Original Paper

Phonetica 2009;66:169–187
DOI: 10.1159/000235659

Received: December 22, 2008
Accepted: July 16, 2009

Perceptual Assimilation and L2 Learning: Evidence from the Perception of Southern British English Vowels by Native Speakers of Greek and Japanese

Angelos Lengeris

Speech Hearing and Phonetic Sciences, UCL, London, UK

Abstract

This study examined the extent to which previous experience with duration in first language (L1) vowel distinctions affects the use of duration when perceiving vowels in a second language (L2). Native speakers of Greek (where duration is not used to differentiate vowels) and Japanese (where vowels are distinguished by duration) first identified and rated the eleven English monophthongs, embedded in /bVb/ and /bVp/ contexts, in terms of their L1 categories and then carried out discrimination tests on those English vowels. The results demonstrated that both L2 groups were sensitive to durational cues when perceiving the English vowels. However, listeners were found to *temporally* assimilate L2 vowels to L1 category/categories. Temporal information was available in discrimination only when the listeners' L1 duration category/categories did not interfere with the target duration categories and hence the use of duration in such cases cannot be attributed to its perceptual salience as has been proposed.

1. Introduction

When perceiving speech, listeners do not weight all aspects of the acoustic signal in the same manner, a process known as *cue-weighting*. For example, while both spectral and durational cues distinguish English /iː/ from /ɪ/, native speakers of English primarily use the former when perceiving the tense-lax /iː/-/ɪ/ contrast [e.g. Hillenbrand et al., 2000]. Although second language (L2) learners are often found to weight the acoustic cues of the target language differently than the native speakers of that language do [e.g. Bohn, 1995; Flege et al., 1997; Iverson et al., 2003], an issue that requires further exploration is the extent to which experience or not with an acoustic cue in their first language (L1) affects the availability of that cue in L2 perception. The present

KARGER

Fax +41 61 306 12 34
E-Mail karger@karger.ch
www.karger.com

0031–8388/09/0663–0169
$26.00/0
Accessible online at:
www.karger.com/pho

Angelos Lengeris
UCL Department of Speech Hearing and Phonetic Sciences
UCL, Chandler House, 2, Wakefield Street
London WC1N 1PF (UK)
Tel. +44 20 7679 4041, Fax +44 207 679 4010,
E-Mail a.lengeris@uclmail.net

study addresses this issue by examining the use of durational cues in the perception of English vowels by native speakers of two languages that employ five relatively similar vowel quality distinctions but differ greatly with regard to the use of duration in contrasting vowels, i.e. Greek and Japanese.

Adults often have difficulties acquiring the vowel system of an L2 [Polka, 1995; Flege et al., 1999; Flege and MacKay, 2004 among others]. Linguistic experience alters perception so that infants born with the ability to discriminate the sounds of any language [Eimas et al., 1971] show, after 6 months of age, reduced sensitivity to non-native vowel contrasts [Kuhl et al., 1992; Polka and Werker, 1994]. Recent work suggests that infants acquire their L1 categories through distribution-based learning [Maye et al., 2002; Maye and Weiss, 2003], a process that sharpens L1 perception but unavoidably *interferes* with L2 learning [Iverson et al., 2003; Kuhl et al., 2006, 2008]. Current models of cross-language speech perception such as the Perceptual Assimilation Model [PAM: Best et al., 1988; Best, 1995] and the Speech Learning Model [SLM: Flege, 1995] aim at modelling the difficulty non-native sounds will pose to the learner based on how the sounds in question perceptually assimilate to the learner's L1 categories [but see Iverson and Evans, 2007, for a different approach that de-emphasizes the role of assimilation in L2 learning].

PAM was originally proposed to account for naïve listeners' perception of *non-native* sounds and has recently been applied to L2 learning [Guion et al., 2000; Best and Tyler, 2007]. According to the model, discrimination of a non-native contrast will be excellent when each sound is assimilated, via detection of cross-language gestural similarities/dissimilarities, to a different L1 category (a Two-Category contrast), while discrimination will be problematic when both sounds are assimilated to the same L1 category equally well or poorly (a Single-Category contrast). When one sound is a better match than the other, discrimination will be moderate (a Category Goodness contrast). When one sound is identified with an L1 category and the other falls between two L1 categories discrimination will be very good (an Uncategorized-Categorized contrast), while in cases where neither sound matches with an L1 category discrimination will vary from poor to good according to how similar the non-native segments are to each other and to native categories (an Uncategorized-Uncategorized contrast).

SLM is concerned with *L2 learning* and particularly with experienced L2 learners. It posits that speech-learning mechanisms remain intact across the life span. The advantage of early over late L2 learners [e.g. Flege et al., 1999; Flege and MacKay, 2004] is attributed to the fact that as the L1 categories develop with age [Hazan and Barrett, 2000] they become more powerful attractors of L2 categories [Flege et al., 2003]. L2 categories are initially classified in terms of L1 categories ('equivalence qualification') based on the perceived phonetic similarity/dissimilarity between the L1 and L2 categories; the learner has thus to detect phonetic differences between the L2 and the closest L1 sound in order to form a new category. SLM differs from PAM in that it does not specify the nature of the cross-language perceived similarity/dissimilarity. Also, while PAM provides discrimination predictions for pairs of non-native sounds, SLM predicts the difficulty listeners will face when learning individual L2 sounds.

One of the difficulties an L2 learner may be faced with when perceiving an L2 contrast is to correctly attend to the acoustic cues that are informative in signalling that contrast. That difficulty is clearly shown in Japanese listeners' perception of the English /r/-/l/ contrast [e.g. Goto, 1971]. Iverson et al. [2003] showed that Japanese listeners pay attention to the non-critical second formant frequency when trying to distinguish English /r/ from /l/ instead of focusing on the third formant onset frequency that is critical to native speakers of English, the reason being that they mainly use the former when perceiving the single Japanese sound perceptually related to English /r/ and /l/, namely /ɾ/. In a vowel study, McAllister et al. [2002] tested the hypothesis that category formation is difficult when based on a phonetic feature not used contrastively in L1 ('feature hypothesis'). The hypothesis, implied in SLM, would predict that success in using durational cues when learning L2 vowels will be related to previous experience with duration in L1 vowel distinctions. The authors compared the perception and production of the Swedish vowel length contrasts by native speakers of Estonian, American English and Latin American Spanish. The results in perception showed that the Estonian participants, who are extremely experienced with duration distinctions in their L1, outperformed the American English participants, who normally use duration as a secondary cue in L1 vowel distinctions, who in turn outperformed the Spanish participants, who do not use duration at all in perceiving L1 vowels. The cross-language differences in the production of Swedish vowels were fewer than the perception differences with the Spanish participants still being consistently less successful in producing the vowels than any other group. These results were seen as confirming the importance of L1 transfer when learning an L2.

However, other studies have demonstrated that listeners remain sensitive to novel acoustic features when perceiving L2 vowels. Bohn [1995] examined the perception of American English vowels by native speakers of German, Spanish and Mandarin. The stimuli were synthetic vowel continua (/e/-/æ/ and /iː/-/ɪ/) that tested the listeners' reliance on spectral and durational cues. Bohn found that duration was predominantly utilized in L2 perception not only by native German speakers, who make use of both spectral and durational cues in L1 to distinguish vowels, but also by native Spanish and Mandarin speakers, neither of which group uses duration in contrasting L1 vowels. To explain this finding, Bohn proposed a desensitization hypothesis. The hypothesis states that, when spectral information is not available (hence the term 'desensitization'), L2 learners will use durational information irrespective of whether duration is used in their L1, as duration is a cue that is acoustically salient and easy to access. Spanish listeners' overreliance on durational cues to differentiate the English /iː/-/ɪ/ contrast has been reported in several subsequent studies [Flege et al., 1997; Morrison, 2002; Escudero and Boersma, 2004; Escudero, 2005; Cebrian, 2006]. Similar results have been reported for native speakers of Korean, who are also inexperienced with the duration feature in L1 vowel distinctions [Flege et al., 1997]. Escudero and Boersma [2004] offered a different explanation for Spanish speakers' preference for durational cues when distinguishing English /iː/ from /ɪ/. They proposed that since Spanish does not employ duration contrastively to signal vowel contrasts, it is easier for Spanish listeners to create a new category (duration) than splitting their already existing (spectral) Spanish /i/ category. Finally, Iverson and Evans [2007] also found that L2 learners make use of intrinsic formant movement and duration when perceiving L2 vowels irrespective of L1 background. Their results comparing Spanish, French, German and Norwegian listeners' perception of the Southern British English vowel system suggest that L2 learning shows a high degree of uniformity in the use of secondary acoustic cues.

The main goal of the present study was to further investigate the use of duration as a perceptual cue in L2 vowel learning. More specifically, it aimed at evaluating the main competing hypotheses in the literature, namely the feature hypothesis [McAllister et al., 2002] and the desensitization hypothesis [Bohn, 1995]. For that purpose, the perception of English vowels by native speakers of Greek and Japanese was examined. Greek and Japanese have similar five-vowel quality systems but differ fundamentally with regard to the use of duration in signalling vowel contrasts. The Modern Greek vowel system contains five vowels /i, e, a, o, u/ and employs no tense-lax or long-short distinctions [for a comprehensive review of studies on Greek vowels, see Arvaniti, 2007]. The Japanese vowel system contains five short (one-mora) and five long (two-morae) vowels /i, e, a, o, u/ and /iː, eː, aː, oː, uː/, respectively (in Standard Japanese the high back vowel is unrounded /ɯ/) [see for example Shibatani, 1990]. The short and the long vowels are almost identical in terms of spectral characteristics with the former being approximately 50% shorter [Shibatani, 1990; Hirata, 2004]. Greek and Japanese are also similar in terms of the simplicity of their syllable structure. In Greek, that takes the form of $C_{(0–3)}VC_{(0–1)}$. Open syllables are much more common than closed ones and the consonants in word-final position are limited to /s/ and /n/. In Japanese, the syllable structure takes the form of $C_{(0–1)}V$, but there is also the possibility of /CVn/ and /Vn/ syllables (with /n/ being a separate mora). The Southern British English vowel system, the target system in this study, is more complex than either the Greek or the Japanese one. It includes eleven monophthongs that can take stress /iː, ɪ, e, æ, ʌ, aː, ɜː, ɒ, ɔː, ʊ, uː/ with some vowels being inherently longer than others [e.g. Giegerich, 1992]. Contrary to what happens in either Greek or Japanese, vowels in all varieties of English are longer before voiced than before voiceless consonants [e.g. Peterson and Lehiste, 1960; House, 1961 for American English; Giegerich, 1992 for Southern British English].

Rather than asking participants to identify English vowels from a synthetic continuum varying in durational and spectral cues, which is a common technique in the L2 perception literature [e.g. Bohn, 1995; Flege et al., 1997; Cebrian, 2006], this study employed two perceptual tasks using natural English vowels: (1) a cross-language perceptual assimilation task and (2) a categorial oddity discrimination task. The use of these two tasks also allowed testing of whether cross-language assimilation patterns predicted discrimination of English vowels as proposed by Best's PAM. All eleven English monophthongs were used as perceptual stimuli with the goal of obtaining representative data on how Greek and Japanese listeners perceive the entire English vowel space. The vowels were placed in two contexts, namely /bVb/ and /bVp/, to evaluate how the shortening of the vowels before a voiceless stop consonant would affect their perception.

2. Recording of English, Greek and Japanese Vowels

Native speakers of English, Greek and Japanese recorded productions of their L1 vowels. English speakers' vowel productions were used as perceptual stimuli. Greek and Japanese speakers' vowel productions where used in order to explain the results concerning the cross-language assimilation patterns (experiment 1) as well as the L2 listeners' discrimination performance (experiment 2).

2.1. English Vowel Stimuli

Three native speakers of Southern British English, all female (mean = 26.4 years, range = 23–30 years), recorded the 11 English monophthongs (/iː, ɪ, e, æ, ʌ, aː, ɜː, ɒ, ɔː, ʊ, uː/) in the sentence *I read ___ on the screen* (speakers were instructed to use the present tense of *read* /riːd/). The vowels were uttered in two consonantal contexts, /bVb/ and a /bVp/, with all the /bVb/ tokens recorded before the

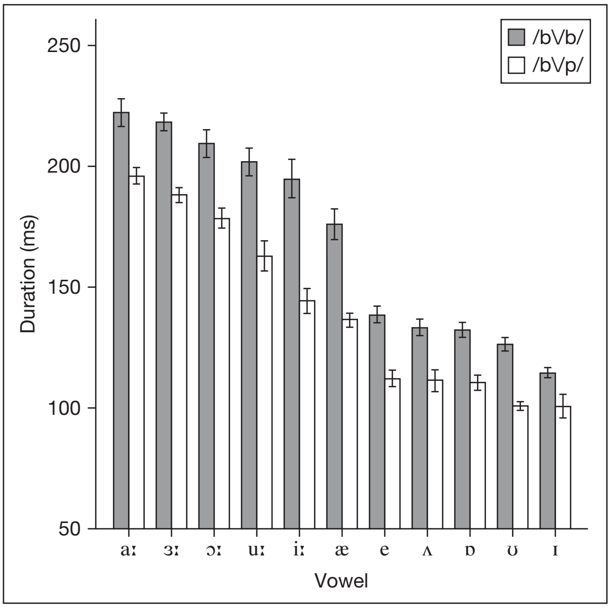

Fig. 1. Mean durations (ms) of the English vowel stimuli used in the study. Error bars represent standard errors of the mean.

/bVp/ tokens. The speakers read each vowel 4 consecutive times at a normal speaking rate, giving a total of 264 tokens (3 speakers × 11 vowels × 2 contexts × 4 repetitions). The recordings took place in an anechoic chamber at the University College London with a sampling rate of 44.1 kHz, using a Sony 60ES DAT recorder with a B&K Sound Level Meter Type 2231 fitted with a 4165 microphone cartridge. The author and a very experienced phonetician, a native speaker of Southern British English, chose the best three tokens for each English vowel (in almost all cases these were the first three tokens). The final number of stimuli was 198 (3 speakers × 11 vowels × 2 contexts × 3 repetitions).

Duration and first (F1) and second (F2) formant frequencies were measured for each vowel. All measurements were made manually using the SFS speech analysis software [Huckvale, 2008]. Duration was measured from spectrograms, from the onset to the offset of periodic energy in F2. F1and F2 frequencies were measured by placing the cursor at the centre of the relatively steady-state region of each vowel. Spectral peaks were then estimated from a 12-pole autocorrelation LPC analysis with a 50-ms rectangular window, and the selection of peaks corresponding to F1 and F2 were verified by visual examination of the spectrogram and an average FFT spectrum of the interval (multiple 50-ms Hamming windowed sections overlapping by 25 ms). The process was also checked by moving the cursor by small amounts to ensure that the peak frequencies were not strongly influenced by selection of a specific time interval. The decision to perform all acoustic analyses manually was made because Greek and Japanese speakers' vowels were between nasal consonants (see section 2.2) thus making the automatic estimations of duration and formant frequencies less reliable. In order to be consistent across languages it was therefore decided to measure English vowels manually too. Mean vowel durations, averaged across speakers and repetitions, are displayed in figure 1 and mean F1 and F2 frequencies, averaged across speakers and repetitions, are displayed in figure 2a (see also table A1 in 'Appendix' for mean F1 and F2 frequencies and standard deviations for all English vowels in two consonantal contexts). A visual inspection of the two figures indicates that the duration of English vowels is clearly affected by consonantal context whereas their F1 and F2 frequencies are very similar across contexts. Vowel durations were submitted to a two-way ANOVA with Vowel (11 levels) and Context (2 levels) as factors. The ANOVA yielded a significant main effect of Vowel [$F(10, 176) = 149.5$; $p < 0.001$], which confirmed that English vowels differ in intrinsic duration, and a significant main effect of Context [$F(1, 176) = 243.7$; $p < 0.001$], which confirmed that English vowels are shorter in /bVp/

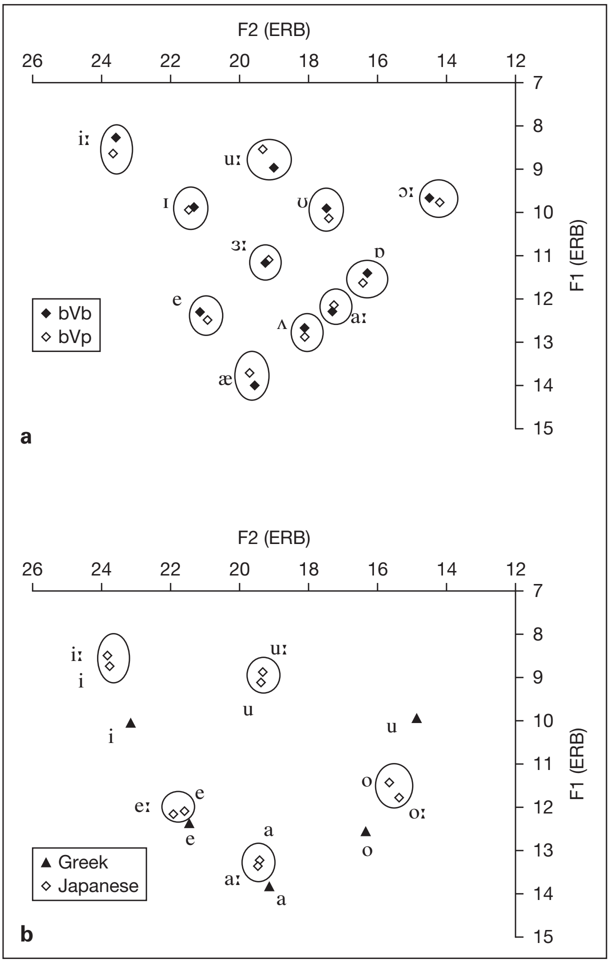

Fig. 2. Mean frequencies (ERB) of the F1 and the F2 of each of the eleven English vowels in /bVb/ and /bVp/ context (**a**) and of each of the five Greek vowels and the ten Japanese vowels in /mVn/ context (**b**). The ellipses surrounding English and Japanese vowels are for illustration purposes only and have no statistical status.

than in /bVb/ context (mean = 142 ms vs. mean = 171 ms). The ANOVA also yielded a significant Vowel × Context interaction [$F(10, 176) = 2.6$; $p < 0.05$], which indicated that some vowels were shorter than others when placed in a /bVp/ context [Giegerich, 1992]. Table 1 (section 3.2) presents the mean duration of each English vowel in both consonantal contexts averaged across speakers and repetitions (standard deviations in parentheses).

2.2. Greek and Japanese Vowels

Three native speakers of Greek (mean = 27 years, range = 26–28 years) recorded their L1 vowels /i, e, a, o, u/ in the sentence [ðʝaˈvazo ___ stin oˈθoni] (‘I read ___ on the screen’). Three native speakers of Japanese (mean = 29.5 years, range = 28–31 years) recorded their L1 vowels /i, iː, e, eː, a, aː, o, oː, u, uː/ in the sentence [gaˈmen ni ___ to ariˈmasɯ] (also ‘I read ___ on the screen’). All 6 speakers were female. In both languages, vowels were uttered in a /mVn/ context. That differed from the /bVb/ and /bVp/ contexts used to elicit the English perceptual stimuli, however, the use of an identical syllable structure (although /n/ would be a separate mora in Japanese) that would be phonologically permissible in both languages was preferred over matching for context across L1 and L2. The speakers read each vowel four times at a normal speaking rate giving 60 tokens for Greek (3 speakers × 5 vowels ×

4 repetitions) and 120 tokens for Japanese (3 speakers × 10 vowels × 4 repetitions). Recordings were made using a digital recorder (MicroTrack 24/96) in a quiet room at a sampling rate of 44.1 kHz. The first three repetitions for each vowel were selected for acoustic analysis (45 vowel tokens for Greek and 90 vowel tokens for Japanese, giving a total of 135 vowel tokens). The author, a native speaker of Greek, and a Japanese speaker judged whether the Greek and Japanese speakers, respectively, had correctly produced the tokens. Vowel duration was measured from spectrograms, taking as vowel onset and offset points the clearly visible changes in the amplitude of upper formants. F1 and F2 were also measured for each vowel as described in the previous section. Mean vowel durations and standard deviations are given in table 1 (section 3.2), averaged across speakers and repetitions. Mean F1 and F2 frequencies are displayed in figure 2b (see also tables A2, A3 in 'Appendix' for mean F1 and F2 frequencies and standard deviations for Greek and Japanese vowels, respectively). As can be seen in figure 2b, although both Greek and Japanese have five vowel qualities /i, e, a, o, u/ in their systems, most of those vowels are phonetically realized differently across languages. The most noticeable difference is, unsurprisingly, that between Greek and Japanese /u/; further, Japanese /i/ is located at a much higher and fronted position than Greek /i/; finally, Japanese /o/ is located between Greek /o/ and /u/ and Japanese /a/ is slightly higher in the vowel space than Greek /a/.

3. Experiment 1: Cross-Language Perceptual Assimilation Task

The purpose of this experiment was to assess how Greek and Japanese listeners classed the English vowels in terms of their native vowels, referred to henceforth as perceptual assimilation. All subjects performed a cross-language identification task with goodness ratings [Best, 1995; Schmidt, 1996; Flege et al., 1997; Cebrian, 2006]. If L2 learners have access only to cues used contrastively in their L1, as proposed by the feature hypothesis, only Japanese listeners' perceptual assimilation of English vowels to their L1 vowel categories should be affected by the context of the vowel stimuli (/bVb/ or /bVp/); context-induced differences in the perceptual assimilation of English vowels to L1 categories might be expressed as changes in percentage of cross-language identification (i.e. an English vowel might assimilate more consistently to an L1 vowel category) and/or similarity ratings (i.e. an English vowel might sound as a better match to the selected L1 category; see section 3.1.3 for details). If on the other hand, the use of duration is a language-independent perceptual strategy based on the salience of duration, as proposed by the desensitization hypothesis, both Greek and Japanese listeners should be affected in how they perceptually assimilate the English vowels to their L1 categories.

3.1. Method

3.1.1. Participants. Thirty-three adult learners of English, all university students, were tested. Eighteen were native speakers of Greek and 15 were native speakers of Japanese. Greek speakers (mean = 23.3 years, range = 18–25 years) were all from Athens, spoke Standard Modern Greek and were tested in Greece. Japanese speakers (mean = 21.5 years, range = 18–24 years) were from Tokyo and surrounding areas, spoke Standard Tokyo Japanese and were tested in London during their 2-week stay for the UCL Summer Course in English Phonetics. The participants had received formal English instruction by L1-accented language instructors in Greece and Japan, respectively, for 10–15 years. Their class level was rather high and relatively uniform across individuals and language groups (e.g. 6.5–7.5 in IELTS, 550–590 in TOEFL, Cambridge FCE), but they had very little, if any, interaction with native speakers of English and none had spent a period of more than 1 month in an English-speaking environment as shown in a language questionnaire completed by the participants before testing. Given the nature of the instruction the participants had received, they can better be defined as *naive* listeners as compared to *experienced* learners [see Best and Tyler, 2007 for an excellent comparison of the characteristics of the two types of listeners as examined in speech perception literature]. All of the listeners that were tested reported normal hearing and no language impairments.

3.1.2. Stimuli. The eleven English vowels (in /bVb/ and /bVp/ contexts) described in section 2.1 were used as perceptual stimuli.

3.1.3. Procedure. Participants were tested individually in quiet rooms using a laptop computer. They were presented the 99 English /bVb/ tokens and the 99 /bVp/ tokens at a comfortable intensity level over high-quality headphones and completed two tasks: a forced-choice cross-language identification task, and a goodness-rating task. They first heard an English token and identified which of their L1 vowel categories sounded closest to that token by clicking on a label on a screen. For Greek listeners, the labels were given in Greek orthography: ‘Ι’ /i/, ‘Ε’ /e/, ‘Α’ /a/, ‘Ο’ /o/, ‘ΟΥ’ /u/. For Japanese listeners, the Roman alphabet was used: ‘I’ /i/, ‘II’ /iː/, ‘E’ /e/, ‘EE’ /eː /, ‘A’ /a/, ‘AA’ /aː/, ‘O’ /o/ ‘OO’ /oː/, ‘U’ /u/, ‘UU’ /uː/. Although it was expected that the Japanese participants would be familiar with the Roman alphabet and its correspondence to Japanese short/long vowel categories, the author made sure before testing that all the Japanese participants understood the labels used. The task was self-paced and once a response was given, the same token was played again and the participants rated its goodness-of-fit to the chosen L1 vowel category using a scale from 1 (totally different) to 7 (identical). The 198 stimuli were blocked by voiced-voiceless coda context, with order of context counterbalanced across listeners and were fully randomized within voiced/voiceless context. Before the test began, a 33-trial practice session (3 speakers × 11 vowels) was presented to familiarize listeners with the procedure. Consonantal context in the practice session was different from the context to be tested first. Written instructions were given in the L1 of each language group.

3.2. Results

The frequency with which an L1 category was selected by the listeners to classify each English vowel was converted to a percentage of total presentations and the mean goodness rating that vowel received as an example of an L1 category was estimated. Mean percentage classification and mean goodness rating were combined into a single metric unit (i.e. the two numbers were multiplied) expressing a ‘fit index’ of each English vowel to an L1 vowel category [Halle et al., 1999; Guion et al., 2000; Iverson and Evans, 2007]. Table 1 presents the L1 vowel that was judged to be perceptually most like each English vowel (as indicated by a higher fit index, see also table A4 in ‘Appendix’ for the most frequent and the second most frequent L1 classification with the relevant goodness ratings). Greek listeners repeatedly assimilated more than one English vowel to the same L1 category. For example, English /aː/, /ɒ/, and /ɔː/ were all identified with Greek /o/ although with varying degrees of fit. Japanese listeners, on the other hand, assimilated each English vowel to a different L1 category, with the exception of English /ɜː/ and /aː/, which were both related to Japanese /aː/. A comparison of how Greek and Japanese listeners spectrally assimilated English vowels to their L1 categories (i.e. when duration is not taken into account for the latter) shows a similar pattern across L2 groups for most English vowels tested. This is not surprising given the similarity, at least at a phonemic level, of the Greek and the Japanese vowel systems. Still, there were two English vowels that Greek and Japanese listeners related to a different L1 vowel quality: English /ɜː/ was judged as having an /e/ quality by Greek listeners and an /a/ quality by Japanese listeners, and English /aː/ was heard as falling between /o/ and /a/ by Greek listeners and as having an /a/ quality by Japanese listeners. Although fit indexes showed that Greek and Japanese listeners judged those two vowels as being poor examples of their L1 categories, indicating that listeners were struggling to comply with the experimental demands, still these differences in assimilation patterns could not have been predicted by an abstract comparison between the phoneme inventories of English with both Greek and Japanese. A spectral comparison of the relevant vowels (fig. 2a, b, see also tables A1–A3 in the ‘Appendix’) offers some insight into why English /ɜː/ and /aː/ were classified differently by Greek and Japanese

Table 1. Assimilation fit of English vowels to L1 vowel categories for Greek and Japanese listeners, and t test results indicating whether English vowels fitted better to L1 categories in the context where the mean duration in L2 was closer to the mean duration in L1. Assimilation fit is expressed by a fit index, a single metric unit combining percentage identifications and goodness ratings. Mean vowel durations (ms) in L2 (in both /bVb and /bVp contexts) and L1 are given. Standard deviations are also given in parentheses

English vowel		mean duration	Greek L1 closest vowel	Greek mean duration	Greek fit index	Japanese L1 closest vowel	Japanese mean duration	Japanese fit index
1.	biːb	195 (22)	i	107 (8)	5.0	iː	182 (18)	4.2[a]
	biːp	145 (9)			5.4[a]			3.8
2.	bɪb	114 (6)			5.4[a]	i	81 (10)	2.7
	bɪp	98 (12)			5.0			2.7
3.	bɛb	138 (10)	e	116 (9)	4.8	e	98 (12)	4.6
	bɛp	112 (9)			4.9			5.1[a]
4.	bɜːb	213 (10)			2.6	aː	209 (16)	1.3
	bɜːp	188 (8)			2.5			1.1
5.	baːb	219 (16)	o	119 (9)	2.2			2.8[a]
	baːp	191 (10)			2.2			2.3
6.	bɒb	132 (7)			4.8	o	78 (9)	3.9
	bɒp	110 (6)			4.8			3.5[b]
7.	bɔːb	211 (17)			2.3	oː	179 (15)	3.0
	bɔːp	178 (13)			2.3			3.1
8.	bæb	179 (18)	a	122 (10)	4.5	a	106 (11)	2.2
	bæp	140 (7)			5.0[a]			3.7[a]
9.	bʌb	133 (10)			2.6			3.7
	bʌp	111 (13)			2.7			3.7
10.	bʊb	132 (9)	u	112 (7)	3.7	u	86 (10)	2.8
	bʊp	104 (5)			3.5			3.3[a]
11.	buːb	202 (16)			3.1	uː	189 (19)	4.4
	buːp	166 (21)			3.5[a]			4.3

[a] $p < 0.05$ higher fit index in the context where vowel duration in L2 was closer to vowel duration in L1.
[b] $p < 0.05$ lower fit index in the context where vowel duration in L2 was closer to vowel duration in L1.

listeners. The reader should keep in mind of course that plotting F1 and F2 frequencies in a two-dimensional space does not take into account other dimensions along which vowels between languages vary (e.g. fundamental frequency, higher formant frequencies and intrinsic formant movement); another complication is that, while consonantal context was kept constant across Greek and Japanese, it differed for English. First, consider the case of English /ɜː/: As can be seen in figure 2b, Japanese /a/ is located slightly higher in the vowel space than Greek /a/ and hence is closer to English /ɜː/, which might explain why Japanese listeners perceived English /ɜː/ as having an /a/ quality. Similarly, Greek /o/ is much closer to English /aː/ than Japanese /o/ is, which might explain why Greek listeners perceived English /aː/ as falling between Greek /a/ and /o/ while for Japanese listeners English /aː/ had an /a/ quality.

To examine whether context affected Greek and Japanese listeners' assimilation patterns, the fit indexes derived for the 11 English vowels were submitted to separate repeated measures two-way ANOVAs for each L2 group, with Vowel (11 levels) and Context (2 levels) as factors. For Greek listeners, the ANOVA yielded a significant main effect of Vowel [$F(10, 170) = 25.3$; $p < 0.001$] and a significant Vowel × Context [$F(10, 170) = 10.1$; $p < 0.001$] interaction. Likewise, for Japanese listeners the ANOVA yielded a significant main effect of Vowel [$F(10, 140) = 11.5$; $p < 0.001$] and a significant Vowel × Context [$F(10, 140) = 9.3$; $p < 0.001$] interaction. The significant effect of Vowel indicated that English vowels varied in their fit to L1 categories for both L2 groups (for Greek listeners range = 2.2–5.4; for Japanese listeners range = 1.1–5.1). The significant Vowel × Context interaction indicated that context affected how well English vowels fitted to L1 categories, but this effect was not uniform across contexts. This initial analysis suggests that not only Japanese but also Greek listeners attend to both spectral and durational cues when perceiving the English vowels.

To further analyse the effect of context on assimilation patterns, paired samples t tests (for Greek listeners each with d.f. = 17; for Japanese listeners each with d.f. = 14) compared the fit indexes derived for each English vowel in two consonantal contexts (significance level set to $p < 0.005$ to correct for multiple comparisons). For Greek listeners, the t tests showed that four English vowels differed in their fit to L1 categories as a function of context: English /iː, æ, uː/ fitted better in /bVp/ context while English /ɪ/ fitted better in /bVb/ context. Table 1 shows that Greek listeners preferred (as indicated by a higher fit index) these four English vowels in the context where the mean vowel duration in L2 was closer to the mean vowel duration in L1: English /iː, uː, æ/ in their 'short' version and English /ɪ/ in its 'long' version. For Japanese listeners, the t tests showed that six English vowels differed in their fit to L1 categories as a function of context: English /iː, aː, ɒ/ fitted better in /bVb/ context and English /e, æ, ʊ/ in /bVp/ context. Again, it seems that once Japanese listeners assigned an English vowel to either a Japanese short or long category, they preferred that vowel in the context where it was closer in duration to that category. The only exception was English /ɒ/, which fitted better to the Japanese short /o/ category in its longer version (i.e. when presented in a /bVb/ context). Although context did not affect the fit indexes for all English vowels it is important to note that most of the vowels that fitted equally well to Greek and Japanese vowel categories across consonantal contexts had either a mean duration that was equally close in either context to the mean duration in L1 (e.g. English /ʌ/ for Greek listeners) or they were generally judged as being 'poor' examples of an L1 category (e.g. English /ɪ/ for Japanese listeners). Regarding the latter case, it seems that if an L2 vowel did not spectrally match an L1 category well, a better fit in duration would not significantly change the listener's identification and/or goodness rating judgement.

4. Experiment 2: Discrimination Task

The purpose of this experiment was to examine Greek and Japanese listeners' discrimination of English vowels in /bVb/ and /bVp/ contexts. Discrimination was assessed by means of a categorial discrimination test often used in L2 perception studies [e.g. Guion et al., 2000; Aoyama et al., 2004; Flege and MacKay, 2004]. According to the feature hypothesis, only Japanese listeners' discrimination of English vowels should be affected by the context in which vowels were presented to the listeners. According to the desensitization hypothesis on the other hand, both L2 groups' discrimination should be affected by consonantal context. Given the cross-language perceptual data obtained in experiment 1, an additional question addressed in this experiment was whether cross-language perceptual assimilation patterns predicted L2 discrimination as proposed by Best's PAM.

4.1. Method

4.1.1. Participants. The Greek and Japanese participants were the same as in experiment 1. Ten English university students (mean = 25.3 years, range = 18–28 years) all born in London were also tested as controls.

4.1.2. Stimuli. The eleven English vowels described in section 2.1 were combined to create nine contrastive vowel pairs: /iː/-/ɪ/, /ɪ/-/e/, /æ/-/ʌ/, /æ/-/aː/, /æ/-/ɜː/, /ʌ/-/aː/, /ɒ/-/ɔː/, /ʊ/-/uː/ and /ɔː/-/uː/. Contrast selection was based on previous findings for Spanish learners of English, whose system is very similar to the Greek one [e.g. Flege et al., 1994, 1997; Iverson and Evans, 2007], and Japanese learners of English [e.g. Strange et al., 1998]. An effort was made to use contrasts that would vary in degree of discrimination difficulty.

4.1.3. Procedure. Greek and Japanese listeners participated in experiment 2 after completing experiment 1 using the same laptop and headphones. In each trial of the categorial discrimination test, listeners were presented three items, each spoken by a different native English speaker. Each contrast was tested by eight 'different' trials that contained an odd vowel category and eight 'catch' trials that contained three tokens of the same vowel category. The participants were instructed to identify the odd item out by clicking '1', '2' or '3' (in the 'different' trials) or 'same' when all the vowel instances were judged to belong to the same category. Consistent with Guion et al. [2000] and Flege and MacKay [2004], participants were also instructed to ignore differences in speakers' voices and to focus on vowel identity. The inter-stimulus interval was 1.2 s and the inter-trial interval was 3 s. To minimize response bias, A' scores [Snodgrass et al., 1985] were computed for each contrast based on *hits*, when the odd item was correctly selected in 'different' trials and *false alarms*, when an item was incorrectly selected in 'catch' trials. If H (hit) = FA (false alarm) then A' = 0.5.

If $H > FA$ then

$$A' = 0.5 + [(H - FA) \times (1 + H - FA)]/[(4 \times H) \times (1 - FA)]$$

and if $H < FA$ then

$$A' = 0.5 - [(FA - H) \times (1 + FA - H)]/[(4 \times FA) \times (1 - H)].$$

A' score of 1.0 indicates perfect discrimination of a contrast, whereas A' score of 0.5 indicates discrimination at chance level. Before the experiment began, a 20-item practice session (20 trials randomly selected) was presented to familiarize listeners with the procedure. Consonantal context in the practice session was different from the context to be tested first. As in experiment 1, written instructions were given in the L1 of each language group.

4.2. Results

Table 2 shows the accuracy with which native speakers of Greek and Japanese discriminated the nine English vowel contrasts in /bVb/ and /bVp/ contexts. A' scores were firstly submitted to separate two-way repeated measures ANOVAs for each L2 group with Contrast (9 levels) and Context (2 levels) as factors. For Greek listeners, the ANOVA yielded a significant main effect of Contrast [$F(8, 136) = 13.5$; $p < 0.001$], which indicated that discrimination scores varied considerably among English contrasts. The ANOVA also yielded a significant effect of Context [$F(1, 17) = 25.2$; $p < 0.001$] demonstrating Greek listeners' sensitivity to vowel duration changes as well as a significant Contrast × Context [$F(8, 136) = 6.6$; $p < 0.001$] interaction showing that

Table 2. Mean duration (ms) of vowels in each English contrast, duration difference between vowels and mean discrimination scores obtained by Greek and Japanese listeners

English contrast					Greek listeners	Japanese listeners
		mean durations (ms)	duration ratio (longer to shorter)	duration difference (ms)	A' score	A' score
1.	biːb vs. bɪb	195 vs. 114	1.71	81	0.79	0.93
	biːp vs. bɪp	145 vs. 98	1.48	47	0.80	0.95
2.	bɪb vs. beb	114 vs. 138	1.21	26	0.99	0.85
	bɪp vs. bep	98 vs. 112	1.14	14	0.97	0.86
3.	bæb vs. bʌb	179 vs. 133	1.35	46	0.74*	0.65*
	bæp vs. bʌp	140 vs. 111	1.26	29	0.51	0.50
4.	bæb vs. baːb	179 vs. 219	1.22	40	0.83	0.62
	bæp vs. baːp	140 vs. 191	1.36	51	0.80	0.71*
5.	bæb vs. bɜːb	179 vs. 213	1.19	34	0.92	0.73
	bæp vs. bɜːp	140 vs. 188	1.34	48	0.96	0.90*
6.	bʌb vs. baːb	133 vs. 219	1.65	86	0.71*	0.78*
	bʌp vs. baːp	111 vs. 191	1.72	80	0.56	0.70
7.	bɒb vs. bɔːb	132 vs. 211	1.60	79	0.83	0.82
	bɒp vs. bɔːp	110 vs. 178	1.62	68	0.84	0.85
8.	bʊb vs. buːb	132 vs. 202	1.53	70	0.70*	0.86
	bʊp vs. buːp	104 vs. 166	1.60	62	0.61	0.88
9.	bɔːb vs. buːb	211 vs. 202	1.04	9	0.74	0.93
	bɔːp vs. buːp	178 vs. 166	1.07	12	0.75	0.92

*p < 0.05 higher discrimination than that obtained in the other context.

the effect of context on discrimination scores was not uniform across contrasts. For Japanese listeners, the ANOVA yielded a significant main effect of Contrast [$F(8, 112) = 12.7$; $p < 0.001$] as well as a significant Contrast × Context interaction [$F(8, 112) = 6.4$; $p < 0.001$], which again showed that the effect of context on discrimination scores was not uniform across contrasts. These results are consistent with the results of experiment 1 and indicate that both Greek and Japanese listeners make use of durational cues when discriminating L2 vowels.

To further explore the effect of context on the discrimination of English vowels by the two L2 groups, paired samples t tests were used comparing the A' scores obtained for each English contrast in /bVb/ and /bVp/ context (significance level set to $p < 0.005$ to correct for multiple comparisons). The paired comparisons showed that context significantly affected the discrimination of three out of nine English contrasts for Greek listeners and four out of nine English contrasts for Japanese listeners (table 2). Greek listeners showed a better discrimination for English /æ/-/ʌ/, /ʌ/-/aː/, and /ʊ/-/uː/ in /bVb/ than in /bVp/ context. Japanese listeners showed a better discrimination for English /æ/-/ʌ/ and /ʌ/-/aː/ in /bVb/ than in /bVp/ context while showing the opposite pattern for English /æ/-/aː/ and /æ/-/ɜː/.

The next step was to see whether context-induced changes in the duration difference between the vowels of those pairs could explain the observed differences in discrimination, in other words whether the context that showed significantly higher discrimination also provided to the listeners more temporal information than the other context did. An inspection of table 2 reveals that in three of those contrasts (/ʌ/-/aː/ and /ʊ/-/uː/ for Greek listeners and /ʌ/-/aː/ for Japanese listeners) that was not the case since the difference in duration between the vowels in those pairs was relatively constant across contexts, i.e. less than 10 ms, which would be unlikely to be noticed by listeners in syllables whose vowels have the durations of those in this study. (In fact the duration ratio of longer to shorter vowel in these contrasts was smaller in the context that showed better discrimination.) Additionally, there was at least one English contrast (/iː/-/ɪ/ for Greek listeners) that did provide to the listeners considerably more temporal information in /bVb/ than in /bVp/ context, i.e. 81 ms vs. 47 ms, respectively (longer to shorter duration ratio 1.71 vs. 1.48), but showed similar discrimination accuracy across contexts. Greek listeners' discrimination was about as accurate in this case as for the /ɒ/-/ɔː/ contrast, which only differed by 11 ms (79 ms vs. 68 ms), and had similar duration ratios (1.60 vs. 1.62). These examples suggest that listeners do not simply compare the vowel durations when trying to distinguish between the two vowels.

Finally, it was investigated whether perceptual assimilation patterns from experiment 1 predicted discrimination accuracy in experiment 2. The nine English contrasts tested in experiment 2 were assigned to PAM categories based on the perceptual assimilation data from experiment 1. First, the cross-language identification percentages were used to decide whether an English vowel was consistently identified with a single L1 category or was heard as falling between two L1 categories (an Uncategorized sound according to PAM). A 60% identification criterion was adopted. [Harnsberger, 2001 discusses the much higher identification criterion of 90%, but this results in most non-native sounds being classed as uncategorized.] When both English vowels in a contrast were identified with the same L1 vowel, paired sample t tests defined whether that contrast would qualify as a Category Goodness contrast (i.e. the fit indexes of the two vowels differed significantly) or a Single-Category contrast (i.e. the fit indexes of the two vowels did not differ significantly, with significance level set to $p < 0.005$).

Figure 3 shows the mean discrimination scores obtained by Greek and Japanese listeners for each assimilation type, averaged over all English vowel contrasts and two consonantal contexts. Native English control listeners obtained excellent discrimination scores (mean = 0.96–0.98) across vowel contrasts and hence their A' scores will not be discussed further. To examine the effect of assimilation type on L2 discrimination, A' scores were submitted to separate one-way ANOVAs for each language group. The effect of assimilation type was significant for both Greek listeners [$F(3, 320) = 267.5$; $p < 0.001$] and Japanese listeners [$F(3, 266) = 123.5$; $p < 0.001$]. Tukey post-hoc comparisons of the four assimilation types showed the following results, which were consistent across language groups: Two-Category contrasts were easier than Uncategorized-Categorized, Category Goodness and Single-Category contrasts, as PAM would predict, with listeners obtaining generally very high scores in those contrasts. Uncategorized-Categorized and Category Goodness contrasts were easier than Single-Category contrasts again as expected with the latter being the most difficult contrasts to discriminate. Although there was a trend of discrimination scores for Uncategorized-Categorized contrasts being higher than Category Goodness contrasts, this difference was not significant (note the large variability in scores after averaging over contrasts and contexts).

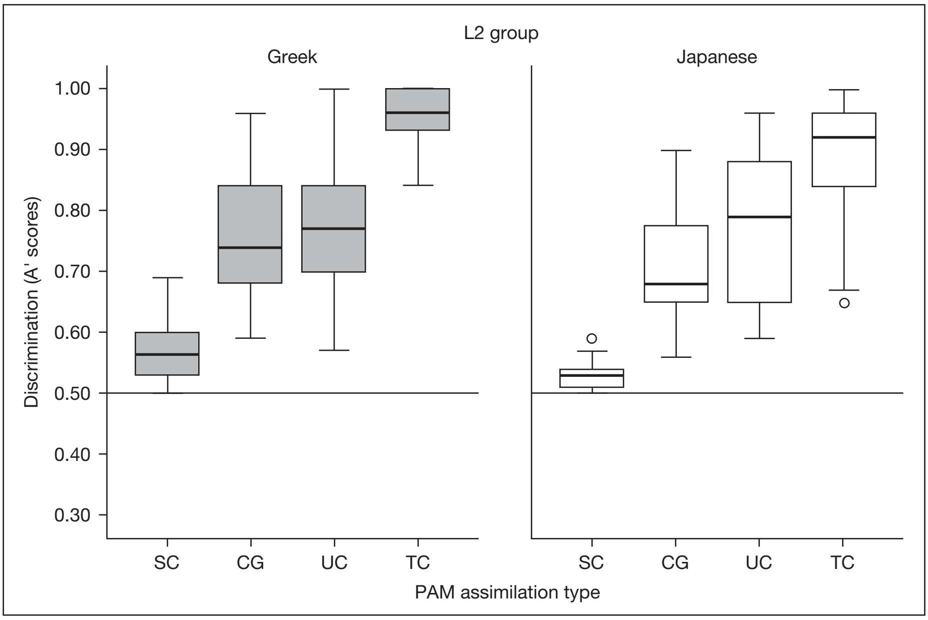

Fig. 3. Boxplots of English vowel discrimination accuracy for Single-Category (SC), Category Goodness (CG), Uncategorized-Categorized (UC), and Two-Category (TC) assimilation types averaged over nine English contrasts and two consonantal contexts by Greek (left panel) and Japanese listeners (right panel). An A' score of 0.5 indicates discrimination at chance level.

5. Discussion and Conclusion

This study examined the extent to which previous linguistic experience with an acoustic cue in L1 affects the use of that cue in L2 perception and, more particularly, how experience with duration in L1 vowel distinctions relates to the use of durational cues when perceiving vowels in an L2. The availability of this particular acoustic cue to listeners with no such L1 experience has been a matter of debate in past years with studies arriving at different conclusions. The two main proposals in the literature are represented by the feature hypothesis, proposed by McAllister et al. [2002], and the desensitization hypothesis, proposed by Bohn [1995]. The former is based on the notion of L1 transfer in L2 learning and posits that L2 learners do not have access to cues that are not used in L1 to signal contrasts. The latter posits that L2 learners are sensitive to durational cues when perceiving L2 vowels irrespective of the status of duration in their L1 and that, in fact, learners tend to rely more on durational than spectral cues when faced with difficult L2 contrasts [Escudero, 2005; Cebrian, 2006 among others]. To evaluate these hypotheses, the perception of English vowels by native speakers of Greek and Japanese was examined. Participants performed two tasks with natural English vowels, an assimilation task with goodness ratings (experiment 1) and a discrimination task (experiment 2). The stimuli in both experiments were /bVb/ and /bVp/ syllables. That way, the effect of vowel duration differentiations introduced by the voicing vs. voicelessness of the stop consonant following the vowel on both L2 groups' performance was tested. The L2 groups were selected on the basis of the similarity of their vowel systems. Both languages have five relatively similar vowel qualities (/i, e, a, o, u/) and are also similar to each other in terms of the simplicity of their syllable structure and phonotactic constraints. The major difference between Greek and Japanese is that only the latter distinguishes vowels on the basis of duration.

The results of experiment 1 showed, firstly, that most English vowels were assimilated to the same spectral categories by the two L2 groups, which confirmed the similarity of the two vowel systems. Two exceptions to this pattern were English /ɜː/ and /aː/, which were related to different L1 spectral categories by Greek and Japanese listeners. This finding shows that the two languages may share the same number of contrastive spectral categories but, as shown in the spectral analysis of the Greek and Japanese vowels, there are differences in the phonetic realization of those categories, which resulted in different patterns of spectral assimilation for the two English vowels. Importantly, the cross-language perceptual assimilation task proved to be able to reveal those subtle acoustic/articulatory differences. Regarding the main question asked by this study, that is whether durational cues are available to listeners with no such L1 experience, the results of experiment 1 indicated a positive answer. This finding seems to be in disagreement with the feature hypothesis and, at first glance, in support of the desensitization hypothesis. However, when looking closer at between-context comparisons conducted for each English vowel separately it was found that English vowels generally fitted better to L1 categories in the context where they resembled more the duration of the spectrally closest L1 vowel. This suggests that L2 learners assimilate both temporally and spectrally L2 vowels to L1 categories and hence there is nothing special about duration as Bohn [1995] has proposed. The observed patterns of temporal assimilation reflect a temporal 'matching' to the L1 categories irrespective of whether the L1 has a phonemic vowel length contrast or not.

Discrimination performance in experiment 2 was generally consistent with the predictions made by Best's [1995] PAM. Greek and Japanese listeners had no difficulty with Two-Category contrasts, had some difficulty with Uncategorized-Categorized and Category Goodness contrasts and found Single-Category contrasts the most difficult to discriminate. The discrimination scores for Uncategorized-Categorized contrasts were somewhat lower than predicted and did not differ significantly to those obtained for Category Goodness contrasts. Guion et al. [2000] report on a similar finding in their data and propose a possible revision of PAM regarding the discriminability of Uncategorized-Categorized contrasts where the uncategorized sound is close in the perceptual space to the categorized one. The results regarding the effect of context on the discrimination of English vowels showed that both Greek and Japanese listeners were sensitive to durational cues. Again, this seems to run contra the feature hypothesis and in favour of the desensitization hypothesis. However, paired comparisons conducted for each English contrast separately indicated that L2 learners were not simply comparing the durations of the two members in a pair when trying to distinguish one from another. There were contrasts which proved to be easier in one context than the other despite the fact that the duration difference between the two vowels was similar across contexts. There were also contrasts where context-induced changes in the duration difference between the two vowels did not result in changes in discrimination performance.

Taken together, the results of experiments 1 and 2 suggest that L2 vowels undergo both *temporal* and *spectral* perceptual assimilation to L1 category/categories and hence duration does not have a special status in L2 vowel perception compared to that of spectral cues. L2 learners who do not exploit duration in L1 may have access to temporal cues in L2 provided that their (single) L1 duration category does not temporally interfere with the perception of a given L2 contrast. The fact that the more often tested English contrast in the literature, namely English /iː/-/ɪ/, does not suffer from L1 temporal interference (at least for listeners with a single duration category for English /iː/ being too long when compared to the typical L1 vowel durations) seems to be the reason for the widespread view that listeners with a single L1 vowel duration category have access to durational cues irrespective of the contrast to be perceived. Seen in this context, the results of this study are compatible with the perceptual interference account [Iverson et al. 2003; Kuhl et al., 2006, 2008] and the current L2 speech perception models [PAM: Best, 1995; SLM: Flege, 1995] that emphasize the role of L1 transfer. It seems that what is transferred is not an increased or decreased temporal acuity, depending on previous experience with duration in vowel distinctions, as the feature hypothesis would predict. Instead, the listeners transfer their L1 temporal pattern, which may impede or aid L2 perception depending on the cross-language temporal relationships. One explanation for the fact that the Latin American Spanish native speakers in the study by McAllister et al. [2002] did not show any sensitivity to durational cues is indeed given by the authors themselves in the discussion of their results. They draw attention to the fact that in their study L2 perception was assessed by means of a word recognition task, which does not exclude the possibility that some of the participants simply did not know whether a word contained a short or a long vowel rather than being unable to distinguish short from long vowels. For listeners with no previous experience with duration in L1 vowel distinctions a more sensitive task may be therefore needed to capture their sensitivity to that acoustic cue. In conclusion, L2 vowel perception is guided by a complicated interplay between spectral and temporal L1-L2 relationships; however, the process of accessing spectral and temporal information in an L2 appears to be working in similar ways across individuals with different L1 backgrounds.

Acknowledgements

I would like to thank Valerie Hazan and Paul Iverson and two reviewers for their useful comments on an earlier version of the manuscript, and Michael Ashby for his help in stimuli selection. Earlier versions of this work were presented at the 16th International Congress of Phonetic Sciences (Saarbrücken, Germany) and at the 8th International Conference of Greek Linguistics (Ioannina, Greece). This work was supported by research grants from the A.G. Leventis Foundation and from the University College London Graduate School.

References

Aoyama, K.; Flege, J.E.; Guion, S.G.; Akahane-Yamada, R.; Yamada, T.: Perceived phonetic dissimilarity and L2 speech learning: The case of Japanese /r/ and English /l/ and /r/. J. Phonet. *32:* 233–250 (2004).

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

Best, C.T.: A direct realist view of cross-language speech perception; in Strange, Speech perception and linguistic experience: Issues in cross-language research, pp. 171–204 (York Press, Baltimore 1995).

Best, C.T.; McRoberts, G.W.; Sithole, N.M.: Examination of perceptual reorganization for nonnative speech contrasts: Zulu click discrimination by English-speaking adults and infants. J. exp. Psychol. hum. Percept. Perform. *14:* 345–360 (1988).

Best, C.T.; Tyler, M.D.: Nonnative and second-language speech perception: Commonalities and complementarities; in Munro, Bohn, Second language speech learning, pp. 13–34 (Benjamins, Amsterdam 2007).
Bohn, O.-S.: Cross-language speech perception in adults: First language transfer doesn't tell it all; in Strange, Speech perception and linguistic experience: Issues in cross-language research, pp. 279–304 (York Press, Baltimore 1995).
Cebrian, J.: Experience and the use of duration in the categorization of L2 vowels. J. Phonet. *34:* 372–387 (2006).
Eimas, P.D.; Siqueland, E.R.; Jusczyk, P.; Vigorito, J.: Speech perception in infants. Science *171:* 303–306 (1971).
Escudero, P.: Linguistic perception and second language acquisition: Explaining the attainment of optimal phonological categorization; PhD diss. Utrecht University, Utrecht (unpublished, 2005).
Escudero, P.; Boersma, P.: Bridging the gap between L2 speech perception research and phonological theory. Stud. second Lang. Acquis. *26:* 551–585 (2004).
Flege, J.E.: Second language speech learning: Theory, findings, and problems; in Strange, Speech perception and linguistic experience: Issues in cross-language research, pp. 233–277 (York Press, Baltimore 1995).
Flege, J.E.; Bohn, O.-S.; Jang, S.: Effects of experience on non-native speakers' production and perception of English vowels. J. Phonet. *25:* 437–470 (1997).
Flege, J.E.; MacKay, I.: Perceiving vowels in a second language. Stud. second Lang. Acquis. *26:* 1–34 (2004).
Flege, J.E.; MacKay, I.; Meador, D.: Native Italian speakers' perception and production of English vowels. J. acoust. Soc. Am. *106:* 2973–2987 (1999).
Flege, J.E.; Munro, M.J.; Fox, R.A.: Auditory and categorical effects on cross-language vowel perception. J. acoust. Soc. Am. *95:* 3623–3641 (1994).
Flege, J.E.; Schirru, C.; MacKay, I.R.A.: Interaction between the native and second language phonetic subsystems. Speech Commun. *40:* 467–491 (2003).
Giegerich, H.: English phonology: An introduction (Cambridge University Press, Cambridge 1992).
Goto, H.: Auditory perception by normal Japanese adults of the sounds 'L' and 'R'. Neuropsychologia *9:* 317–323 (1971).
Guion, S.G.; Flege, J.E.; Akahane-Yamada, R.; Pruitt, J.C.: An investigation of current models of second language speech perception: The case of Japanese adults' perception of English consonants. J. acoust. Soc. Am. *107:* 2711–2724 (2000).
Halle, P.A.; Best, C.T.; Levitt, A.: Phonetic vs. phonological influences on French listeners' perception of American English approximants. J. Phonet. *27:* 281–306 (1999).
Harnsberger, J.D.: On the relationship between identification and discrimination of non-native nasal consonants. J. acoust. Soc. Am. *110:* 489–503 (2001).
Hazan, V.; Barrett, S.: The development of phonemic categorisation in children aged 6 to 12. J. Phonet. *28:* 377–396 (2000).
Hillenbrand, J.M.; Clark, M.J.; Houde, R.A.: Some effects of duration on vowel recognition. J. acoust. Soc. Am. *108:* 3013–3022 (2000).
Hirata, Y.: Effects of speaking rate on the vowel length distinction in Japanese. J. Phonet. *32:* 565–589 (2004).
House, A.S.: On vowel duration in English. J. acoust. Soc. Am. *33:* 1174–1182 (1961).
Huckvale, M.: Speech filing system tools for speech research (SFS). University College London. Retrieved March 16, 2008, from http://www.phon.ucl.ac.uk/resource/sfs/.
Iverson, P.; Evans, B.G.: Learning English vowels with different first-language vowel systems: Perception of formant targets, formant movement, and duration. J. acoust. Soc. Am. *122:* 2842–2854 (2007).
Iverson, P.; Kuhl, P.K.; Akahane-Yamada, R.; Diesch, E.; Tohkura, Y.; Kettermann, A.; Siebert, C.: A perceptual interference account of acquisition difficulties for non-native phonemes. Cognition *87:* 47–57 (2003).
Kuhl, P.K.; Conboy, B.T.; Coffey-Corina, S.; Padden, D.; Rivera-Gaxiola, M.; Nelson, T.: Phonetic learning as a pathway to language: New data and native language magnet theory expanded (NLM-e). Phil. Trans. R. Soc. *363:* 979–1000 (2008).
Kuhl, P.K.; Stevens, E.; Hayashi, A.; Deguchi, T.; Kiritani, S.; Iverson, P.: Infants show a facilitation effect for native language phonetic perception between 6 and 12 months. Dev. Sci. *9:* 13–21 (2006).
Kuhl, P.K.; Williams, K.A.; Lacerda, F.; Stevens, K.N.; Lindblom, B.: Linguistic experience alters phonetic perception in infants by 6 months of age. Science *255:* 606–608 (1992).
Maye, J.; Weiss, D.J.: Statistical cues facilitate infants' discrimination of difficult phonetic contrasts. 27th Annu. Boston Univ. Conf. on Lang. Dev., 2003.
Maye, J.; Werker, J.; Gerken, L.: Infant sensitivity to distributional information can affect phonetic discrimination. Cognition *82:* 101–111 (2002).
McAllister, R.; Flege, J.E.; Piske, T.: The influence of L1 on the acquisition of Swedish quantity by native speakers of Spanish, English and Estonian. J. Phonet. *30:* 229–258 (2002).
Morrison, G.S.: Effects of L1 duration experience on Japanese and Spanish listeners' perception of English high front vowels; MA diss. Simon Fraser University, Burnaby, British Columbia (unpublished, 2002).
Peterson, G.E.; Lehiste, I.: Duration of syllable nuclei in English. J. acoust. Soc. Am. *32:* 693–703 (1960).
Polka, L.: Linguistic influences in adult perception of non-native vowel contrasts. J. acoust. Soc. Am. *97:* 1286–1296 (1995).
Polka, L.; Werker, J.F.: Developmental changes in perception of nonnative vowel contrasts. J. exp. Psychol. hum. Percept. Perform. *20:* 421–435 (1994).

Schmidt, A.M.: Cross-language identification of consonants. Part 1: Korean perception of English. J. acoust. Soc. Am. *99:* 3201–3211 (1996).
Shibatani, M.: The languages of Japan (Cambridge University Press, Cambridge1990).
Snodgrass, J.G.; Levy-Berger, G.; Haydon, M.: Human experimental psychology (Oxford University Press, New York 1985).
Strange, W.; Yamada, R.A.; Kubo, R.; Trent, S.A.; Nishi, K.; Jenkins, J.J.: Perceptual assimilation of American English vowels by Japanese listeners. J. Phonet. *26:* 311–344 (1998).

Appendix

Table A1. Mean F1 and F2 frequencies (Hz) of English vowels in /bVb/ and /bVp/ context with standard deviations in parentheses

English	F1	F2	F1	F2
	/bVb/		/bVp/	
iː	327 (20)	2,666 (158)	350 (18)	2,690 (145)
ɪ	432 (38)	2,039 (132)	436 (37)	2,078 (121)
e	628 (66)	1,996 (111)	646 (71)	1,948 (116)
ɜː	531 (50)	1,586 (87)	526 (55)	1,571 (75)
æ	800 (72)	1,648 (62)	771 (66)	1,677 (64)
ʌ	665 (75)	1,377 (93)	683 (88)	1,378 (77)
aː	629 (55)	1,244 (71)	614 (67)	1,235 (68)
ɒ	551 (30)	1,091 (66)	569 (36)	1,109 (61)
ɔː	417 (31)	859 (75)	426 (30)	826 (72)
ʊ	434 (22)	1,271 (92)	452 (29)	1,260 (77)
uː	371 (23)	1,538 (101)	344 (19)	1,601 (111)

Table A2. Mean F1 and F2 frequencies (Hz) of Greek vowels in /mVn/ context with standard deviations in parentheses

Greek	F1	F2
	/mVn/	
i	447 (32)	2,528 (125)
e	638 (61)	2,073 (102)
a	783 (83)	1,562 (92)
o	655 (55)	1,099 (71)
u	439 (24)	901 (69)

Table A3. Mean F1 and F2 frequencies (Hz) of Japanese vowels in /mVn/ context with standard deviations in parentheses

Japanese	F1	F2
	/mVn/	
i	359 (25)	2,724 (147)
iː	342 (26)	2,740 (153)
e	610 (53)	2,106 (106)
eː	617 (46)	2,190 (109)
a	720 (79)	1,621 (77)
aː	734 (88)	1,629 (83)
o	553 (43)	1,006 (54)
oː	584 (41)	964 (63)
u	380 (23)	1,610 (124)
uː	366 (27)	1,601 (131)

Table A4. Most frequent and second most frequent percentage classification of English vowels in terms of Greek and Japanese vowel categories with the relevant goodness ratings assigned

SBE	/bVb/						/bVp/					
	modal response			2nd response			modal response			2nd response		
	Greek listeners											
iː	i	100	5.0	–	–	–	i	100	5.4	–	–	–
ɪ	i	100	5.4	–	–	–	i	100	5.0	–	–	–
e	e	97	5.0	–	–	–	e	92	5.2	i	4	2.0
ɜː	e	87	3.0	o	9	2.5	e	77	3.2	o	14	2.4
æ	a	95	4.7	o	5	1.5	a	95	5.2	o	3	1.0
ʌ	a	62	4.2	o	36	5.1	a	66	4.1	o	30	4.3
aː	o	57	4.0	a	43	3.6	o	54	4.1	a	46	3.9
ɒ	o	97	5.0	a	3	1.0	o	97	5.0	–	–	–
ɔː	o	55	4.1	u	45	3.8	o	52	4.1	u	48	3.8
ʊ	u	92	4.0	o	4	1.8	u	84	4.2	o	8	2.4
uː	u	82	3.8	i	14	2.5	u	92	3.8	i	8	1.9
	Japanese listeners											
iː	iː	80	5.3	i	14	3.9	iː	74	5.2	i	23	4.6
ɪ	i	63	4.2	e	33	3.3	i	61	4.4	e	37	3.7
e	e	95	4.9	eː	3	4.3	e	95	5.4	eː	4	3.7
ɜː	aː	58	2.4	a	19	1.7	aː	48	2.3	a	28	1.4
æ	a	64	3.1	aː	33	3.3	a	87	4.3	aː	11	3.0
ʌ	a	90	4.1	o	7	2.0	a	89	4.2	o	4	3.0
aː	aː	64	4.3	a	26	2.8	aː	62	3.7	a	19	3.3
ɒ	o	84	4.7	a	9	2.5	o	86	4.1	a	13	2.7
ɔː	oː	69	4.4	o	20	3.9	oː	71	4.4	o	20	3.3
ʊ	u	81	3.5	uː	9	3.9	u	83	4.0	uː	9	4.1
uː	uː	85	5.2	u	10	4.7	uː	86	5.0	u	12	4.5

SBE = Southern British English.