# Learning English vowels with different first-language vowel systems II: Auditory training for native Spanish and German speakers[a)]

Paul Iverson and Bronwen G. Evans
*Division of Psychology and Language Sciences, University College London, Chandler House, 2 Wakefield Street, London WC1N 1PF, United Kingdom*

(Received 13 August 2008; revised 11 March 2009; accepted 12 May 2009)

This study investigated whether individuals with small and large native-language (L1) vowel inventories learn second-language (L2) vowel systems differently, in order to better understand how L1 categories interfere with new vowel learning. Listener groups whose L1 was Spanish (5 vowels) or German (18 vowels) were given five sessions of high-variability auditory training for English vowels, after having been matched to assess their pre-test English vowel identification accuracy. Listeners were tested before and after training in terms of their identification accuracy for English vowels, the assimilation of these vowels into their L1 vowel categories, and their best exemplars for English (i.e., perceptual vowel space map). The results demonstrated that Germans improved more than Spanish speakers, despite the Germans' more crowded L1 vowel space. A subsequent experiment demonstrated that Spanish listeners were able to improve as much as the German group after an additional ten sessions of training, and that both groups were able to retain this learning. The findings suggest that a larger vowel category inventory may facilitate new learning, and support a hypothesis that auditory training improves identification by making the application of existing categories to L2 phonemes more automatic and efficient.
© *2009 Acoustical Society of America.* [DOI: 10.1121/1.3148196]

PACS number(s): 43.71.Hw, 43.71.Es [RSN] Pages: 866–877

## I. INTRODUCTION

One could imagine that the task of learning a second-language (L2) vowel system would be fundamentally different for adults with small and large native-language (L1) vowel systems. For example, novice learners are thought to apply their existing L1 categories to perceive L2 phonemes (e.g., Best, 1995; Best *et al.*, 2001; Trubetzkoy, 1969). This can create ambiguity for individuals who have a small number of L1 vowels, because there are likely to be situations when multiple L2 vowels are assimilated into the same L1 category, making them sound the same (e.g., Spanish speakers hearing English /i/ and /ɪ/ as the same as Spanish /i/; Escudero and Boersma, 2004; Flege *et al.*, 1997; Iverson and Evans, 2007; Morrison, 2002). Moreover, individuals with smaller vowel inventories may use fewer dimensions to distinguish L1 vowels (e.g., only F1 and F2), and need to become sensitive to other aspects (e.g., quantity, diphthongalization, and nasalization) to distinguish vowels in the L2 (e.g., see Bohn, 1995; Bohn and Flege, 1990; Gottfried and Beddor, 1988; Iverson and Evans, 2007; McAllister *et al.*, 2002).

Individuals with large and complex L1 vowel systems may likewise have an early advantage in L2 vowel perception, but their large numbers of categories may make further learning difficult. For example, Flege's (1995, 2003) Speech Learning Model (SLM) claims that L1 and L2 vowels exist in the same phonological space, and learning a new vowel is harder when it is close to an existing category. Individuals with larger L1 vowel systems may therefore have a relatively crowded vowel space that interferes with the formation of new categories, whereas those with smaller L1 vowel spaces may have more room to learn, although it is not clear whether individuals with fewer categories actually have more "uncommitted" vowel space (e.g., Meunier *et al.*, 2003). Moreover, individuals with smaller vowel inventories likely have more incentive to learn given that they have more initial difficulties with L2 vowels.

The available evidence, however, suggests that individuals with large and small L1 vowel systems may learn L2 vowel systems similarly. The types of L1-L2 interactions described above have been well established for individual vowel contrasts such as /i/-/ɪ/ (e.g., Flege *et al.*, 1997; Flege *et al.*, 2003), but we recently found that these kinds of local interactions did not produce fundamentally different ways in which individual Spanish, French, German, and Norwegian listeners learn the English vowel system (Iverson and Evans, 2007). Our study took an individual difference approach, comparing English vowel perception and category representations among listeners with a wide range of L1s, but did not examine learning within individuals (e.g., training or longitudinal study). There were large overall differences in how accurately the language groups recognized English vowels, with lower scores for listeners with smaller L1 vowel systems (i.e., Spanish and French) and higher scores for those with larger L1 vowel systems (i.e., German and Norwegian). However, the acoustic cues that they used were the same; all

[a)] Part I: Iverson, P., Evans, B. G. (**2007**). "Learning English vowels with different first-language vowel systems: Perception of formant targets, formant movement, and duration," J. Acoust. Soc. Am. **122**, 2842–2854.

groups relied on primary acoustic cues such as F1/F2 target formant frequencies, as well as more fine-grained cues such as formant movement and duration, even though Spanish and French vowels do not contrast in formant movement and duration whereas German and Norwegian vowels do (see also Bohn, 1995).

Moreover, our results also suggested that individuals with small and large L1 vowel systems both had learned aspects of the English vowel inventory (Iverson and Evans, 2007). The subjects completed a vowel space mapping task in which they found best exemplars for vowels in their L1 and L2 (English), and all language groups had systematic differences between their L1 and L2 vowels. For example, the Spanish vowel space had five best exemplars with little formant movement or duration contrast, but the L1 Spanish speakers chose best exemplars for English that were markedly different, with formant movement and duration contrast, as well as a larger number of distinct categories. The Norwegian vowel space was larger (22 vowels), but there were still differences in the vowels that they chose for English; their English /ɪ/ vowel, for example, had a spectrum like that of native English speakers even though Norwegians assimilate English /ɪ/-/i/ into a Norwegian /i/-/iː/ contrast that is made purely with duration. All groups (Spanish, French, German, and Norwegian) exhibited similar amounts of learning, even though L1 assimilation judgments indicated that this learning was not completely necessary for Germans and Norwegians. That is, nearly all English vowels were assimilated into a unique L1 category in German and Norwegian, so these listeners could have, in theory, simply used their existing L1 vowels when listening to English. We thus found no evidence that the larger L1 vowel spaces interfered with new learning.

Such cross-language comparisons are difficult because one cannot completely match the learning experience of different subject groups. For example, even if one could find Spanish and German speakers with identical amounts and ages of English classroom instruction, they could differ, for example, in the type of instruction they received, their exposure to English outside of the classroom, and their individual motivations to learn. The approach of the present study was to control for experience by giving both groups the same amount of auditory training, to further compare the learning of English vowels by individuals with L1 vowel systems that are small and large; Spanish has five vowels with no diphthongs or duration contrast (e.g., Delattre, 1965; Flege, 1989; Stockwell and Bowen, 1965) whereas German has 18 vowels with diphthongs and duration contrast (e.g., Delattre, 1965; Strange *et al.*, 2005). The aim was to examine how their vowel recognition accuracy, L1 assimilation, and vowel space mapping differed before and after training, in order to evaluate whether the Spanish and German L1 vowel spaces made these individuals learn differently.

Several recent studies have adapted the high-variability phonetic training method (Logan *et al.*, 1991) to vowel stimuli, for the purpose of training Japanese adults on English vowels (Lambacher *et al.*, 2005; Nishi and Kewley-Port, 2007), English adults on the Japanese vowel-length contrast (Hirata *et al.*, 2007; Tajima *et al.*, 2008), and English adults on German non-low vowels (Kingston, 2003). Training Japanese adults on monophthongal English vowels has been successful, improving performance by 16–25 percentage points (Lambacher *et al.*, 2005; Nishi and Kewley-Port, 2007; see also Kingston, 2003), whereas training English adults on Japanese vowel-length contrasts has generally yielded smaller degrees of improvement that do not always generalize to untrained phonetic contexts and speaking rates (Hirata *et al.*, 2007; Tajima *et al.*, 2008). Most training protocols have trained listeners using closed-set responses (e.g., long vs short) and small numbers of vowels (e.g., 5). However, Nishi and Kewley-Port (2007) suggested that training on larger sets is more effective overall than concentrating on only the most difficult vowels; Japanese adults who were trained on a set of nine English vowels had broad improvement for all nine vowel categories, but those that were trained on only the three most difficult vowels improved only for these three vowels. The present study trained adults on an even larger set: 14 English vowels including diphthongs.

Previous work on English vowel training has embedded the vowels in CVC contexts (Lambacher *et al.*, 2005; Nishi and Kewley-Port, 2007). The present study instead trained listeners on real English minimal-pair words, to increase the range of phonetic variability and the naturalness of the training materials. Very few minimal pairs can be found that can span the set of 14 English vowels, so we divided this vowel space into four subsets based on cluster analyses of previous vowel confusion data by L2 speakers of English (Iverson and Evans, 2007). For example, listeners could hear the word *pet* and be given the response alternatives *pet, part, pat*, and *putt*. They were thus trained on a relatively large set of vowels, but were given response alternatives that were restricted to a subset of words that they would be expected to confuse.

Training improvements can be difficult to compare across different performance levels because we do not fully understand the underlying mechanisms. For example, there is no way of knowing whether a subject who improves from 20% to 40% recognition accuracy has actually learned the same amount as an individual who improves from 70% to 90%, because we do not know exactly what people are learning and how this translates to identification accuracy. In order to avoid this issue, we selected Spanish and German speakers so that they were matched in terms of their pre-test English vowel identification accuracy. Spanish speakers would normally be expected to be worse than Germans at English vowels, given their small L1 vowel system (Iverson and Evans, 2007). To help equalize this difference, we tested Spanish speakers in London (i.e., regular exposure to English) and tested German speakers in Germany who had little exposure to English outside of the classroom and media. The groups thus differed somewhat in English exposure, but were the same in terms of how well they recognized English vowels.

Both groups of subjects were given the same battery of pre and post tests. They were tested on /b/-V-/t/ words and talkers that were not part of the training set, in order to evaluate their degree of training improvement. Subjects were also tested in terms of L1 assimilation because it is thought that novice learners, at least, perceive vowels in terms of their native-language phonology (e.g., Best, 1995); we wished to evaluate whether these assimilation patterns could predict recognition accuracy as well as explain improvements in training. Subjects were additionally tested using a vowel space mapping procedure, in which they found best exemplars for English vowels in a large 5-dimensional vowel space that included F1 and F2 target frequencies, F1 and F2 formant movement, and duration (see Iverson and Evans, 2007; Iverson *et al.*, 2006). This evaluated how their underlying notions of what vowels sound good in English changed with training; our previous work demonstrated that individuals whose best exemplars are closer to those of L1 English speakers are also more accurate at recognizing natural recordings of English vowels. Experiment 1 evaluated performance immediately after training. Experiment 2 evaluated retention, as well as the effect of additional training.

## II. EXPERIMENT 1: AUDITORY-PHONETIC TRAINING

### A. Method

#### *1. Subjects*

A total of 33 subjects were initially tested (17 Spanish and 16 German). Pre-test English vowel identification accuracy (see Procedure, Sec. II A 4) ranged somewhat lower for Spanish speakers (30%–83%) than for Germans (42%–89%). In order to match the groups on this measure, the three highest-accuracy German subjects (scores of 87%–89%) and four of the five lowest-accuracy Spanish subjects (scores of 30%–50%) were dropped from the data analysis, creating two groups of 13 subjects each; one relatively low-accuracy Spanish subject (41%) was retained in the study to provide a match for the lowest-accuracy German subject (42%). In these matched groups, the ranges of identification accuracy scores were 41%–83% (mean 67%) for Spanish speakers and 42%–86% (mean 68%) for German speakers. This matching made the interpretation of group differences clearer, but it should be noted that the significant statistical differences reported here (see Results, Sec. II B) remained significant even when all of the original 33 subjects were included.

The Spanish subjects in the matched group were all tested in London and had 1–72 months of experience living in English-speaking countries (median 18 months). They were 21–40 years old (median 27 years), and began learning English when they were 6–34 years old (median 14 years). They came from several countries (Spain, Mexico, Columbia, Peru, Ecuador, Cuba, and Venezuela) but all had a standard Spanish five-vowel system.

The German subjects in the matched group were tested in Potsdam, Germany, and none had lived in English-speaking countries. They were 19–38 years old (median 25 years), and began learning English when they were 9–15 years old (median 12 years). The subjects were predominantly from the Brandenburg region of Germany, and none had non-standard German vowel systems.

The two groups were thus quite different in terms of the length of experience living in English-speaking countries. They were also slightly different in terms of the age of first instruction (median 12 years for German and 14 years for Spanish), even though their median duration of English use (age at test minus age of first instruction) were the same (13 years). In order to evaluate whether these differences could affect the extent that listeners benefited from training, the degree of training improvement (post- minus pre-test identification accuracy; see Results, Sec. II B) was compared using Pearson correlations to the experience and age of first instruction measures, separately for each language group. None of these correlations was significant, $p > 0.05$, suggesting that the age at which the subjects began learning English or the amount of time they spent living in English-speaking countries did not substantially affect the main experimental results.

In order to evaluate English language skills independent of their perceptual abilities, all subjects were given the written grammar portion of the *Oxford Placement Test I* (Allan, 1992). The two language groups did not differ significantly on this measure, $p > 0.05$. The average percentages of correct answers were 68% for Spanish speakers and 59% for Germans, indicating that the subjects predominantly had a lower-intermediate level of English competence (i.e., a functional, but not fluent, command of English).

#### *2. Apparatus*

The pre and post tests were conducted in quiet rooms, with stimuli played over headphones at a user-controlled comfortable level, and computers (PC and PDA) producing the stimuli and collecting responses. All training was conducted by subjects on their own; they borrowed PDAs, or the training software was installed on their own laptops. The training software created password-protected log files that the subjects could not access, so that we could verify that they completed all sessions.

All stimulus recordings were made in an anechoic chamber with 44 100 16-bit samples per second, and later downsampled to 11 025 samples per second.

#### *3. Stimuli*

*a. Training*

Recordings of English words were made from five speakers of British English, two male and three female. The words were groups of minimal pairs selected by dividing 14 British English vowels into four clusters: /ɛ/, /ɑ/, /a/, /ʌ/ (e.g., *pet, part, pat, putt*); /i/, /ɪ/, /aɪ/, /eɪ/ (e.g., *feel, fill, file, fail*); /u/, /əʊ/, /ɔ/ (e.g., *was, woes, wars*); and /u/, /aʊ/, /ɜ/ (e.g., *shoot, shout, shirt*). The clusters were selected by conducting a hierarchical cluster analysis on previous English vowel identification data by native Spanish and German speakers (Iverson and Evans, 2007); the first three clusters comprised vowels that were mutually confusable by many listeners, and the last cluster (i.e., /u/, /aʊ/, /ɜ/) was formed of remainders (i.e., vowels that were not strongly clustered with others). There were 10 sets of minimal pair words for each for the 4 clusters, for a total of 140 words. Each speaker recorded each word twice. All words were displayed to speakers one at a time in a random order during the recording session, to avoid list intonation.

#### b. Pre/post tests

Recordings of English /b/-V-/t/ words were made from two speakers of British English, one male and one female. Neither of these speakers and none of the words were used in the training corpus, such that all pre/post tests measured generalization to new stimuli. The speakers read the words *beat* /i/, *bit* /ɪ/, *bet* /ɛ/, *Burt* /ɜ/, *bat* /a/, *Bart* /ɑ/, *bot* /ɒ/, *but* /ʌ/, *bought* /ɔ/, *boot* /u/, *bait* /eɪ/, *bite* /aɪ/, *bout* /aʊ/ and *boat* /əʊ/; English vowels that would create non-words in the /b/-V-/t/ context (e.g., /ʊ/) were not included in the study. Four repetitions of these 14 English words were recorded for each talker, for a total of 112 stimuli.

A large set of synthesized stimuli from a previous study (Iverson and Evans, 2007) was used to map best exemplars. The stimuli were synthesized /b/-V-/t/ words embedded in a naturally spoken sentence frame (*Say_again*, spoken by a male speaker of British English), including the /b/ burst and the /t/ stop gap from the natural recording. The vowel stimuli were created using the cascade branch of a Klatt synthesizer (Klatt and Klatt, 1990). The synthesis parameters were chosen so that the synthesized vowel approximated the original vowel in the natural carrier sentence in terms of F0, amplitude contours, and spectrum. The stimuli primarily varied F1, F2, and duration, with some covaried variation in F3 (F3 was normally fixed to 2500 Hz, but was raised to be 200 Hz greater than F2 whenever F2 was greater than 2300 Hz). The F1 and F2 formant frequencies changed linearly from the beginning to the end of the vowel, and there were no additional consonantal formant transitions. F1 frequency was restricted so that it had a lower limit of 5 ERB (Glasberg and Moore, 1990) and an upper limit of 15 ERB. F2 frequency was restricted so that it had a lower limit of 10 ERB, was always at least 1 ERB higher than F1, and had an upper limit defined by the equation $F2=25-(F1-5)/2$. The stimuli were synthesized in advance with a 1 ERB spacing of the vowel space, and with 7 log-spaced levels of duration (54, 75, 104, 144, 200, 277, and 383 ms), for a total of 109 375 individual stimuli. The ERB and log-duration transforms allowed us to efficiently distribute the stimuli with regard to perception.

### 4. Procedure

#### a. Training

There were 5 sessions of high-variability phonetic training consisting of 225 trials of vowel identification with feedback, and an initial 14-trial practice session. The training sessions were run with no more than one session per day, and the entire course of training was completed over 1–2 weeks. The duration of each session was approximately 45 min. There was a different talker each session, and all subjects heard the talkers in the same order.

On each trial, subjects heard a stimulus word and clicked on three or four minimal-pair alternatives (depending on vowel; see stimulus description). For example, they could hear *slit* and be asked whether it sounded like *sleet, slit, slight*, or *slate*. The stimulus word was played before the response alternatives were shown, with the intent that the initial recognition of the word would be open set (e.g., not primed by the response alternatives), even though they gave a closed-set response. Every response word was accompanied by a more common word that had the same vowel (e.g., *seed, sit, night*, and *eight*) in case the response word was unfamiliar to the subjects. These example words were the same whenever that vowel appeared as a response, and the subjects were shown these example words during the initial instruction of the experiment.

If subjects gave a correct response, they saw “Yes!” on the computer screen accompanied by a cash register sound, then heard the word one more time. If subjects gave a wrong response, they saw “Wrong” on the computer screen accompanied by two tones with descending pitch, heard the correct word played, then heard a four-stimulus alternating series of the correct word, the incorrect response, the correct word, and the incorrect response. For example, if the stimulus word was *slit* and they clicked on the *sleet* response, they would hear an alternating series of *slit, sleet, slit*, and *sleet* so that they could learn the contrast between these two words.

Within each 225-trial training session, the first 70 trials were 5 repetitions of the 14 vowels in a random order, the next 85 trials were chosen adaptively based on the subject’s errors, and the last 70 trials were 5 repetitions of the 14 vowels in a random order. This design ensured that all subjects were trained on all vowels at the beginning and end, while allowing some of the training to be customized to fit the needs of each individual subject. The adaptive trials were selected randomly, with the selection probability of an individual vowel being weighted by combining the proportions of misses and false alarms for that vowel. That is, the probability of a vowel being selected increased when it was identified incorrectly, or when that vowel was chosen incorrectly as a response when another stimulus had been played.

The stimulus words on each trial were chosen randomly for each vowel. That is, if the trial was intended to have an /i/ stimulus, the computer program randomly chose one of the ten minimal-pair stimulus words that had this vowel. This random selection was blocked, such that each of the ten minimal-pair word sets was used once before the list was recycled.

#### b. Pre/post tests

##### i. Vowel identification

Subjects heard natural recordings of English /b/-V-/t/ words and gave a closed-set identification response (all 14 words as response options). To give their response, they mouse clicked on a button which listed the stimulus word (e.g., *bot*) as well as a common English word that had the same vowel (e.g., *hot*). Prior to starting the experiment, they heard the speaker read a short story (i.e., *The North Wind and The Sun*) in order to familiarize them with the talker. They were shown the word response alternatives and were able to ask questions if they were unsure which vowels were indicated. The experiment was run twice (once for each talker), with four repetitions of each of the 14 vowels in a random order.

##### ii. L1 assimilation

Subjects heard natural recordings of English /b/-V-/t/ words and identified which of their own L1 vowels sounded closest to the vowel in the word that they heard. They were told that even though these were English vowels, they should

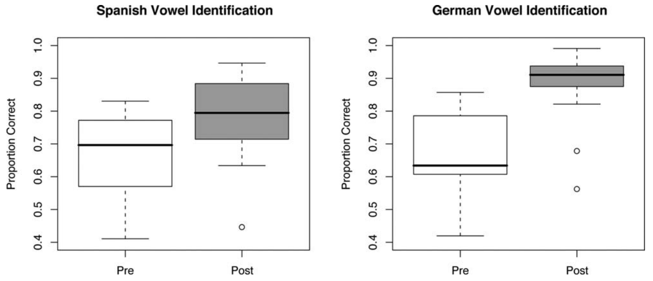

FIG. 1. Boxplots of the proportion correct vowel recognition in Experiment 1, pre and post training for Spanish and German speakers. Boxplots represent the quartile ranges of the scores, with outliers marked by circles.

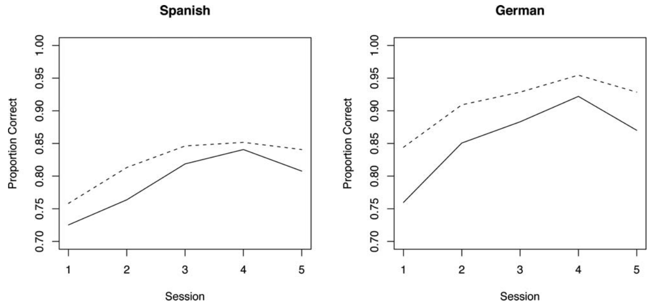

FIG. 2. Mean proportion correct at the start (solid line) and end (dashed line) of each training session in Experiment 1.

be classified as if they were listening to an L1 English speaker who was trying to speak their language. After each identification, they mouse clicked on a graphical continuum to rate whether this stimulus was *close* or *far away* from this L1 vowel category. The experiment was run twice (once for each talker), with two repetitions of each of the 14 vowels in a random order.

*iii. Vowel-space mapping*

On each trial, subjects saw an English /b/-V-/t/ word on the computer screen (e.g., *bot*), as well as a more common word that had the same vowel (e.g., *hot*), and heard a stimulus (synthesized /b/-V-/t/ embedded in a natural carrier sentence). They rated on a continuous scale how far away the /b/-V-/t/ that they heard was from being a good exemplar of the printed word. Their ratings were given by mouse clicking on a continuous bar presented on a computer screen.

A goodness optimization procedure (Evans and Iverson, 2004, 2007; Iverson and Evans, 2003, 2007; Iverson *et al.*, 2006) was used to iteratively change the stimuli that subjects heard on each trial, to search through the multidimensional stimulus space for good exemplars of each vowel. The full procedure will not be described here (see Iverson and Evans, 2007), but it involved simplifying the dimensionality of the search by finding best exemplars along straight-line paths that cut through the five-dimensional space, and efficiently choosing stimuli along each path so that they would be likely to be near to good exemplars (e.g., weighting the stimulus selection based on the subjects' previous responses). There were a total of seven search vectors and five trials per vector for each vowel. That is, subjects were able to find best exemplars after 35 trials, despite the large stimulus set (109 375 stimuli) and wide range of possible acoustic values available to subjects.

## B. Results and discussion

### 1. Vowel identification

Figure 1 displays the vowel recognition accuracy for Spanish and German speakers before and after training. The subjects had been matched to minimize the pre-test differences between Spanish and German listeners, which is reflected in the similar pre-test boxplots. The post-test scores demonstrated improvement with training for both groups. However, the Spanish speakers improved relatively modestly (average 10 percentage points), while the German speakers improved twice as much (average 20 percentage points) and began to reach ceiling performance. These differences were confirmed with a repeated-measures analysis of variance (ANOVA) on arc-sin transformed scores. Specifically, there was a significant within-subject effect of pre/post, $F(1,24)=105.7$, $p<0.001$, demonstrating an overall improvement with training, and a significant interaction of pre/post and language group, $F(1,24)=14.7$, $p<0.001$, demonstrating that the two groups learned to different degrees; there was no main effect of language group, $p>0.05$. The results thus suggest that the relatively crowded vowel space of German speakers actually may have made vowel learning easier, rather than providing interference.

Hierarchical cluster analysis of the pre-test data revealed that Spanish subjects most frequently confused /i/-/ɪ/, /a/-/ʌ/-/ɑ/, and /ɒ/-/ɔ/; Germans most frequently confused /ɛ/-/a/-/ʌ/, and /ɑ/-/əʊ/-/ɒ/-/ɔ/-/aʊ/. After training, the two language groups improved both for these difficult clusters and for words that were not as strongly clustered. That is, there was a general improvement in vowel identification rather than a pattern of improvement that was markedly stronger or weaker for individual pairs. The improvement of Germans for /ɑ/-/əʊ/-/ɒ/-/ɔ/-/aʊ/ is notable because these words crossed the minimal-pair clusters used in the training. That is, Germans decreased their confusions for pairs like /ɒ/-/ʊ/ even though their forced-choice responses during training did not directly contrast these vowels.

In order to examine the improvement during the course of training, the accuracy for the first and last 28 trials (i.e., two repetitions of each vowel) was calculated for each training session (see Fig. 2). At the beginning of the first training session, the averages for Spanish and German listeners were relatively similar (0.73 and 0.76, respectively), but the Germans improved more on successive sessions. Compared across sessions, the results suggest that subjects may have approached asymptotic performance toward the end of the training session (i.e., the curve begins to flatten). However, the speakers occurred in the same order for each subject, so it cannot be determined, for example, whether the dip in performance on session 5 occurred because the subjects lost concentration or because the speaker that subjects heard in that session was difficult. Also, subjects continued to improve within each session even for sessions 4 and 5, suggesting that they were still learning in these sessions. To evaluate these observations statistically, a repeated-measures ANOVA analysis was conducted on arc-sin transformed scores. There was a significant main effect of session, $F(4,44)=11.8$, $p$

TABLE I. Assimilation of English vowels into L1 vowels, before and after training. The table lists the closest L1 vowel, with the percentage of trials chosen and its average perceived similarity (0 different to 1 same) listed in parentheses.

| English vowel | Pre-test Spanish | Post-test Spanish | Pre-test German | Post-test German |
|---|---|---|---|---|
| i | i (100, 0.90) | i (100, 0.90) | i (72, 0.87) | i (95, 0.93) |
| ɪ | i (85, 0.88) | i (96, 0.88) | ɪ (84, 0.91) | ɪ (100, 0.94) |
| eɪ | e (100, 0.79) | e (100, 0.79) | ɛ: (40, 0.80) | ɛ: (50, 0.76) |
| aɪ | a (62, 0.81) | a (77, 0.79) | aɪ (88, 0.86) | aɪ (95, 0.92) |
| ɛ | e (96, 0.92) | e (96, 0.90) | ɛ (76, 0.88) | ɛ (82, 0.93) |
| a | a (92, 0.91) | a (100, 0.90) | a: (28, 0.84) | ɛ: (41, 0.77) |
| ʌ | a (81, 0.91) | a (88, 0.89) | a (68, 0.84) | a (82, 0.90) |
| ɑ | a (65, 0.87) | a (85, 0.83) | a: (68, 0.78) | a: (100, 0.88) |
| əʊ | o (100, 0.82) | o (100, 0.80) | o (72, 0.73) | o (77, 0.78) |
| ɒ | o (100, 0.91) | o (100, 0.88) | ɔ (96, 0.83) | ɔ (95, 0.91) |
| ɔ | o (100, 0.87) | o (100, 0.87) | o (88, 0.86) | o (95, 0.81) |
| aʊ | a (77, 0.77) | a (92, 0.78) | aʊ (96, 0.84) | aʊ (100, 0.92) |
| ɜ | e (46, 0.81) | e (62, 0.79) | ø (52, 0.70) | ø (55, 0.77) |
| u | u (100, 0.89) | u (96, 0.87) | u (52, 0.80) | u (68, 0.84) |

$<0.001$, indicating that subjects improved as training progressed, and an interaction between session and language, $F(4,44)=2.8$, $p=0.036$, indicating that the Spanish and German speakers improved at different rates. There was a significant main effect of beginning/end, $F(1,11)=29.4$, $p<0.001$, indicating that listeners were better for the trials at the end of each session than they were at the start of each session, but there were no significant interactions with language or session, $p>0.05$. There was also a main effect of language, $F(1,148)=23.5$, $p<0.001$, indicating that Germans were more accurate overall than were Spanish speakers.

### 2. L1 assimilation

Table I lists how the stimuli assimilated into L1 categories before and after training. The patterns of assimilations only roughly correspond to the confusions made in identification. For example, the confusion of /ɒ/-/ɔ/ by Spanish speakers during the pre test makes sense because both sounded like /o/ in Spanish, but the assimilation ratings also predict that /əʊ/ should have been as frequently confused with these phonemes. The German /a/-/əʊ/-/ɒ/-/ɔ/-/aʊ/ confusions are poorly predicted by assimilation, because each of these vowels assimilated into different German categories. Part of this discrepancy could have been caused by the fact that these assimilation patterns were not very consistent (e.g., /ɑ/ assimilated most strongly into German /a:/, but did so only on 68% of the trials before training), indicating that there was some variability in how these stimuli were perceived.

The post-test results indicate that the assimilations may have changed somewhat after training, such that the closest L1 category was chosen more consistently. It is thus possible that the ability of these subjects to identify the English phonemes correctly was related to how consistently listeners assimilated them into their L1 categories. However, there is no established method for translating assimilation ratings into predictions of identification accuracy. If listeners were literally perceiving these stimuli in terms of their L1 categories, a plausible decision model for the identification task could be (1) listeners identify the English vowel in terms of an L1 category and (2) give an English response based on what English vowel usually sounds like that L1 category (e.g., a maximum-likelihood decision). For example, if a German listener was played English /ʌ/ and perceived it as closest to German /a/, it would be logical for them to give an answer of English /ʌ/ because that vowel most often sounds like German /a/. However, if a German listener was played English /ʌ/ and perceived it as closest to German /a:/, it would be logical for them to give a response of /ɑ/ (i.e., make an error) because /ɑ/ most often sounds like German /a:/. Although this model may be simplistic, one advantage of this decision model is that it can be easily translated into predictions from confusion matrix data. That is, each assimilation response on every trial can be translated into a maximum-likelihood English response (e.g., replacing all German /a/ assimilations with English /ʌ/), and then a predicted proportion of correct responses can be calculated.

This decision model was applied to the assimilation data for individual subjects, and compared to their actual identification accuracy scores. For German subjects before training, the predictions were very close. The average predicted accuracy (69%) was close to the obtained average accuracy in the identification experiment (68%), and there was no significant difference between these scores in a paired t-test, $p>0.05$. Moreover, individual differences in the predicted and obtained scores were significantly correlated, $r=0.69$, $p=0.009$. That is, individuals who were more accurate at identifying these phonemes also were more consistent at assimilating them to the closest German category. It is thus plausible that Germans before training were simply giving identification responses based on their assimilation into L1 categories.

However, this correspondence for Germans was weaker after training. The average predicted accuracy (77.6%) increased because of the increasing consistency of assimilation ratings, but the obtained identification accuracy increased more (87.8%). The two measures were significantly different, $t(10)=-2.72$, $p=0.021$, and the individual differences were no longer significantly correlated, $r=0.40$, $p>0.05$. It thus appears that training may have changed the extent to which the German subjects relied on L1 assimilation.

This assimilation model does not work at all for the Spanish subjects. Given that Spanish has only 5 vowels, the maximum performance that they could achieve would be 35.7% correct (i.e., 5 of the 14 English vowels correct). The decision model predicted that nearly all Spanish subjects would achieve this maximum before and after training, but their actual performance was much higher, averaging 66.8% before and 77.1% after training. The only way that Spanish subjects could plausibly have obtained these levels of performance was if they had been using additional categories or cues that they had learned for English, not just by assimilating these vowels into their Spanish categories.

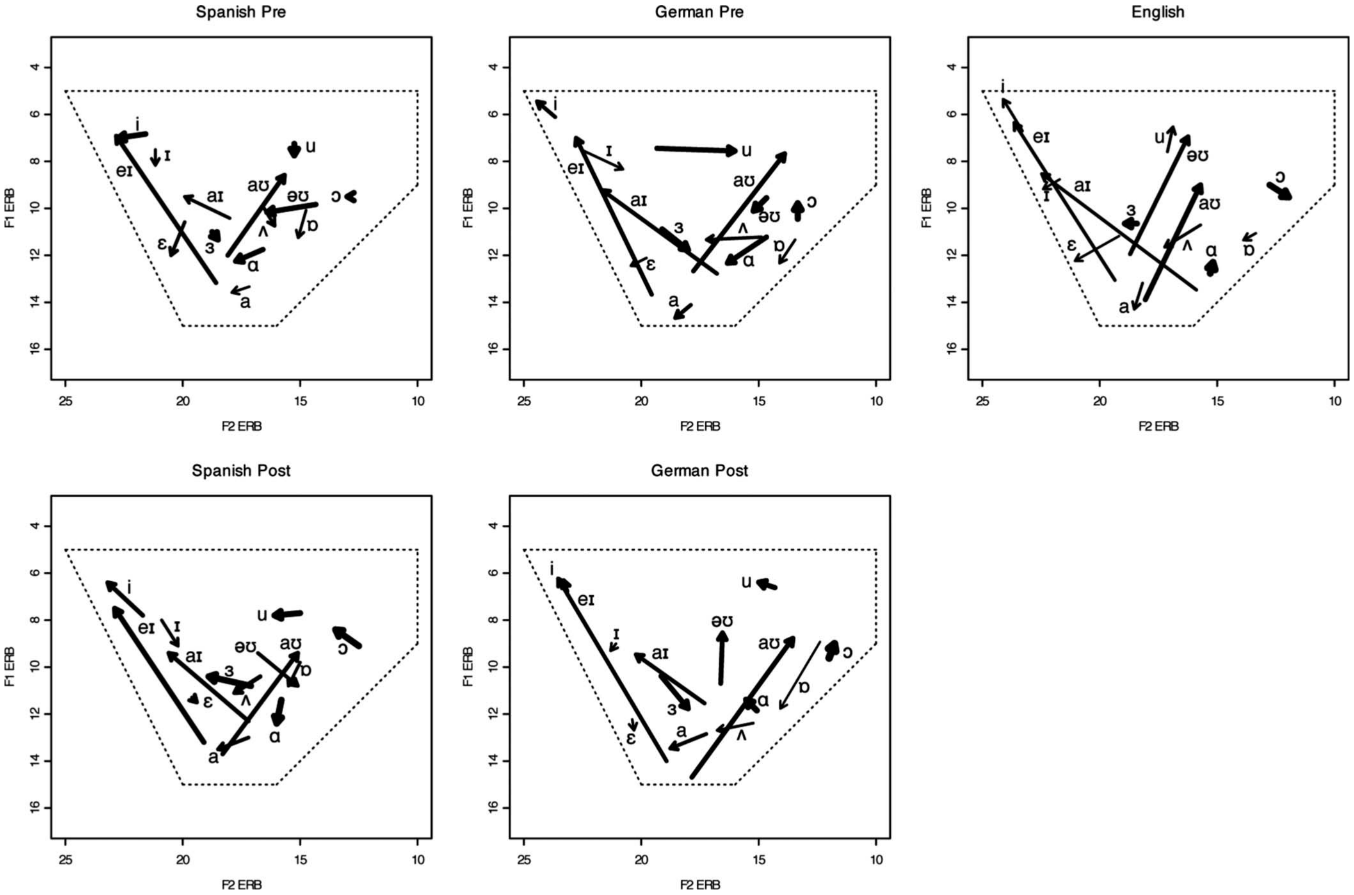

FIG. 3. Average best exemplar locations of English vowels in Experiment 1, for Spanish and German speakers pre and post training, and for L1 English speakers (from Iverson and Evans, 2007). Each vowel best exemplar is represented by a line from the starting to the ending first and second formant frequencies (i.e., indicating formant movement). Duration is indicated by the weight of the line, with thicker lines for longer vowels.

### 3. Vowel space mapping

Figure 3 displays the average best exemplars for Spanish and German speakers, as well as for English speakers from a previous study (Iverson and Evans, 2007). Although all language groups had generally similar vowel spaces (e.g., vowels in roughly the right locations, with appropriate degrees of formant movement), there were apparent differences. For example, Spanish and German speakers both had little contrast between /ʌ/ and /a/, and little format movement for /əʊ/, compared to English speakers.

In order to calculate how close the Spanish and German speakers were to the English averages before and after training, the distance was calculated between each individual's best exemplars and the average best exemplars for L1 English subjects (see Fig. 4). These distances were calculated separately for F1/F2 location, formant movement, and duration. The F1/F2 location accuracy was measured by averaging the beginning and ending frequency of each vowel for F1 and F2, giving a two-dimensional F1/F2 coordinate for that vowel with no formant movement. The Euclidean distance (i.e., root mean square) was then calculated between the F1/F2 locations of each individual's English best exemplars and the L1 English averages. Formant movement accuracy was measured by subtracting the F1/F2 location values above, so that each vowel had a vector representing the direction and magnitude of F1/F2 formant movement, with the center of each line passing through zero (i.e., normalizing the vowel's location in the vowel space). As above, Euclidean distances between these formant movement vectors were measured for each individual's vowels and the L1 English averages. Duration accuracy was quantified by calculating the average absolute-value difference between the durations of each individual's best exemplars and those of the L1 English averages.

Repeated-measures ANOVA analyses were conducted separately for each acoustic distance measure. For the measures of F1/F2 location and duration, there were no significant main effects or interactions of pre/post or language, $p > 0.05$; there was no evidence that listeners improved their vowel representations in these respects. There was a significant main effect of pre/post on formant movement, $F(1,16)=8.94$, $p=0.009$, although there was no main effect or interactions with language, $p > 0.05$. Although the subjects thus improved in the formant movement of their best exemplars after training, this improvement was small (averaging 0.4 ERB), and the individual differences in this improvement were not significantly correlated with changes in identification accuracy, $r=0.30$, $p > 0.05$.

Despite the fact that the accuracy of best exemplars did not improve much with training, identification accuracy was significantly correlated with individual differences in F1/F2 location accuracy, $r=-0.37$, $p=0.027$, and formant move-

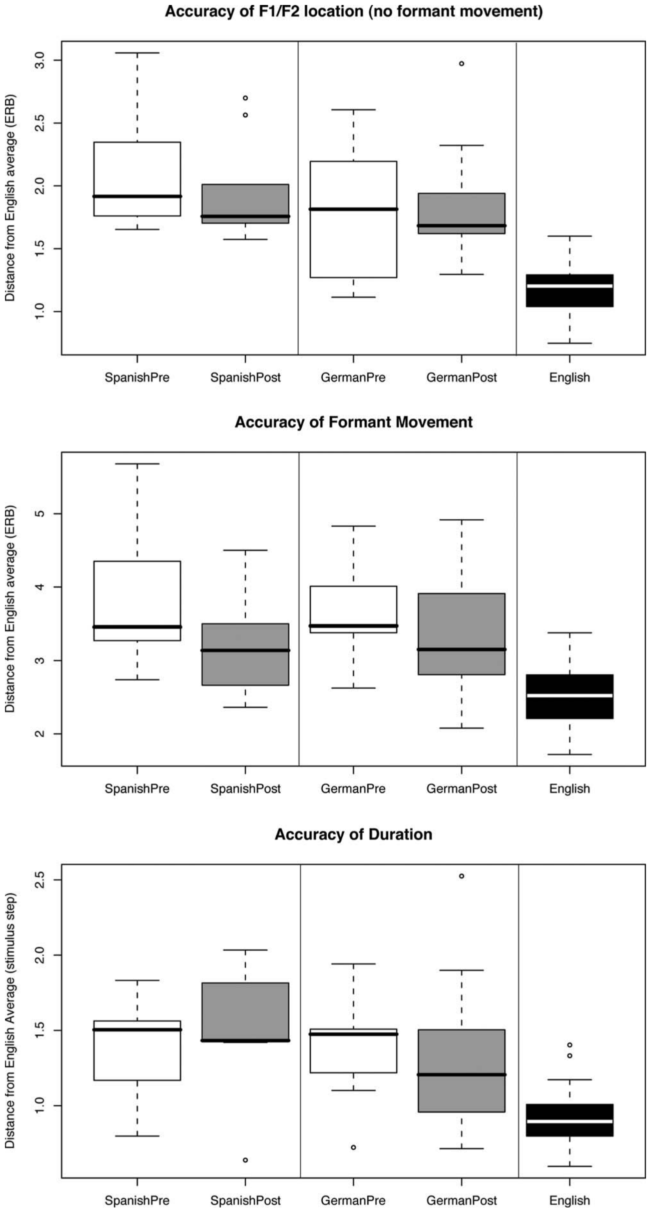

FIG. 4. Boxplots of the accuracy of each individual's best exemplar locations in Experiment 1 for Spanish and German speakers before and after training, and for L1 English speakers (from Iverson and Evans, 2007). The distances are listed separately for the accuracy of F1/F2 location, formant movement, and duration.

ment accuracy, $r=-0.41$, $p=0.014$, but not duration accuracy, $r=-0.05$, $p>0.05$. That is, individuals who were more accurate with their best exemplars in terms of F1/F2 location and formant movement tended to be better at identifying natural vowels, in accord with our previous study (Iverson and Evans, 2007).

## III. EXPERIMENT 2: ADDITIONAL TRAINING AND RETENTION

There were two main unexpected results of Experiment 1. First, Spanish speakers learned less than the Germans. Their training results in each session began to flatten toward the later sessions, indicating that they may have been reaching an upper limit of learning, although the subjects continued to improve within each session. It is thus not clear whether the smaller improvements for Spanish speakers occurred because they reached an L1-related upper limit on their recognition accuracy (at least with this training method), or whether they simply learned slower than the Germans and could benefit from additional training. Second, the improvements in vowel identification accuracy for both groups were not accompanied by corresponding changes in best exemplars. One interpretation of this result is that the improvements in training may have been superficial. For example, the training could have served as a short term refresher of what they had learned previously about English vowels (e.g., in school), rather than producing long-term changes in their categorization processes. This hypothesis could help explain why Spanish subjects learned less; they were already using English while living in London, so they could have benefited less from "refreshing" than did Germans, who had not been using English recently.

Experiment 2 was conducted to investigate these possibilities. The Spanish speakers in Experiment 1 were contacted an average of 4 months later and asked to complete an additional ten sessions of training. The German subjects were contacted 1 year later and were asked to complete only the identification task. The timing of these tests was based on circumstance, rather than being a planned aspect of the experimental design, and we tested whatever subjects in Experiment 1 were willing to participate (i.e., all subjects rather than only the matched subgroups). However, this design still met the goals of evaluating whether both groups of listeners were able to retain their improvement over long intervals, and whether Spanish listeners could further improve with additional training.

### A. Method

#### *1. Subjects*

The subjects were 9 Spanish and 11 German speakers. They were recruited from the group of subjects who participated in Experiment 1, including 2 Spanish subjects who had participated in a pilot version of the training protocol which was similar to the final version but had a more random selection of stimuli.

The Spanish subjects were tested 2–6 months (mean 4 months) after they completed Experiment 1. They completed all pre/post tests and a ten-session course of training, as described below. The German subjects were tested 1 year later, and completed only the vowel identification test.

#### *2. Procedure*

The pre/post tests were the same as in Experiment 1. The training protocol was expanded to ten sessions by recording stimuli from five additional speakers (three male and two female), but was the same in all other respects.

### B. Results and discussion

#### *1. Vowel identification*

The identification results (Fig. 5) suggest that both Spanish and German speakers retained their training improvements in Experiment 1. For Spanish speakers, the mean accuracy in the Experiment 1 post test (0.76) was similar to

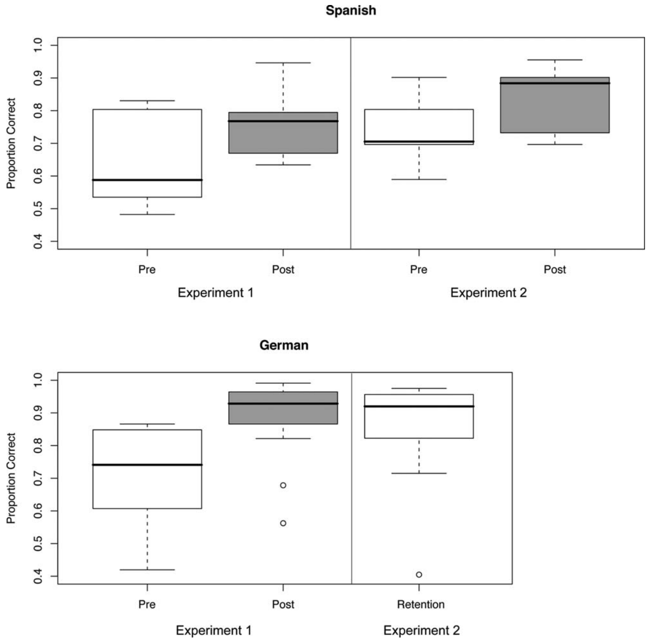

FIG. 5. Boxplots of the proportion correct vowel recognition in Experiments 1 and 2. The Experiment 1 data are displayed only for the subset of subjects who participated in both experiments.

that obtained 4–5 months later (0.73), and there was no significant difference between these scores, $t(8)=-0.92$, $p>0.05$. However, the conclusion that there was retention is tempered by the fact that there was also no significant difference between their accuracy in the Experiment 1 pre test (mean=0.64) and the Experiment 2 pre test, $t(8)=-2.04$, $p=0.076$. The retention results for German subjects are clearer; their accuracy 1 year after training (mean=0.85) was not significantly different from their Experiment 1 post-test accuracy (mean=0.88), $p>0.05$, and was significantly greater than their Experiment 1 pre-test accuracy (mean=0.71), $t(10)=6.50$, $p<0.001$.

Spanish speakers significantly improved from the additional training in Experiment 2, raising their accuracy from a mean of 0.73 at the pre test to 0.84 at the post test, $t(8)=-5.24$, $p<0.001$. This result demonstrates that they had not reached a ceiling in performance in Experiment 1, and were able to achieve similar levels of performance as German subjects had in Experiment 1 after they were given additional training. Spanish speakers may thus take longer to learn English vowels than do German speakers, but still have the same basic capacity to learn.

### *2. Vowel space mapping*

Despite the fact that Spanish speakers had more training, Fig. 6 demonstrates that the accuracy of their vowel space mapping did not improve. There were no significant differences between pre and post tests in terms of how close their best exemplars were to the average best exemplars for English, in terms of F1/F2 location, formant movement, or duration, $p>0.05$. The results were thus much the same as in Experiment 1.

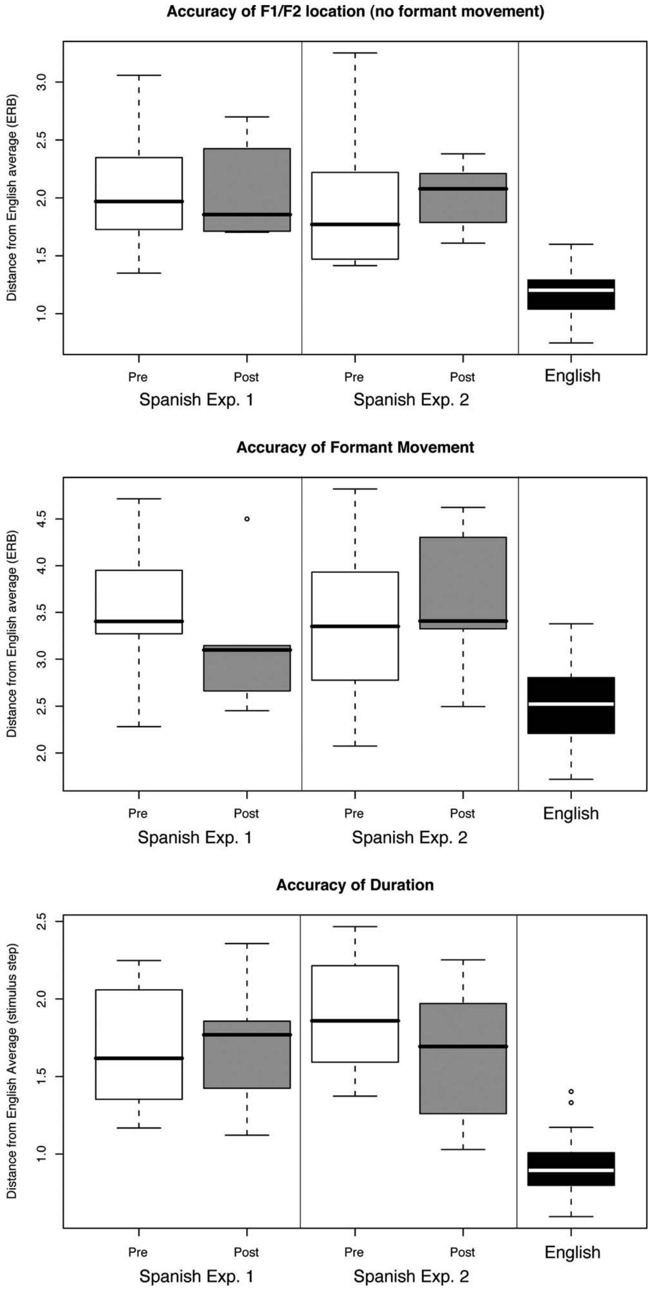

FIG. 6. Boxplots of the accuracy of each individual's best exemplar locations in Experiments 1 and 2, for Spanish before and after training, and for L1 English speakers (from Iverson and Evans, 2007). The distances are listed separately for the accuracy of F1/F2 location, formant movement, and duration.

## IV. GENERAL DISCUSSION

The results demonstrate that German and Spanish speakers learn at different rates given auditory training; German speakers improved twice as much, on average, as Spanish speakers after five sessions of training, although Spanish speakers attained similar levels of performance after completing an additional ten training sessions. One of our working hypotheses based on SLM (Flege, 1995) had been that learning would be more difficult for the Germans; their crowded L1 vowel space should have left less room for learning new vowel categories. However, the German speakers learned more easily than did the Spanish speakers, supporting our previous evidence of learning by individuals with large L1 systems (Iverson and Evans, 2007). Although this result may thus appear to be contrary to SLM, there was little evidence that German subjects actually formed new categories during the experiment (i.e., no improvement in the accuracy of best exemplars); without category learning there would be no expectation for there to be L1 category interference.

How did identification accuracy improve without corresponding changes in best exemplars? One could initially question the methods used in this study, raising doubts about whether subjective notions of best exemplars are related to the processes underlying vowel recognition. However, the validity of the best exemplar task is well supported by previous work; the accuracy of best exemplars has been significantly correlated with individual differences in the accuracy of English vowel recognition by L2 speakers (i.e., Experiment 1, and more strongly with the wider range of subjects tested in Iverson and Evans, 2007), vowel production and speech-in-noise recognition by English speakers with different accents (Evans and Iverson, 2004, 2007), vowel recognition by cochlear implant users (Iverson *et al.*, 2006), and with the identification of English /r/-/l/ by Japanese speakers (Hattori and Iverson, 2009).

The task used to map best exemplars has relatively low processing demands, in that listeners are able to replay the stimulus and only have to listen for one word at a time. In the present study, it assessed whether listeners had an accurate subjective notion of what vowels sound good in English. However, it is one thing to "know" what vowels sound good in English and another to put this knowledge in practice. For example, the recognition of vowels in natural speech involves more variability in the stimulus (e.g., related to talker differences or phonetic environment), more possible responses, and only a single time to hear the stimulus; listeners must rapidly encode the phonetic information that is relevant to the L2 categorization. Likewise, the perception and production of real-world speech (e.g., in a conversation) involves yet a higher processing load (i.e., less opportunity to focus only on the phonetic content of speech when listeners need to concentrate on meaning). Our hypothesis based on the present results is that auditory training improves the ability of subjects to apply their existing category knowledge to natural variable speech (i.e., both L1 and L2 category knowledge), without changing this knowledge itself. That is, auditory training makes the categorization process more efficient and automatic in a way that is long lasting, but does not generally change the representation of the categories (e.g., use of cues).

The original high-variability phonetic training papers hypothesized that training was successful because it changed attentional weights for phonetic dimensions, causing listeners to attend more to aspects of the stimuli that can distinguish categories over a range of talkers and phonetic environments and less to irrelevant variability (e.g., Lively *et al.*, 1993; Logan *et al.*, 1991). High variability training was thought to generalize better than training with single talkers or small stimulus sets, because, in part, the higher variability taught listeners which cues were most robust. However, there has been no evidence that this reweighting of cues occurs due to this kind of training. For example, Iverson *et al.* (2005) trained Japanese adults on English /r/-/l/ using stimuli that had been signal processed in different conditions to alter the cues that they heard (e.g., enhancing differences along F3 or manipulating the variability of secondary dimensions such as F2 and transition duration). Japanese adults learned in all conditions, but did not change the cues that they used to match the acoustic manipulations of the conditions. Instead, they appeared to become more consistent at labeling stimuli as English /l/ when the stimuli became acoustically similar to a Japanese flap (i.e., short closure, short transition; e.g., see Hattori and Iverson, 2009), even when this decision conflicted with the acoustic information that they had received in training. That is, listeners became better at applying their existing L1 knowledge to this task, without necessarily improving in the cues that they used for /r/ and /l/. Similarly, Heeren and Schouten (2008) found that Dutch speakers can be trained to identify the Finnish /t/-/t:/ contrast, but do this without obtaining the peak in discrimination sensitivity at the category boundary that is typical of native Finnish speakers for these stimuli, suggesting that auditory training may not produce pervasive changes in the way that individuals perceive these phonetic contrasts.

There is some evidence that laboratory based training can alter the use of acoustic cues, but this work has all involved restricted stimulus sets or entirely novel categories. For example, Francis and colleagues found in several studies that listeners change perceptual weightings for stimuli following identification training, but this has occurred for relatively homogenous sets of synthesized speech (i.e., no simulated talker differences; Francis *et al.*, 2000, 2007) and for a novel Korean stop-consonant contrast produced by a single talker (Francis and Nusbaum, 2002). Likewise, Holt and Lotto (2006) trained listeners to learn a novel category for non-speech sounds (frequency-modulated tones) and found that listeners developed cue weightings that were affected by the distribution of the training stimuli (i.e., relying less on dimensions that were more acoustically variable). It is thus clear that individuals can, in principle, learn the cues of new categories and re-weight the cues of existing categories when given laboratory-based auditory training. Although this work may still be useful for understanding auditory categorization, there is little evidence that changes in cue weightings are the primary learning mechanism in typical high-variability phonetic training studies, where individuals have previous experience with the categories and are trained on natural speech from multiple talkers.

High-variability phonetic training may thus be more effective than training with less variable stimulus sets, because the stimulus variability trains the process of applying categories to real speech. That is, listeners prior to auditory training typically have some idea about what the various categories sound like, and training makes them more efficient at applying this knowledge to situations in which they do not know exactly what they will hear from trial to trial, and when there is phonetic variability that is irrelevant to the categorization judgement. Whereas training with less variable stimulus sets (e.g., Lively *et al.*, 1993) simply does not train this kind of ability; listeners are able to learn to categorize small stimulus sets without improving in a more general ability that can be applied to other stimuli. If high-variability training improves the application of category knowledge to variable speech more than it alters cue weightings, it may not be necessary for the variability in the training set to be fully natural in order for performance to improve (e.g., Iverson *et al.*, 2005).

In the present study, German listeners may thus have learned faster than Spanish listeners because they had more pre-training category knowledge to apply to the task. For example, the Germans had more L1 vowel categories, and the analyses of assimilation ratings of Experiment 1 indicated that part of their improvement with English vowels may have occurred because they became more consistent at assimilating them into German categories. Both German and Spanish speakers likely had some English category knowledge prior to training, which would also be expected to become applied more automatically after training. One could suspect that the English language experience of Spanish speakers could have made them benefit less from training, because they may have already been using their category knowledge more efficiently due to their daily experience with using English. However, there was no correlation between their amount of time living in English-speaking countries and their improvement due to training in Experiment 1, and Experiment 2 demonstrated that they had a capacity to learn if given more time (i.e., they had not reached a ceiling for this type of learning). Moreover, we have recently examined the role of experience more directly (comparing the English vowel perception of L1 French speakers who were living in France or the UK), and found that individuals who used English every day improved from training as much as individuals who had little spoken English experience (Iverson and Preece-Pinet, 2008).

Although high-variability phonetic training may primarily improve recognition by making existing categorization processes more efficient, it is clear that long-term exposure to phonetic categories (e.g., when living in an English-speaking environment) creates changes to the cues that individuals use. For example, differences between subjects in their best exemplars of English vowels are predictive of vowel identification ability (Iverson and Evans, 2007), and L1 Spanish speakers can weight English vowel duration differently depending on how duration is used by the English accent that they primarily hear (Escudero and Boersma, 2004). Therefore, although high-variability phonetic training improves identification performance, it may not provide a full simulation for the kinds of changes in phonetic perception that occur during longer term L2 language learning.

## ACKNOWLEDGMENTS

The authors are grateful to Anke Sennema for subject recruitment and hosting the data collection at University of Potsdam (Germany). This research was funded by Grant No. RES-000-23-0838 from the UK Economic and Social Research Council.

Allan, D. (**1992**). *Oxford Placement Tests 1*, Oxford University Press, Oxford, UK.

Best, C. T. (**1995**). "A direct realist view of cross-language speech perception," in *Speech Perception and Language Experience: Issues in Cross-Language Research*, edited by W. Strange (York, Baltimore) pp. 171–204.

Best, C. T., McRoberts, G. W., and Goodell, E. (**2001**). "American listeners' perception of nonnative consonant contrasts varying in perceptual assimilation to English phonology," J. Acoust. Soc. Am. **1097**, 775–794.

Bohn, O.-S. (**1995**). "Cross-language speech perception in adults: First language transfer doesn't tell it all," *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*, edited by W. Strange (York, Baltimore), pp. 279–304.

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

Delattre, P. (**1965**). *Comparing the Phonetic Features of English, French, German, and Spanish* (Harrap & Co., London).

Escudero, P., and Boersma, P. (**2004**). "Bridging the gap between L2 speech perception research and phonological theory," Stud. Second Lang. Acquis. **26**, 551–585.

Evans, B. G., and Iverson, P. (**2004**). "Vowel normalization for accent: An investigation of best exemplar locations in northern and southern British English sentences," J. Acoust. Soc. Am. **115**, 352–361.

Evans, B. G., and Iverson, P. (**2007**). "Plasticity in vowel perception and production: A study of accent in young adults," J. Acoust. Soc. Am. **121**, 3814–3826.

Flege, J. E. (**1989**). "Differences in inventory size affect the location but not the precision of tongue positioning in vowel production," Lang Speech **32**, 123–147.

Flege, J. E. (**1995**). "Second language speech learning: Theory, findings, and problems," *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*, edited by W. Strange (York, Baltimore), pp. 233–277.

Flege, J. E. (**2003**). "Assessing constraints on second-language segmental production and perception," *Phonetics and Phonology in Language Comprehension and Production: Differences and Similarities*, edited by A. Meyer and N. Schiller (Mouton de Gruyter, Berlin).

Flege, J. E., Bohn, O.-S., and Jang, S. (**1997**). "The effect of experience on nonnative subjects' production and perception of English vowels," J. Phonetics **25**, 437–470.

Flege, J. E., Schirru, C., and MacKay, I. R. A. (**2003**). "Interaction between the native and second language phonetic subsystems," Speech Commun. **40**, 467–491.

Francis, A. L., and Nusbaum, H. C. (**2002**). "Selective attention and the acquisition of new phonetic categories," J. Exp. Psychol. Hum. Percept. Perform. **28**, 349–366.

Francis, A. L., Baldwin, K., and Nusbaum, H. C. (**2000**). "Effects of training on attention to acoustic cues," Percept. Psychophys. **62**, 1668–1680.

Francis, A. L., Nusbaum, H. C., and Fenn, K. (**2007**). "Effects of training on the acoustic phonetic representation of synthetic speech," J. Speech Lang. Hear. Res. **50**, 1445–1465.

Glasberg, B. R., and Moore, B. C. J. (**1990**). "Derivation of auditory filter shapes from notched-noise data," Hear. Res. **47**, 103–138.

Gottfried, T., and Beddor, P. S. (**1988**). "Perception of spectral and temporal information in French vowels," Lang Speech **31**, 57–75.

Hattori, K., and Iverson, P. (**2009**). "English /r/-/l/ category assimilation by Japanese adults: Individual differences and the link to identification accuracy," J. Acoust. Soc. Am. **125**, 469–479.

Heeren, W. F. L., and Schouten, M. E. H. (**2008**). "Perceptual development of phoneme contrasts: How sensitivity changes along acoustic dimensions that contrast phoneme categories," J. Acoust. Soc. Am. **124**, 2291–2302.

Hirata, Y., Whitehurst, E., and Cullings, E. (**2007**). "Training native English speakers to identify Japanese vowel length contrast with sentences at varied speaking rates," J. Acoust. Soc. Am. **121**, 3837–3845.

Holt, L. L., and Lotto, A. J. (**2006**). "Cue weighting in auditory categorization: Implications for first and second language acquisition," J. Acoust. Soc. Am. **119**, 3059–3071.

Iverson, P., and Evans, B. G. (**2003**). "A goodness optimization method for investigating phonetic categorization," in the Proceedings of the 15th International Conference of Phonetic Sciences, Barcelona, Spain.

Iverson, P., and Evans, B. G. (**2007**). "Learning English vowels with different first-language vowel systems: Perception of formant targets, formant movement, and duration," J. Acoust. Soc. Am. **122**, 2842–2854.

Iverson, P., and Preece-Pinet, M. (**2008**). "Training English vowels for French speakers with varying English experience," J. Acoust. Soc. Am. **123**, 3734.

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

Iverson, P., Smith, C. A., and Evans, B. G. (**2006**). "Vowel recognition via cochlear implants and noise vocoders: Effects of formant movement," J. Acoust. Soc. Am. **120**, 3998–4006.

Kingston, J. (**2003**). "Learning foreign vowels," Lang Speech **46**, 295–349.

Klatt, D. H., and Klatt, L. C. (**1990**). "Analysis, synthesis, and perception of voice quality variations among female and male talkers," J. Acoust. Soc. Am. **87**, 820–857.

Lambacher, S., Martens, W., Kakehi, K., Marasinghe, C., and Molholt, G. (**2005**). "The effects of identification training on the identification and production of American English vowels by native speakers of Japanese," Appl. Psycholinguist. **26**, 227–247.

Lively, S. E., Logan, J. S., and Pisoni, D. B. (**1993**). "Training Japanese listeners to identify English /r/ and /l/. II: The role of phonetic environment and talker variability in learning new perceptual categories," J. Acoust. Soc. Am. **94**, 1242–1255.

Logan, J. S., Lively, S. E., and Pisoni, D. B. (**1991**). "Training Japanese listeners to identify English /r/ and /l/: A first report," J. Acoust. Soc. Am. **89**, 874–886.

Meunier, C., Frenck-Mestre, C., Lelekov-Boissard, T., and Le Besnerais, M. (**2003**). "Production and perception of foreign vowels: does the density of the system play a role?" in the Proceedings of the 15th International Congress of Phonetic Sciences, Barcelona, Spain.

McAllister, R., Flege, J. E., and Piske, T. (**2002**). "The influence of the L1 on the acquisition of Swedish vowel quantity by native speakers of Spanish, English and Estonian," J. Phonetics **30**, 229–258.

Morrison, G. (**2002**). "Perception of English /i/ and /ɪ/ by Japanese and Spanish listeners: Longitudinal results," in *Proceedings of the North West Linguistics Conference 2002*, edited by G. S. Morrison and L. Zsoldes (Simon Fraser University Linguistics Graduate Student Association, Burnaby, BC, Canada), pp. 29–48.

Nishi, K., and Kewley-Port, D. (**2007**). "Training Japanese listeners to perceive American English vowels: Influence of training sets," J. Speech Lang. Hear. Res. **50**, 1496–1509.

Stockwell, R. P., and Bowen, J. D. (**1965**). *The Sounds of English and Spanish* (University of Chicago Press, Chicago).

Strange, W., Bohn, O.-S., Nishi, K., and Trent, S. (**2005**). "Contextual variation in the acoustic and perceptual variation of North German and American English vowels," J. Acoust. Soc. Am. **118**, 1751–1762.

Tajima, K., Kato, H., Rothwell, A., Akahane-Yamada, R., and Munhall, K. G. (**2008**). "Training English listeners to perceive phonemic length contrasts in Japanese," J. Acoust. Soc. Am. **123**, 397–413.

Trubetzkoy, N. S. (**1969**). *Principles of Phonology*, translated by C. A. M. Baltaxe (University of California Press, Berkeley).