Ab initio perceptual learning of foreign language sounds: Spanish consonant acquisition by Chinese learners

Jian Gong a, b, *, María Luisa García Lecumberri b, Martin Cooke b, c

a School of Foreign Languages, Jiangsu University of Science and Technology, 2 Mengxi Road, 212003 Zhenjiang, China
b Language and Speech Laboratory, Universidad del País Vasco, Paseo de la Universidad 5, 01006 Vitoria, Spain
c Ikerbasque (Basque Science Foundation), Spain

Article info
Article history:
Received 12 January 2016
Received in revised form 17 March 2017
Accepted 19 March 2017

Keywords:
Phonetic training
Consonant perception
Spanish
Chinese
Second language sound acquisition

Abstract
High-variability phonetic training is effective in the acquisition of foreign language sounds. Previous studies have largely focused on small sets of contrasts, and have not controlled for the quantity of prior or simultaneous exposure to new sounds. The current study examined the effectiveness of phonetic training in full-inventory foreign language consonant acquisition by listeners with no previous exposure to the language. Chinese adult listeners underwent an intensive training programme, bracketed by tests that measured both assimilation of foreign sounds to native categories, and foreign category identification rates and confusions. Very rapid learning was evident in the results, with initial misidentification rates halving by the time of the mid-test, and continuing to fall in subsequent training sessions. Changes as a result of training in perceptual assimilation together with improved identifications and reduced response dispersion suggest an expansion of listeners’ native categories to accommodate the foreign sounds and an incipient process of foreign language category formation.

1. Introduction

Adult language learners have problems in perceiving and producing certain non-native phonemes. Together with age of acquisition and first language (L1) influence, the amount of foreign language (FL) exposure is crucial in learners’ FL perception and production. However, FL exposure can be difficult to quantify. In immersion settings, length of residence is normally used as a proxy (Flege & Liu, 2001) since it is an easily measurable variable, although not necessarily highly predictive (DeKeyser & Larson-Hall, 2005), perhaps, as Flege (2009) suggests, because length of residence in itself does not necessarily reflect an individual’s real amount or quality of input in the target language. Length of formal FL instruction maybe similarly unreliable. Increasingly, particularly for English, learners get exposed to the FL outside the classroom through music, films, internet and conversations. Under these circumstances, quantifying the total amount of exposure is challenging (Flege, 2009).

Training is one way of controlling and enhancing the amount of L2 input. High-variability phonetic training is effective in improving adults’ foreign language perception (e.g., Lengeris & Nicolaidis, 2015; Lively, Logan, & Pisoni, 1993; Nishi & Kewley-Port, 2007). These studies have also shown that the use of a relatively small amount of high-variability speech material forces adult listeners to filter out natural variation (such as that produced by positional allophones and individual speaker

* Corresponding author. School of Foreign Languages, Jiangsu University of Science and Technology, 2 Mengxi Road, 212003 Zhenjiang, China.
E-mail address: j.gong@just.edu.cn (J. Gong).

http://dx.doi.org/10.1016/j.system.2017.03.018
0346-251X/© 2017 Elsevier Ltd. All rights reserved.
differences) and construct more robust phonetic categories for the target language, as attested by rapid improvement in their FL phoneme identification performance and generalisation to other contexts/voices. Most high-variability training studies have been carried out with experienced listeners who had been learning the target language for some time before the experiment, resulting in a large degree of variability in the amount and form of input prior to training. Even once training has started, the amount of additional exposure is difficult to control, especially in an immersion setting. Studies with adult listeners employing English as the target language further complicate quantification of prior exposure given the widespread use of English in the media. Prior knowledge and concurrent exposure are thus confounding factors in controlled studies of the role of input in phonetic category development.

An alternative approach is to train listeners on artificial language sounds or on languages previously unknown to listeners (e.g., Holt & Lotto, 2006). Among the few studies which have focused on the acquisition of entirely unknown phonetic categories, most have been based on a small set of contrasts. Goelstani and Zatorre (2004) trained English speaking listeners to identify a Hindi dental—retroflex contrast. Similarly Pruitt, Jenkins, and Strange (2006) trained American English speakers and Japanese speakers to identify this contrast. Hirata, Whitehurst, and Cullings (2007) focused on training native English speakers to identify five Japanese vowel length contrasts. In the current study, we also investigate the acquisition of unknown phonetic categories through training, mainly as a way of controlling learners’ competence and input as well as the intervention of higher-order linguistic processing. We include the full consonantal inventory of the target language (Spanish) to better observe mutual influences amongst phonetic categories within the Spanish inventory and across the two languages’ sound systems.

More recently, high-variability phonetic training has been extended to a larger set of contrasts, mostly involving English vowels: Japanese learners, 9 vowels (Nishi & Kewley-Port, 2007); Spanish and German learners, 14 vowels (Iverson & Evans, 2009); French learners, 14 vowels (Iverson, Pinet, & Evans, 2012). Greek learners of English in a study by Lengeris and Nicolaidis (2015) did use consonants, but only for 7 exemplars. To the authors’ knowledge, no studies have used the full set of consonants from an unknown target language, and consequently we are left with an incomplete picture of the mutual consonant confusions that can occur at an early stage of acquisition. The primary purpose of the current study is to examine, in a setting with tight control over input and prior exposure to the target language, the evolution of FL sound acquisition for a full consonant inventory, following the high-variability training paradigm. Here, we investigate the case of Chinese listeners exposed to high-variability training in an unknown language, viz. Spanish.

Few studies have investigated the perception of Spanish consonants for non-native listeners. Rose (2010, 2012) examined how native English learners of Spanish discriminate and assimilate four Spanish consonants (/r, r, t, d/) to English categories, finding Spanish /r/ to be the most difficult contrast for native English speakers. Spanish /r/ was assimilated to English /r/ most of the time, Spanish /r/ was assimilated to /t/ and /d/, and Spanish /d/ was assimilated to /l/, /d/ and /r/. More studies investigated the production of Spanish consonants by non-native learners. Diaz Campos (2004) compared the production of Spanish word–initial stops, intervocalic fricatives, word–final laterals and palatal nasals by native English speakers in L2 and FL settings. Face and Menke (2009) analyzed how native English learners of Spanish produce Spanish voiced plosives in intervocalic positions. Chen (2007) compared plosive productions by native Spanish speakers and Chinese L2 learners. Gonzalez Lopez and Counselman (2013) focused on the production of Spanish voiceless plosives by English speakers.

L1 influence is probably the single strongest factor in FL sound acquisition (Best, 1995; Flege, 1993, pp. 233–277; Kuhl, 1993) but a learner’s phonological system is dynamic and can, in principle, change in its treatment of FL and L1 sounds (Best & Tyler, 2007; Heeren & Schouten, 2008). A learner’s interpretation of the FL phonological system and FL to L1 assimilation patterns may also change as a function of experience (Major, 2001; Wu, Munro, & Wang, 2014). Changes in L1 vowel production have been observed as a result of short-term training on two unknown non-native vowels (Kartushina, Hervais-Adelman, Frauenfelder, & Golestani, 2016). In a cross-language vowel mapping experiment, Bundgaard-Nielsen, Best, and Tyler (2011) found that those Japanese learners of English with more English experience, as measured by a larger vocabulary size, had more concentrated English to Japanese assimilation patterns (i.e., individual English vowels were assimilated to fewer Japanese categories) than learners with less English experience. Bundgaard-Nielsen et al. (2011) claim that more experienced Japanese learners of English were ‘reattuning and rephonologizing’, by stretching their L1 categories to accommodate the FL vowel system, or through the formation of new categories for FL sounds. Similarly, in a study of French to English vowel assimilation, Levy (2009) found that native American English speakers with no French experience demonstrated a scattered assimilation pattern for French mid-front rounded /œ/, while those with extensive experience had more concentrated assimilation patterns for /œ/. Interestingly, similar assimilation concentrations were reported in a perceptual training study (Iverson & Evans, 2009), where native Spanish and German (L1) speakers’ English (L2) vowel assimilation selections became more consistent on those target sounds with the largest assimilation percentages after high-variability auditory training. To explore the possibility that training can induce rephonologization, we examine here changes in assimilation patterns following intensive exposure.

The current study focuses on two research questions. First, is high-variability phonetic training effective ab initio in the challenging setting where all FL categories are present? Second, do listeners’ non-native to native assimilation patterns change as a function of experience through intensive training? To ensure tight control over FL input prior to and during the

---

1 According to Bundgaard-Nielsen et al. (2011), rephonologization means to “establish an L2 phonology by modification of or addition to the learner’s existing L1 phonological system.”
experiment, the target language chosen (Spanish) was one in which participants had no previous exposure and were unlikely to encounter outside the training context. Chinese listeners underwent an intensive 16-part training programme, bracketed by tests measuring performance on the FL and FL-to-L1 assimilations. Consequently, the precise amount of exposure to the sounds of the target language was measurable, and changes in consonant confusions could be monitored.

2. The Spanish and Chinese consonant systems

Chinese is described as having 24 consonant phonemes /pʰ, tʰ, kʰ, tʃʰ, tsʰ, ts, ʃ, ɹ, θ, s, f, s̩, ɕ, x, m, n, ɲ, j, l, w/ (Lee & Zee, 2003). Spanish has 19 consonant phonemes /p, b, t, d, k, g, j, f, ð, s, ɾ, x, m, n, ɲ, l, r, ʎ/ although the status of /ɻ/ is debated (Gil Fernandez, 2007; Hualde, 2005).

A clear difference between the two languages lies in the affricates. Chinese has a large group of affricate sounds contrasting in aspiration /tsʰ, ts, ʃ, ɹ/ plus the palatalised (un)aspirated pair, while Spanish has only one affricate phoneme /tʃ/ and, allophonically, a voiced palatal affricate /ʃʃ/. The Spanish voiceless palatal affricate /ʃʃ/ is close to Chinese /ʃw/.

Chinese has three pairs of plosives, contrasting in aspiration. Spanish also has three plosive pairs, but distinguished by voicing. Chinese plosives are all voiceless while Spanish plosives are all unaspirated, so Spanish /p, t, k/ are similar to the Chinese voiceless unaspirated /p, t, k/ as are the plosive realisations of Spanish /b, d, g/. Chinese listeners find it hard to distinguish between the two types of Spanish plosives (Chen, 2007). Spanish /b, d, g/ are often realised as continuants /β, δ, γ/ (Hualde, 2005).

Spanish nasals /m, n/ are similar to their Chinese counterparts. The Spanish palatal nasal /ɲ/ resembles the realisation of Chinese /n/ when followed by high front vowels.

Spanish and Chinese lack voiced fricative phonemes. Labiodental /ʃ/ and velar /x/ in both languages are similar. Some varieties of Spanish (including the one in the current study) have an interdental fricative /θ/, Spanish /s/ has several possible dialectal realisations. In the variety of Spanish used here the main realisations are alveolar (apical or laminal). Chinese has two different phonemes located in these articulatory regions: the dental fricative /s/ and the apical-post-alveolar/retroflex /ʃ/.

Chinese has two approximant phonemes (/ɻ, ɭ/) and two glides whose status is controversial (Lin, 2007). Chinese /ʃ/ is an apical post-alveolar approximant (Lee & Zee, 2003) sometimes described as retroflex (Lin, 2005), which may also have fricative realisations. Spanish has two contrastive rhotic phonemes, the alveolar tap /ɾ/ and the alveolar trill /ɾ̩/, neither of which has close counterparts in Chinese. Spanish canonically has two lateral phonemes but the palatal lateral /ɻ/ has been merged with the palatal approximant /ʃʃ/ in most accents. The realisation of this latter sound varies between approximant /ʃʃ/ and fricative /ʃ/ and affrimate /ʃʃ/ (Gil Fernandez, 2007; Hualde, 2005). The problems which Chinese speakers face when learning other languages such as Spanish or English (Lee & Xue, 2011) may be related to the actual realisations of the rhotics plus the more restrictive phonotactic distributions of liquids in Chinese.

3. Methods

3.1. Overview

The study employed three tasks arranged in sequence (Table 1): (i) Perceptual assimilation: a category assimilation task in which listeners hear Spanish consonants and identify them using Chinese categories; (ii) Spanish consonant identification: a forced-choice test of Spanish consonant identification; and (iii) Training: a forced-choice procedure involving Spanish consonant identification with feedback on incorrect responses. Tasks (i) and (ii), in that order, constituted a test block, which was carried out at three points in the training sequence: prior to training (pre-test), halfway through training (mid-test) and following training (post-test). Speech materials used in the test block were identical for the pre-, mid- and post-tests.

3.2. Participants

Twenty native Mandarin Chinese listeners, 11 females and 9 males, participated in the study. All were students at the Southern Medical University, China (age: 20–23, mean: 20.5). To avoid possible dialectal influences on listeners’ FL perception (Gong, Cooke, & Garcia Lecumberri, 2010), all participants were from north China and spoke with a Northern Mandarin dialect. All were studying medical courses and had received formal instruction in English for 10.3 years on average. None had studied Spanish or any other Romance language, and no participant had lived outside China. No participant reported hearing
problems. Listeners were paid for their participation. One listener was excluded due to a technical problem during the post-test.

An additional cohort of eight listeners with a similar profile to those in the main study (native Mandarin, no knowledge of any FL other than English, mean age 23.6, paid) formed a control group, enabling a measure of possible influences of factors other than training with feedback (e.g., external factors or familiarity effects from the test itself). Since such effects were deemed to be measurable by the point of the mid-test, the control group listeners undertook only the pre- and mid-tests, with a gap of two days matching the experimental group.

3.3. Materials

All tasks employed consonants in vowel-consonant-vowel (VCV) tokens.

3.3.1. Spanish

The Spanish VCV corpus, collected for the current study, contains 18 Spanish consonant phonemes /p, b, t, d, k, g, tʃ, f, ð, s, x, m, n, l, r, j/. Consonant /k/ was excluded since most young Spanish speakers do not differentiate it from /j/ (Hualde, 2005). VCVs with all 9 combinations of the 3 corner vowels /a, i, u/ in initial and final positions were recorded with both initial and final vowel stress, leading to 324 combinations (18 consonants × 9 vowel contexts × 2 stress patterns). 16 male native talkers with northern Spanish accents living in the vicinity of Vitoria produced the entire corpus. Speakers were instructed to produce the VCV tokens at a normal speaking rate. Tokens with initial and final stress were collected in separate recording blocks. Recordings took place in a sound-attenuated recording studio in the Phonetics Laboratory at the University of the Basque Country. Recording was controlled by a custom MATLAB program which presented the next token for the talker to utter on a computer screen in response to a keypress from the talker. Recordings were made using an AKG model 4500 microphone and a Motu 8pre A-D converter at 44.1 kHz, later downsampled to 25 kHz and high-pass filtered to remove components below 50 Hz. Tokens were automatically endpointed and subsequently manually checked and adjusted where necessary to remove leading and trailing silence. Two native judges screened the entire corpus to identify mis-pronunciations and noisy tokens, resulting in the removal of 253 tokens. Tokens from 10 speakers were used in the training sessions, while material from the remaining 6 speakers was used for the FL and assimilation tests.

3.3.2. Chinese

Chinese control tokens came from a VCV corpus collected for an earlier English-Chinese mapping study (Gong et al., 2010). This corpus contains tokens for 24 Chinese consonants [pʰ, p, tʰ, t, kʰ, k, tsʰ, ts, tʃʰ, tʃ, w, f, s, j, ɾ, x, m, n, j, l, j, w] in the same 9 vowel contexts employed in the Spanish corpus. The status of the palatal consonants in the Chinese phonemic inventory is controversial (Lin, 2007) but we decided to include them following Lee and Zee (2003). The glides [j, w] were included because the phonological status of [j] is equally controversial in Spanish but it is nevertheless a very frequent sound, and for [w] our experience showed that it is a frequent candidate in Chinese learners’ confusions for certain Spanish consonants such as Spanish /b/ and /g/, which are often assimilated to Chinese /w/. Speakers produced both initial and final vowels in the first lexical tone. Since lexical stress is not a feature of Chinese, the corpus does not differentiate the two stress types. Seventeen male speakers produced the Chinese corpus, generating 3672 tokens in all, reduced to 3331 after screening.

3.4. Tasks

Tasks were carried out in a quiet computer laboratory at the Southern Medical University in China. Participants were tested simultaneously. Stimuli were presented using AC 97 sound cards and SALAR A522 headphones. Stimulus presentation and response collection was controlled by custom MATLAB programs. All stimuli were normalised to have equal RMS energy prior to presentation. Participants were allowed to adjust the volume to a comfortable listening level.

3.4.1. Perceptual assimilation

The perceptual assimilation task required listeners to classify Spanish consonants (plus Chinese controls) into Chinese categories and to rate their category goodness (Best, 1995), the latter used as a measure of conformity to the L1 prototype (Iverson et al., 2003). Stimuli were 360 Spanish VCVs and 48 Chinese VCV control tokens. Spanish VCVs were made up of 10 each of initial and final stress for each of the 18 Spanish consonants, randomly sampled from 6 talkers. VCV stimuli had different initial vowels chosen from /a, i, u/ but the final vowel was always /a/ to give a relatively constant coarticulatory context. Chinese VCV control tokens (2 exemplars for each of the 24 consonants) were used to check whether the task was understood and to provide a native reference (Guion, Flege, Akahane-Yamada, & Pruitt, 2000) for the ratings given to Spanish stimuli. An additional 30 Chinese VCVs formed practice stimuli. On average, vowels in Chinese tokens were somewhat longer than their Spanish counterparts (initial: 129 vs. 113 ms, final: 158 vs. 127 ms). To avoid providing cues that might enable listeners to identify control tokens, initial vowels longer than 100 ms and final vowels longer than 110 ms were truncated by applying a linear ramp to the onset or offset portion, providing a better match between the Chinese and Spanish VCVs. Since a previous study (Gong et al., 2010) found that Latin alphabet representations of Chinese phonetic categories produced orthographic biases on perceived categories, listeners selected their response from an onscreen grid which displayed Chinese characters with the corresponding consonant in initial position. An exception was /ŋ/, which was presented in final position to
meet phonotactic constraints. Listeners were not informed about the language of the sounds they would hear. On hearing each token, listeners first clicked on the category in the grid, then moved a slider bar to rate the category goodness of fit. Endpoints of the slider were labelled in Chinese as ‘bad exemplar’ and ‘good exemplar’. Responses were real-valued in the range 0–100. The whole test lasted about 30 min.

3.4.2. Spanish consonant identification

For the Spanish consonant identification test listeners identified Spanish VCVs using Spanish consonant categories. A different subset of the Spanish corpus from that employed in the perceptual assimilation task was used. The Spanish consonant identification test contained 360 VCVs, 10 of each stress for each of the 18 consonants derived by random sampling from 5 speakers, using all nine vowel contexts. The interface was similar to the perceptual assimilation test. Latin alphabet letters were used to represent Spanish consonant categories due to a clear orthographic-phonemic correspondence. To avoid a possible orthographic influence from Chinese Pinyin characters, the symbol ‘h’ was used to represent Spanish sound /x/. No instruction about the Spanish consonant system was given prior to the test: listeners learned the mapping from Spanish consonants to orthography solely as a result of feedback given to wrong answers during the subsequent training sessions. It was felt that any instructions on the phoneme-grapheme correspondence in Spanish would constitute a form of training and for our purposes it was necessary to ensure that listeners approached the training from a completely naive initial state.

3.4.3. Training sessions

During each of the 16 training sessions listeners identified Spanish VCVs as one of 18 Spanish consonant categories, receiving feedback on incorrect responses. VCV tokens were derived from the 10 speakers of the training portion of the corpus. Each training session consisted of 180 tokens, 10 for each of 18 Spanish consonants (5 each for initial and final stress). All 9 types of vowel context were used in order to increase exposure to coarticularatory variants. No token was repeated during 16 training sessions, leading to a total exposure to 2880 different tokens, 160 for each consonant. Training sessions occurred on different days from test sessions. Each of the four training days was composed of four training sessions, for a total of 16 training sessions. Each training session lasted approximately 10 min. The interface was identical to that used in the FL test. For incorrect responses, a button with the correct answer was highlighted, and listeners had to listen to the stimulus exactly once before proceeding to the next token. Hence, the total amount of exposure was controlled, with a limited amount of additional exposure for incorrect tokens.

4. Results

Percentages are reported in the text but converted to rationalised arcsine units (Studebaker, 1985) for statistical analysis.

4.1. Assimilation test

Chinese control tokens were identified correctly 94.5, 94.2 and 93.5% of the time in the pre-, mid- and post-tests. A one-way repeated-measures ANOVA with a factor of test time confirmed that these differences are not significant ($F(2, 36) = 0.38, p = 0.69$). Goodness ratings for the L1 tokens reached 87.4 (on a scale of 0–100) in the pre-test, falling to 84.5 and 83.4 in the mid- and post-tests. Here, the effect of test time approached significance ($F(2, 36) = 3.0, p = 0.06, \eta^2_p = 0.14$).

Table 2 summarises mappings and goodness ratings for the Spanish tokens in pre-, mid- and post-tests. Only those Chinese consonant categories reported more than 5% of the time are included. To test for an overall training effect on assimilation patterns, a linear mixed-effects analysis was performed on the RAU-transformed percentages from the dominant assimilation mappings, using lme4 (Bates, Maechler, & Bolker, 2012). The model included one fixed factor, Training (pre-test vs. post-test, with pre-test as the reference level), and by-subject and by-consonant random slopes adjusted for Training were included as random effects. The analysis indicates a significant effect of Training [$\beta = 9.863, SE = 4.521, t = -2.182, p = 0.042 < 0.05$].

In the pre-test Spanish consonants /p, t, k, f, x, m, n, l/ were predominantly categorised to single Chinese categories with very high percentages (Table 2). Separate linear mixed-effects analysis on these sounds with Training as a fixed factor and with by-subject and by-consonant random slopes adjusted for Training showed no significant Training effect for this group of sounds [$\beta = 1.664, SE = 1.716, t = -0.97, p = 0.361$]. For the remaining 10 Spanish consonants, a separate linear mixed-effects analysis confirmed a significant Training effect [$\beta = 16.421, SE = 7.266, t = -2.260, p = 0.046 < 0.05$]. Separate linear mixed-effects analyses were conducted on the individual assimilation mappings, with Training as a fixed factor and with by-subject intercept as a random effect; an asterisk indicates significant effects ($p < 0.05$) in Table 2.

Similar linear mixed-effects analyses were conducted on goodness ratings, with Training as the fixed factor, and by-subject and by-consonant random slopes adjusted for Training as random effects. The results showed no overall significant change of goodness ratings before and after training for the 18 Spanish consonants [$\beta = 2.343, SE = 3.436, t = -0.682, p = 0.502$]. Separate linear mixed-effects analyses indicate no goodness rating change after training for the group of sounds whose assimilations did not change significantly, i.e. /p, t, k, f, x, m, n, l/ [$\beta = 3.296, SE = 3.204, t = -1.029, p = 0.316$], nor for those that did /b, d, g, jf, ʊ, s, n, r, r, j/ [$\beta = 1.942, SE = 4.752, t = -0.409, p = 0.69$]. Indeed, there is no systematic correlation between changes in assimilation and goodness ratings from pre-test to mid-test or post-test [pre-mid: $r = 0.06, p = 0.78$; pre-post: $r = 0.11, p = 0.59$].
4.1.1. Initial assimilation patterns

In the pre-test, the Spanish consonants /p, t, k, f, x, m, n, l/ were predominantly categorised to single Chinese categories with similar frequencies and goodness ratings as Chinese control sounds, indicating that Chinese listeners perceived these sounds as being virtually the same as their native counterparts. Spanish sounds /ʧ, ɕ, j/ were also largely categorised to single Chinese categories but slightly less frequently and with a lower goodness rating, suggesting that these Spanish sounds are less similar to their Chinese counterparts. Spanish /ɲ/ and /r/, although in the main also categorised to a single Chinese category, had much lower goodness ratings. The remaining Spanish sounds /b, d, ɡ, s, ɾ/ showed more dispersed assimilation patterns, frequently being placed into two or more Chinese categories or receiving lower goodness ratings.

4.1.2. Changes in assimilation following training

As noted above, the assimilation patterns of those Spanish consonants with very similar Chinese counterparts showed no change as a result of training and, particularly for the three voiceless plosives, a similar pattern to that obtained for the L1 control consonants is observed. However, changes in assimilation patterns for the Spanish voiced plosives were striking and statistically significant.

Further assimilation pattern changes are present for Spanish consonants that had a weaker L1 correspondence. The number of assimilation targets with significant response proportions for the sounds /d, g, j, s, ɬ/ was reduced (e.g., /ɡ/ changes

<table>
<thead>
<tr>
<th>FL stimulus</th>
<th>L1 response</th>
<th>Assimilation (%)</th>
<th>Goodness (0–100)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>pre</td>
<td>mid</td>
</tr>
<tr>
<td>p</td>
<td>p</td>
<td>99</td>
<td>98</td>
</tr>
<tr>
<td>b</td>
<td>p</td>
<td>32</td>
<td>66</td>
</tr>
<tr>
<td>w</td>
<td>t</td>
<td>28</td>
<td>77</td>
</tr>
<tr>
<td>t</td>
<td>t</td>
<td>94</td>
<td>96</td>
</tr>
<tr>
<td>d</td>
<td>t</td>
<td>18</td>
<td>34</td>
</tr>
<tr>
<td>ts</td>
<td>l</td>
<td>30</td>
<td>16</td>
</tr>
<tr>
<td>l</td>
<td>j</td>
<td>99</td>
<td>98</td>
</tr>
<tr>
<td>k</td>
<td>k</td>
<td>24</td>
<td>57</td>
</tr>
<tr>
<td>g</td>
<td>k</td>
<td>18</td>
<td>14</td>
</tr>
<tr>
<td>x</td>
<td>l</td>
<td>21</td>
<td>8</td>
</tr>
<tr>
<td>l</td>
<td>w</td>
<td>29</td>
<td>8</td>
</tr>
<tr>
<td>w</td>
<td>η</td>
<td>10</td>
<td>4</td>
</tr>
<tr>
<td>η</td>
<td>η</td>
<td>91</td>
<td>90</td>
</tr>
<tr>
<td>η</td>
<td>η</td>
<td>91</td>
<td>90</td>
</tr>
<tr>
<td>η</td>
<td>η</td>
<td>91</td>
<td>90</td>
</tr>
<tr>
<td>f</td>
<td>f</td>
<td>99</td>
<td>97</td>
</tr>
<tr>
<td>θ</td>
<td>s</td>
<td>85</td>
<td>89</td>
</tr>
<tr>
<td>θ</td>
<td>f</td>
<td>7</td>
<td>2</td>
</tr>
<tr>
<td>θ</td>
<td>f</td>
<td>7</td>
<td>2</td>
</tr>
<tr>
<td>s</td>
<td>s</td>
<td>53</td>
<td>43</td>
</tr>
<tr>
<td>f</td>
<td>j</td>
<td>41</td>
<td>53</td>
</tr>
<tr>
<td>f</td>
<td>j</td>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td>x</td>
<td>x</td>
<td>99</td>
<td>99</td>
</tr>
<tr>
<td>m</td>
<td>m</td>
<td>99</td>
<td>98</td>
</tr>
<tr>
<td>n</td>
<td>n</td>
<td>100</td>
<td>99</td>
</tr>
<tr>
<td>n</td>
<td>n</td>
<td>87</td>
<td>95</td>
</tr>
<tr>
<td>η</td>
<td>η</td>
<td>7</td>
<td>4</td>
</tr>
<tr>
<td>η</td>
<td>η</td>
<td>7</td>
<td>4</td>
</tr>
<tr>
<td>l</td>
<td>l</td>
<td>99</td>
<td>97</td>
</tr>
<tr>
<td>r</td>
<td>t</td>
<td>36</td>
<td>51</td>
</tr>
<tr>
<td>r</td>
<td>l</td>
<td>59</td>
<td>38</td>
</tr>
<tr>
<td>r</td>
<td>l</td>
<td>84</td>
<td>50</td>
</tr>
<tr>
<td>j</td>
<td>j</td>
<td>91</td>
<td>92</td>
</tr>
<tr>
<td>j</td>
<td>η</td>
<td>8</td>
<td>7</td>
</tr>
</tbody>
</table>

4.1.1. Initial assimilation patterns

In the pre-test, the Spanish consonants /p, t, k, f, x, m, n, l/ were predominantly categorised to single Chinese categories with similar frequencies and goodness ratings as Chinese control sounds, indicating that Chinese listeners perceived these sounds as being virtually the same as their native counterparts. Spanish sounds /ʧ, ɕ, j/ were also largely categorised to single Chinese categories but slightly less frequently and with a lower goodness rating, suggesting that these Spanish sounds are less similar to their Chinese counterparts. Spanish /ɲ/ and /r/, although in the main also categorised to a single Chinese category, had much lower goodness ratings. The remaining Spanish sounds /b, d, g, s, ɾ/ showed more dispersed assimilation patterns, frequently being placed into two or more Chinese categories or receiving lower goodness ratings.

4.1.2. Changes in assimilation following training

As noted above, the assimilation patterns of those Spanish consonants with very similar Chinese counterparts showed no change as a result of training and, particularly for the three voiceless plosives, a similar pattern to that obtained for the L1 control consonants is observed. However, changes in assimilation patterns for the Spanish voiced plosives were striking and statistically significant.

Further assimilation pattern changes are present for Spanish consonants that had a weaker L1 correspondence. The number of assimilation targets with significant response proportions for the sounds /d, g, j, s, ɬ/ was reduced (e.g., /ɡ/ changes

Table 2

FL to L1 percentage assimilations (standard errors in parentheses) and goodness of fit ratings in the pre-, mid- and post-tests. Each Spanish consonant listed in the FL stimulus column maps to one or more Chinese consonant response categories. Only those L1 response categories selected more than 5% of the time are listed. Goodness ratings range from 'bad exemplar'(0) to 'good exemplar'(100). Asterisk indicates significant Training effect (p < 0.05).
from 5 to 2 assimilation target sounds), although competing alternatives remain strong and the goodness ratings still show that none of the categories are considered more than average exemplars of a native sound (goodness ratings between 50 and 60%). A related pattern of assimilation change was seen for /ʃ/, which significantly strengthened its pull on the FL category, although the goodness ratings show no loss of sensitivity to differences between the FL and the L1.

The reverse pattern is also visible, where the near absolute dominance of a primary target is reduced following training with an increase of dispersion or reduction in dominance, as seen for /tʃ/ and /ɾ/). Training also led to a switch in the primary assimilation target for some sounds (e.g., /b, r, s/). In the case of /s/ this change is in the less obvious direction where an initially good match becomes a secondary target; again, the goodness ratings do not reflect the reduction in favour of the initial category. Overall, a change in assimilation pattern is not accompanied by a change in goodness ratings in the same direction, perhaps indicating that assimilation pattern changes are not affecting the core or prototypical values of most L1 categories.

4.1.3. **Interim discussion**

The assimilation test (FL sounds’ perceptual assimilation to L1 categories) revealed a dispersion of classification for some FL categories, indicative of a mismatch of phonetic features used in the two languages to define their phonetic categories as well as listeners’ sensitivity to FL category internal variation (Iverson & Evans, 2009; Levy, 2009). For example, Spanish /s/ was classified as either Chinese /s/ or /ʃ/, two categories differentiated in Chinese as laminal vs. apical and denti-alveolar vs. post-alveolar/retroflex (Lee & Zee, 2003). In the region where speakers were recorded, Spanish /s/ has apico- or lamino-alveolar realisations, which are not contrastive in Spanish but which would make the sound lean more towards one or other Chinese category depending on the particular token.

The greatest assimilatory dispersion was found for the Spanish voiced plosives. In the VCV environment in which they were recorded, lenition often leads to realisation as approximants and less frequently fricatives (Hualde, 2005). Native Spanish speakers ignore these intracategory variants, but Chinese listeners may have been perceptually aware of the variation and tried to classify these realisations into separate categories. Sensitivity to allophonic variation as reflected in L1 assimilation diversity has been found in previous studies for FL vowels (Iverson & Evans, 2009; Levy, 2009).

With regards to our second research question, we did find that non-native to native assimilation patterns changed as a result of intensive training. For some sounds, training increased the level of uncertainty either by increasing assimilation to a less similar category (e.g., assimilation from Spanish /s/ to Chinese /ʃ/) or by reducing the dominance of the initially most frequently chosen category (e.g., Spanish /ɾ/ to Chinese /ɾ/). This behaviour is indicative of an increased sensitivity to differences within the FL categories and also of an increased sensitivity to differences between FL categories and those of the L1, a necessary first step in the acquisition process. According to Kuhl’s NLM model (Kuhl, 1993), L1 categories act as magnetic prototypes and the perceptual space around them is perceptually shrunk. FL sounds which fall within a prototype’s area of influence are not distinguished from it. In our results, the pre-test shows that some sounds are initially absorbed by the prototype but that exposure enhances awareness of the characteristics of certain FL sounds - including variants - which distinguish them from the L1. Consequently, the perceptual space around the prototype is altered, decreasing its pull on some FL sounds. The newly-perceived characteristics also lead FL sounds to fall into the perceptual space of alternative L1 prototypes, increasing the dispersion of assimilations.

Spanish consonants with very similar Chinese counterparts showed no change in assimilation pattern as a result of training, supporting the idea that the nearer the FL sound is to the L1 prototype, the more likely it is to be totally absorbed by it, and as a consequence any FL-L1 differences will be hard to perceive (Kuhl, 1993).

Goodness ratings did not change significantly, and there were no significant correlations between the assimilation changes and the category goodness changes from pre-to mid-test or from pre-to post-tests. Thus, training affected the way Chinese listeners classified some Spanish sounds but these classification changes were often not accompanied by goodness rating changes. This may indicate that training was capable of modifying L1 category boundaries in order to accommodate FL sounds, but did not affect the L1 nucleus or prototype for a category.

4.2. **Spanish consonant identification**

4.2.1. **Effect of exposure to test materials**

To evaluate any possible effect of pure exposure as opposed to training on Spanish consonant identification, we compared the control group’s pre- and mid-test scores (Fig. 1). The increase from 43.8% in the pre-test to 45.8% in the mid-test is not statistically-significant [t(7) = 1.50; p = 0.18], suggesting that exposure itself does not produce the very large gains observed when feedback is provided.

4.2.2. **Consonant identification rate**

Fig. 1 (upper) summarises mean correct consonant identifications during training and testing. There was a significant effect of test time [F(1.47, 26.4) = 233, p < 0.001, η² = 0.93 after correcting for violation of sphericity] with differing identification performance at all test points [p < 0.001].

Significant learning is evident with a rapid initial improvement from a pre-test value of 46% to around 58% in the first training session. This may be partly due to listeners learning the correct labels for sounds such as /p, t, k/ with a conflicting symbol/sound relationship between the two languages, although considerable improvements were also observed in the first
session for other sounds (e.g., /r, p, ɲ, q, ʧ/). Performance continued to improve at a slower rate throughout the experiment, with post-test performance significantly higher than in the mid-test.

4.2.3. Individual listeners

Mean correct consonant identification scores in the test and training sessions for each participant are shown in Fig. 2. The striking feature of this data is the inter-individual similarity of the overall progress of learning with time throughout the experiment. Apart from subject 15, all participants had a similar pre-test score. Consistent with the overall performance shown in Fig. 1, all individuals demonstrated rapid learning in the early training sessions followed by a more gradual improvement in later sessions.

4.2.4. Consonant identification rates

Fig. 3 compares the individual consonant identification rates in pre-, mid- and post-tests, while Fig. 4 depicts consonant confusion matrices. Consonants such as /f, x, m, n, l/, categorised to unique Chinese counterparts in the assimilation test, were identified at near ceiling level in the pre-test and therefore had little scope for improvement during training. Identification rates for /n/ and /l/ showed a slight deterioration. In contrast, the three voiceless plosives /p, t, k/, which were very well mapped on to supposedly similar Chinese counterparts in the assimilation pre-test, were very poorly identified initially but showed a dramatic increase in scores from around 1% in the pre-test to 77% in the mid-test. This improvement may be due to label learning. Equally interesting are those sounds which experienced little or no improvement through training or did not reach high correct identifications: /s/ was quite resistant to training with a non statistically-significant improvement from 72% to 79%, while scores for /b, d, ɡ/ actually deteriorated after training.

Although improvement may not be apparent in the number of correct classifications, it can be seen in Fig. 4 that the degree of dispersion is considerably reduced for some of these sounds, a sign of acquisition. In the case of the voiced plosives /b, d, ɡ/, their approximant realisations led listeners to a great amount of initial uncertainty. In PAM’s terms (Best, 1995), these sounds were probably poor exemplars of several native categories. Through training they appear to be forming predominantly plosive categories. However, the sounds still seem to be seen as exemplars of two different categories, the voiced and voiceless plosives, as is the case when Chinese listeners acquire plosives from other languages contrasting in voicing (see, e.g., Flege, 1989).

4.2.5. Response times

Response times (RTs) are used as a measure of speech processing difficulty (Cutler & Norris, 1979). Mean RTs in each test and training session are summarised separately for correct and incorrect responses in the lower panel of Fig. 1. Response time was measured from the offset of stimulus presentation to the point at which the listener selected their response. Mean identification scores across training and test sessions are moderately negatively correlated with RTs $r(359) = -0.33$, 

![Fig. 1. Upper: mean Spanish consonant identification rates in pre-, mid- and post-tests and training sessions. Lower: mean response times plotted separately for correct and incorrect responses. Error bars represent ±1 standard error.](image-url)
Fig. 2. Individual listeners' mean Spanish consonant identification performance in pre-, mid- and post-tests and training sessions.
for correct responses and weakly positively-correlated for incorrect responses $[r(359) = 0.12, p < 0.05]$ i.e., RTs decreased as the proportion of incorrect responses decreased. The mean RT for both correct and incorrect responses decreased through the training period, ending at around two-thirds of its starting value. RTs for incorrect items were 38% longer initially than those for correct items, with the differential increasing to 44% by the end of the training period. The mean time for correct responses also fell for the FL tests, though by a smaller amount. No overall decrease in RT was observed for those stimuli identified incorrectly in FL tests, and indeed RT increased in the mid-test for incorrect responses. RTs and identification scores per consonant computed across all test and training sessions are strongly negatively correlated $[r(340) = -0.81, p < 0.001]$ for stimuli identified correctly, but not significantly so for those identified incorrectly $[r(336) = -0.1, p = 0.062]$.

4.2.6. Interim discussion

Our first research question asked whether high-variability phonetic training is effective for naive learners when all FL categories are present. Our consonant identification data provide a clear answer: the 2.6-fold decrease in identification errors from pre-to post-test demonstrates a substantial training effect. Similar patterns across individuals suggests that training was effective for all participants, perhaps reflecting common learning strategies. Since high levels of individual variation are the norm in foreign language acquisition studies, even for homogeneous cohorts, we speculate that this outcome is due both to the use of a target language for which the entire listener group can be considered to be naive learners and to the tight control on target language exposure throughout the training process. Another factor might be homogeneity of linguistic background: all were from the same L1 dialectal region and all had a similar foreign language (English) learning history.

In the pre-test, predictably, Chinese listeners classified almost all Spanish voiceless plosives as their voiced counterparts (Fig. 4). Phonetically, Spanish voiceless plosives are most similar to Chinese unaspirated plosives, but there could also have been an orthographic component to the classification pattern. In Chinese, the letters ‘b’, ‘d’ and ‘g’ are the graphemes used in the Pinyin system to represent the Chinese voiceless unaspirated counterparts of Spanish /p, t, k/.

It is notable that for /n/ and to a lesser extent /l/, two phonetic categories which were near to ceiling in the pre-test, training had an adverse effect with scores actually falling in the post-test. This attrition-like pattern is a stage in the typical U-shaped behaviour observed in many areas of FL and L1 acquisition as a consequence of over-generalisation, as categories in the process of being established (here /ɲ/ and /ɾ/ respectively) attract some of the stimuli for those categories which had initially higher identification scores. A final case of apparent deterioration was /b/ whose initial moderate identification rate decreased after training. However, the initial dispersion of chosen categories virtually disappeared, to be focused instead on two competing targets. We interpret this as a step in reattunement of their L1 categories to accommodate the FL ones and FL category formation (“rephonologizing”; Bundgaard-Nielsen et al., 2011), inasmuch as listeners are beginning to disregard some of the variation and be more consistent in their choices (Bundgaard-Nielsen et al., 2011; Iverson & Evans, 2009; Levy, 2009).

For some sounds, identification rates reached only moderate levels after training. As mentioned above, /d/ and /ɡ/ again provide evidence of category formation or reattunement, since initial perceptual dispersion became more focused after training. A particularly interesting case is /ɾ/, which had a majority (but incorrect) identification as /l/ in the pre-test, probably...
due to the predominance of L1 influence in the initial stages, as suggested by Major (2001). Training made listeners aware of the differences between these two sounds so that /l/ was abandoned as a predominant target. However, the sound that is now perceived as different is still in the process of being acquired and hence displays a variety of responses reflecting influences other than the L1 (Major, 2001). Alternatively, the dispersion may indicate that it is perceived as a poor exemplar of several categories (Best, 1995), notably the two alveolar plosives as well as the tap and the lateral. Presumably, further training would result in more focused responses such as those found for /d/ and /ɡ/ in the post-test, and, ultimately, a majority of correct identifications.

Interestingly, unlike identification scores, which improved more between the pre- and mid-test, the decrease in response times was larger after the mid-test. While some of the initial decrease may be due to task/symbol familiarisation, the substantial RT decreases seen in the latter part of the training regime presumably reflect increasing confidence in category identification, even after the point at which identification scores show only marginal improvement. We speculate that RT decreases after the mid-test indicate gradual formation of more robust categories. Training regimes which terminate on reaching some criterion level of category identification performance may be less effective than those which also consider asymptotic behaviour of response time as a stopping criterion.

Fig. 4. Consonant responses in the pre- (top), mid- (middle) and post-test (lower). The area of each square is proportional to the identification rate.
5. General discussion

The current study investigated the effect of phonetic training on Chinese listeners’ identification of a full set of Spanish consonants in a context in which prior and concurrent FL exposure was tightly-controlled. Overall, FL consonant identification improved substantially throughout the training period. Unlike much previous work which has focused on one or two contrasts, here learners were faced with more response alternatives and were less able to focus their acquisitional effort. Our results demonstrate that high-variability phonetic training is effective in ab initio learning of the full consonant inventory (research question 1). As Bundgaard-Nielsen et al. (2011) suggest, the ‘whole system approach’ permits analysis of the full range of confusions which may not be apparent in more restricted forced choice tasks. Additionally, this methodology allows us to observe the evolution of confusion patterns.

Some of the FL phoneme identifications suggest the following developmental pattern. Initially, the wrong sound altogether is typically chosen (e.g., in the current study /t, n, p, t, k/), presumably due to L1 sound-label associations. In this instance exposure may produce a perceptual awareness of a mismatch, but uncertainty ensues if the just-noticed phonetic cues do not point to a single other category (as for /n, p, t, k/) and therefore responses become dispersed, as for /f/ here. Alternatively, the sound may be perceived from the outset to be sufficiently different from any of the L1 sounds – ‘uncategorised sound’ or ‘deviant exemplar’ in Best’s framework (Best, 1995). This produces uncertainty in its classification and thus identifications show dispersion amongst several categories (e.g., /θ, η, b, d, g, r/). Such dispersion is not necessarily a reflection of the dispersion seen in the FL to L1 assimilation patterns since /θ, η, r, f/ show little FL-L1 dispersion. Gradually, training leads to a strengthening of the correct category and reduction in the number and weight of the other categories (/θ, d, g, r/) leading to a focus on a reduced number of dominant categories (/b, g/) and predominance of the correct category choice (/n, t/). Finally, the learning process homes in on the correct category with very few incorrect choices (e.g., /r/ and /f/). While arguing for the logical ordering of acquisitional stages outlined above, note that sounds might start the process at any of the stages or skip (e.g., there is evidence that /n, p, t, k/ skip the dispersion phase). In the case of the voiceless plosives, confusions could to some extent be due to orthographic influences but for /n/ there is clearly only one competing category which gets gradually abandoned.

The FL to L1 assimilation patterns revealed changes after training, with generally less dispersion and more concentrated assimilations in the post-test and in some cases (e.g., /bl/) changes in the primary L1 targets (research question 2). These assimilation pattern changes are consistent with previous assimilation studies which focused on comparing learners with different levels of experience (Bundgaard-Nielsen et al., 2011; Levy, 2009), or before and after phonetic training (Iverson & Evans, 2009). Bundgaard-Nielsen et al. (2011) claimed that the assimilation concentration is evidence in support of what in the PAM-L2 model (Best & Tyler, 2007) is termed ‘re-phonologization’— establishment of L2 phonology by expanding or modifying the L1 phonological system— among the more experienced listeners, driven by the linguistic pressure of an increasing vocabulary size. However, our results suggest that linguistic pressure may not be a necessary condition for re-phonologization, since high-variability training using nonsense tokens revealed indications of an incipient FL phonological system.

Listeners started to expand their phonetic categories to accommodate FL sounds very early on in the training. At the same time, listeners’ L1 goodness ratings for FL sounds did not change systematically following the assimilation pattern changes, suggesting that L1 prototype nuclei were not greatly affected by training.

Some of the early fast improvement can be ascribed to new label learning for similar sounds, especially for the Spanish voiceless plosives, but rapid improvements in identification scores for other consonants are unlikely to be due solely to labelling and most sounds continued to improve during the experiment, including the most difficult sounds for Chinese listeners (e.g., /d, t/). The process of learning labels for a new language’s categories is an important part of the category formation process. Nevertheless, our L1 assimilation results suggest that Chinese listeners were also fine-tuning their newly-established non-native sound recognition mechanisms, indicating a more profound learning process than simply mapping sounds to labels. Fast perceptual learning was also reported in previous L1 studies (Clarke & Garrett, 2004) in which native listeners adapted to foreign accented speech after training with only a few tokens. Although L1 perceptual learning and adaptation benefits from a listener’s rich lexical knowledge of their native language (Norris, McQueen, & Cutler, 2003), the current study suggests that rapid learning also takes place for adults in the initial stages of FL phonological acquisition. Even without previous experience of the target language, learners were able to quickly implement what they learned from limited training material and possibly also from their existing L1 phonetic knowledge to improve their performance in non-native category identification.

The current study demonstrates a clear pattern of rapid perceptual learning in the earlier training sessions, consistent with previous L2 consonant training studies in that identification improvements in early training sessions were more significant than in later sessions (Lively et al., 1993; Pruitt et al., 2006). The current study extends these findings to full inventory consonant training, which simulates a more naturalistic situation in language sound learning, and is likely to be more challenging than training on a preselected subset due to the possibility of more scattered confusions. Similar rapid perceptual learning of larger sets of sounds has also been reported in L2 vowel training studies (Iverson & Evans, 2009; Nishi & Kewley-Port, 2007). In the current study the improvement in identification rates became more gradual by training sessions 4—5. Although there are many differences between individual studies, it is interesting to note that in Pruitt et al. (2006), a study with 12 training sessions, the reduction in rate of identification improvement occurs at around training session 3 to 4, while in Nishi and Kewley-Port (2007) this occurred at day 2 to day 5 during the 9-day training for most participants.
Major (2001) argues that L1 influence is the strongest factor in sound acquisition at the outset of L2 learning, and the results for our listener group with homogeneous L1 background and L2 learning history support this idea. Learners at later stages of acquisition are characterised by the heterogeneity of their language learning histories, so that even in controlled training studies, previous L2 experience is a factor in individual outcomes. The fact that training using the full set of consonant contrasts was highly-effective for all learners also highlights the potential pedagogical value of intensive phonetic training at the outset of FL learning and suggests it is a useful paradigm to employ in language teaching simultaneously with, or even preceding, instruction on other linguistic skills.

6. Conclusions

Chinese learners with no previous or concurrent exposure to Spanish showed very significant improvements in full inventory consonant identification as a result of an intensive high-variability training programme. Rapid learning was evident, with substantial decreases in error rate and reductions in response time. Learners showed a homogenous pattern of improvement, highlighting the essential role of FL exposure isolated from individual learner variables. Changes in assimilation patterns indicated an expansion in native categories to incorporate FL sounds, and relatively static goodness ratings suggest that L1 category prototypes were little altered by training.

Acknowledgements

The authors acknowledge support from the Jangsul Social Science Fund grant No. 14YCYC001, the Humanities and Social Science Fund of the Ministry of Education grant No. 16YJC740020, the Spanish MICIN grant FFI 2012-31597 for the DiaceX Project and from the Basque Government Consolidados Laslab grant.

References

Bates, D. M., Maechler, M., & Bolker, B. (2012). lme4: Linear mixed-effects models using S4 classes. R package version 0.999999-0.


