The effect of native vowel processing ability and frequency discrimination acuity on the phonetic training of English vowels for native speakers of Greek

Angelos Lengeris[a)] and Valerie Hazan
Department of Speech Hearing and Phonetic Sciences, University College London, Chandler House, 2 Wakefield Street, London WC1N 1PF, United Kingdom

(Received 5 February 2010; revised 5 October 2010; accepted 6 October 2010)

The perception and production of nonnative phones in second language (L2) learners can be improved via auditory training, but L2 learning is often characterized by large differences in performance across individuals. This study examined whether success in learning L2 vowels, via five sessions of high-variability phonetic training, related to the learners' native (L1) vowel processing ability or their frequency discrimination acuity. A group of native speakers of Greek received training, while another completed the pre-/post-tests but without training. Pre-/post-tests assessed different aspects of their L2 and L1 vowel processing and frequency acuity. L2 and L1 vowel processing were assessed via: (a) Natural English (L2) vowel identification in quiet and in multi-talker babble, and natural Greek (L1) vowel identification in babble; (b) the categorization of synthetic English and Greek vowel continua; and (c) discrimination of the same continua. Frequency discrimination acuity was assessed for a nonspeech continuum. Frequency discrimination acuity was related to measures of both L1 and L2 vowel processing, a finding that favors an auditory processing over a speech-specific explanation for individual variability in L2 vowel learning. The most efficient frequency discriminators at pre-test were also the most accurate both in English vowel perception and production after training. © *2010 Acoustical Society of America*. [DOI: 10.1121/1.3506351]

PACS number(s): 43.71.Hw, 43.71.Es [RSN] Pages: 3757–3768

I. INTRODUCTION

Adults often have difficulty in learning nonnative vowels especially when the vowel inventories in the first (L1) and second language (L2) are very different. For example, native Greek speakers struggle to distinguish English /iː/ from /ɪ/ because they lack such a contrast in their L1 and instead have a single vowel category /i/ in the acoustic/perceptual space occupied by the two English vowels (Lengeris, 2009); see also Cebrian (2006), Flege *et al*. (1997), and Iverson and Evans (2009) for a similar finding concerning Spanish learners of English. At the same time, there is evidence that individuals who come from the same L1 background and who have similar profiles vary (a) in their ability to perceive and pronounce novel speech sounds in the L2 (e.g., Jilka, 2009) and (b) in the degree to which they respond to auditory training of L2 sounds in laboratory settings (e.g., Hazan *et al*., 2005, 2006; Golestani and Zatorre, 2009). This study investigated whether such individual differences are related to differences in L1 vowel processing ability or frequency discrimination acuity.

A variety of factors may determine the success of L2 phoneme learning: These include the relationship between the segmental inventory of the L1 and the L2 (e.g., Best, 1995; Flege, 1995), the acoustic (Polka, 1991) and visual (Hazan *et al*., 2006) salience of the novel phonetic contrast itself, the age of learning an L2 (e.g., Flege *et al*., 1999b), the length of residence in an L2 setting (e.g., Flege *et al*., 1997a; Flege and Liu, 2001), and the degree of ongoing L1 use (e.g., Flege and MacKay, 2004). Factors unrelated to L1/L2 experience such as motivation to learn and language learning aptitude are also likely to affect the success of L2 phoneme learning (see Bongaerts *et al*., 1997; Moyer, 1999, examining the role of motivation on degree of global foreign accent). Even when language-related factors are controlled, individuals are often found to differ in their ability to learn the sounds of an L2.

A few recent studies have examined individual differences in both initial L2 perception acuity and the impact of auditory training, focusing on how various phonetic measures of L1 assimilation, L2 category structure, and L2 identification are related to each other. With regard to the individual differences in initial L2 perception, Iverson *et al*. (2008) tested the perception of the English /w/-/v/ distinction by native Sinhala, German, and Dutch speakers and showed that participants' identification, goodness ratings, and discrimination of synthetic English stimuli were related with their identification accuracy of natural English /w/-/v/ recordings. Hattori and Iverson (2009) found that individual differences in native Japanese speakers' representation of F3 in their best exemplar locations (i.e., perceptual representations) of /r/ and /l/ were predictive of natural English /r/-/l/ identification, rather than their /r/-/l/ assimilation to Japanese /ɾ/. In a study on the contribution of visual cues to L2 perception, Hazan *et al*. (2006) showed substantial variability both in the identification of nonnative consonant contrasts in participants with similar language profiles and in the relative use of auditory and visual cues prior to training. With respect to vowel learning, Iverson and Evans (2007b, 2009) found

a)Author to whom correspondence should be addressed. Electronic mail: a.lengeris@uclmail.net

that individual differences in best exemplar F1 and F2 locations of English vowels for Spanish, French, Norwegian, and German learners of English were correlated with their identification accuracy of natural English vowels. Other studies looking at the neural correlates of individual differences in L2 learning report a relationship between learning and anatomical and functional differences in the learner's brain (e.g., Golestani *et al.*, 2002; Golestani *et al.*, 2007; Diaz *et al.*, 2008).

One way to characterize individual differences in L2 learning is to employ auditory training and examine the degree to which participants improve with training. Several training protocols have been used in the past (for a review, see Bradlow, 2008). The most commonly used training protocols emphasize the importance of exposing listeners to naturally produced highly variable minimal pairs contrasting the target sounds in multiple environments. This approach has been shown to improve the perception of consonants (e.g., Logan *et al.*, 1991; Lively *et al.*, 1993; Pruit *et al.*, 2006; Hazan *et al.*, 2005) and vowels (e.g., Nishi and Kewley-Port, 2007, 2008; Iverson and Evans, 2009). Improvement is retained several months after training (Lively *et al.*, 1994; Bradlow *et al.*, 1999; Iverson and Evans, 2009) and transfers to speech production for consonants (Bradlow *et al.*, 1997; Hazan *et al.*, 2005) and vowels (Lambacher *et al.*, 2005). Most of the above studies have looked at group improvement; when individual data are examined, in addition to large pretest variability, improvement after training can range from no improvement to very significant gains across individuals (Bradlow *et al.*, 1997, 1999).

Listeners may differ not only in their ability to learn new sounds in an L2 but also in their ability to process speech in their native language. Although such L1 variability is expected not to be easily noticed in everyday life given the redundancy of acoustic cues in the speech signal, differences across individuals become evident when their perceptual system is stressed, specifically when listening to speech in noise (Surprenant and Watson, 2001; Kidd *et al.*, 2007) or when analytical tasks are used, for example, measuring acoustic cue weighting for consonants (Hazan and Rosen, 1991) and discrimination accuracy for vowels (Kewley-Port, 2001; Gerrits and Schouten, 2004). Given that individuals may vary in processing relevant acoustic cues for the perception of L1 speech, we hypothesized that this may influence their ability to learn novel sound distinctions ("L1 phonetic" hypothesis). Current cross-language/L2 models such as the perceptual assimilation model (PAM) (Best, 1995; Best and Tyler, 2007), the speech learning model (SLM) (Flege, 1995, 2003), and the native language magnet (NLM) model (Kuhl, 2000; Kuhl *et al.*, 2008; Iverson *et al.*, 2003) agree that experience with the native language interferes with L2 learning. For example, the SLM attributes age effects to age-related changes in robustness of L1 phonetic categories. That is, as children grow up they become more committed to their L1 categories, which results in the difficulty adult learners are faced with when learning an L2 (e.g., Flege *et al.*, 2003; Walley and Flege, 1999). If this L1/L2 "trade-off" is extended to adult L2 learners, individuals with relatively poorly defined L1 categories (shallower identification slopes, better within-category discrimination) might prove to be better at learning L2 vowels.

Similarly, research in auditory abilities has shown that listeners vary in their performance on psychoacoustic tasks (Johnson *et al.*, 1987; Surprenant and Watson, 2001; Kidd *et al.*, 2007; see also Carroll, 1993 for a comprehensive review of factor-analytic studies of human cognitive abilities, based on 38 studies published before 1993); it is possible that this may impact their learning of new distinctions ("auditory processing" hypothesis). Wong and Perrachione (2007) and Lee *et al.* (2007) showed that auditory pitch ability, as measured using nonspeech stimuli, can predict success in the use of pitch patterns in lexical identification in a tone language by L2 learners; however, since pitch is a shared acoustic feature of music and tone perception, one may question whether such a link is specific to the acquisition of tone languages or may relate to general L2 learning ability. It is also important to note that the two hypotheses mentioned above, namely the L1 phonetic hypothesis and auditory processing hypothesis, are not mutually exclusive, i.e., auditory/frequency discrimination ability may underlie both L1 and L2 vowel processing.

The main aim of the present study was to assess whether individual differences in English vowel learning are related to listeners' perception of vowels in their native language or to their frequency discrimination acuity. Participants were trained on their perception of English vowels using a training program that included real English words and that has been shown to successfully improve English vowel perception and production for Spanish and German speakers (Iverson and Evans, 2009).

Both the trainees and another group of Greek speakers who received no training completed a large battery of pre-/post-tests which assessed different aspects of their processing of vowels in their L1 and L2 as well as their frequency discrimination acuity. L2 vowel processing was assessed using four types of tests. First we tested their ability to correctly identify natural English (L2) vowels. Second, the same natural tokens were presented in a background of multi-talker babble in order to examine whether learning is robust enough to transfer to a situation that resembles real-world communication. Third, we investigated the categorization of synthetic vowel continua for two English vowel contrasts (/iː/-/ɪ/ and /æ/-/ʌ/), which are difficult for native Greek speakers due to their different phonological status in the L1 (Lengeris, 2009); both measures of consistency of categorization (identification slope) and category boundary (phoneme boundary) were obtained. Fourth, we investigated participants' ability to discriminate the same two L2 vowel contrasts, using an adaptive procedure to obtain a measure of just noticeable difference. These last two tests were to probe whether category representation, rather than merely identification accuracy, changed as a result of the training. To evaluate the predictions of the "L1 phonetic" hypothesis that speakers with less robust L1 vowel categories (shallower identification slopes) will be better at learning L2 vowels, participants were also tested on their identification and discrimination of pairs of synthetic vowels in Greek, chosen within the same area of the vowel space as the English pairs

(/i/-/e/ and /a/-/o/). An identification task with natural Greek vowels in multi-talker babble was used as another test of L1 category robustness that might reveal individual differences in the processing of L1 vowels. Finally, to test the "auditory processing" hypothesis that speakers with better discrimination ability would also be better at learning L2 vowels, a frequency discrimination task using nonspeech stimuli (single F2 formant) was presented. In order to evaluate any transfer of the perception training to L2 vowel production, recordings of the participants' production of English vowels made at the pre- and post-test sessions were evaluated by English listeners and measured acoustically.

II. METHOD

A. Participants

A total of 28 speakers of Standard Modern Greek were tested; 18 completed a vowel training program and ten served as controls, i.e., performed the pre-/post-tests but received no training (to evaluate the learning effect that would come from test repetition). The trainees had a mean age of 23 yr (range = 18–35 yr) and the controls had a mean age of 26 yr (range = 18–42 yr). The majority of participants (24/28) were recruited from two English language schools in Athens, Greece; all of the participants had 10–12 yr of formal English instruction but had very little, if any, interaction with native English speakers and none had spent a period of more than 1 month in an English-speaking environment. Their proficiency level was moderately high and relatively uniform across individuals [e.g., Cambridge First Certificate in English (FCE), Cambridge Certificate in Advanced English (CAE)]. All participants passed a pure-tone hearing screening at frequencies from 250 to 4000 Hz at 20 dB hearing level (HL) and were paid for their participation.

B. Stimuli

1. Natural vowels

Tests including natural vowel tokens investigated the perception of a range of natural English (L2) and Greek (L1) vowels both in quiet and in background babble to assess vowel processing in the L2 and L1. Digital recordings of all natural stimuli were made in an anechoic chamber at a sampling rate of 44 100 16-bit samples per second. Two Greek speakers, one male and one female, read three times in random order the five Greek vowels /i, e, a, o, u/ embedded in /pVs/ words. Two Southern British English speakers, one male and one female, read three times in random order ten English vowels /iː, ɪ, e, ɜː, æ, ʌ, ɑː, ɒ, ɔː, uː/ embedded in /bVt/ words. A further Southern British English speaker recorded the same words, so that we could test the generalization of learning. The best two repetitions of each vowel were chosen for each speaker. Ideally, consonantal context should be kept constant across languages; however, the /pVs/ context is one of the very few contexts that create a minimally contrastive set of words in Greek.

2. Synthetic vowels

Identification and discrimination tests of synthetic vowel continua in English (L2) and Greek (L1) were used to assess categorization and category representation in the L2 and consistency of categorization in the L1. The synthetic stimuli consisted of two Greek and two English vowel continua embedded within natural Greek and English words spoken by a Greek and an English native speaker, respectively. The Greek continua ranged from /i/ to /e/ and from /a/ to /o/; the English continua ranged from /iː/ to /ɪ/ and from /æ/ to /ʌ/. These vowels constitute two pairs of vowel contrasts that cover similar areas in the acoustic/perceptual space across languages (English /iː/-/ɪ/ vs Greek /i/-/e/ and English /æ/-/ʌ/ vs Greek /a/-/o/, see Lengeris, 2009) and were selected so that we could compare learners' performance across languages. Greek speakers in Lengeris (2009) had less difficulty in perceiving English /æ/-/ʌ/ than /iː/-/ɪ/ and we expected a similar pattern of results with our synthetic stimuli. The Greek /i/-/e/ continuum was placed in a /pVta/ context and the Greek /a/-/o/ continuum was placed in a /pVte/ context (there is no minimal pair in Greek contrasting all four vowels in either context). The English /iː/-/ɪ/ and /æ/-/ʌ/ continua were embedded in a /bVt/ context. Since Greek voiceless stops /p/, /t/, /k/ are unaspirated in all positions, (e.g., Arvaniti, 2007; Botinis, *et al*., 2000) and English voiced stops /b/, /d/, /g/ are phonetically realized as voiceless in initial position (Docherty, 1992), the selection of these contexts ensured that participants would be tested on phonetically similar consonantal contexts across languages.

The synthetic vowels were created using a Klatt synthesizer (Klatt and Klatt, 1990) in a cascade configuration with a sampling rate of 11 025 Hz. Each vowel continuum had 51 stimuli varying in 50 equal steps in terms of F1 and F2 onset and offset frequencies and duration. The rest of the synthesis parameters were kept constant across vowels. These were the F3, F4, and F5 frequencies (2500, 3500, and 4500 Hz), the formant bandwidths (B1 = 100, B2 = 180, B3 = 250, B4 = 300, and B5 = 550), the tilt (TL = 0 dB slope), and the open quotient (OQ = 60%). The synthesized vowels matched the natural ones in terms of F_0 and amplitude. The endpoint values for the first and second formant values and duration for the Greek and English continua are given in Table I. To examine the effect of stimulus duration on the perception of English vowel continua by Greek speakers, we created two versions of the English /iː/-/ɪ/ continuum: In the *natural duration* condition, /iː/ had a duration of 110 ms and /ɪ/ had a duration of

TABLE I. Endpoint values for onset and offset F1 and F2 frequencies (Hz) and duration (ms) for each synthetic vowel continuum.

Synthetic continuum	Endpoint	Onset F1 (Hz)	Offset F1 (Hz)	Onset F2 (Hz)	Offset F2 (Hz)	Duration (ms)
Greek /i/-/e/	/i/	330	236	2265	2518	55
	/e/	577	518	1739	2084	
Greek /a/-/o/	/a/	848	814	1196	1464	65
	/o/	543	494	749	825	
English /iː/-/ɪ/ *nat*.	/iː/	247	182	2527	2785	110
	/ɪ/	364	392	1986	2170	70
English /iː/-/ɪ/ *neut*.	/iː/	247	182	2527	2785	90
	/ɪ/	364	392	1986	2170	
English /æ/-/ʌ/	/æ/	701	809	1458	1521	85
	/ʌ/	574	651	1011	1233	

70 ms while in the *neutralized duration* condition, vowel duration was 90 ms, a duration intermediate to that of the *natural duration* condition.

3. Nonspeech continuum

The discrimination test involving nonspeech tokens was used to assess frequency discrimination acuity. The nonspeech continuum consisted of a single formant which varied in center-frequency from 1250 to 1500 Hz and thus was a nonspeech analog to a vowel second formant. Token duration was 150 ms, the formant had a bandwidth of 100 Hz, and F_0 was set at a constant value of 120 Hz, resembling the pitch of a male speaker. The nonspeech continuum had 51 stimuli varying in equal steps in terms of the formant frequency. The nonspeech continuum therefore shared similar acoustic properties with the vowel continua (i.e., harmonic structure, similar duration, and pitch) without being perceived as speech.

C. Procedure

The pre-/post-tests were carried out in quiet rooms in a single session lasting about 1.5 h. All participants were tested on the same computer using Sennheiser HD 433 headphones (Buckinghamshire, UK) at a comfortable listening level set by each individual. To reduce the possibility of the nonspeech continuum being treated as speech, this was the first task that was run. All tasks with English vowels preceded those with Greek vowels. All tasks are listed in Table II in their order of presentation.

Training was run at the participants' homes; the training software was installed on their laptops/desktops and the training sessions were done in a quiet room via Sennheiser HD 433 headphones. Stimuli were presented at a comfortable listening level set by each individual. The details of each training session (e.g., participant details, day and time of completion) were automatically logged in a password-protected file that was not accessible to the participants to ensure that they completed all sessions.

1. Perceptual training

The training software, stimuli, and procedures were the same as in Iverson and Evans (2009). The training stimuli were English words containing 14 English vowels (all ten vowels used in the pre-/post-tests and four additional diphthongs) spoken by five Southern British English speakers (two males and three females). Given that very few minimal-pair sets in English contrast all 14 vowels, Iverson and Evans (2007a) arranged the vowels in four minimal-pair groups /iː/, /ɪ/, /aɪ/, /eɪ/ (e.g., *peel*, *pill*, *pile*, *pale*), /uː/, /aʊ/, /ɜː/ (e.g., *blues*, *blouse*, *blurs*), /ɒ/, /əʊ/, /ɔː/ (e.g., *stock*, *stoke*, *stork*), and /e/, /ɑː/, /æ/, /ʌ/ (e.g., *mesh*, *marsh*, *mash*, *mush*) after conducting a hierarchical cluster analysis on identification data by Spanish and German speakers; the first three groups contained vowels which were problematic for both language groups and the last group contained the remaining vowels. Given the similarity of the Greek and the Spanish vowel systems, Greek speakers were expected to face similar difficulties with these four vowel groups. Each vowel group contained ten sets of minimal-pair words (i.e., there were 140 different words in the training stimuli) which were recorded twice by each English speaker, thus ensuring that the training stimuli were highly variable.

The trainees completed five training sessions (identification with feedback) on separate days over a 2-week period. Each session lasted about 45 min and consisted of 225 trials with a different speaker on each day. A short session consisting of 14 trials was given before training to familiarize trainees with the procedure. Training was partly adaptive; the first 70 trials were five random repetitions of the 14 English vowels, the next 85 were based on the participants' errors (i.e., a vowel that would prove difficult would be tested more times than a vowel that would prove easy) and the last 70 trials were again five random repetitions of the 14 English vowels. On each trial testing an English vowel, a set of minimal-pair words containing that vowel was chosen in a random order with the restriction that all ten sets were presented once before a set was used again. During training, on each trial, the learners heard an English word and chose one of three or four candidates (accompanied by a common word with the same vowel) as displayed on a computer screen. If the target word was correctly identified "Yes!" was displayed on the screen, a cash register sound was heard, and the target word was repeated once. If the target word was misidentified "Wrong" was displayed on the screen, two

TABLE II. Pre-/post-test battery completed by all Greek speakers. The tasks are listed in their order of presentation to participants.

Task	Stimulus
(1) Nonspeech discrimination	1250–1500 Hz continuum
(2) English natural vowel identification in quiet	/bVt/ words
(3) English natural vowel identification in noise	/bVt/ words (SNR = −4 dB)
(4) English vowel production	/bVt/ words
(5) English synthetic vowel identification	i. /biːt/-/bɪt/ *natural duration* continuum
	ii. /bæt/-/bʌt/ continuum
	iii. /biːt/-/bɪt/ *neutralized duration* continuum
(6) English synthetic vowel discrimination	Same as above
(7) Greek natural vowel identification in noise	/pVs/ words (SNR = −10 dB)
(8) Greek synthetic vowel identification	i. /pita/-/peta/ continuum
	ii. /pate/-/pote/ continuum
(9) Greek synthetic vowel discrimination	Same as above

beeps were heard, and both the target and the (incorrectly) chosen word were repeated twice.

2. Pre-/post-test battery

a. Natural English vowel identification in quiet and in noise. The English /bVt/ words were presented within a ten-alternative forced-choice task. Participants clicked on one of the options displayed on a computer screen (English orthographic labels accompanied by a common English word with the same vowel). In the noise condition, multi-talker babble was played simultaneously with the /bVt/ words at a signal-to-noise ratio (SNR) of -4 dB.[1] For each condition, there were 40 presentations in the pre-test (2 speakers $\times$ 10 vowels $\times$ 2 repetitions) and 60 in the post-test (3 speakers $\times$ 10 vowels $\times$ 2 repetitions), as a "new" speaker was included to test generalization of the learning. Presentation was blocked by speaker and vowels were fully randomized within each block. Before testing both in quiet and noise, listeners heard all of the words spoken by one speaker once (the initial presentation of words for the noise session was presented with noise), together with their orthographic labels.

b. Natural Greek vowel identification in noise. The Greek /pVs/ words were presented within a five-alternative forced-choice task. Participants clicked on one of the options displayed on a computer screen (Greek orthographic labels). There were 20 presentations in the pre-/post-tests (2 speakers $\times$ 5 vowels $\times$ 2 repetitions). Presentation was blocked by speaker and vowels were fully randomized within each block. The multi-talker babble was played simultaneously with the /pVs/ words at a SNR of -10 dB.[2] Before testing, listeners heard all of the words spoken by one speaker presented with noise, together with their orthographic labels.

c. Synthetic vowel identification. A two-alternative forced-choice task was used to assess the identification of L1 and L2 synthetic vowels. After hearing a Greek or an English word containing a synthetic vowel from a continuum, participants identified the vowel by clicking on a button displaying the stimulus word as well as a picture representing that word (to reduce any effects of orthography). Stimuli were presented using an interleaved adaptive procedure that made effective use of a relatively small number of presentations. Two independent adaptive tracks started at opposite ends of the continuum (each time the choice of track was random) and estimated the point on the continuum where the stimuli were labeled as a given word 71% of the time using a two-down/one-up rule (Levitt, 1971). To prevent listeners from continuously hearing ambiguous stimuli, 20% of the trials were endpoint stimuli. Testing ended after seven reversals on each track or 50 trials. For each listener and vowel continuum, logistic regression was used to obtain a best-fit sigmoid function and estimates of the boundary and slope were calculated from the fitted coefficients. Identification boundary defines the point in the continuum where the two vowel responses are equally probable, i.e., the phoneme boundary, and the identification slope measures identification consistency.

d. Synthetic vowel and nonspeech discrimination. A three-alternative forced-choice task was used to assess L1, L2, and nonspeech discrimination. The L1/L2 discrimination test used the same stimuli as the identification tests described above. In the discrimination test, three frogs appeared on the screen "saying" one word that contained a vowel from the continuum or, in the case of nonspeech discrimination, one stimulus from the nonspeech continuum. Participants were told that two of the words (or the nonspeech stimuli) were the same and one was different and were asked to indicate the odd one out by clicking the appropriate frog which could be in any of the three positions. Feedback was provided in the form of a tick or an x mark above the selected frog. A method of "standard" was used against which the other stimuli were compared. The standard was one endpoint of the continuum (the first vowel in each vowel continuum and the 1250 Hz endpoint in the nonspeech continuum). A three-down/one-up rule was used (Levitt, 1971), which found the just noticeable difference (jnd), i.e., the stimulus that could be discriminated from the standard 79% of the time. Testing ended after seven reversals on each track or 50 trials and the mean of the last four reversals defined the jnd. The inter-stimulus interval was 250 ms.

e. English vowel production. The participants read from a screen each of the ten English /bVt/ words they had previously identified. Recordings were made using a MicroTrack 24/96 digital recorder at a sampling rate of 44.1 kHz. In order to get a quality rating for vowel production by Greek speakers, two Southern British English speakers first identified each vowel from a forced-choice set of ten English categories and then, after hearing the same vowel once more, rated its goodness in a scale from 1 (very bad example) to 7 (excellent example). Each English speaker performed 560 judgments (28 speakers $\times$ 10 vowels $\times$ pre- and post-tests) with vowels fully randomized.

III. RESULTS AND DISCUSSION

A. Does perceptual training improve the identification of L2 natural vowels in quiet and in noise?

Figure 1 displays the English vowel identification accuracy for the trained group (upper panels) and the control group (lower panels) in quiet and in noise before and after training. Independent *t* tests with Group as a between-subject variable confirmed that the two groups did not differ in their initial performance in quiet or in noise, $p > 0.05$. A repeated-measures analysis of variance (ANOVA) on identification scores with Group (trained, control) as a between-subject factor and Noise condition (quiet, noise) and Test (pre-test, post-test, generalization) as within-subject factors showed a significant main effect of Group, $F(1,26) = 10.9$, $p < 0.01$, of Noise condition, $F(1,26) = 229.4$, $p < 0.001$ and of Test, $F(2,52) = 31.6$, $p < 0.001$. There was also a significant Test $\times$ Group interaction $F(2,52) = 13.9$, $p < 0.001$ which was explored through simple effect tests. The simple effect of Test was significant for the trained group $F(2,34) = 48.6$, $p < 0.001$ but not for the control group, $p > 0.05$. Pairwise comparisons (Bonferroni adjusted) showed that, across noise conditions, the trained group improved from

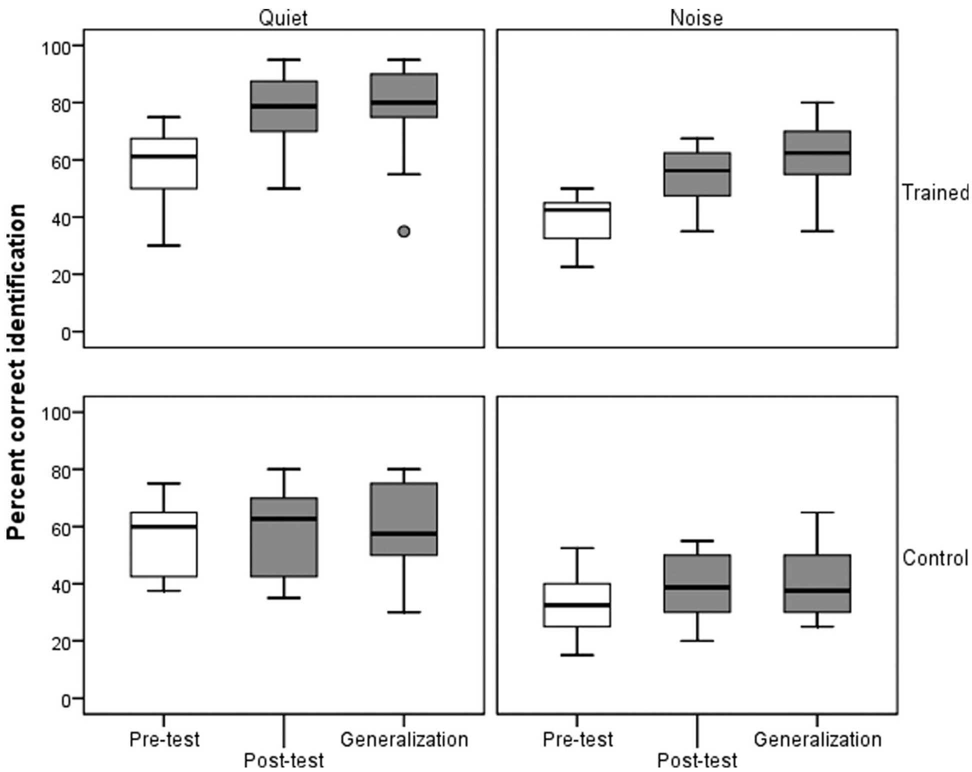

FIG. 1. Boxplots showing identification accuracy for English vowels by the trained (upper panels) and the control (lower panels) group of native Greek speakers in quiet (left panels) and in noise (right panels) in the pre-test, the post-test, and the generalization test. Whiskers extend to at most 1.5 times the interquartile range of the box, with outliers (i.e., more than 1.5 times the interquartile range away from the median) marked by circles.

pre-test (48.6% correct) to post-test (65.9% correct) and generalization test (70.2% correct), demonstrating a transfer of learning both to a new English speaker and to speech-in-noise conditions. Pearson correlations on identification scores for the trained group ($n = 18$) showed that identification in quiet was correlated with that in noise, both before, $r = 0.7$, $p < 0.01$, and after training, $r = 0.59$, $p < 0.01$, indicating that individuals who were better at perceiving English vowels in quiet were also better at doing so in noise.

B. Does perceptual training improve the categorization of L1 and L2 synthetic vowel continua?

A repeated-measures ANOVA on identification boundary locations for five vowel continua (two Greek and three English continua), with Group (trained, control) as a between-subject factor and Test (pre-test, post-test) and Vowel continuum (five levels) as within-subject factors showed no main effects or interactions of Group, Test, or Vowel continuum, $p > 0.05$, suggesting that training or test repetition had no effect on the position of the vowel boundary for any of the continua.

Figure 2 displays the slope of the identification functions for the same vowel continua in the pre-/post-tests for the trained and the control groups (upper and lower panels, respectively). The slope measures the steepness of the identification function and is thus an index of identification consistency: The steeper the slope, the more consistently a vowel is labeled. As expected, the participants were much more consistent in their identification of their native Greek vowels than they were in their identification of English vowels ($M = 0.5$ vs $M = 0.1$). Given the large differences in the responses to the L1 and L2 vowels both in terms of identification consistency and range of scores, separate repeated-measures ANOVAs on identification slopes were performed for each language with Group (trained, control) as a between-subject factor and Test (pre-test, post-test) and Vowel continuum (two levels for Greek and three levels for English) as within-subject factors. For Greek, there were no significant main effects or interactions of Group, Test, or Vowel continuum, $p > 0.05$, suggesting that training or test repetition did not change identification consistency. For English, there was a significant main effect of Test, $F(1,26) = 7.8$, $p < 0.01$ and a significant Test $\times$ Group interaction, $F(1,26) = 6.4$, $p < 0.01$. *Post-hoc t* tests showed that the trained group had significantly steeper identification slopes in the post-test than in the pre-test for /iː/-/ɪ/ *natural duration* ($M = 0.19$ vs $M = 0.10$, respectively) and for /iː/-/ɪ/ *neutralized duration* ($M = 0.18$ vs $M = 0.09$, respectively) but not for /æ/-/ʌ/ ($M = 0.17$ vs $M = 0.16$, respectively) whereas identification slopes for the control group did not change after training for any of the three English vowel continua. These findings show a transfer of learning to the consistency of categorization of synthetic vowel continua. Pearson correlations on identification slopes showed no significant correlations either within or between L1 and L2, $p > 0.05$ in the pre-test; however, after training there were significant correlations between identification slopes for /iː/-/ɪ/ *natural* and *neutralized duration*, $r = 0.52$, $p < 0.05$, between /iː/-/ɪ/ *natural duration* and /æ/-/ʌ/, $r = 0.5$, $p < 0.05$, and between /iː/-/ɪ/ *neutralized duration* and /æ/-/ʌ/, $r = 0.63$, $p < 0.01$, i.e., individuals showed consistently strong or poor identification ability (steep or shallow identification slopes, respectively) following the vowel training program.

C. Does perceptual training improve the discrimination of L1 and L2 vowel and nonspeech continua?

Since four (spectrally) different synthetic vowel continua were used (and hence the acoustical/perceptual difference between the endpoints in each continuum differed), before

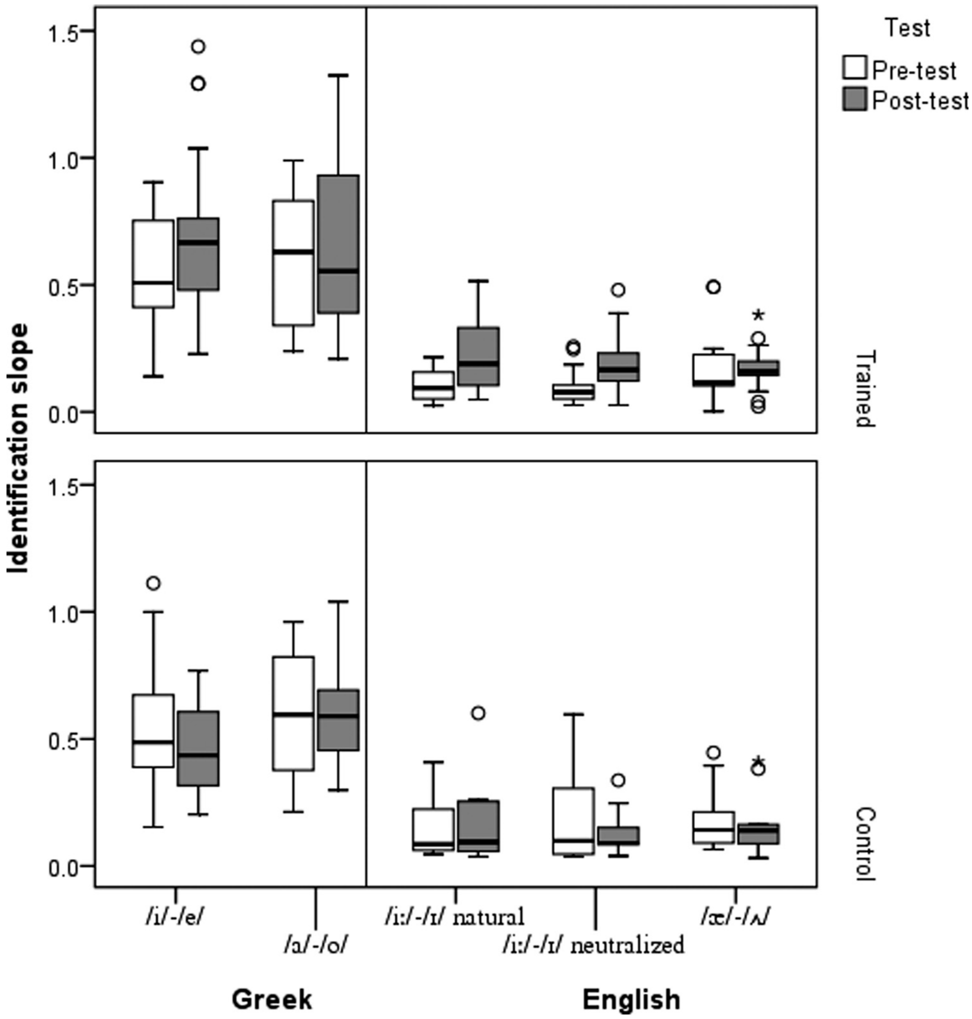

FIG. 2. Boxplots showing identification slopes for Greek and English vowel continua by the trained (upper panel) and the control (lower panel) group of native Greek speakers in the pre-test and the post-test. Whiskers extend to at most 1.5 times the interquartile range of the box, with outliers marked by circles and asterisks.

making any comparisons between continua, the Euclidean distance (Hz) between the endpoints (x_1, y_1) and (x_2, y_2) was calculated for each continuum (after taking the center/mean of F1 and F2 movement for each endpoint) and the jnd from the fixed reference was estimated. The endpoints of the nonspeech continuum did not entail formant movement and so the Euclidean distance (Hz) between the endpoints (x_1, y_1) and (x_2, y_2) was calculated and the jnd from the fixed reference was estimated. Figure 3 displays the jnd's (Hz) for all five vowel continua and the nonspeech continuum for the trained (upper panel) and the control (lower panel) group before and after training. A repeated-measures ANOVA on jnds with Group (trained, control) as a between-subject factor and Test (pre-test, post-test) and Continuum (six levels[3]) as within-subject factors yielded a significant effect of Continuum, $F(5,130) = 26.3$, $p < 0.001$ and no effect of Test, Group, or interactions, $p > 0.05$, suggesting that training did not change subjects' discrimination. Pairwise comparisons (Bonferroni adjusted) showed that, across groups and tests, Greek speakers showed better discrimination for Greek /i/-/e/ (jnd = 156 Hz) than for English /iː/-/ɪ/ *natural* and *neutralized duration* (jnd = 255 and 262 Hz, respectively) but did not differ in their discrimination for Greek /a/-/o/ (jnd = 144 Hz) compared to that for English /æ/-/ʌ/ (jnd = 130 Hz); this demonstrates an L1 advantage over the two versions of English /iː/-/ɪ/ but no such advantage over English /æ/-/ʌ/. Nonspeech discrimination for the Greek speakers was at the same levels (jnd = 152 Hz) as their discrimination for Greek /i/-/e/ and /a/-/o/. Pearson correlations on jnd's showed that almost all measures correlated with each other in pre-/post-tests (Table III). Importantly, nonspeech discrimination correlated with all vowel pairs across L1/L2 and pre-/post-tests, demonstrating that participants were consistently "strong" or "poor" discriminators of stimuli varying in frequency.

D. Does perceptual training improve English vowel production?

The English vowels produced by the Greek speakers in the trained and control groups in the pre-/post-tests were presented to native English speakers, and identification accurary scores were obtained (see Fig. 4). A two-way repeated-measures ANOVA on identification scores with Group (trained, control) as a between-subject factor and Test (pre-test, post-test) as a within-subject factor yielded a significant main effect of Group, $F(1,26) = 8.4$, $p < 0.01$, of Test $F(1,26) = 40.4$, $p < 0.001$ and a significant Test × Group interaction, $F(1,26) = 23.8$, $p < 0.001$. The simple effect of Test was significant for the trained group, $F(1,17) = 73.3$, $p < 0.001$ (from 61.9% to 75.8% correct) but not for the control group, $p > 0.05$ (from 60.75% to 61.2% correct). Similarly, a two-way repeated-measures ANOVA on goodness ratings given by

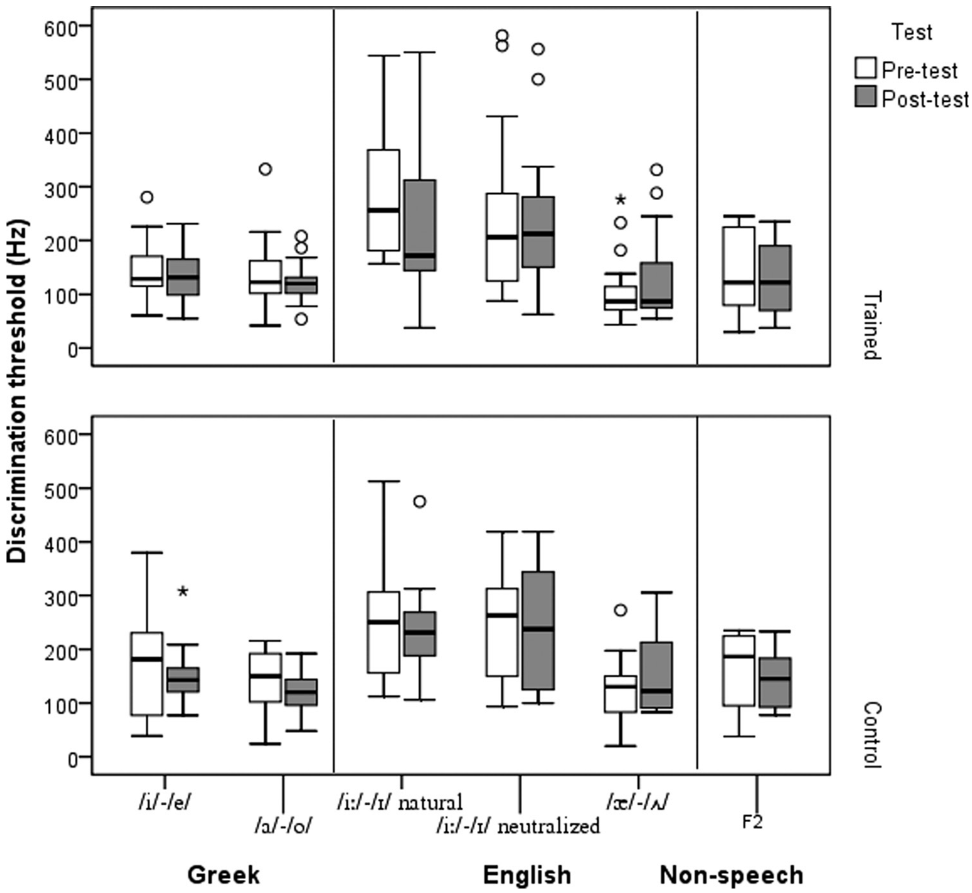

FIG. 3. Boxplots showing discrimination thresholds (Hz) for Greek, English, and nonspeech continua by the trained (upper panel) and the control (lower panel) group of native Greek speakers in the pre-test and the post-test.

native listeners yielded a significant main effect of Group, $F(1,26) = 13.6$, $p < 0.001$, of Test $F(1,26) = 9.5$, $p < 0.01$ and a significant Test × Group interaction, $F(1,26) = 38.3$, $p < 0.001$. The simple effect of Test was significant for the trained group, $F(1,17) = 48.8$, $p < 0.001$ (from 3.6 to 4.1) but not for the control group, $p > 0.05$ (from 3.3 to 3.2). These results indicate that English vowels produced by Greek speakers were not only identified with greater rates of accuracy after training but also received higher goodness ratings.

The English vowels produced by the trained group were measured acoustically using the speech filing system (SFS)

TABLE III. Correlations (r) among discrimination tasks for five synthetic vowel continua and a nonspeech continuum in the pre-/post-tests for the trained group of Greek speakers.

	Greek /i/-/e/	English /iː/-/ɪ/ *nat*.	English /iː/-/ɪ/ *neut*.	Greek /a/-/o/	English /æ/-/ʌ/	F2 only
Pre-test						
Greek /i;/-/e/	1					
English /iː/-/ɪ/*nat*.	0.38*	1				
English /iː/-/ɪ/*neut*.	0.39*	0.67**	1			
Greek /a/-/o/	0.38*	0.25	0.44*	1		
English /æ/-/ʌ/	0.45*	0.54**	0.71**	0.43*	1	
F2 only	0.55**	0.62**	0.54**	0.47*	0.72**	1
Post-test						
Greek /i/-/e/	1					
English /iː/-/ɪ/*nat*.	0.51*	1				
English /iː/-/ɪ/*neut*.	0.33	0.8**	1			
Greek /a/-/o/	0.48*	0.65**	0.74**	1		
English /æ/-/ʌ/	0.61**	0.76**	0.66**	0.64**	1	
F2 only	0.46*	0.71**	0.46*	0.65**	0.68**	1

*$p < 0.05$.
**$p < 0.01$.

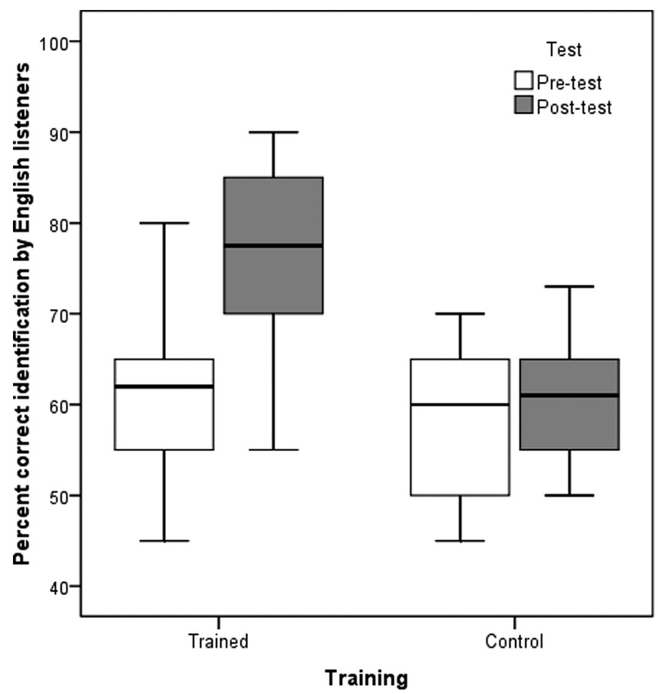

FIG. 4. Boxplots showing identification accuracy for English vowels produced by the trained and the control group of native Greek speakers in the pre-test and the post-test as judged by native English listeners.

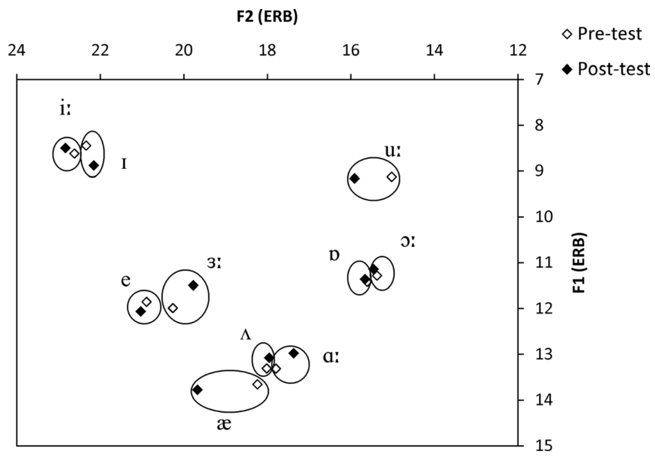

FIG. 5. Mean F1 and F2 frequencies of English vowels produced by the trained group of native Greek speakers with the formant frequencies in Hz converted to the equivalent rectangular bandwidth (ERB) scale (Glasberg and Moore, 1990). The ellipses surround pre-/post-test production of English vowels and have no statistical status.

speech analysis software (Huckvale, 2008). F1 and F2 frequencies were estimated automatically from a linear predictive coding (LPC) analysis with 12 coefficients below 5 kHz and cross-checked from an average fast Fourier transform (FFT) spectrum when the LPC analysis failed to produce reasonable values and are plotted in the vowel space (Fig. 5). Before training, English vowels were arranged into five clusters, suggesting that Greek speakers used their five native categories /i/, /e/, /a/, /o/, and /u/ in their English vowel production.[4] After training, there was much less overlap of English vowels, especially in the high-front area of /iː/ and /ɪ/, the mid-front/central area of /ɜː/ and /e/, and the low area of /æ/, /ɑː/, and /ʌ/, confirming that the trainees learned to differentiate English vowels in their speech production following the perceptual training.

E. Are there correlations in performance across tasks?

To examine whether individual differences in L2 vowel processing are related to differences in L1 vowel processing ability or frequency discrimination acuity, for each participant we calculated an L2 identification boundary (ID BOUND), an L2 identification slope (ID SLOPE), and an L2 discrimination (DISCR) score (by averaging performance over the three English synthetic vowel continua). Similarly, L1 ID BOUND, L1 ID SLOPE, and L1 DISCR scores were calculated by averaging performance over the two Greek synthetic vowel continua. As can be seen in Table IV showing the correlations obtained before and after training, natural English vowel identification across noise conditions correlated, in pre-/post-tests, with L1 DISCR and frequency discrimination. One important difference between pre- and post-tests is that only in the latter was L2 vowel production (combined percent correct identification and goodness ratings by English listeners) correlated with both L1 DISCR and nonspeech discrimination, i.e., the most efficient discriminators were judged to produce more native-like English vowels. Confirming the link between L2 vowel perception and production following perceptual training, contrary to

TABLE IV. Correlations (r) among L1, L2, and nonspeech tasks in the pre-/post-tests for the trained group of Greek speakers using composite scores. For each participant, we calculated an L2 identification boundary (L2 ID BOUND), an L2 identification slope (L2 ID SLOPE), and an L2 discrimination (L2 DISCR) score by averaging performance over the three English synthetic vowel continua. Similarly, L1 ID BOUND, L1 ID SLOPE, and L1 DISCR scores were calculated by averaging performance over the two Greek synthetic vowel continua. Discrimination accuracy is expressed by lower discrimination thresholds and so an inverse correlation between, for example, L2 vowel identification and L1 DISCR indicates that successful discriminators were also better in identifying L2 vowels.

L2 tasks	L1 Natural id noise	L1 ID BOUND	L1 ID SLOPE	L1 DISCR	F2 discrimination
Pre-test					
L2 Natural id quiet	−0.04	−0.17	−0.13	−0.52**	−0.47*
L2 Natural id noise	−0.01	0.21	−0.22	−0.51**	−0.45*
L2 ID BOUND	−0.27	0.11	0.03	0.19	0.26
L2 ID SLOPE	−0.24	−0.23	0.12	−0.21	−0.26
L2 DISCR	−0.31	−0.12	0.32	0.63**	0.73**
L2 production	−0.22	−0.16	0.12	−0.31	−0.15
Post-test					
L2 Natural id quiet	0.08	−0.17	0.11	−0.63**	−0.50*
L2 Natural id noise	0.04	−0.27	0.07	−0.48*	−0.43*
L2 ID BOUND	−0.36	0.22	0.23	0.02	−0.01
L2 ID SLOPE	0.15	−0.09	0.18	−0.30	−0.23
L2 DISCR	−0.08	−0.06	−0.28	0.76**	0.74**
L2 production	−0.01	0.37	0.11	−0.44*	−0.65**

*$p < 0.05$.
**$p < 0.01$.

what was found in the pre-test, L2 vowel production was correlated with both identification of English vowels in quiet, $r = 0.44$, $p < 0.05$ and in noise, $r = 0.65$, $p < 0.01$.

Looking closer at the effect of training for individuals, there was a negative correlation between pre-test English vowel identification averaged over noise conditions and degree of improvement relative to pre-test accuracy, $r = -0.6$, $p < 0.01$, indicating that those individuals who performed poorly in the pre-test improved more than those who performed well in the pre-test, a finding that cannot be attributed to a ceiling effect since the highest score obtained was 81.25% correct. At the same time, a positive correlation between pre- and post-test English vowel identification averaged over noise conditions, $r = 0.52$, $p < 0.05$, demonstrates that individuals who were better before training were also better after training, despite having improved less than those who performed poorly in the pre-test. With respect to English vowel production, there was a negative correlation, $r = -0.52$, $p < 0.05$ between pre-test English vowel production and degree of improvement relative to pre-test performance. However, there was no significant correlation between pre- and post-test English vowel production, $r = 0.28$, $p > 0.05$.

A final issue addressed was whether successful learners could be identified before training. Post-test English vowel identification averaged over noise conditions was correlated with pre-test L1 DISCR, $r = 0.55$, $p < 0.05$, L2 DISCR, $r = 0.56$, $p < 0.05$, and nonspeech discrimination, $r = 0.55$, $p < 0.05$. Similarly, post-test English vowel production accuracy was correlated with pre-test L1 DISCR, $r = 0.52$, $p < 0.05$, L2 DISCR, $r = 0.68$, $p < 0.01$, and nonspeech discrimination, $r = 0.65$, $p < 0.01$. These results demonstrate that the most efficient discriminators at the pre-test stage were also the most accurate both in English vowel perception and production after training.

IV. GENERAL DISCUSSION

This study examined whether individual variability in L2 vowel learning is related with learners' processing of L1 vowels ("L1 phonetic" hypothesis) and/or frequency discrimination acuity ("auditory processing" hypothesis). To this end, native Greek speakers were given a battery of tests assessing L1 and L2 vowel identification and discrimination, and frequency discrimination acuity, both before and after receiving phonetic training for English vowels. The results first replicated the finding that high-variability phonetic training significantly improves by about 20% points the identification of L2 vowels (Nishi and Kewley-Port, 2007, 2008; Iverson and Evans, 2009) and that learning transfers to L2 vowel production (Lambacher *et al.*, 2005) as judged by native English speakers and confirmed by an acoustic analysis of the English vowels produced by Greek speakers. Our findings provide novel information concerning the effect of phonetic training on speech-in-noise L2 perception. As expected, L2 vowel perception was significantly poorer in noise than in quiet (Cutler *et al.*, 2004; Mayo *et al.*, 1997; Iverson and Evans, 2007b) across pre-/post-tests. Still, training in quiet significantly improved perception in noise by about 15% points, adding to the existing evidence for the effectiveness of the high-variability approach to training.

Training improved learners' categorization consistency (steeper slopes) of the English /iː/-/ɪ/ continuum but not for /æ/-/ʌ/. It may be that five sessions of training can only improve categorization of synthetic stimuli to a certain degree (note that post-test /iː/-/ɪ/ categorization consistency was about the same as pre-/post-test /æ/-/ʌ/ consistency, a finding that confirmed the prediction that the latter contrast would suffer less from L1 spectral interference than the former). Even though training improved vowel identification, it did not lead to trainees being able to better discriminate the same contrast; these results are in line with those of Heeren and Schouten (2008) who successfully trained native Dutch speakers in identifying the Finish /t/-/tː/ contrast but found that trainees did not improve in their discrimination of the same contrast. The present work differs from Heeren and Schouten (2008) in terms of both the type of L2 contrasts tested (vowels vs consonants) and of the type of discrimination task used (adaptive vs nonadaptive), and so the replication of the lack of effect of identification-based auditory training on discrimination ability is noteworthy. Further, Iverson and Evans (2009) found that despite improving in English identification accuracy, Spanish and German speakers did not improve in their English vowel space mapping after perceptual training, i.e., their best exemplar locations for English vowels were not closer to the target vowels post-training. This led the authors to conclude that high-variability phonetic training may be more effective than low-variability training because stimulus variability trains listeners to better apply L2 categories to real speech but that it does not actually change category representation. Our results seem to agree with this view: Training improves vowel identification and makes categorization more consistent, i.e., boundaries get sharper but it does not appear to change category representation as the phoneme boundaries do not change and discrimination does not improve.

The main objective of this study was to investigate whether we could relate the effect of training in individuals to their frequency discrimination acuity or to their native (L1) vowel processing. Despite the relatively small number of participants, we found several significant correlations across L1, L2, and nonspeech tasks. Not only were L1, L2, and frequency discrimination correlated (across vowel pairs and pre-/post-tests) but also the composite L1 DISCR, L2 DISCR, and frequency discrimination scores were correlated with identification accuracy for natural L2 vowels in quiet and in noise (across pre-/post-tests). That is, individuals with better frequency discrimination acuity for synthetic vowels in L1 and L2 and nonspeech stimuli were better at identifying natural L2 vowels both before and after training. Our results also showed that natural English vowel identification and English vowel production in the post-test were correlated with L1, L2, and frequency discrimination in the pre-test, suggesting that those individuals who showed lower discrimination thresholds when tested for the first time achieved better scores at English vowel perception and production after perceptual training, thus supporting a perception-production link (Flege, 1999; Flege *et al.*, 1997a, 1999a).

One issue to consider is whether these correlations simply reflect the fact that certain individuals show a particular aptitude at laboratory-based perceptual tests, linked, for example, to good selective attention or paired association learning (Buckner and Wheeler, 2001). This is an issue that affects all perception studies but is particularly sensitive for studies such as that investigate cross-test correlations. Although cognitive tests that provide measures of, for example, selective attention, short-term memory, and attentional switching would have been valuable, as these factors may all be considered to potentially affect the performance on our vowel tests, their correlations with L2 learning have typically been shown to be weak. For example, Aliaga-Garcia *et al.* (2010) found that participants with better phonological short-term memory showed better pre-/post-test scores in an L2 vowel training study, but the effect was weak. Hazan and Kim (2010) obtained correlations (at $p < 0.05$ level) between measures of attentional switching and paired association learning and phonetic ability in an L2 consonant learning task, but again the effects were not strong.

The results appear to showcase an underlying auditory acuity component for L2 speech processing and support the "auditory processing" hypothesis over the "L1 phonetic" hypothesis. There was no evidence that individuals with less robust L1 categories, as shown by shallower categorization slopes and poorer vowel discrimination, were more flexible in terms of learning new categories; in other words, there was no evidence of an inverse correlation between L1 ID SLOPE and natural English vowel perception or L2 ID SLOPE as the "L1 phonetic" hypothesis would predict. Previous studies have failed to find a connection between speech and nonspeech processing. However, as noted in Surprenant and Watson (2001), speech and nonspeech are typically measured using tasks that tap into different processing abilities; speech ability is measured via recognition-in-noise tasks whereas nonspeech ability is measured using discrimination tasks that require analytic listening. The authors proposed that nonspeech tasks that require more global listening may be more appropriate for the prediction of individual variability in speech perception. Rather than using nonspeech tasks requiring more global listening, this study employed more analytical speech tasks and found a connection between nonspeech processing and vowel processing both in L1 and L2.

These findings do not reject, of course, the role of L1 interference in L2 speech learning as has been acknowledged by current cross-language models. At a group level, there was a clear effect of L1 experience on L2 vowel perception and production; Greek speakers showed much shallower identification slopes and lower discrimination accuracy for synthetic English vowels and had difficulties in identifying natural English vowels in quiet and in noise and in English vowel production. The results of this study show that, while L1 experience affects L2 processing, some individuals are better in using spectral/acoustic information to overcome L1 biases.

ACKNOWLEDGMENTS

This study was funded by a doctoral studentship to the first author by the A. G. Leventis Foundation. We are grateful to Paul Iverson and Bronwen Evans for providing the software and materials used in the best exemplars and the training experiment and to Stuart Rosen and Souhila Messaoud-Galusi for their help with the software used to run the tasks with synthetic speech. We also thank Paul Iverson for his constructive comments on a previous draft of this paper.

[1]The level of noise was determined after running a short pilot test with four native Greek speakers who had just moved to London to study and whose level of English experience was thus comparable to that of the speakers that would be tested in Greece. After trying different SNRs (from -2 to -6 dB), a SNR of -4 dB was chosen that yielded percent correct identification accuracy of about 40%.

[2]The level of noise was determined after running a short pilot test with five native Greek speakers where different SNRs were tried (from -4 to -12 dB). At a SNR of -10 dB percent correct identification accuracy was about 75%. That was a different level of accuracy to that selected for English. However, the level of noise required to obtain an intelligibility level of 40% in Greek would be so high that it would increase the possibility of the two tasks tapping into different processing abilities.

[3]Although the nonspeech continuum differed from the speech continua in that only F2 frequency changed, it was decided to include it in the analysis.

[4]To confirm that Greek speakers imposed, at least with respect to spectral distinctions, their five-vowel system on English vowel production, separate two-way repeated-measures ANOVAs were run on F1 and F2 frequencies with Group (trained, control) as a between-subject factor and Vowel (ten vowels) as a within-subject factor. For F1, the ANOVA yielded a significant effect of Vowel, $F(9,234) = 221.6$, $p < 0.001$ and no effect of Group or Vowel $\times$ Group interaction, $p > 0.05$. Pairwise comparisons (Bonferroni adjusted) showed the following pattern: /iː/, /ɪ/ and /uː/ < /e/, /ɜː/, /ɒ/ and /ɔː/ < /æ/, /ʌ/ and /ɑː/. For F2, the ANOVA yielded a significant effect of Vowel, $F(9,234) = 217.1$, $p < 0.001$ and no effect of Group or Vowel $\times$ Group interaction, $p > 0.05$. Pairwise comparisons (Bonferroni adjusted) showed the following pattern: /iː/ and /ɪ/ > /e/ and /ɜː/ > /æ/, /ʌ/ and /ɑː/ > /ɒ/, /ɔː/ and /uː/.

Aliaga-Garcia, C., Mora, J. C., and Cerviño-Povedano, E. (**2010**). "Phonological short-term memory and L2 speech learning in adulthood," in *Proceedings of the 6th International Symposium on the Acquisition of Second Language Speech, New Sounds 2010*, Poznań, Poland, pp. 19–24.

Arvaniti, A. (**2007**). "Greek phonetics: The state of the Art," J. Greek Linguist. **8**, 97–208.

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

Best, C. T., and Tyler, M. D. (**2007**). "Nonnative and second-language speech perception: Commonalities and complementarities," in *Language Experience in Second Language Speech Learning: In Honor of James Flege*, edited by M. Munro and O.-S. Bohn (John Benjamins, Amsterdam), pp. 13–34.

Bongaerts, T., Van Summeren, C., Planken, B., and Schils, E. (**1997**). "Age and ultimate attainment in the pronunciation of a foreign language," Stud. Second Lang. Acquis. **19**, 447–465.

Botinis, A., Fourakis, M., and Prinou, I. (**2000**). "Acoustic structure of the Greek stop consonants," Glossologia **11–12**, 167–199.

Bradlow, A. R. (**2008**). "Training non-native language sound patterns: Lessons from training Japanese adults on the English /r/-/l/ contrast," in *Phonology and Second Language Acquisition*, edited by J. G. Hansen Edwards and M. L. Zampini (John Benjamins, Amsterdam), pp. 287–308.

Bradlow, A. R., Akahane-Yamada, R., Pisoni, D. B., and Tohkura, Y. (**1999**). "Training Japanese listeners to identify English /r/ and /l/: Long-term retention of learning in perception and production," Percept. Psychophys. **61**, 977–985.

Bradlow, A. R., Pisoni, D. B., Akahane-Yamada, R., and Tohkura, Y. (**1997**). "Training Japanese listeners to identify English /r/ and /l/: IV. Some effects of perceptual learning on speech production," J. Acoust. Soc. Am. **101**, 2299–2310.

Buckner, R., and Wheeler, M. (**2001**). "The cognitive neuroscience of remembering," Nat. Rev. Neurosci. **2**, 624–634.

Carroll, J. B. (**1993**). "Abilities in the domain of auditory reception," in *Human Cognitive Abilities: A Survey of Factor-Analytic Studies* (Cambridge University Press, New York), pp. 364–393.

Cebrian, J. (**2006**). "Experience and the use of duration in the categorization of L2 vowels," J. Phonetics **34**, 372–387.

Cutler, A., Andrea, W. A., Smits, R., and Cooper, N. (**2004**). "Patterns of English phoneme confusions by native and non-native listeners," J. Acoust. Soc. Am. **116**, 3668–3678.

Diaz, B., Baus, C., Escera, C., Costa, A., and Sebastian-Galles, N. (**2008**). "Brain potentials to native phoneme discrimination reveal the origin of individual differences in learning the sounds of a second language," Proc. Natl. Acad. Sci. U.S.A. **105**(42), 16083–16088.

Docherty, G. (**1992**). *The Timing of Voicing in British English Obstruents* (Foris, Berlin), pp. 1–227.

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

Flege, J. E. (**1999**). "The relation between L2 production and perception," in *Proceedings of the 14th International Congress of Phonetic Sciences*, San Francisco, pp. 1273–1276.

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), pp. 319–355.

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

Flege, J. E., Frieda, E. M., and Nozawa, T. (**1997b**). "Amount of native-language (L1) use affects the pronunciation of an L2," J. Phonetics **25**, 169–186.

Flege, J. E., and Liu, S. (**2001**). "The effect of experience on adults' acquisition of a second language," Stud. Second Lang. Acquis. **23**, 527–552.

Flege, J. E., and Mackay, I. R. A. (**2004**). "Perceiving vowels in a second language," Stud. Second Lang. Acquis. **26**, 1–34.

Flege, J. E., Mackay, I. R. A., and Meador, D. (**1999a**). "Native Italian speakers' production and perception of English vowels," J. Acoust. Soc. Am. **106**, 2973–2987.

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.

Flege, J. E., Yeni-Komshian, G. H., and Liu, H.-M. (**1999b**). "Age constraints on second language acquisition," J. Mem. Lang. **41**, 78–104.

Gerrits, E., and Schouten, M. E. H. (**2004**). "Categorical perception depends on the discrimination task," Percept. Psychophys. **66**, 363–376.

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

Golestani, N., Pause, T., and Zatorre, R. J. (**2002**). "Anatomical correlates of learning novel speech sound," Neuron **35**, 997–1010.

Golestani, N., and Zatorre, R. J. (**2009**). "Individual differences in the acquisition of second language phonology," Brain Lang. **109**, 55–67.

Golestani, N., Molko, N., Dehaene, S., LeBihan, D., and Pallier, C. (**2007**). "Brain structure predicts the learning of foreign speech sounds," Cereb. Cortex **17**, 575–582.

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.

Hazan, V., and Kim, Y. H. (**2010**). "Can we predict who will benefit from computer-based phonetic training?" in *Online Proceedings of the INTERSPEECH 2010 Satellite Workshop on Second Language Studies: Acquisition, Learning, Education and Technology (L2WS 2010)*, Tokyo, Japan. Available at http://www.gavo.t.u-tokyo.ac.jp/L2WS2010/papers/L2WS2010_P2-06.pdf

Hazan, V., and Rosen, S. (**1991**). "Individual variability in the perception of cues to place contrasts in initial stops," Percept. Psychophys. **49**, 187–200.

Hazan, V., Sennema, A., Faulkner, A., Ortega-Llebaria, M., Iba, M., and Chung, H. (**2006**). "The use of visual cues in the perception of non-native consonant contrasts," J. Acoust. Soc. Am. **119**, 1740–1751.

Hazan, V., Sennema, A., Iba, M., and Faulkner, A. (**2005**). "Effect of audiovisual perceptual training on the perception and production of consonants by Japanese learners of English," Speech Commun. **47**, 360–378.

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.

Huckvale, M. (**2008**). "Speech filing system tools for speech research (SFS)," from http://www.phon.ucl.ac.uk/resource/sfs (Last viewed March 2008).

Iverson, P., and Evans, B. G. (**2007a**). "Auditory training of English vowels for first language speakers of Spanish and German," in *Proceedings of the 16th International Congress of Phonetic Sciences*, Saarbrucken, Germany, pp. 1625–1628.

Iverson, P., and Evans, B. G. (**2007b**). "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 Evans, B. G. (**2009**). "Learning English vowels with different first-language vowel systems II: Auditory training for native Spanish and German speakers," J. Acoust. Soc. Am. **126**, 866–877.

Iverson, P., Kuhl, P. K., Akahane-Yamada, R., Diesch, E., Tohkura, Y., Kettermann, A., and Siebert, C. (**2003**). "A perceptual interference account of acquisition difficulties for non-native phonemes," Cognition **87**, 47–57.

Iverson, P., Ekanayake, D., Hamann, S., Sennema, A., and Evans, B. G. (**2008**). "Category and perceptual interference in second-language phoneme learning: An examination of English /w/-/v/ learning by Sinhala, German, and Dutch speakers," J. Exp. Psychol. Hum. Percept. Perform. **34**, 1305–1316.

Jilka, M. (**2009**). "Talent and proficiency in language," in *Language Talent and Brain Activity*, edited by G. Dogil and S. Reiterer (Mouton De Gruyter, Berlin), pp. 1–16.

Johnson, D. M., Watson, C. S., and Jensen, J. K. (**1987**). "Individual differences in auditory capabilities," J. Acoust. Soc. Am. **81**, 427–438.

Kewley-Port, D. (**2001**). "Vowel formant discrimination II: Effects of stimulus uncertainty, consonantal context and training," J. Acoust. Soc. Am. **110**, 2141–2155.

Kidd, G. R., Watson, C. S., and Gygi, B. (**2007**). "Individual differences in auditory abilities," J. Acoust. Soc. Am. **122**, 418–435.

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.

Kuhl, P. K. (**2000**). "A new view of language acquisition," Proc. Natl. Acad. Sci. U.S.A. **97**, 11850–11857.

Kuhl, P. K., Conboy, B. T., Coffey-Corina, S., Padden, D., Rivera-Gaxiola, M., and Nelson, T. (**2008**). "Phonetic learning as a pathway to language: New data and native language magnet theory expanded (NLM-e)," Philos. Trans. R. Soc. B **363**, 979–1000.

Lambacher, S. G., Martens, W. L., Kakehi, K., Marasinghe, C. A., 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.

Lee, J., Perrachione, T., Dees, T., and Wong, P. C. M. (**2007**). "Differential effects of stimulus variability and learners' pre-existing pitch perception ability in lexical tone learning by native English speakers," in *Proceedings of the 16th International Congress of Phonetic Sciences*, Saarbrucken, Germany, pp. 1589–1592.

Lengeris, A. (**2009**). "Perceptual assimilation and L2 learning: Evidence from the perception of Southern British English vowels by native speakers of Greek and Japanese," Phonetica **66**, 169–187.

Levitt, H. (**1971**). "Transformed up-down methods in psychoacoustics," J. Acoust. Soc. Am. **49**, 467–477.

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.

Lively, S. E., Pisoni, D. B., Yamada, R. A., Tohkura, Y., and Yamada, T. (**1994**). "Training Japanese listeners to identify English /r/ and /l/. III. Long-term retention of new phonetic categories," J. Acoust. Soc. Am. **96**, 2076–2087.

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.

Mayo, H. L., Florentine, M., and Buus, S. (**1997**). "Age of second-language acquisition and perception of speech in noise," J. Speech Lang. Hear. Res. **40**, 686–693.

Moyer, A. (**1999**). "Ultimate attainment in L2 phonology," Stud. Second Lang. Acquis. **21**, 81–108.

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.

Nishi, K., and Kewley-Port, D. (**2008**). "Nonnative speech perception training using vowel subsets: Effects of vowels in sets and order of training," J. Speech Lang. Hear. Res. **51**, 1480–1493.

Polka, L. (**1991**). "Cross-language speech perception in adults: Phonemic, phonetic, and acoustic contributions," J. Acoust. Soc. Am. **89**, 2961–2977.

Pruit, J. S., Jenkins, J. J., and Strange, W. (**2006**) "Training the perception of Hindi dental and retroflex stops by native speakers of American English and Japanese," J. Acoust. Soc. Am. **119**, 1684–1696.

Surprenant, A. M., and Watson, C. S. (**2001**). "Individual differences in the processing of speech and nonspeech sounds by normal-hearing listeners," J. Acoust. Soc. Am. **110**, 2085–2095.

Walley, A. C., and Flege, J. (**1999**). "Effect of lexical status on children's and adults' perception of native and non-native sounds," J. Phonetics **27**, 307–332.

Wong, P. C. M., and Perrachione, T. K. (**2007**). "Learning pitch patterns in lexical identification by native English-speaking adults," Appl. Psycholinguist. **28**, 565–585.