Auditory training of English vowels for first-language speakers of Spanish and German

Paul Iverson and Bronwen G. Evans

Department of Phonetics and Linguistics, University College London, London, UK
p.iverson@ucl.ac.uk, b.evans@ucl.ac.uk

ABSTRACT

This study compared how first-language Spanish and German speakers learn English vowels via computer-based auditory training. Spanish has fewer vowels than German, and thus Spanish speakers may have more unused room in their vowel space for new category learning. However, our results demonstrated that Germans improved twice as much (20 percentage points) as Spanish speakers (10 percentage points) following 5-sessions of training on English vowels (high-variability identification with feedback). The results suggest that the large first-language vowel inventory of German speakers facilitates rather than interfere with new learning.

Keywords: second-language learning, phonetic training, vowel perception

1. INTRODUCTION

At the level of individual phonemes, it has been well established that first-language (L1) categories affect the ease of learning second-language (L2) phonemes as an adult. For example, novice L2 learners are thought to assimilate new phonemes into their existing L1 categories [1]. Pairs of L2 phonemes can thus be difficult to distinguish when they are both similar to the same L1 category, because they sound the same with regard the L1 phonological system (e.g., Spanish learners of English perceive both English /i/ and /ɪ/ as sounding like the Spanish /i/ [4]). L2 phonemes are thought to be easiest to learn when they are far from existing L1 categories [2]; such phonemes can be learned without changing any L1 categories, but the L1 and L2 categories often must be merged when the phonemes are closer [3].

There has been little work, however, on whether these interactions at the level of individual categories extend to learning entire L2 phonological systems. The task of learning English vowels, for example, may be fundamentally different for individuals whose L1 vowel inventory is small (e.g., Spanish, which has 5 vowels, no duration contrast, and no vowel formant movement) than for individuals whose L1 vowel inventory is larger and more complex (e.g., German, which has 15 monophthongs in 7 tense-lax pairs, and 3 diphthongs). When individuals begin learning English, Spanish speakers would be likely to have more difficulty than Germans, because the smaller number of vowels in Spanish makes it more likely that multiple English vowels will assimilate into the same L1 category. However, learning could actually be expected to be easier for Spanish speakers. That is, there may be more unused areas of the vowel space for Spanish listeners (e.g., more opportunity to learn new vowels without interfering with existing L1 categories), but German listeners already have a vowel space that is relatively dense (e.g., any new L2 category is more likely to be similar to an existing L1 category).

The present study examined the interaction of L1 and L2 vowel systems by comparing how native Spanish and German speakers respond to auditory training for English vowels. One difficulty with such a comparison is that among L2 English learners with comparable amounts of experience, Spanish speakers tend to be less accurate at L2 vowel recognition than Germans, presumably due to the assimilation patterns described above. Such accuracy differences can complicate the interpretation of training experiments, because it is open to interpretation whether, for example, an improvement from 50% to 60% correct is equivalent to an improvement from 85% to 95% correct. To solve this problem, we matched the language groups based on pre-training vowel recognition accuracy rather than based on experience. That is, we attempted to recruit Germans who had less experience with English than did the Spanish subjects, and then further screened the subjects based on their pre-test English vowel identification scores.

Subjects completed a 5-session course of high variability phonetic training (identification with feedback) with a different talker each day, and the vowels presented in multiple real-word minimal pairs. They were given a large battery of pre/post training tests, although English vowel recognition (different talkers and words than in the training set) is only reported here.

2. Method

2.1. Subjects

A total of 26 subjects were tested, 13 Spanish and 13 German. Spanish subjects were tested in London. They were 21-40 years old (median 27 years), began learning English when they were 6-34 years old (median 14 years), and had 1-72 months experience of living in English-speaking countries (median 18 months). German subjects were tested in Potsdam, Germany. They were 19-38 years old (median 25 years), began learning English when they were 9-15 years old (median 11.5 years), and had no experience of living in English-speaking countries. All subjects were screened so that they were matched across groups in terms of English vowel identification accuracy, and all subjects had no known hearing or learning impairments.

2.2. Stimuli and Apparatus

The pre/post test stimuli consisted of 14 bVt words, covering the whole of the English vowel space (i.e., bite /aɪ/, bait /eɪ/, bet /ɛ/, beat /i/, bit /ɪ/, bart /ɑ/, bert /ɜ/, bout /aʊ/, boat /əʊ/, bot /ɒ/, bought /ɔ/, boot /u/, bat /a/, but /ʌ/). The stimuli were produced by 2 native speakers of Standard Southern British English (SSBE), 1 male and 1 female.

The training corpus was recorded by 5 native SSBE speakers (2 male, 3 female); none of these were the same as the speakers of the test stimuli. The stimuli formed sets of minimal pairs that previous data had shown were confused by native Spanish and German listeners. The sets divided the vowel space into 4 groups: /i, ɪ, aɪ, eɪ/, /u, aʊ, ɜ/, /ɒ, əʊ, ɔ/ and /ɛ, ɑ, a, ʌ/. Ten sets of minimal pairs were constructed for each of these sets of vowels, giving a total of 140 target words.

Recordings were made in an anechoic chamber using a calibrated microphone. Stimuli were recorded with 44 100 16-bit samples per second and were later downsampled to 22 050 samples per second. Subjects were tested and trained using Pocket PCs and headphones.

2.3. Procedure

Before and after training, subjects were tested in their identification of the isolated bVt words using a closed-set identification task (all 14 vowels as response options). On each trial, subjects heard a stimulus word and clicked on one of 14 buttons that listed the stimulus word they thought they had heard. Subjects completed 56 trials (4 repetitions of each of the 14 test words).

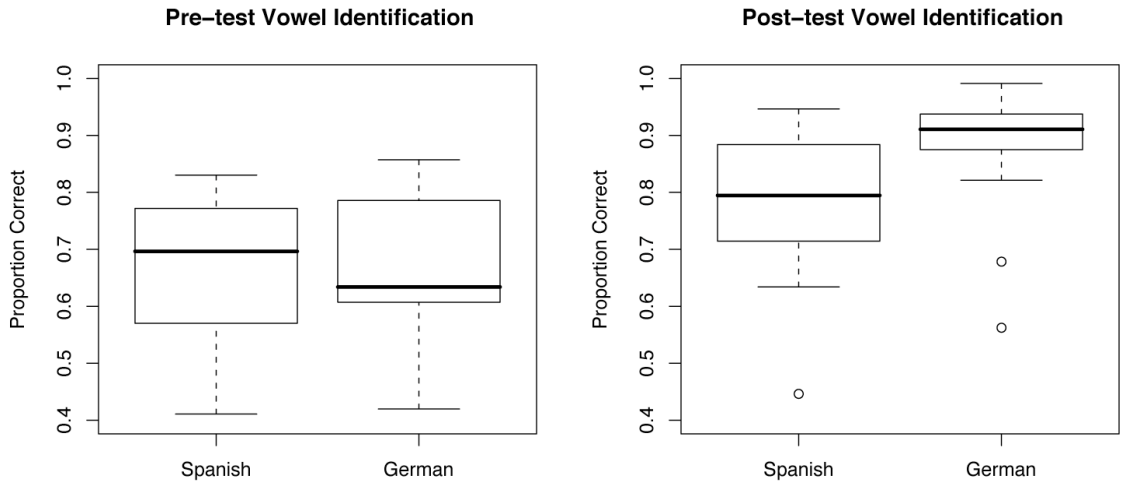
The training comprised 5 sessions, each lasting approximately 45 minutes. Subjects completed only 1 session each day, with a different talker each day, and completed the entire training over 5-10 days.

Each training session comprised 225 trials of identification with feedback. On each trial, subjects heard an English word (e.g., *peel*), saw the response alternatives printed on the screen (i.e., all the minimal pairs from the set, eg., *peel*, *pill*, *pail*, *pile*), and clicked on the word that they thought they heard. If they identified the word correctly, they saw "Correct!" on the screen, heard a cash register sound and then heard the stimulus again. If they identified the word incorrectly, they saw "Wrong" on the screen, heard two beeps, and then heard the correct stimulus, followed by two repetitions of the correct stimulus and the word they had chosen. The words were highlighted on the screen as they were played. At the end of each training session, subjects' overall score (percentage correct) was printed on the screen.

3. Results

Figure 1 displays boxplots of pre- and post-training English vowel identification accuracy for L1 Spanish and German speakers. Subjects were selected to be matched on this measure before training, and thus Spanish and German speakers had very similar ranges of pre-training scores. After training, Spanish speakers appeared to learn to a moderate degree (average 10 percentage point improvement), but German speakers improved twice as much (average 20 percentage point improvement). Moreover, the between-subject variability became reduced for German speakers after training, suggesting that they were approaching an upper limit in their performance.

Figure 1: Boxplots of English vowel identification accuracy, for Spanish and German speakers before and after training.



These differences in learning were confirmed using a two-way ANOVA with language as a between-subjects factor and pre/post as a within-subjects factor. There was a significant effect of pre/post, $F(1,24) = 105.7$, $p < 0.001$, demonstrating that the subjects as a whole improved with training. There was no significant main effect of language, $p > 0.05$, but there was a significant interaction of language and pre/post, $F(1,24) = 14.7$, $p < 0.001$, confirming that the groups improved differently following training.

Figure 2 displays hierarchical cluster analyses based on the vowel confusions that listeners made before and after training. The results demonstrate that, before training, Spanish speakers made the most confusions between the pairs *beat*/*bit*, *bot*/*bought*, and *bat*/*but*. After training, they continued to confuse *beat*/*bit* and *bat*/*but*, but improved on *bot*/*bought*. Germans most often confused, before training, *bat*/*bet* (and *but* to a lesser extent) as well as *bart*/*bought* (and *bout*, to a lesser extent). After training, Germans improved for all vowels. This fits the analysis of average recognition accuracy. That is, German learners had a pervasive pattern of learning, but Spanish speakers improved most for just a single contrast.

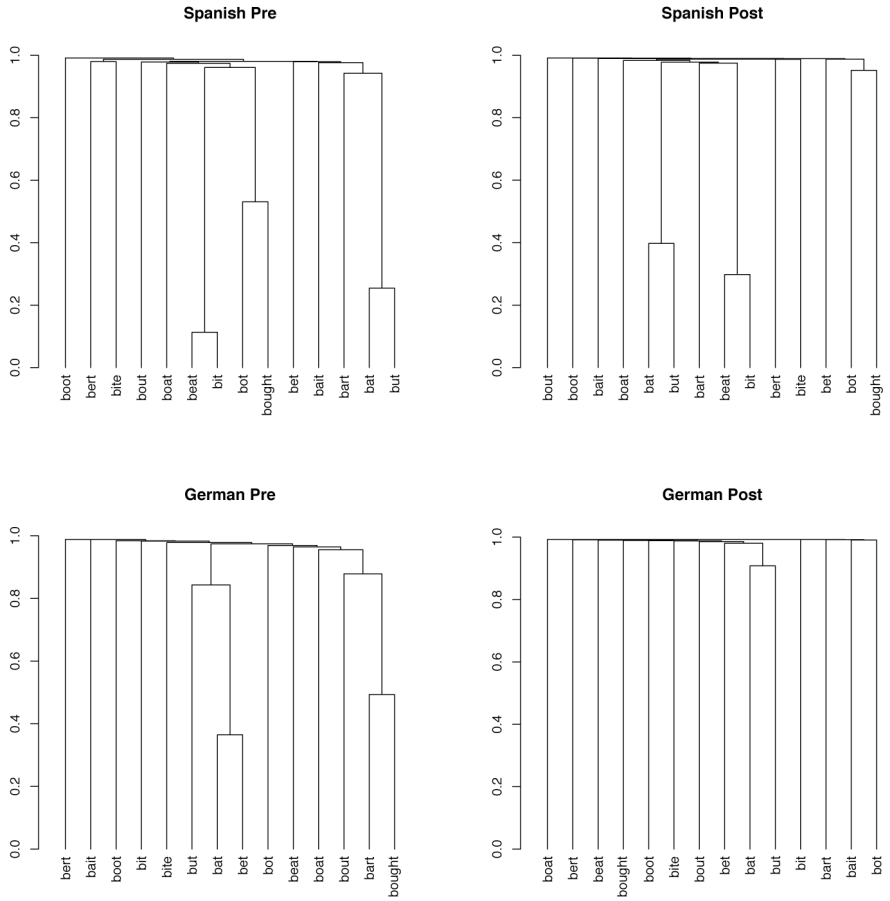
4. Discussion

The results demonstrate that L1 Germans speakers are able to learn English vowels faster than L1 Spanish speakers who had similar pre-test identification accuracy. Based on assimilation models [2, 3], we had predicted that the dense L1 vowel space of Germans listeners would have interfered with learning, because there would be less unused space for new categories. However, it appears that their dense L1 vowel space had the opposite effect. It is possible that a larger vowel inventory simply makes individuals more sensitive to gradient categorical differences between vowels. For example, when a Spanish speaker hears the English word *bat* they can only hear how close the vowel is to a single L1 vowel category (/a/), whereas Germans can compare it to several similar L1 vowels (/aː/, /a/, /ɛ/, /ɛː/). That is, Spanish and German speakers may have similar low-level sensitivity to variation in formant frequencies, but German speakers are able to judge L1 category similarity along more vectors. It is possible that this greater sensitivity to category differences facilitates new category learning.

It is possible too that differences in English experience could account for the differences in performance. Our Spanish speakers were tested when they were immersed in an English speaking environment, while our Germans were living in Germany and using German in most daily conversation. Although this meant that our Spanish speakers were likely more motivated to learn to improve their English communication, it is possible that they were already operating near a ceiling in performance. That is, they may have

already learned all of the relatively easy English categories (e.g., diphthongs), and could improve only by learning distinctions that are fundamentally harder for them (e.g., /i-/ɪ/). Germans, instead, were at a level in which they could still improve on English categories that were relatively easy for them (e.g., *but*).

Figure 2: Hierarchical cluster analysis of vowel errors before and after training, for Spanish and German speakers



Although this qualifies the differences between Spanish and German listeners, the fact alone that Germans are able to learn substantially with training is significant. L1 assimilation is thought to be a relatively automatic process that is used by novice listeners [1], and Germans have enough L1 categories so that almost all English vowels sound distinct with regard to German. However, relatively inexperienced learners of English are far from being perfect. It is possible that the L1 assimilation process itself requires adjustment and learning as an individual adapts to an L2 phonological system.

5. REFERENCES

[1] Best, C. T. 1995. A direct-realist view of cross-language perception. In W. Strange (ed.) Speech Perception and Linguistic Experience: Issues in Cross-Language Research. Baltimore, MD: New York Press, 171-204.

[2] 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. Baltimore, MD: New York Press, 233-277.

[3] 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: Differences and Similarities. Berlin: Mouton de Gruyter, 319-355.

[4] Flege, J. E., Bohn, O-S., & Jang, S. 1997. The effect of experience on nonnative subjects' production and perception of English vowels, Journal of Phonetics 25, 437-470.