Escudero, 2005, see also Escudero, 2000), which in contrast to the models mentioned above addresses the possibility that L2 speech sound perception may develop in steps, and that adult listeners may differ in both their perception of L2 speech sounds (see the individual differences in cue use mentioned above) and the way in which this perception develops. Escudero (2000, 2005) explicitly posits successive developmental stages with differential cue weightings for Spanish listeners who learn the English vowels /i/ (as in 'beat') and /ɪ/ (as in 'bit'). Specifically, Escudero proposes the following stages: (0) no distinction between the two vowels, (1) use of duration to distinguish them, (2) a main reliance on duration with a subtle use of spectral cues, and (3) a main focus on spectral cues with an additional use of duration, which is in accordance with native speaker performance. Morrison (2008) suggests an extra stage between 0 and 1. In this stage $\frac{1}{2}$, listeners use spectral cues to classify the vowels as 'good' or 'bad' examples of Spanish /i/, while they also start using durational differences, which are not distinctive in Spanish. Given Spanish learners' difficulty to perceive spectral differences between both English /i/ and /ɪ/ and Dutch /ɑ/ and /aː/, and the

tendency, in both cases, to resort to the use of duration (for English /i/–/ɪ/: Escudero, 2000, 2005; Morrison, 2008; for Dutch /ɑ/–/aː/: Escudero et al., 2009), we expect to find listening strategies that are roughly similar to the ones suggested by Escudero (2000, 2005) and Morrison (2008).

1.2. Latent class modeling

The statistical technique that we use to identify types of listening strategies is based on latent class regression analysis (Huang & Bandeen-Roche, 2004). It is an increasingly popular method for identifying groups of participants with similar latent (i.e., non-overt) individual characteristics in a statistically reliable way. For instance, the technique has been used to study children's reasoning strategies (Bouwmeester, Sijtsma, & Vermunt, 2004) and Japanese women's gender-role attitudes (Yamaguchi, 2000). Also, it has been used to identify groups among psychotic patients (Schmitz, Malla, Norman, Archie, & Zipursky, 2007) and to pinpoint the sources of knowledge in Artificial Grammar Learning (Visser, Raijmakers, & Pothos, 2009). In this paper, we introduce the technique to the field of speech perception and its development.

The proposed analysis detects groups of listeners with the same listening strategy within the experimental groups. Thus, it is more fine-grained than a standard group analysis, in which individual deviations from group patterns are not accounted for. At the same time, it goes beyond describing the strategy for each listener separately by highlighting similarities between individuals with the same listening strategy.

Previous research has investigated individual strategies and the clustering of individuals separately. For instance, Chandrasekaran et al. (2010), who examined the effect of native American English speakers' cue weighting of pitch height and direction on their ability to learn Mandarin lexical tone, divided listeners in 'good' and 'poor' learners on the basis of performance scores, before analyzing group differences in the use of specific cues. Escudero and Boersma (2004) first examined listening strategies per individual and then listed the number of listeners who utilized each type of strategy. Unlike these previous proposals, we follow Morrison (2007, 2008, 2009) in that we cluster listeners' strategies with a *statistical* technique, but we do so in a more far-reaching way. This is because Morrison's hierarchical cluster analysis still requires the researcher to choose the number of groups. In contrast, the latent class regression analysis groups participants *simultaneously* with the extraction of the strategies. The strategies are represented by a latent variable in the model and are not defined a priori. We use a common model selection technique (Section 2.3) to determine the optimal, most parsimonious number of strategies within the group, which makes the method statistically more robust.

Crucially, the applied method for strategy detection does not use performance scores to assign a participant to a class with a certain listening strategy. Rather, it extracts listening strategies through determining the degree to which acoustic cues predict an individual's vowel classifications, regardless of the correctness or incorrectness of the responses. Thus, an acoustic vowel dimension that is a statistically significant predictor of a participant's classifications is considered a cue that he or she used, and consequently a significant part of that listener's strategy. Because the outcome variable, i.e., an individual's vowel classification (Section 2.3), is categorical, we applied logistic regression models, which have many advantages compared to ANOVA techniques (Jaeger, 2008). Since the proposed analysis relies on cues rather than accuracy, it is specifically suited for our purpose of determining what is learned in the distributional learning process.

2. Method

2.1. Participants

The present study included 150 adult native speakers of Spanish ($M=36.8$ years, *Range*=19–60 years; 123 female and 27 male), who were living in the Netherlands at the time of testing, and had arrived in the Netherlands after the age of 15 years. They were divided into three groups of 50 each: the Enhanced, Bimodal, and Music groups. All these participants completed a pre-test, a training phase and a post-test. Only the training phase differed per group. The Bimodal and Enhanced groups listened to vowel distributions (Section 2.2.2), while the Music group (or control group) was exposed to classical music.

These Spanish-speaking participants had enrolled in a longitudinal project on the perception of Dutch vowels, which included a larger participant pool ($N=500$) and was led by the second author. They had all taken part in the first session of the longitudinal project six months earlier. During this first session, participants in the Music group had performed the same pre- and post-tests and had listened to the same classical music as in the present study, while participants in the Bimodal and Enhanced groups had only performed the pre-test, and had not received any training.[1] In the first session, assignment to groups had been random. In the present study, which reports results of the second session, participants were assigned to the Bimodal and Enhanced groups while considering their first-session pre-test scores, which were matched with those of the Music group. Other than that, assignment to the two training groups was random.

Table 1 lists each group's age at the time of testing (AaT), age of arrival (AoA), length of residence (LoR) in the Netherlands, and Dutch proficiency score, i.e., the level of general comprehension of Dutch as measured by the language comprehension component of the Dialang test (www.dialang.org; Alderson & Huhta, 2005). The groups were not significantly different in any of these measures (LoR: $F(2,149)=.52$, $p=.60$; AaT: $F(2,149)=1.6$, $p=.20$; AoA: $F(2,149)=1.2$, $p=.29$ and Dutch proficiency $F(2,148)=.34$, $p=.71$). Additionally, the median and range for AaT was comparable across groups: Enhanced: 36 (range: 21–56), Bimodal: 34 (22–55), and Music: 37 (19–60). All participants reported normal hearing.

Further, just as in the larger longitudinal project where out of 500 registrations only 50 were from men, the number of female participants in the present study (38, 41 and 44 in the Enhanced, Bimodal and Music groups respectively) was larger than that of male participants (12, 9 and 6 respectively). In Section 3, we examine whether our results are representative of both men and women.

Unlike Escudero et al. (2011), we also included an age-matched group of 25 adult native speakers of Dutch ($M=32$ years, Range=18–60 years; 21 female). These Dutch natives performed the same test as the Spanish listeners but only once, and they received no training. We will compare the Dutch results for this single test to both the pre- and post-tests that Spanish listeners performed in order to assess these listeners' L2 development after training.

[1] Results of the Music group's first session (i.e., of the pre- and post-test) are reported in Escudero et al. (2011). Results of the Bimodal and Enhanced group's first session (i.e., of the pre-test only) are reported in Escudero & Wanrooij (2010).

Table 1
Mean age at testing (AaT), age of arrival (AoA) and length of residence (LoR) in the Netherlands (in years), and Dutch proficiency score (see text), per Spanish group. Standard deviations are given between parentheses.

Group	AaT	AoA	LoR	Dutch proficiency
Enhanced	37.3 (8.0)	31.9 (6.9)	5.4 (5.0)	3.9 (2.2)
Bimodal	35.0 (8.7)	29.9 (7.0)	5.2 (5.4)	4.2 (2.2)
Music	38.0 (9.0)	31.7 (7.2)	6.3 (6.8)	4.0 (2.1)

Table 2
Average F1, F2, f0 (in Hz) and duration (in milliseconds) of the X stimuli in the XAB-test. Standard deviations are given between parentheses.

Vowel	F1		F2		f0		Duration	
	Females	Males	Females	Males	Females	Males	Females	Males
/ɑ/	719 (100)	584 (99)	1239 (168)	1156 (127)	223 (50)	154 (24)	93 (13)	94 (24)
/aː/	923 (75)	652 (144)	1552 (107)	1424 (98)	183 (36)	132 (18)	216 (43)	204 (14)

2.2. Stimuli and procedure

2.2.1. Test

Spanish and Dutch listeners performed a forced-choice classification task in an XAB format, designed to assess classification performance of Dutch /ɑ/ and /aː/. To promote classification rather than discrimination, the inter-stimulus-interval (ISI) between the three stimuli in each trial (i.e., X, A and B) was chosen to be relatively long (1.2 s) (Van Hesse & Schouten, 1999; Werker & Logan, 1985), and the X stimuli were chosen to be natural tokens containing much variability, as explained below.

Prior to performing the XAB-task, participants had a practice session of five trials where it was ascertained that they heard the stimuli well and that they understood the task. None of the listeners demonstrated hearing problems or failed to correctly identify the vowels in this practice session. As mentioned above (Section 2.1), only the Spanish listeners performed the XAB task a second time after training, i.e., they had a pre- and a post-test. The test procedure, which was the same as in Escudero et al. (2011), was as follows. In each trial, listeners heard a natural token of /ɑ/ or /aː/ (the X stimulus), followed by two synthetic response options (A and B). There were 20 unique X stimuli for each vowel, which were a subset of the vowels reported in Adank et al.'s (2004) corpus and which were produced by 10 male and 10 female speakers of Standard Northern Dutch in an /s–V–s/ context. The average fundamental frequency (f0), first formant (F1), second formant (F2) and duration of the X stimuli are listed in Table 2, for females and males separately.

Unlike in Escudero et al. (2011), where each X stimulus was presented once and the response options were randomly ordered, we included two repetitions of each X stimulus by counterbalancing the response options. Thus, our XAB task included 80 trials ($=20$ unique X stimuli $\times 2$ vowels $\times 2$ repetitions). The two response options A and B were synthetic stimuli (created using the Praat program of Boersma & Weenink (2011)), because the acoustic properties had to be compatible with those of the training stimuli (Section 2.2.2). They were based on typical tokens of /ɑ/ and /aː/ (Pols et al., 1973), with F1-values of 687 and 770 Hertz (Hz) and F2-values of 1104 and 1303 Hz respectively, which five Dutch natives had judged as better exemplars of the Dutch vowels than tokens generated using Adank et al.'s (2004) values (Escudero & Wanrooij, 2010). For both response options, the duration was 140 ms and f0 fell from 150 to 100 Hz, which represents a male voice (e.g., Hollien, Dew, & Philips, 1971).

The task was self-paced: listeners were told that the next trial would only appear after their response. They were encouraged to respond as quickly as possible and were asked to guess if uncertain. Also, they were told that they could take a short break (available every 20 trials) if needed. Spanish and Dutch listeners took approximately 7 min to complete the task.

2.2.2. Training

Only the Spanish listeners were presented with the training phase. The training stimuli and procedure, which were the same as in Escudero et al. (2011), were as follows. The stimuli during the training phase differed across Spanish groups: Bimodal and Enhanced listeners heard, respectively, bimodal and enhanced training distributions of the Dutch vowel contrast /ɑ/ – /aː/, while the Music group listened to instrumental classical music. The goal of the bimodal and enhanced training was to expose participants to the spectral difference between Dutch /ɑ/ and /aː/. Because Spanish listeners tend to classify /ɑ/ and /aː/ on the basis of their duration while ignoring their spectral differences (Section 1), the training stimuli differed from one another only in the spectral values for F1, F2 and F3 (the third formant) and not in duration. Table 3 lists the F1 and F2 values for each of the eight stimuli in the bimodal and enhanced training distributions separately, which were synthesized in the computer program Praat (Boersma & Weenink, 2011).

The endpoint values (i.e., stimulus 1 and 8 in the table) of the bimodal distribution were similar to the average production values of Dutch /ɑ/ (stimulus number 1) and /aː/ (stimulus number 8), as measured by Pols et al. (1973). The endpoint values of the enhanced distribution were calculated as the average production of /ɑ/ minus one standard deviation (stimulus 1) and the average production of /aː/ plus one standard deviation (stimulus 8). The standard deviations were based on Pols et al. (1973). In each distribution, the steps between consecutive values were approximately equal on the psychoacoustic ERB scale (Bimodal: 0.1 ERB for F1, 0.2 ERB for F2; Enhanced: 0.4 ERB for F1 and F2). F3 was calculated for each stimulus as the stimulus' F2 plus 1000 Hz. All training stimuli had an f0 that fell from 150 to 100 Hz, and a duration of 140 ms. The table also shows the frequency of presentation for each training stimulus. There were 128 stimuli in total, which were presented with an ISI of 750 ms, for a total training duration of less than 2 min. The Music group listened to classical music for the same time.

Before the training phase, all participants were told that they would perform another test afterwards. Listeners in the Enhanced and Bimodal groups were instructed to listen to the training vowels carefully, while listeners in the Music group were asked to relax while listening to the classical music.

Table 3
F1 and F2 values (in Hz) and frequency of presentation for each stimulus in the enhanced and bimodal training distributions (Escudero et al., 2011).

Token number	1	2	3	4	5	6	7	8
Token frequency	8	32	16	8	8	16	32	8
Enhanced								
F1	600	637	675	714	755	797	840	885
F2	1000	1055	1112	1171	1233	1296	1362	1430
Bimodal								
F1	700	713	726	740	753	767	781	795
F2	1115	1144	1174	1204	1235	1266	1298	1330

2.3. Statistical analysis

A traditional comparison of mean accuracy across groups served to demonstrate the same distributional training results as in Escudero et al. (2011) and thus to demonstrate the validity of our data for the subsequent analysis of listening strategies, i.e., specific uses of acoustic cues in perception, in each group. To identify listening strategies, we used *latent class regression* (LCR) analysis (Huang & Bandeen-Roche, 2004), as mentioned in Section 1. LCR analysis explains correlations between responses to different items by introducing a *latent* variable. This variable is nominal, which indicates the existence of a number of different types (*classes*) of behavior rather than a dimension on which people vary continuously. Furthermore, a finite number of types of behavior, each with a unique set of *regression* coefficients (and intercepts), is assumed.

We identified the five most important acoustic components for the classification of the natural vowel productions (i.e., the X stimuli) that were presented in the XAB task: duration, F1, F2, F3, and f0. Correct classification needed to be based primarily on F1, F2, duration or a combination of these cues (Section 1), and secondarily on higher formants such as F3, which adds subtle information but cannot be used as a single cue to distinguish the two vowels. Further, f0 could not be used to classify the vowels correctly, because it is not a cue for vowel identity.

When participants took only duration into account when classifying the vowels, their listening strategy was confined to the use of this cue. We described such a listening strategy with a binomial regression model, i.e., with a binomially distributed dependent variable and multiple predictors. The dependent variable was the number of times a participant chose the category /aː/ for each specific X stimulus. Since every specific X stimulus was presented twice, the number of times a participant opted for response /aː/ when presented with a token of /ɑ/ or /aː/ was 0, 1 or 2. Note that we thus modeled the categorization of stimuli and not the accuracy of the categorization (Section 1.2). The predictors were the five acoustic components of the vowels mentioned above.[2]

In a standard regression analysis, the same regression coefficients apply to each participant. In LCR analysis, the same regression coefficients apply only to members of the same latent group. It is important to note that group membership is not a manifest variable (i.e., an observable variable) but is assigned only after fitting the LCR model to the data. The specified LCR model had the following form:

$$L(y_i) = \mu_c + \beta_{Dc}D + \beta_{\mathrm{f0}c}\mathrm{f0} + \beta_{\mathrm{F1}c}\mathrm{F1} + \beta_{\mathrm{F2}c}\mathrm{F2} + \beta_{\mathrm{F3}c}\mathrm{F3}$$
$$c = 1...N_c, \quad i = 1...n \tag{1}$$

Here y_i is the number of times (0, 1, or 2) that a specific X stimulus was classified as /aː/ and L is the standard link function[3] for a binomial regression model (Jaeger, 2008; McCullagh & Nelder, 1989), i.e., the logit function $\log p/(1-p)$, where p is the mean of the binomial distribution; μ_c is the intercept of latent class c; parameters β_{Dc}, $\beta_{\mathrm{f0}c}$, $\beta_{\mathrm{F1}c}$, $\beta_{\mathrm{F2}c}$, and $\beta_{\mathrm{F3}c}$ are the regression coefficients for latent class c; N_c is the number of latent classes; and n is the number of participants. The value of the intercept is a measure of the bias in responding /ɑ/ or /aː/. Because the absolute value is not easy to interpret, we will calculate the bias for each latent class after fitting the model. The regression coefficients indicate how much the logit of the probability of answering /aː/ changed with a one-unit change in the predictor. Note that the regression parameters are not normalized, so that the absolute values are still interpretable given the different ranges for each predictor.

Exploratory LCR models with an increasing number of latent classes were fitted to the Spanish groups' pre-test and post-test classification data and to the Dutch natives' classification data in their single test. To establish the optimal number of latent classes in each condition, we used the Bayesian Information Criterion (BIC; Schwarz, 1978).[4] The BIC is commonly used to compare non-nested competing models, in this case models with an increasing number of latent classes (see Lin and Dayton (1997), for details on the specific uses of BIC in latent class models). The BIC provides a trade-off between goodness of fit (the log likelihood) and the number of parameters in the model. For each added latent class, seven extra parameters are estimated, namely, the intercept and regression coefficients of that class (in our case five regression coefficients for the five predictors), and the proportion of participants that it contains. Lower values for BIC denote better models in which goodness of fit and parsimony are balanced. After fitting the model to the data, each individual participant was assigned to a class. To this end, the posterior probabilities of participants' responses were calculated given each latent class of the model. Subsequently, each participant was assigned to the latent class with the largest likelihood for that participant's data. For fitting models to the data, we used the statistical R-package of FlexMix (Leisch, 2004; see also Grün and Leisch (2007), for an example of fitting mixtures of logistic regressions in R).

3. Results

Table 4 shows the group results for the Dutch and Spanish listeners, which are given in accuracy percentages, i.e., the percentage of time listeners correctly classified the 80 test stimuli. The Dutch accuracy was substantially higher than that in all Spanish groups for both the pre- and the post-tests, which confirms previous Dutch results on the same task (Escudero & Wanrooij, 2010), and thus ascertains that the stimuli and the response options were good examples of the Dutch vowels /ɑ/ and /aː/. The Dutch accuracy also shows that the task was relatively difficult, since Dutch listeners did not score at ceiling.

[2] We used logarithmic scales for the five acoustic cues to account for the fact that the human ear is better at discriminating small differences in shorter durations and lower frequencies than in longer durations and higher frequencies (e.g., Allan & Gibbon, 1991; Kewley-Port & Watson, 1994; Stevens, Volkmann, & Newman, 1937).

[3] The link function provides the relationship between the linear predictor and the mean of the distribution.

[4] The BIC is defined as minus 2 times the log likelihood of the model, plus the number of parameters times ln(N), with N being the number of participants.

Table 4
Mean Spanish (pre- and post-test) and Dutch (single-test) accuracy percentages. Standard deviations are given between parentheses.

Test	Enhanced	Bimodal	Music	All Spanish	Dutch
Pre-test	60.4 (11.7)	60.4 (12.2)	61.7 (11.1)	60.8 (11.6)	83.1 (9.6)
Post-test	67.1 (13.5)	64.2 (14.5)	63.7 (13.3)	65.0 (13.7)	–

To investigate if our results for the Spanish participants were similar to those of Escudero et al. (2011), we ran a mixed design analysis with Test as a within-subjects factor (pre-test vs. post-test accuracy) and Group as a between-subjects factor (Bimodal, Enhanced and Music). The results revealed no main effect of Group ($F(2, 147)=0.20$, $p=.82$), which supports the homogeneity of the groups, and a main effect of Test ($F(1, 147)=29.70$, $p<.001$), which indicates that the improvement between pre- and post-test shown in Table 4 is statistically significant. Further, the analysis yielded a significant Test × Group interaction ($F(2, 147)=3.12$, $p=.047$), which indicates that some group(s) improved more than others.

Posthoc *t*-tests on difference scores (i.e., post- minus pre-test accuracy percentages, as shown in Table 4) using Tukey's HSD revealed that the Enhanced group improved more than the Music group (difference=4.63%, with a 95% Confidence Interval, CI=+0.21..+9.04%, $p=.038$), and that the differences in improvement between the Bimodal and Enhanced groups, and between the Bimodal and Music groups were not significant ($ps>.05$). These results are the same as those reported in Escudero et al. (2011). Further, to test whether each group improved significantly in the post-test as compared to the pre-test, the difference score of each group was compared to 0 (which represents no improvement) in a one-sample t-test. Again in accordance with Escudero et al., a significant improvement was found for the Enhanced group (6.63% with CI=+4.05..+9.20%, $t(49)=5.17$, $p<.001$), and not for the Music group (2.00% with CI=–0.50..+4.50%, $t(49)=1.61$, $p=.12$). Unlike in Escudero et al., there was also a significant improvement for the Bimodal group (3.83% with CI=+0.97..+6.68%, $t(49)=2.69$, $p=.010$).

We also examined whether pre-test accuracy and difference scores ($n=150$) were significantly correlated with Spanish listeners' LoR, AaT, AoA and Dutch proficiency (Section 2.1) using non-parametric correlations (Spearman's ρ). There was a significant correlation between *pre-test accuracy* and both AaT ($\rho=-.19$, $p=.023$) and AoA ($\rho=-.23$, $p=.005$), indicating that the younger participants were when they performed the task and the younger they were when they arrived in the Netherlands, the higher their accuracy at pre-test. There was no significant correlation between pre-test accuracy and LoR or Dutch proficiency (both $ps\geq.71$).

Further, there was no significant correlation between *difference scores* and AaT, AoA or Dutch proficiency (all $ps\geq.13$). Difference scores were significantly correlated with LoR ($\rho=.17$, $p=.033$).

3.1. Listening strategies before distributional training

Table 5 summarizes the optimal latent class regression models for Spanish learners' pre-test and Dutch natives' single test. It contains the identified classes per group, and the cues that each class used, i.e., their listening strategy. In the regression model the cues are the predictors (Section 2.3). None of the Spanish and Dutch classes exhibited a response bias to /ɑ/ or /aː/ (one-sample $ts<2.2$, $ps>.05$).[5]

It can be observed that each Spanish group had two latent classes: one with the majority of participants with low mean accuracy (hence "low performers"), and the other with the minority of participants with high accuracy (hence "high performers"). These two pre-test classes per group confirm the equality of the groups at pre-test and are also visible in Fig. 1, left column, which shows the number of participants (*y*-axis) for each accuracy percentage (*x*-axis). The figure clearly shows that most, if not all, low performers (black bars) indeed had lower accuracy than high performers (white bars).

There was a strong correlation between the accuracy percentage obtained in each Spanish class and the number of cues used: Spearman's $\rho=.88$, p(one-tailed)[6] $=.011$. Thus, not surprisingly, Spanish learners of the Dutch contrast /ɑ/–/aː/ tend to score higher when they use more cues. Low performers used two cues, namely duration and either F1 (in Enhanced and Music) or F3 (in Bimodal), while high performers used three, namely duration and a combination of F1 and F2. Overall, all six Spanish classes used duration, five classes used F1, three used F2, one used F3, and none used f0, which suggests that Spanish listeners tend to favor certain cues above others. Interestingly, high performers not only used more cues than low performers, but also tended to use cues more intensely, as reflected by their betas (i.e., the regression coefficients in the model; Section 2.3). For example, Table 5 shows that low performers had duration betas of 0.91, 0.71 and 1.25, while high performers had duration betas of over 4.

Because our participant group contained a larger number of females than males, we examined whether the division into low and high performers in the pre-test was representative of both women and men. For this, we counted the number of low and high performers who were female (94 low and 29 high performers) versus male (16 low and 11 high performers). A chi-square test showed no significant difference in listening strategies between the sexes ($\chi(1)=3.34$, $p=.068$).

Dutch natives also had two different listening strategies: half of them focused on three cues (duration, F1 and F3) and had moderate accuracy ($M=75.3\%$), while the other half used four cues (duration, f0, F1 and F2) and had very high accuracy ($M=91.5\%$). A comparison of the Spanish and Dutch performance shown in Table 5 suggests that Spanish high performers approximated the Dutch natives who performed moderately well.

3.2. Listening strategies after distributional training

The Spanish post-test classes are shown in Table 6, where it can be observed that the post-test yielded three, four and two classes in the Enhanced, Bimodal and Music groups respectively.

Similarly to the pre-test, significant cues for classes with 60% or lower accuracy did not include a combination of F1 and F2 and the maximum number of cues was two, while learners in classes with 70% or higher accuracy used at least three cues including duration, F1 and F2. Classes with 80% or higher accuracy also included F3. Again, a strong correlation was found between accuracy and the number of cues identified for a class:

[5] As mentioned in Section 2.3, the number of /aː/ responses for any specific stimulus /ɑ/ or /aː/ could be 0, 1 or 2. For the response bias analysis, we thus used the null hypothesis that the average number of /aː/ responses in each class was 1.

[6] The significance test is one-tailed because we expect a positive correlation between the number of predictors and vowel classification accuracy.

Table 5
Spanish (pre-test) and Dutch (single test) classes, including number of participants per class (*N*), their mean accuracy, statistically significant predictors (Cues), estimated regression coefficients (Betas) and *p*-values. *D*=duration.

Group	Class	*N*	Accuracy (SD)	Cues	Beta (SE)	*p*-Value
Spanish Enhanced	1	33	53.2 (6.2)	D	0.71 (.33)	.032
				F1	1.36 (.55)	.013
	2	17	74.4 (5.4)	D	4.16 (.58)	<.0001
				F1	4.00 (.85)	<.0001
				F2	5.58 (1.54)	.00028
Spanish Bimodal	1	39	55.1 (7.1)	D	0.91 (.30)	.0028
				F3	2.22 (.94)	.019
	2	11	79.1 (6.9)	D	4.39 (.69)	<.0001
				F1	4.09 (1.11)	.00023
				F2	10.08 (2.12)	<.0001
Spanish Music	1	38	56.6 (6.6)	D	1.25 (.31)	<.0001
				F1	1.93 (.51)	.00015
	2	12	78.0 (4.6)	D	4.96 (.69)	<.0001
				F1	7.14 (1.13)	<.0001
				F2	4.81 (1.90)	.012
Dutch	1	13	75.3 (6.1)	D	5.21 (.66)	<.0001
				F1	4.14 (.96)	<.0001
				F3	6.04 (1.90)	.0015
	2	12	91.5 (3.3)	D	8.58 (.87)	<.0001
				F0	–5.28 (1.66)	.0015
				F1	6.15 (1.63)	.00016
				F2	14.80 (2.95)	<.0001

Spearman's $\rho = .89$, p(one-tailed)[7] = .001, which indicates that when Spanish learners focus on more cues, accuracy of classification of /ɑ/ and /a:/ increases. Duration was also the most consistently used cue (8 out of 9 classes), followed by F1 (8 classes), F2 (5 classes), F3 (2 classes) and f0 (1 class). Also as in the pre-test, learners with higher accuracy appeared to use cues more intensely, i.e., they had higher betas, than those with lower accuracy. For instance, duration betas ranged between 0.88 and 2.83 for classes with accuracy below 60%, while they were between 4.68 and 8.72 for classes with higher accuracy.

When comparing the Spanish post-test classes in Table 6 to those of the Dutch single test in Table 5, we observe that more than 20 percent (11 out of 50) of the learners in the Enhanced group ended up using the same cues (duration, f0, F1, and F2) as half of the Dutch natives (12 out of 25), but the Dutch had a higher accuracy (70.1% versus 91.5%). This difference may be due to a more efficient use of duration and F2 in the Dutch natives, as reflected by their higher betas. Remarkably, one class of four Bimodal listeners obtained similar accuracy (93.4%) as the best performing Dutch class, despite the fact that they used a different strategy than the Dutch.

Finally, one-sample *t*-tests for each post-class (alpha = .0056, 05/9 tests) showed that a bias toward the /a:/ response developed in Bimodal class 2 ($M = 1.43$, CI = +1.22..+1.64, $t(5) = 5.35$, $p = .0031$) and Enhanced class 2 ($M = 1.34$, CI = +1.24..+1.44, $t(10) = 7.70$, $p < .001$).

3.3. Improvement with training

A comparison of Table 5 (pre-test) and Table 6 (post-test) shows that after training an increase in number of classes is only observed for the Enhanced (from 2 to 3) and Bimodal (from 2 to 4) groups. Also, while the Music group has the *same* listening strategies in both tests, listening strategies typically changed after distributional training. These observations suggest that distributional training, and not listening to music, diversified listening strategies. Furthermore only after distributional training, Spanish listeners came closer to the Dutch listening strategies and accuracy (Section 3.2).

Fig. 1 illustrates how pre-test performance relates to post-test class membership, as follows. In both the pre-test column (Fig. 1, left) and the post-test column (Fig. 1, right) black bars represent *pre-test* low performers and white bars *pre-test* high performers. Post-test classes are numbered from worst- (1) to best-performing (2 and above). It can be observed that pre-test low and high performers tended to move to the worst and best performing post-test classes respectively, as shown by the higher number of black and white bars in the right column for low and high post-test accuracy respectively.

Specifically, in the *Enhanced group*, out of the 33 pre-test low performers (who used duration and F1 in the pre-test) 21 listeners (64%) moved to the worst-performing post-test class 1 (duration only), 7 (21%) to post-test class 2 (duration, f0, F1 and F2) and 5 (15%) to post-test class 3 (duration, F1, F2 and F3). Out of the 17 Enhanced pre-test high performers (who used duration, F1 and F2 in the pre-test), 1 (6%) moved to post-test class 1 (duration only), 4 (24%) to post-test class 2 (duration, f0, F1 and F2) and 12 (71%) to post-test class 3 (duration, F1, F2 and F3). In the *Bimodal group*, out of the 39 pre-test low performers (who used duration and F3 in the pre-test) 20 (51%) moved to post-test class 1 (no cues), 6 (15%) to post-test class 2 (duration, F1), 12 (31%) to post-test class 3 (duration, F1 and F2) and 1 (3%) to post-test class 4 (duration, F1, F2 and F3). Out of the 11 Bimodal high performers (who used duration, F1 and F2 in the pre-test) 8 (73%) retained the same strategy in post-test class 3, while 3 (27%) moved to post-test class 4 (duration, F1, F2 and F3). In the *Music group*, out of the 38 pre-test low performers (who used duration and F1 in the pre-test) 31 (82%) retained the same strategy in post-test class 1 and 7 (18%) moved to post-test class 2 (duration, F1 and F2), while out of the 12 pre-test high performers (who used duration, F1 and F2 in the pre-test), 2 (17%) moved to post-test class 1 (duration and F1) and 10 (83%) retained the same strategy in post-test class 2.

[7] See the previous note.

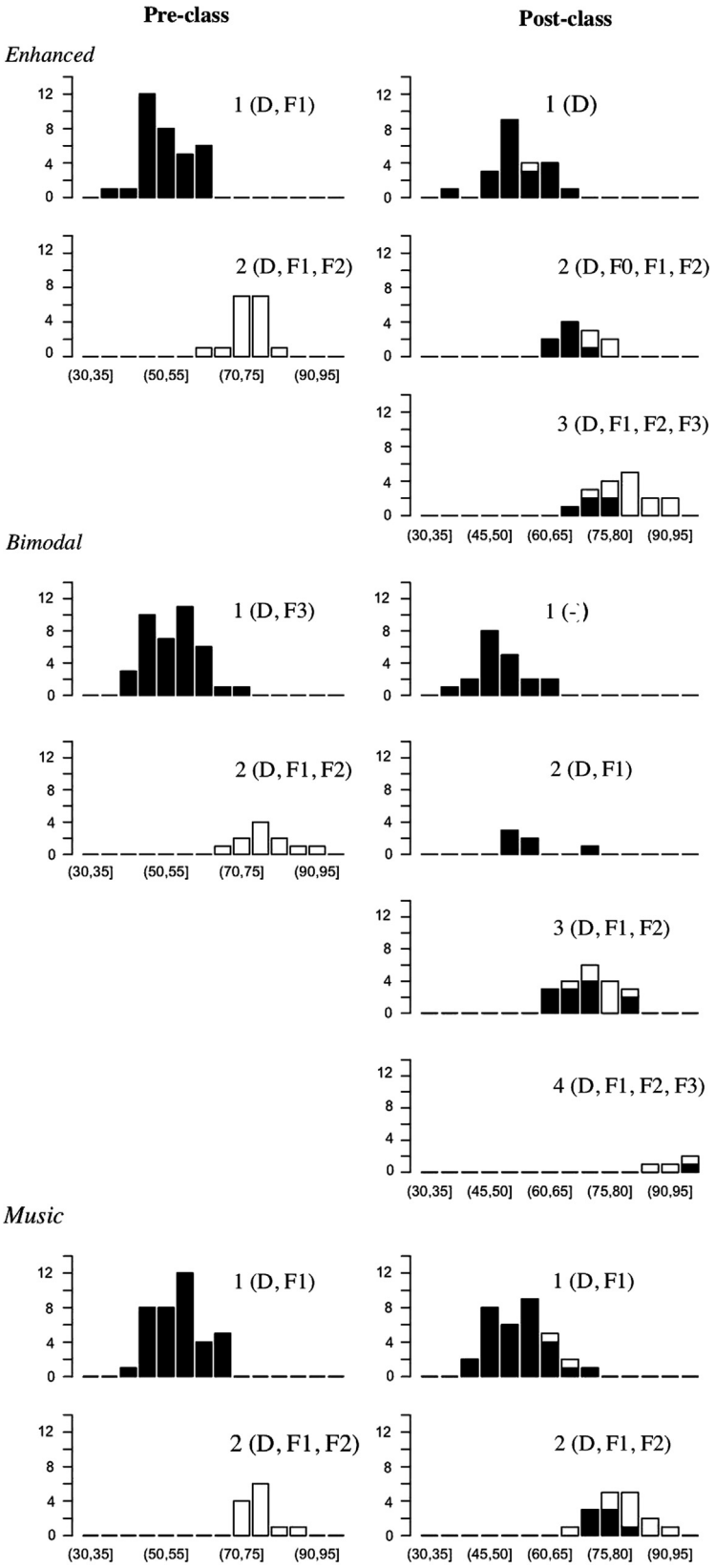

Fig. 1. Histograms showing the number of Spanish learners (*y*-axis) for each accuracy percentage (*x*-axis) per pre-test class (left) and post-test class (right) in each group (Enhanced, Bimodal, and Music). For each class the listening strategy (one or more of the acoustic cues duration, F1, F2, F3 and f0) is given. In both the pre-test and the post-test column black bars represent *pre-test* low performers and white bars *pre-test* high performers.

Fig. 1 also illustrates that if listeners used new cues after training, these cues were always F1 and/or F2 for pre-test low performers, while pre-test high performers, who continued using F1 and F2, also used F3. Some Enhanced listeners also started to use f0. To quantify new cue use more precisely, we counted the number of pre-test low and high performers who started to use new relevant cues (i.e., the primary cues F1 and F2, and the

Table 6
Spanish post-test classes with the same variables as in Table 5.

Group	Class	N	Accuracy (SD)	Cues	Beta (SE)	*p*-value
Spanish Enhanced	1	22	54.6 (6.6)	D	0.88 (.40)	.030
	2	11	70.1 (6.0)	D	4.68 (.71)	<.0001
				F0	3.23 (1.33)	.015
				F1	7.08 (1.34)	<.0001
				F2	5.84 (2.16)	.0067
	3	17	81.2 (6.8)	D	6.23 (.60)	<.0001
				F1	4.31 (.96)	<.0001
				F2	5.01 (1.69)	.0031
				F3	–5.59 (2.12)	.0085
Spanish Bimodal	1	20	51.4 (6.2)	–	–	–
	2	6	57.7 (7.3)	D	2.83 (.85)	.00091
				F1	3.04 (1.44)	.035
	3	20	73.1 (6.1)	D	4.03 (.50)	<.0001
				F1	5.53 (.80)	<.0001
				F2	3.64 (1.50)	.015
	4	4	93.4 (4.1)	D	8.72 (1.64)	<.0001
				F1	13.66 (3.95)	.00054
				F2	41.73 (8.9)	<.0001
				F3	–22.40 (7.64)	.0034
Spanish Music	1	33	55.5 (7.1)	D	1.15 (.33)	.00051
				F1	1.29 (.54)	.017
	2	17	79.6 (6.0)	D	5.50 (.57)	<.0001
				F1	5.54 (.92)	<.0001
				F2	6.28 (1.65)	.00014

Table 7
Number of low and high performers in the pre-test, who started to use F1, F2 and/or F3 after training (new users) versus those who did not (others).

	Low performers			High performers		
	Enhanced	Bimodal	Music	Enhanced	Bimodal	Music
New-cue users	12	19	7	12	3	0
Others	21	20	31	5	8	12

secondary subtle cue F3)[8] versus those who did not, which are listed in Table 7. It can be inferred from the table that 36.4% (12 listeners) of the pre-test low-performers who were trained in the Enhanced condition began using F1 and/or F2 in the post-test, as compared to 48.7% (19 listeners) in the Bimodal group and only 18.4% (7 listeners) in the Music group. Further, 70.6% (12 listeners) of the pre-test high performers in the Enhanced group started using F3 after training, versus only 27.3% (3 participants) in the Bimodal group. In the Music group none of the pre-test high performers started using new cues.

Two chi-square tests, for pre-test low and high performers separately, showed significant group (Enhanced, Bimodal and Music) differences (low: $\chi(2)=7.88$, $p=.019$, high: $\chi(2)=15.63$, $p<.001$). For *pre-test low performers*, post-hoc chi-square tests showed that more Bimodal than Music listeners started using F1 and/or F2 ($\chi(1)=7.90$, $p=.005$), and that there was no significant difference between the Bimodal and Enhanced groups in this respect ($\chi(1)=1.11$, $p=.29$). Thus, for pre-test low performers, enhanced training did not significantly improve the use of F1 and/or F2 more than bimodal training. For *pre-test high performers*, post-hoc chi-square tests demonstrated that more Enhanced than Bimodal listeners started using F3 after training ($\chi(1)=5.04$, $p=.025$). Thus, for pre-test high performers enhanced training was more effective for learning to use F3 than bimodal training. Further, for the Enhanced group, relatively more pre-test high than low performers started using new cues after training ($\chi(1)=5.27$, $p=.022$), indicating that the enhanced training was more effective for pre-test high than low performers. In the Bimodal group a comparison between pre-test low and high performers was not significant ($\chi(1)=1.60$, $p=.21$).

Interestingly, Fig. 1 also shows that listeners started to use new cues in a certain order, viz., duration, F1, F2 and F3. That is, listeners who started to use F1 after training, always continued to use duration, those who started using F2 also started or continued to use duration and F1, and those who started to use F3, also started or continued to use duration, F1 and F2.

Recall that duration was the only cue used by all pre-test classes (Section 3.1). Table 8 shows how many pre-test low and high performers in each group (Enhanced, Bimodal and Music) increased their use of duration after training versus those who did not. An increase was reflected in a higher beta for duration in the post- as compared to the pre-test. For *pre-test low performers*, a chi-square test showed that the groups differed in this respect ($\chi(2)=45.04$, $p<.001$). In post-hoc chi-square tests, the number of low performers in the pre-test, who increased their use of duration after training was larger in the Enhanced than Music and Bimodal groups (both $\chi(1)>20.71$, $ps<.001$), and in the Bimodal than Music groups ($\chi(1)=7.90$, $p=.005$). For *pre-test high performers* post-hoc Fisher Exact tests showed that fewer Bimodal than Enhanced ($p<.001$) and Music ($p=.012$) listeners increased their use of duration. In sum, across low and high performers listeners increased their use of duration after enhanced training in particular. Notice that the numbers in Tables 7 and 8 are similar. In fact, all listeners who started using new cues also increased their use of duration.

As in Section 3.1, we examined possible sex differences in our results. Specifically, we examined whether men and women differ in their ability to use new cues after training. For this, we counted the number of new-cue users versus other participants, who were female (37 new-cue users and 86

[8] Including the irrelevant cue f0 in the analysis strengthens the significance values reported and does not change the main findings.

Table 8
Number of low and high performers in the pre-test, who increased their use of duration after training versus those who did not.

	Low performers			High performers		
	Enhanced	Bimodal	Music	Enhanced	Bimodal	Music
Increased use	33	19	7	16	3	10
Others	0	20	31	1	8	2

others) and male (16 new-cue users and 11 others). A chi-square test showed a significant difference between men and women ($\chi(1)=8.25$, $p=.005$). Additionally, we examined the sex distribution of new-cue users versus others in post-hoc chi-square tests for pre-test low and high performers separately. For *pre-test low performers*, there was no significant difference in the ability to use new cues after training between men (7 new-cue users, 9 others) and women (31 new-cue users, 63 others; $\chi(1)=0.70$, $p=.402$). For *pre-test high performers*, the post-hoc test showed that men (9 new-cue users, 2 others) were more likely to use new cues after training than women (6 new-cue users, 23 others; Fisher Exact Test: $p=.001$).

4. Discussion

The present study confirmed Escudero et al.'s results (2011) in two ways. First, our new group of Spanish learners that was exposed to an enhanced distribution of the Dutch vowel contrast /ɑ/ – /aː/ (the Enhanced group) classified the Dutch vowels significantly better after than before training, and the control group exposed to classical music (the Music group) did not. Second, this improvement for the Enhanced group was greater than that for the Music group. Unlike Escudero et al., Spanish learners who were exposed to a bimodal distribution of the contrast (the Bimodal group) also improved significantly in the post- as compared to the pre-test. Our findings confirm that distributional vowel training, with enhanced distributions in particular, leads to improvement in the classification of difficult L2 contrasts. This result allowed us to pursue our main objective of identifying listeners' strategies and examining the effect of bimodal versus enhanced training on the different strategy types, which will be discussed below.

We found a negative correlation between Spanish listeners' age at testing and pre-test accuracy and also between age of arrival and pre-test accuracy. This is in line with earlier observations for the influence of age of L2 learning on speech perception (e.g., Flege, MacKay, & Meador, 1999) and on production (see Piske, MacKay, & Flege, 2001, for a review). Further, neither higher general comprehension of Dutch nor longer exposure to Dutch as reflected in the length of residence in the Netherlands were significantly related to higher pre-test perception accuracy. Although a number of previous studies have shown an effect of these two factors on L2 sound perception (e.g., Escudero et al., 2009; Flege et al., 1997), others have failed to find these effects (e.g., Cebrian, 2006; Escudero & Wanrooij, 2010). For the second factor (amount of exposure) this discrepancy in outcomes is probably due to the unreliability of length of residence as a measure of the amount of exposure to the target language (e.g., Moyer, 2009; Piske et al., 2001). It is a poor measure when, for instance, learners have little contact with native speakers or when the quality of the new language input is bad (e.g., Moyer, 2009). Nevertheless, if length of residence in the current study reflected the participants' amount of exposure to Dutch, the observed significant relation between length of residence in the Netherlands and improvement after training could be interpreted as a sign that our distributional training facilitated perceptual learning that had started outside the lab.

The latent class analysis of listening strategies indicated a split in initial listening strategies between listeners who did not focus on the critical combination of F1 and F2, and listeners who did. As expected, the former (“pre-test low performers”) had relatively low and the latter (“pre-test high performers”) relatively high accuracy. After the training phase, listeners in the control group did not change strategies, while listeners in the Bimodal and Enhanced groups diversified their strategies. Improvers among the pre-test low performers started to use F1 and/or F2, while pre-test high performers refined their strategies mainly by adding the subtle secondary cue F3. Further, the outcomes revealed no significant difference between bimodal and enhanced training in learning to use F1 and/or F2 for pre-test low performers, while pre-test high performers profited more from enhanced than bimodal training for learning to include F3 in their listening strategies. This shows the importance of looking beyond group results, which can be considerably affected by group composition. The results for pre-test high performers extend previous research, which shows that enhanced differences in critical acoustic cues can facilitate learning by directing listeners' attention to these cues (e.g., Iverson et al., 2005; Jamieson & Morosan, 1986; Kondaurova & Francis, 2010). Because the usefulness of enhanced training was particularly evident in pre-test high performers' new use of F3 in the post-test, it seems that for listeners who are already attentive to the critical cues, enhancement may facilitate attention to additional, more subtle cues.

Further, our participant groups had mainly female participants. Although in our lab we had not observed sex differences in vowel perception earlier (e.g., sex differences in the data of Escudero and Chladkova (2010) could not be found), Obleser, Eulitz, Lahiri, and Elbert (2001) report a larger left-hemispheric activity for women than for men when listening to vowels. Even though this observation does not necessarily mean that men and women use different acoustic cues when listening to vowels, we explored whether women and men showed different listening strategies and learning behavior. We did not find sex differences in pre-test listening strategies, and in the ability to use new cues after training for pre-test low performers. However, we found that among pre-test high performers (who were already using F1 and F2 in the pre-test) men were more likely to start using F3 after training than women. The precise meaning of this observed sex difference is not clear and should be examined in future research.

Listeners who used new cues after training simultaneously increased their use of duration (Section 3.3). This may be a sign of *cue integration*, i.e., the use of both duration and formant frequencies for vowel perception, as predicted in the L1 distributional learning model of Boersma, Escudero, and Hayes (2003), which was more explicitly formulated and extended in Escudero (2005). The model predicts that, in building a phonological contrast, learners initially use a single cue (e.g., relating a certain duration to a phonological category “short”) and then start to integrate additional cues (e.g., also relating an F1 with a certain frequency value to a phonological category “short”) on the basis of their correlational distributions. Listeners in the current study may have been in the process of relating a relatively low F1 and/or F2 that they heard during training to the short duration (for /ɑ/, or the high F1 and F2 to the long duration for /aː/) that they were already able to use before the training. Longitudinal studies are needed to confirm this cue integration pattern, but if it indeed takes place in development, it is remarkable that it can surface after only 2 min of training.

Some listeners in the Enhanced group started to use f0 after training, which may be related to their response bias to /aː/ (Section 3.2). This is because the average f0 of the natural test stimuli, both for the male and female voices, was somewhat lower for /aː/ than for /ɑ/ (Table 2, Section 2.2.1), and thus

more similar to the male voice of the response options. Given that f0 is not relevant for determining vowel identity and that the response options did not differ in this cue, this new strategy was likely to have hampered listeners' performance. Indeed, the average accuracy for the Enhanced pre-test high performers decreased when they started to use f0 (compare Table 6 Enhanced post-test class 2 and Table 5 Enhanced pre-test class 2), while the average higher accuracy for the Enhanced pre-test low performers who started to use this cue could be based entirely on their new use of F2 and their increased use of duration (Table 6 Enhanced post-test class 2 versus Table 5 Enhanced pre-test class 1).

Further, listeners tended to adopt cues in the order duration, F1, F2 and F3. That is, classes that started to use F1 always continued using duration, classes that started to use F2 also started to use or continued using duration and F1, and classes that started to use F3 also started to use or continued using duration, F1 and F2. In other words, although the analysis of listening strategies after training could have identified several other logically possible strategies (such as F2 alone or F1, F2 and F3), it yielded only four strategies, namely (1) duration, (2) duration and F1, (3) duration, F1 and F2, and (4) duration, F1, F2 and F3. With respect to vowel formants, the observed order seems to reflect a ranking from most to least salient, since lower formants have higher amplitudes in the acoustic signal than higher formants (Klatt, 1980) and differences between two vowels in lower formant frequencies are somewhat easier to discriminate than those in higher formant frequencies (Kewley-Port & Watson, 1994). Possibly because of this perceptual difference listeners started using F1 before F2, despite a larger difference in F2 than in F1 between [ɑ] and [aː] (e.g., the difference between the average natural [ɑ] and [aː] stimuli in the test was 1.76 ERB in F2 and 1.47 ERB in F1). The perceptual difference between F1 and F2 may be related to the larger number of distinctions between vowels in the F1 dimension (three levels in Spanish) than in the F2 dimension (two levels in Spanish) that is observed in the vowel inventories of the world's languages (Ladefoged & Maddieson, 2007).

As for duration, it is not certain whether it is intrinsically more salient as a cue than formants. Spanish listeners in almost all pre- and post-strategy types used duration, despite the fact that they do not use it to distinguish Spanish vowels. This finding is in line with these listeners' attested tendency to resort to duration in order to compensate for their failure to use differences in formant frequencies between non-native vowels (e.g., Escudero & Boersma, 2004; Escudero et al., 2009), and shows that this cue must be fairly accessible. Since duration is also used consistently in non-native speech perception by speakers of other languages than Spanish without native durational differences (e.g., Iverson & Evans, 2007), it has been suggested that the cue is relatively easy to parse for humans in general (Bohn, 1995). Alternatively, the accessibility of duration can stem from the absence of a phonemic contrast along this acoustic dimension in the listeners' L1, as suggested by Escudero and Boersma (2004). Specifically, Escudero & Boersma propose that, when presented with a distribution of speech sounds differing in duration, speakers of languages without phonemic contrasts along this dimension can form durational categories 'from scratch', without interference from existing L1 categories.

Nevertheless, if saliency is indeed the driving force underlying the order in which listeners started to use new cues, this order suggests that in a two-minute distributional training not only the frequency of presentation *across* stimuli affects perception, but also the relative saliency of the acoustic components *within* the presented stimuli. With exposure to a language where the distributional properties of an acoustic cue do not contain linguistically relevant information, it seems that listeners can learn to ignore such a cue, even if it is acoustically salient or accessible. For instance, Spanish listeners without L2 experience do not use duration to distinguish native vowels (e.g., Morrison, 2008). Future research is needed to unravel the precise dynamics between saliency and frequency in distributional learning over a longer time span.

Regarding the nature of development in distributional learning, we expected to find roughly the same developmental stages as posited by Escudero (2000, 2005) and Morrison (2008), as discussed in the Introduction. Although we can only ascertain the existence of these stages with longitudinal data, they can indeed be related to the identified listening strategies. Low performers in the pre-test can be interpreted to be in stage 0, because they could not distinguish /ɑ/ and /aː/ and used duration only slightly. The majority of pre-test low performers started to use duration more intensely after distributional training, which signals a transition to stage 1. Most of them simultaneously started to use F1 (and F2), which corresponds to a transition to stage 2. Moreover, pre-test high performers, who used duration, F1 and F2 in the pre-test, could have started in stage 2 or 3, where listeners attend primarily to F1 and F2, as native listeners do. Indeed, the accuracy of the best-performing Spanish classes (in pre- and post-test) was similar to that of native speakers. This is in line with previous research by Díaz et al. (2012), which showed that in categorization tasks L2 listeners' performance may well reach native-speaker levels. Spanish learners came closer to native speakers' listening strategies after exposure to distributional training as opposed to classical music.

Interestingly, our approach of focusing on the *content* of what was learned rather than on *attained accuracy* also made it possible to detect progress that was not associated with high performance scores. For instance, the majority of the Enhanced low performers in the pre-test, who turned to duration exclusively after the training and who continued performing badly in the post-test (21 listeners, see Section 3.3), could still have progressed from stage 0 to stage 1 because duration, which was introduced in stage 1, was irrelevant for distinguishing the response options. Also, the bias toward /aː/ in the post-test of some Bimodal pre-test low performers who continued to perform rather poorly, could reflect Morrison's developmental stage $\frac{1}{2}$, where listeners classify vowels as good or bad examples of Spanish /a/. It is conceivable that the Spanish learners in this group only labeled the tokens that were acoustically furthest away from the Spanish vowel /a/ as Dutch /ɑ/.

Importantly, Escudero (2000, 2005) and Morrison (2008) did not view development as necessarily discrete jumps from one stage to another, while we implicitly assumed such categorical transitions because we modeled the listening strategies as distinct types. The current data show that cues can be adopted one by one, as reflected in the strategy types duration, duration-F1, duration-F1–F2 and duration-F1–F2–F3, and that the use of cues can be intensified (or weakened), as reflected by the beta coefficients. However, the clear increase in accuracy when using more cues (i.e., when comparing classes in Fig. 1, accuracy seems to increase dramatically when a cue is added) suggests that the actual transition between stages is categorical rather than gradual. Longitudinal studies are needed to ascertain the developmental stages shown in the present data and their categorical nature.

In sum, we have demonstrated that distributional vowel training can help learners to improve their classification of difficult non-native contrasts. We show that the changes in perceptual cue use after training are related to participants' listening strategies before training. Latent class regression analysis is a way to identify such strategies. The strategies identified here can be related to previously reported developmental stages for Spanish learners of English and Dutch vowels, which suggests that our method can shed light on the development of second language speech perception.

Acknowledgments

This research was initiated and supported by grant 275.75.005 from the Netherlands Organization for Scientific Research (NWO) awarded to the second author. Research assistants for participant recruitment and testing were also supported by NWO grant 016.024.018 awarded to Paul Boersma. The first author´s work was supported by NWO grant 277-70-008 awarded to Paul Boersma. The third author´s work was supported by NWO grant

452-06-008. The second and third authors' work was also supported by a grant from the priority program Brain & Cognition of the University of Amsterdam.

References

Adank, P., Van Hout, R., & Smits, R. (2004). An acoustic description of the vowels of Northern and Southern Standard Dutch. *Journal of the Acoustical Society of America*, *116*(3), 1729–1738.

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

Allan, L. G., & Gibbon, J. (1991). Human bisection at the geometric mean. *Learning and Motivation*, *22*, 39–58.

Best, C. (1994). The emergence of native-language phonological influences in infants: A perceptual assimilation model. In: J. Goodman, & H. Nusbaum (Eds.), *The development of speech perception: The transition from speech sounds to spoken words* (pp. 167–224). Cambridge, MA: MIT Press.

Boersma, P., Escudero, P., & Hayes, R. (2003). Learning abstract phonological from auditory phonetic categories: An integrated model for the acquisition of language-specific sound categories. *Proceedings of the 15th International Congress of Phonetic Sciences*, 1013–1016.

Boersma, P., & Weenink, D.Praat: Doing phonetics by computer. (2011). ⟨http://www.praat.org⟩. Accessed 10.12.2007).

Bohn, O. S. (1995). Cross-language speech perception in adults. First language transfer doesn't tell it all. In: W. Strange (Ed.), *Speech perception and linguistic experience: Theoretical and methodological issues in cross-language speech research* (pp. 275–300). Timonium, MD: York Press.

Bohn, O. S., & Flege, J. E. (1990). Interlingual identification and the role of foreign language experience in L2 vowel perception. *Applied Psycholinguistics*, *11*, 303–328.

Bouwmeester, S., Sijtsma, K., & Vermunt, J. K. (2004). Latent class regression analysis to describe cognitive developmental phenomena: An application to transitive reasoning. *European Journal of Developmental Psychology*, *1*, 67–86.

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

Cebrian, J. (2006). Experience and the use of non-native duration in L2 vowel categorization. *Journal of Phonetics*, *34*, 372–387.

Chandrasekaran, B., Sampath, P. D., & Wong, P. C. M. (2010). Individual variability in cue-weighting and lexical tone learning. *Journal of the Acoustical Society of America*, *128*(1), 456–465.

Cheour, M., Čeponiené, R., Lehtokoski, A., Luuk, A., Allik, J., Alho, K., et al. (1998). Development of language-specific phoneme representations in the infant brain. *Nature Neuroscience*, *1* (5), 351–353.

Cristià, A., McGuire, G. L., Seidl, A., & Francis, A. L. (2011). Effects of the distribution of acoustic cues on infants' perception of sibilants. *Journal of Phonetics*, *39*, 388–402.

Díaz, B., Mitterer, H., Broersma, M., & Sebastián-Gallés, N. (2012). Individual differences in late bilinguals' L2 phonological processes: From acoustic-phonetic analysis to lexical access. *Learning and Individual Differences*, http://dx.doi.org/10.1016/j.lindif.2012.05.005.

Escudero, P. (2000). *Developmental patterns in the adult L2 acquisition of new contrasts: The acoustic cue weighting in the perception of Scottish tense/lax vowels by Spanish speakers*. Scotland, UK: University of Edinburgh [Unpublished Master's thesis].

Escudero, P. (2005). *Linguistic perception and second-language acquisition: Explaining the attainment of optimal phonological categorization*. Utrecht University: LOT Dissertation Series 113 [Doctoral dissertation].

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

Escudero, P., Benders, T., & Wanrooij, K. (2011). Enhanced bimodal distributions facilitate the learning of second language vowels. *Journal of the Acoustical Society of America*, *130*(4), EL206–EL212.

Escudero, P., & Boersma, P. (2004). Bridging the gap between L2 speech perception research and phonological theory. *Studies in Second Language Acquisition*, *26*, 551–585.

Escudero, P., & Chladkova, K. (2010). Spanish listeners' perception of American and Southern British English vowels: Different initial stages for L2 development. *Journal of the Acoustical Society of America*, *128*, EL254–EL260.

Escudero, P., & Wanrooij, K. (2010). The effect of L1 orthography on non-native vowel perception. *Language and Speech*, *53*(3), 343–365.

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

Flege, J. E. (2002). Interactions between the native and second-language phonetic systems. In: P. Burmeister, T. Piske, & A. Rohde (Eds.), *An integrated view of language development: Papers in honor of henning wode* (pp. 217–243). Trier, Germany: WVT.

Flege, J. E. (2003). Assessing constraints on second-language segmental production and perception. In: A. Meyer, & N. Schiller (Eds.), *Phonetics and phonology in language comprehension and production* (pp. 319–355). Berlin: Mouton de Gruyter.

Flege, J. E., Bohn, O.-S., & Jang, S. (1997). Effects of experience on non-native speakers' production and perception of English vowels. *Journal of Phonetics*, *25*, 437–470.

Flege, J. E., MacKay, I. R. A., & Meador, D. (1999). Native Italian speakers' perception and production of English vowels. *Journal of the Acoustical Society of America*, *106*(5), 2973–2987.

Flege, J. E., & MacKay, I. R. A. (2004). Perceiving vowels in a second language. *Studies in Second Language Acquisition*, *26*, 1–34.

Giezen, M., Escudero, P., & Baker, A. (2010). Use of acoustic cues by children with cochlear implants. *Journal of Speech, Language and Hearing Research*, *53*(6), 1440–1457.

Goto, H. (1971). Auditory perception by normal Japanese adults of the sounds "l" and "r". *Neuropsychologia*, *9*, 317–323.

Grün, B., & Leisch, F. (2007). Fitting finite mixtures of generalized linear regressions in R. *Computational Statistics and Data Analysis*, *51*, 5247–5252.

Guenther, F. H., & Gjaja, M. N. (1996). The perceptual magnet effect as an emergent property of neural map formation. *Journal of the Acoustical Society of America*, *100*, 1111–1121.

Gulian, M., Escudero, P., & Boersma, P. (2007). Supervision hampers distributional learning of vowel contrasts. *Proceedings of the international congress of phonetic sciences* (pp. 1893–1896), *Saarbrücken*, *Germany*.

Hayes-Harb, R. (2007). Lexical and statistical evidence in the acquisition of second language phonemes. *Second Language Research*, *23*(1), 65–94.

Hollien, H., Dew, D., & Philips, P. (1971). Phonational frequency ranges of adults. *Journal of Speech and Hearing Research*, *14*, 755–760.

Huang, G.-H., & Bandeen-Roche, K. (2004). Building an identifiable latent class model with covariate effects on underlying and measured variables. *Psychometrika*, *69*, 5–32.

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

Iverson, P., Hazan, V., & Bannister, K. (2005). Phonetic training with acoustic cue manipulations: A comparison of methods for teaching English /r/–/l/ to Japanese adults. *Journal of the Acoustical Society of America*, *118*, 3267–3278.

Iverson, P., & Evans, B. G. (2007). Learning English vowels with different first-language vowel systems: Perception of formant targets, formant movement and duration. *Journal of the Acoustical Society of America*, *122*(5), 2842–2854.

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, *59*(4), 434–446.

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

Kewley-Port, D., & Watson, C. S. (1994). Formant-frequency discrimination for isolated English vowels. *Journal of the Acoustical Society of America*, *95*(1), 485–496.

Klatt, D. H. (1980). Software for a cascade/parallel formant synthesizer. *Journal of the Acoustical Society of America*, *67*(3), 971–995.

Kondaurova, M., & 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. *Journal of Phonetics*, *38*, 569–587.

Kuhl, P., Andruski, J., Chistovich, I., Chistovich, L., Kozhevnikova, E., Ryskina, V., et al. (1997). Cross-language analysis of phonetic units in language addressed to infants. *Science*, *227* (5326), 684–686.

Lacerda, F. (1995). The perceptual-magnet effect: An emergent consequence of exemplar-based phonetic memory. In K. Elenius, & P. Branderyd(Eds.) *Proceedings of the XIIIth international congress of phonetic sciences*, *Vol. 2*. (pp. 140–147), *Stockholm*, Stockholm: KTH and Stockholm University.

Ladefoged, P., & Maddieson, I. (2007). *The sounds of the world's languages* (10th ed.). Malden, MA/Oxford, UK/Carlton, Australia: Blackwell Publishing.

Leisch, F. (2004). FlexMix: A general framework for finite mixture models and latent class regression in R. *Journal of Statistical Software*, *11*(8).

Lin, T. H., & Dayton, C. M. (1997). Model selection information criteria for nonnested latent class models. *Journal of Educational and Behavioral Statistics*, *22*(3), 249–264.

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

Maye, J., & Gerken, L. A. (2000). Learning phonemes without minimal pairs. In: C. Howell (Ed.), *BUCLD 24 Proceedings* (pp. 522–533). Somerville, MA: Cascadilla Press.

Maye, J., & Gerken, L. A. (2001). Learning phonemes: How far can the input take us?. In: A. H.-J. Do (Ed.), *BUCLD 25 Proceedings* (pp. 480–490). Somerville, MA: Cascadilla Press.

Maye, J., Weiss, D., & Aslin, R. (2008). Statistical phonetic learning in infants: Facilitation and feature generalization. *Developmental Science*, *11*(1), 122–134.

Maye, J., Werker, J. F., & Gerken, L. A. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition*, *82*(3), B101–B111.

Mayr, R., & Escudero, P. (2010). Explaining individual variation in L2 perception: Rounded vowels in English learners of German. *Bilingualism: Language and Cognition*, *13*(3), 279–297.

McCullagh, P., & Nelder, J. A. (1989). *Generalized linear models* (2nd ed.). Boca Raton, Florida: Chapman & Hall/CRC.
Miyawaki, K., Strange, W., Verbrugge, R., Liberman, A. M., Jenkins, J. J., & Fujimura, O. (1975). An effect of linguistic experience: The discrimination of [r] and [l] by native speakers of Japanese and English. *Perception and Psychophysics*, *18*(5), 331–340.
Morrison, G. S. (2007). Logistic regression modelling for first- and second-language perception data. In: M. J. Solé, P. Prieto, & J. Mascaró (Eds.), *Segmental and prosodic issues in romance phonology* (pp. 219–236). Amsterdam: John Benjamins.
Morrison, G. S. (2008). L1-Spanish speakers' acquisition of the English /i/–/ɪ/ contrast: Duration-based perception is not the initial developmental stage. *Language and Speech*, *51*, 285–315.
Morrison, G. S. (2009). L1-Spanish speakers' acquisition of the English /i/–/ɪ/ contrast II: Perception of vowel inherent spectral change. *Language and Speech*, *52*, 437–462.
Moyer, A. (2009). Input as critical means to an end: Quantity and quality of experience in L2 phonological attainment. In T. Piske, & M. Young-Scholten (Eds.), *Input matters in SLA* (pp. 159–174). Bristol (UK)/Buffalo (USA)/Toronto (Canada): Multilingual Matters.
Obleser, J., Eulitz, C., Lahiri, A., & Elbert, T. (2001). Gender differences in functional hemispheric asymmetry during processing of vowels as reflected by the human brain magnetic response. *Neuroscience letters*, *314*, 131–134.
Oxford, R., Nyikos, M., & Ehrman, M. (1988). Vive la Différence? Reflections on sex differences in use of language learning strategies. *Foreign Language Annals*, *21*(4), 321–329.
Piske, T., MacKay, I. R. A., & Flege, J. E. (2001). Factors affecting degree of foreign accent in an L2: A review. *Journal of Phonetics*, *29*, 191–215.
Pols, L. C. W., Tromp, H. R. C., & Plomp, R. (1973). Frequency analysis of Dutch vowels from 50 male speakers. *Journal of the Acoustical Society of America*, *53*, 1093–1101.
Schmitz, N., Malla, A., Norman, R., Archie, S., & Zipursky, R. (2007). Inconsistency in the relationship between duration of untreated psychosis (DUP) and negative symptoms: Sorting out the problem of heterogeneity. *Schizophrenia Research*, *93*, 152–159.
Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, *6*(2), 461–464.
Stager, C. L., & Werker, J. F. (1997). Infants listen for more phonetic detail in speech perception than in word-learning tasks. *Nature*, *388*, 381–382.
Stevens, S. S., Volkmann, J., & Newman, E. B. (1937). A scale for the measurement of the psychological magnitude pitch. *Journal of the Acoustical Society of America*, *8*, 185–190.
Sundberg, U. (2001). Consonant specification in infant-directed speech. Some preliminary results from a study of Voice Onset Time in speech to one-year-olds. In *Working Papers 49* (pp. 148–151). Department of Linguistics, Stockholm University.
Sundberg, U., & Lacerda, F. (1999). Voice onset time in speech to infants and adults. *Phonetica*, *56*(3–4), 186–199.
Uther, M., Knoll, M. A., & Burnham, D. (2007). Do you speak E-NG-L-I-SH? A comparison of foreigner- and infant-directed speech. *Speech Communication*, *49*, 2–7.
Van Hesse, A. J., & Schouten, M. E. H. (1999). Categorical perception as a function of stimulus quality. *Phonetica*, *56*, 56–72.
Visser, I., Raijmakers, M. E. J., & Pothos, E. M. (2009). Individual strategies in artificial grammar learning. *The American Journal of Psychology*, *122*(3), 293–308.
Werker, J. F., & Logan, J. S. (1985). Cross-language evidence for three factors in speech perception. *Perception and Psychophysics*, *37*, 35–44.
Werker, J. F., & Tees, R. C. (2002). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development*, *25*, 121–133 (First published in 1984: *Infant Behavior and Development*, 7, 49–63).
Yamada, R. A. (1995). Age and acquisition of second language speech sounds: Perception of American English /r/ and /l/ by native speakers of Japanese. In: W. Strange (Ed.), *Speech perception and language experience: Issues in cross-language research* (pp. 305–320). Baltimore, MD: York Press.
Yamaguchi, K. (2000). Multinomial logit latent-class regression models: An analysis of the predictors of gender-role attitudes among Japanese women. *American Journal of Sociology*, *105*(6), 1702–1740.
Yoshida, K. A., Pons, F., Maye, J., & Werker, J. F. (2010). Distributional phonetic learning at 10 months of age. *Infancy*, *15*(*4*), 420–433.