The relationship between native allophonic experience with vowel duration and perception of the English tense/lax vowel contrast by Spanish and Russian listeners

Maria V. Kondaurova
Program in Linguistics, Purdue University, 100 North University Street, West Lafayette, Indiana 47907

Alexander L. Francis[a)]
Program in Linguistics and Department of Speech, Language and Hearing Sciences, Purdue University, 500 Oval Drive, West Lafayette, Indiana 47907

(Received 7 November 2007; revised 5 September 2008; accepted 19 September 2008)

Two studies explored the role of native language use of an acoustic cue, vowel duration, in both native and non-native contexts in order to test the hypothesis that non-native listeners' reliance on vowel duration instead of vowel quality to distinguish the English tense/lax vowel contrast could be explained by the role of duration as a cue in native phonological contrasts. In the first experiment, native Russian, Spanish, and American English listeners identified stimuli from a *beat/bit* continuum varying in nine perceptually equal spectral and duration steps. English listeners relied predominantly on spectrum, but showed some reliance on duration. Russian and Spanish speakers relied entirely on duration. In the second experiment, three tests examined listeners' use of vowel duration in native contrasts. Duration was equally important for the perception of lexical stress for all three groups. However, English listeners relied more on duration as a cue to postvocalic consonant voicing than did native Spanish or Russian listeners, and Spanish listeners relied on duration more than did Russian listeners. Results suggest that, although allophonic experience may contribute to cross-language perceptual patterns, other factors such as the application of statistical learning mechanisms and the influence of language-independent psychoacoustic proclivities cannot be ruled out. © *2008 Acoustical Society of America.* [DOI: 10.1121/1.2999341]

PACS number(s): 43.71.Hw, 43.71.Es, 43.71.An [AJ] Pages: 3959–3971

I. INTRODUCTION

A variety of recent research suggest that native phonetic (allophonic) experience is one factor that can influence the perception of non-native speech sounds (Best and Strange, 1992; Halle, *et al.* 1999; Jenkins and Yeni-Komshian, 1995) and improve learning (Jamieson and Moroson, 1986; McClaskey *et al.* 1983; Pisoni *et al.*, 1982). Both the *type* (Pruitt *et al.*, 2006) and the *amount* (McAllister *et al.*, 2002) of such experience have been proposed to contribute to the success of perception of non-native speech contrasts.

The aims of the present study are (a) to determine whether native Spanish and Russian listeners, who differ in the reported role of allophonic vowel duration in their native languages, differ in the way they employ vowel duration in the perception of the English tense and lax front unrounded vowel contrast, which does not exist in either native language, and (b) to determine whether any differences in the two groups' observed pattern of reliance on duration in this cross-language context can be related to the degree to which they employ vowel duration as a cue in their native language. Vowel duration is not a phonologically contrastive feature in either Spanish or Russian. However, the two languages are reported to differ in terms of the degree to which each relies on vowel duration as an acoustic cue in the perception and/or production of a variety of phonological contrasts including stress and voicing.

In English, tense and lax front unrounded vowels are differentiated along two independent acoustic dimensions: spectrum (vowel quality, related mainly to the first three formant frequencies: F1, F2, and F3) and duration (vowel length). While native English listeners rely predominantly on spectral properties with vowel duration playing a secondary role (Ainsworth, 1972; Mermelstein, 1978; Hillenbrand *et al.*, 2000), there is contradictory evidence on what acoustic cues native Spanish listeners employ for the perception of this contrast and why they are able to use these cues. While some studies suggest that native Spanish listeners are not significantly different from native English listeners in their ability to use spectral cues (Flege *et al.*, 1997), most studies have demonstrated that, unlike native English listeners, Spanish listeners rely exclusively or primarily on vowel duration for the identification of this contrast (Bohn, 1995; Escudero 2006; Escudero and Boersma, 2004).

Typically, the use of spectral properties by native Spanish listeners, when it is observed, is explained by phonological and phonetic experience with this acoustic property in the Spanish vowel system (Fox *et al.*, 1995). However, the widely demonstrated reliance on duration poses a problem due to the absence of phonological and allegedly very limited allophonic experience with vowel duration by native Spanish listeners (Hammond, 2001; Harris, 1969).

[a)]Electronic mail: francisa@purdue.edu

Russian reportedly differs from Spanish in terms of the degree to which vowel duration serves as an acoustic correlate of phonological contrasts. For example, in Russian, lexical stress is a phonologically distinctive feature, in the sense that two words may differ only in terms of stress placement. For example (accent precedes the stressed syllable): ˈслуги /ˈslugi/ "servants" (noun, nominative plural)-слуˈги /sluˈgi/ "a servant" (noun, genetive singular); ˈволос /ˈvolos/ "a hair" (noun, nominative singular)-воˈлос /voˈlos/ "hair" (noun, genetive plural). There are also three levels of stress in Russian that are realized primarily in terms of different degrees of vowel duration (Bondarko, 1998). In stressed syllables vowels are long. In a syllable immediately preceding the stressed syllable vowels are shorter, and in all other unstressed syllables the vowel length is further reduced with an average ratio of 4:2:1 for stressed, prestress, and unstressed, respectively (Bondarko, 1998).

In Spanish, however, both stressed and unstressed syllables have the same linguistic length (Hammond, 2001), and although some researchers have reported finding duration differences between productions of stressed and unstressed vowels (Diaz-Campos, 2000; Manrique and Signorini, 1983), perceptual studies have shown that vowel duration in itself is a poor predictor of the location of lexical stress in Spanish (Llisterri, *et al.*, 2003). Thus, it appears that native Russian listeners may have more phonetic experience than native Spanish listeners with vowel duration as an acoustic correlate of stress in their native language, and this difference in experience might affect the two groups' reliance on this feature in the identification of cross-language vowel contrasts.

On the other hand, the two languages are also quite similar with respect to the use of duration in a variety of segmental contexts. At the segmental level, in both Spanish and Russian, vowels are longer before voiced than before voiceless consonants (Chen, 1970; Mendoz *et al.*, 2003; Pye, 1986; Zimmerman and Sapon, 1958). Furthermore, both Spanish and Russian vowels have intrinsically different lengths (Bolla, 1981; Bondarko, 1998; Mendoz *et al.*, 2003): In both languages, lower vowels (e.g., [ɛ]) are longer than higher vowels (e.g., [i]).[1] Thus, some evidence suggests that Spanish and Russian listeners might differ from one another in their perception of the English tense-lax vowel contrast on the basis of differences in their allophonic experience with duration (i.e., for lexical stress). On the other hand, the two groups might be similar in performance, if Russian listeners' experience is not sufficiently greater or more effectively transferred than that of Spanish speakers.

In the present study we first examined Spanish and Russian listeners' weighting of duration compared to spectral cues in the perception of the English high front tense/lax vowel contrast. We used a two-dimensional stimulus continuum ranging from a familiar speech sound [i] to an unfamiliar one, English [ɪ], to test the prediction that Russian listeners would rely more heavily on duration cues than Spanish listeners because of Russian listeners' greater native allophonic experience with vowel duration, and we compared the performance of both groups to that of native speakers of English. Based on the results of this experiment, we subsequently examined the degree to which these same listeners relied on duration in three allophonic contexts in nonsense words based on the phonology of their native language. Thus, the focus of the second experiment is different from the first, exploring the importance of vowel duration in native phonological contrasts in order to evaluate the hypothesis that Spanish and Russian listeners make differential use of vowel duration natively, and to determine whether it is possible to relate the pattern of results shown in the first experiment to listeners' use of duration as a cue in their native language.

II. EXPERIMENT 1

A. Method

1. Participants

Participants were native speakers of American English (10 women, and 6 men), Spanish (8 women, and 10 men), and Russian (11 women, and 8 men). American English participants were undergraduate and graduate students at Purdue University, who grew up in the Midwestern United States with minimal exposure to other languages as determined by a pretest questionnaire. The Spanish and Russian participants were all graduate students, postdoctoral fellows, and/or instructors at Purdue University.

Nine Spanish participants were from Spain (Madrid, Barcelona, and Cordoba areas) and nine were from Latin American countries (Mexico, Peru, and Colombia). Out of 19 Russian participants, 2 were born and grew up in Ukraine and 3 in Belarus, the rest were born and grew up in the Russian Federation. All 19 Russian participants identified themselves as native monolingual Russian speakers.

As determined by a background questionnaire, no Spanish or Russian participant had experience living in a nonnative language environment for more than a period of 2 weeks prior to arrival in the USA except three Russian listeners who lived in Austria, Hungary, and France for one, three and four years. All participants reported no history of speech or hearing disability and passed a standard hearing test using an M 120 Beltone Audiometer (pure tone audiometry, binaural, frequencies tested: 500 Hz at 20 dB Hearing Level (HL), 1, 2, 4, 6, and 8 kHz at 15 dB HL) prior to the experiment. All participants were paid for their participation in the experiment. For demographic and language background data on Spanish, Russian, and English participants in experiments 1 and 2 see Table I.

2. Stimuli

Stimuli for experiment 1 were a set of 81 syllables of the form /bVt/ (where V is a vowel ranging from [i] to [ɪ]) created according to the following procedure. A 34-year-old male native speaker of a Midwestern dialect of American English produced several examples of the words *beat* and *bit* in isolation. All tokens were recorded using a digital audio tape recorder (Sony TCD-D8), a digital audio tape (Sony PDP-95C PRO-DAT Plus), and a hypercardioid microphone (Audio-Technica D1000HE) in a sound-isolated booth (IAC, Model No. 403A). The recorded tokens were redigitized at 10 kHz and peak amplitude normalized using the PRAAT

TABLE I. Demographic and language background data on Spanish, Russian, and English participants in experiments 1 and 2. M—male, F—female; AOA—age of arrival in the USA; LOR—length of residence in the USA; EFL age—age of starting English as a foreign language education in home countries; EFL period—period of learning English as a foreign language in home countries; start of US education—age when first exposed to formal education in the USA (only includes participants with more than 0 month educational experience); US education—period of formal education in the USA; ESL in US—period of English as a second language education in the USA; motivation—self-rated motivation level in using English; imitation—self-rated ability to imitate English sounds; ESL home—self-rated frequency of use of English at home; ESL social—self-rated frequency of use of English in social settings; and ESL work—self-rated frequency of use of English at work.

	Spanish	Russian	English
No. of participants	18	19	16
Sex	10 M; 8 F	8 M; 11 F	6 M; 10 F
Age (yrs.) (std)	29.2 (6.6)	30.2 (7.1)	21.3 (2)
AOA (yrs.) (std)	25 (5.8)	26.2 (6)	
LOR (yrs.) (std)	3.7 (3.9)	3.8 (3.8)	
EFL age (yrs.) (std)	10.4 (4.0)	9.5 (3.7)	
EFL period (yrs.) (std)	10.7 (6.0)	10.5 (4.4)	
Start of US education (yrs.) (std)	25.2 (6.3)	23.6 (4.09)	
US education (yrs.) (std)	2.9 (2.4)	1.4 (2.0)	
ESL in US (yrs.) (std)	0.1 (0.3)	0.1 (0.2)	
Motivation (ten point scale)	8.7 (1.7)	8.3 (2.1)	
Imitation (ten point scale)	6.7 (1.8)	5.7 (2.1)	
ESL home (ten point scale)	4.8 (3.9)	3.7 (3.2)	
ESL social (ten point scale)	7.3 (2.2)	7.9 (1.7)	
ESL work (ten point scale)	7.7 (2.7)	9.5 (1.0)	

4.1.21 program running on a Dell Optiplex/Windows XP computer with a Sound Blaster Live! sound card.

From this set one token of beat and bit each was selected using the criterion that there be no abrupt changes in formant movement throughout the vowel, no abrupt changes in fundamental frequency, and minimal extraneous noise (to facilitate resynthesis), and having formant and duration values similar to those described by Hillenbrand *et al.*, (1995). The average formant frequencies and duration of the vowel in the beat token were F1: 326 Hz, F2: 2056 Hz, F3: 2943 Hz, and F4: 3389 Hz, duration 171 ms, while those for the bit token were F1: 458 Hz, F2: 1876 Hz, F3: 2523 Hz, and F4: 3388 Hz, duration 125 ms.

The first four formant center frequencies of the beat and bit tokens were extracted using the standard Linear Predictive Coding (LPC) analysis settings implemented in PRAAT 4.2.21. These values were converted to Mel and nine values were calculated for each formant ranging in equal Mel steps between the beat and bit values. For each step and each formant, the original formant contours were preserved. Starting with the original beat token, chosen because it was a marginally superior recording, one version of the syllable was resynthesized for each of the nine sets of formant frequency steps following methods described in the PRAAT 4.2.21 manual entry for source-filter resynthesis. From each of the nine steps along the vowel quality continuum, a continuum ranging in vowel duration from 275 to 35 ms was created ((30 ms/step). The endpoints, 275 and 35 ms, were chosen to allow for nine duration steps to be equal to the number of vowel quality steps (nine steps) in order to avoid the possibility that listeners prefer to rely on the dimension with the least variability (Bohn, 1995). Each step along the duration continuum was approximately equivalent to one just-noticeable difference (JND) for English listeners (25 ms, following Klatt, 1976), as was each step along the spectrum continuum based on F1 values (11–14 Hz, or 3%–5% of the formant center frequency following Kewly-Port and Watson (1994).

3. Procedure

Experiment 1 was generated using E-PRIME Version 1.1 (Schneider *et al.*, 2002). All participants were seated in individual cubicles, equipped with a Dell Optiplex/Windows XP computer, and a Model RB-620 response pad (Cedrus Corporation). Stimuli were presented via a Soundblaster Live! soundcard through headphones (Sennheiser, HD 25-1) at a comfortable listening level of 60–65 dBA. Experiment 1 was completed on the first day of testing.

On each trial listeners heard one stimulus (up to 550 ms) with a simultaneous visual presentation of two possible answers, beat and bit. The possible answers remained on the screen until participants pressed a button on the response pad corresponding to either beat (left button) or bit (right button). Then there was a short pause (250 ms) followed by a blink of the screen (250 ms) to indicate the start of a new trial. Participants were instructed to respond as quickly as possible, but each trial was self-paced with no limit on time to respond. There were 81 stimuli repeated five times each in each of the two blocks (810 trials total). All trials within a block were presented in random order. The average running time for both blocks was 45 min. Before the experiment began, there was a practice session consisting of ten trials with naturally recorded tokens of the words *sheep* and *ship*. The structure of the practice session was identical to the actual experiment but responses were not recorded.

4. Analysis

(a) *Identification functions.* Identification (ID) functions were calculated as the proportion of [ɪ] responses to all nine stimuli sharing a given duration (duration continuum) or spectrum (spectrum continuum) value. ID functions were calculated as average group responses.

(b) *Discriminant function analysis*. Discriminant function analysis (StatSoft, Inc., 2003) uses a linear discriminant equation to determine the degree to which each variable (spectrum, and duration) contributes to the stimulus token's membership in a particular response category (tense or lax) in terms of Wilks' lambda coefficient. Wilks' lambda coefficient ranges from 0 to 1: The closer it is to 0, the stronger a particular variable contributes to the predicted membership. A coefficient close to 1 indicates little or no contribution.

(c) *Multiple regression and simple t-tests*. Forward, stepwise multiple regression ($F>1$ to enter) was employed to identify correlations between factors known to affect degree of foreign accent in L2 to estimate foreign language

proficiency and the relative weighting given to spectrum versus duration (Piske *et al.*, 2001; McAllister *et al.*, 2002; Ylinen *et al.*, 2005). Factors (predictor variables) were (see Table I): (1) age of participant, (2) age of arrival (AOA) in the USA, (3) length of residency (LOR) in the USA, (4) age of starting English as a foreign language (EFL) education in the home country, (5) period of EFL in the home country, (6) period of English as a second language (ESL) study in the USA after arrival, (7) age of starting formal education in the USA, (8) period of formal education in the USA, (9) self-estimated degree of motivation in using ESL, (10) self-estimated degree of imitation abilities, (11) self-estimated degree of using ESL at home, (12) self-estimated degree of using ESL in a social environment, and (13) self-estimated degree of using ESL at work.

In order to investigate the contribution of these factors to performance in experiment 1, the changes in the proportion of lax (bit) responses between the low and high ends of ID functions for both the spectrum and duration continua were calculated for each subject by subtracting the proportion of bit responses given at step 9 from the proportion of bit responses given at step 1. The larger the value along a given dimension, the greater the reliance on that dimension. Using these values, a spectrum-to-duration ratio (Bondarko, 1977; Escudero and Boersma, 2004) was calculated for each participant by dividing the change along the spectrum dimension by the change along the duration dimension. Then multiple regression analysis was run between the ratio and each factor. In addition, because the number of factors was large compared to the size of the number of samples (participants) in the analysis, simple t-tests on these factors were run comparing group means to confirm that the groups were matched in English language proficiency.

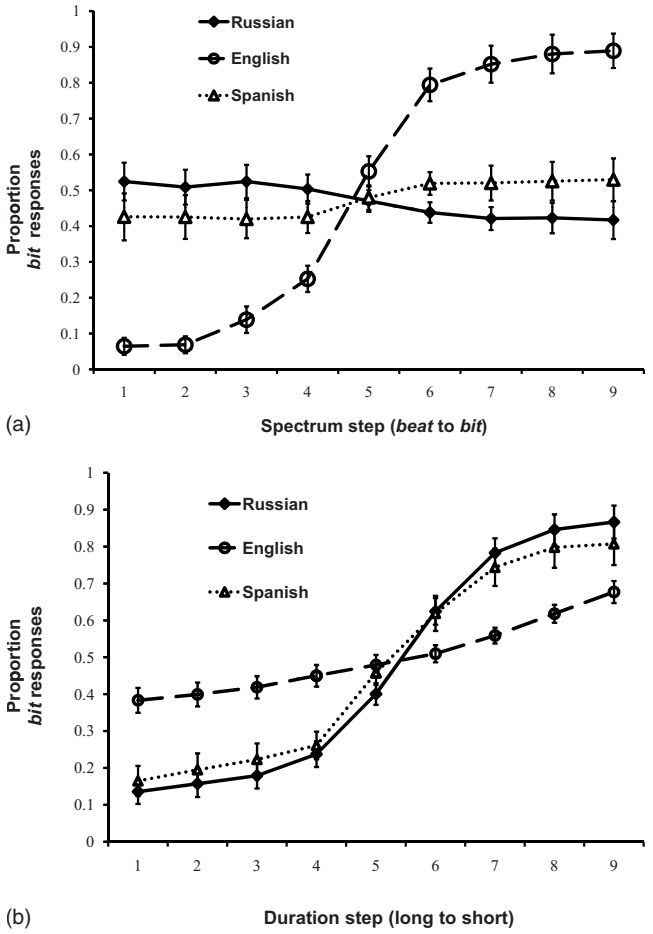


FIG. 1. Experiment 1: Proportion of bit responses to all stimuli averaged over spectrum (a) and duration (b). The error bars indicate standard error of the mean.

B. Results

Analysis using t-tests of independent means of the two groups' scores on each of the 13 factors previously shown to relate to performance on cross-language perceptual tests showed no significant difference between Spanish and Russian scores for 12 of the 13 factors, the exception being self-estimated degree of using ESL at work (Russian mean 9.5 and Spanish mean 7.7). This score notwithstanding, these results suggest that the two groups were generally well matched in English language experience.

ID functions for spectrum and duration are shown in Figs. 1(a) and 1(b), respectively. Separate but identical mixed factorial analyses of variance (ANOVAs) were carried out on spectrum and duration with one between group factor (group: English, Russian, Spanish) and one within group factor (steps 1 and 9). The results showed no significant effect of group [spectrum, $F(2,50)=0.07$, $p=0.93$; duration, $F(2,50)=1.33$, $p=0.274$], but a significant effect of step [spectrum, $F(2,50)=21.44$, $p<0.001$; and duration: $F(2,50)=167.61$, $p<0.001$] and a significant interaction between step and group [spectrum, $F(2,50)=21.99$, $p<0.001$; duration: $F(2,50)=9.30$, $p<0.001$]. These results suggest that there were differences between the English, Russian, and Spanish listeners in terms of how they used both spectral and duration properties.

Post hoc analysis (Tukey HSD, all significance levels reported at $p<0.05$ level or better) for the *spectrum* continuum demonstrated that only English listeners showed a significant difference between the proportion of bit responses at step 1 (means: English 0.06, Russian 0.52, and Spanish 0.42) and step 9 (means: English 0.88, Russian 0.42, and Spanish 0.53). For the *duration* continuum, all three groups showed a significant difference between the proportion of bit responses at step 1 (means: English 0.38, Russian 0.13, and Spanish 0.16) and step 9 (means: English 0.68, Russian 0.87, and Spanish 0.81). This suggests that English listeners used both duration and spectrum in classifying stimuli as beat or bit, while Russian and Spanish listeners used only duration.

Moreover, the pattern of the three groups' use of duration was also different. Comparing the three groups' performance at the endpoints of the duration continuum showed that English listeners gave significantly more bit responses at step 1 than Russian or Spanish listeners (means: English 0.38, Russian 0.14, and Spanish 0.16), with no significant difference between Russian and Spanish participants (contrary to predictions). Similarly, English listeners gave significantly fewer bit responses at step 9 (means: English 0.68,

Russian 0.87, and Spanish 0.81), but again there was no significant difference between Russian and Spanish (again, contrary to predictions). These results suggest that Russian and Spanish listeners relied more strongly on duration than did English listeners when classifying stimuli as beat or bit.

Similar results were obtained from the discriminant function analysis. For English listeners, a function that includes both spectrum and duration in the model was found to be significant, $F(2,78)=112.94$, $p<0.001$, with a partial Wilks's lambda of 0.26 for spectrum and 0.93 for duration. These results suggest that, for English listeners, both dimensions contributed to their classification of stimuli as beat or bit, but spectrum contributed more to their decision than did duration.

For Russian and Spanish listeners the discriminant function was significant with only one variable, duration, in the model: Russian: $F(2,78)=117.00$, $p<0.001$; Spanish, $F(2,78)=113.58$, $p<0.001$. The partial Wilks' lambda for spectrum was considerably closer to 1 for both groups (Russian 1.00 and Spanish 0.99) than was the value for duration (Russian 0.25 and Spanish 0.26). Again, these results corroborate the findings from the identification function analysis, showing that duration was the only property used by Russian and Spanish listeners to differentiate beat from bit, and that spectrum was not a useful predictor for either nonnative group (in contradiction to initial predictions that Russian listeners might rely more on duration than Spanish listeners).

In order to identify any effect of factors known to affect individual differences in cross-language perception of acoustic cues, forward stepwise multiple regression analysis was used to compare each individual's spectrum-to-duration ratio with the 13 individual factors described above. The final model was significant with six factors (period of EFL in the home country, AOA, self-estimated degree of using ESL at home, self-estimated degree of using ESL at work, age of starting EFL education in the home country, and period of formal education in the USA), $F(6,30)=3.94$, $p=0.005$. However, only one of these factors, period of EFL in the home country, was significant in its own right, $F=13.32$, $p=0.001$, $R=0.53$.

C. Discussion

English listeners used both spectrum and duration to identify tense and lax vowels, although, consistent with previous studies, spectral properties were weighted more heavily than duration (Hillenbrand *et al.*, 2000). In contrast, Spanish and Russian listeners employed only vowel duration, in agreement with previous findings for Spanish (Bohn, 1995; Escudero, 2006; Escudero and Boersma, 2004) and expanding the set of languages whose speakers have documented difficulty perceiving this contrast in a native English-like manner. There are a variety of possible explanations for these results, including those based on properties of the experimental design and those based on existing theories of cross-language perception and phonetic learning.

1. Experiment design considerations

Before interpreting these results further, it is important to note that the Spanish and Russian results might at first appear to suggest the effects of an experimental artifact, namely, the much greater range of stimuli distributed along the duration as opposed to the spectral dimension. Along the spectral dimension, stimuli ranged from the formant frequencies of a natural [i] to those of a natural [ɪ], but in duration the range was considerably exaggerated, especially toward the short end of the continuum. This was done in order to avoid the possibility that listeners might rely on a dimension with fewer steps (Bohn, 1995). On the other hand, it is possible that exaggerating duration but not spectral range encouraged listeners to preferentially attend to the exaggerated dimension alone (Holt, and Lotto, 2006). While such an explanation is possible, three observations mitigate against it. First, in the present experiment, the duration continuum ranged from 35 to 275 ms in 30 ms step intervals, a range of approximately ten JNDs (Klatt, 1976) determined that the JND for vowel duration is 25 ms for native English listeners. The spectrum continuum also spanned a range of ten JNDs, with F1 ranging from 326 to 443 Hz. Thus, although the duration range was exaggerated with respect to the natural span between the two categories, it was not *perceptually* enhanced with respect to the spectrum dimension. Second, Bohn (1995) and Flege *et al.* (1997) varied duration and spectrum within only their natural ranges and still found that Spanish listeners used duration as a primary dimension for classifying English tense and lax vowels. Finally, in the present study, native English listeners were not affected by the exaggerated duration range, as their pattern of response was quite similar to that found in previous studies (Hillenbrand *et al.*, 2000), suggesting that the design of the present experiment did not, in itself, influence listeners' cue weighting strategies.

2. Theoretical interpretations

The lack of a difference between Spanish and Russian listeners in terms of their use of duration in the perception of English [i] and [ɪ] might suggest that neither group is transferring their native experience with vowel duration to the perception of English [i] and [ɪ], instead resorting to the use of duration either on a language-independent psychoacoustic basis (Bohn, 1995) or as a consequence of applying residual first-language acquisition strategies (Escudero and Boersma, 2004). On the other hand, it is also possible that previously reported differences in the importance of duration as an acoustic cue in Russian (Bolla, 1981; Bondarko,1998) as compared to Spanish (Hammond, 2001; Hualde, 2005) may be overstated, perhaps because they are based strongly on acoustic analyses, demonstrating that the two languages exhibit differences in the way durational properties are produced, rather than on perceptual analyses, which would show the degree to which speakers of the two languages actually *use* duration as a cue when it is available to them.

3. Individual differences

It is also possible that the group similarities observed here actually obscure more significant patterns of individual

differences based on such variables as general English proficiency or experience with and education in the L2, or even the specific English dialects to which the learner is exposed (Escudero and Boersma, 2004). Although no independent test for English language proficiency was employed in the present study, we analyzed 13 factors (see Table I) known to affect degree of foreign accent in L2 and that are used widely in the literature to estimate foreign language proficiency (Piske *et al.*, 2001; McAllister *et al.*, 2002; Ylinen *et al.*, 2005). The two groups (Spanish and Russian) did not differ from one another according to 12 of the 13 measured variables suggesting that the two groups were well matched in terms of English proficiency. However, we did determine that, on an individual basis, the longer a participant had taken EFL instruction in their home country, the greater their reliance on spectrum over duration in the present task. This suggests that, whatever differences there might have been in terms of, for example, the different English dialects to which each listener was previously exposed, these were not as significant a factor as simply the amount of EFL education at home.

III. EXPERIMENT 2

In order to determine whether native Spanish and Russian listeners differ in their use of vowel duration as a perceptual cue to phonological contrasts in their native languages, we examined two cases in which vowel duration appears as an acoustic correlate of a phonological contrast, one suprasegmental and one segmental, in both languages as well as in English.

At the segmental level, vowels are longer before voiced than before voiceless consonants in English, Spanish, and Russian (Chen, 1970; Pye, 1986). This difference is considerably smaller in Spanish and Russian than it is in English, suggesting that the linguistic significance of allophonic vowel duration differences may be lower in these two languages than in English (Chen, 1970, Pye, 1986). In English, this difference in duration has been shown to function as an acoustic cue to voicing: A longer vowel cues the perception of voicing in the following stop consonant, while a shorter vowel cues the perception of voicelessness (Crowther and Mann, 1992; Raphael, 1972). Thus, in this case English listeners serve as a control group against which to measure the degree of Spanish and Russian listeners' reliance on vowel duration in the identification of final consonant voicing.

At the suprasegmental level, English and Russian stressed vowels are longer than unstressed vowels and these duration differences are sufficient to cue the perception of both English and Russian stress contrasts (Bondarko, 1998; Fry, 1955, 1958), although duration is arguably a weaker cue for English speakers [compared to fundamental frequency, vowel reduction, and intensity, (Beckman, 1986), while it is considered the strongest cue to stress differences for Russian speakers (Bondarko, 1998)]. In Spanish, stressed vowels are slightly longer than unstressed ones but in general syllable durations are much more equal than in English or Russian (Hammond, 2001).

Based on these findings, we would predict that Spanish and Russian listeners should perform comparably when using vowel duration to classify postvocalic stop consonant voicing contrasts, but they should show a weaker reliance on this cue than do English listeners. At the same time, when classifying stress contrasts, Russian and English listeners should show a much stronger reliance on duration cues than Spanish listeners, and Russian listeners may show a greater reliance on this cue than do English listeners. On the other hand, the results of the first experiment suggest that, perhaps, Spanish and Russian listeners are not all that different in terms of their native language experience with vowel duration. In this case, we would expect to find no differences between Spanish and Russian listeners' reliance on vowel duration in these native language contrasts: Both should be less reliant than English speakers in the consonant voicing contrast and equally reliant on duration in the lexical stress contrast.

A. Method

1. Participants

Participants were the same as in experiment 1.

2. Stimuli

Stimuli for experiment 2 consisted of three sets: *fápa/fapá, lapko/labko*, and *lapo/labo* created according to the following procedure. A 30-year-old male native speaker of Chilean Spanish[2] produced several examples of the words *fápa, fapá, lapko, labko*, and *lapo, labo* in isolation. All tokens were recorded using the same procedures as in experiment 1 except that the initial recording was made using a Marantz Professional solid state recorder (PMD660) directly to wav format (sampling frequency 22 050 Hz), eliminating the need to redigitize the initial recordings for subsequent processing. From the set of produced words, one token each of *fápa, fapá, lapko, labko*, and *lapo, labo* was selected using the same criteria as in experiment 1.

Evaluating postconsonantal voicing using stimuli that are acceptable nonwords in both Russian and Spanish is particularly difficult, due to a number of phonological rules involving stop consonant voicing in both languages. Stimuli ending in the target consonant (e.g., CVC syllables) were not useable due to a process of word-final voicing neutralization in Russian. On the other hand, voiced stop consonants in intervocalic position (e.g., VCV) are not typically acceptable in Spanish, due to a process of intervocalic spirantization. Ultimately, it was decided to use two different types of stimuli, *lapko/labko*, and *labo/lapo* in order to have one block (*lapko/labko*) in which the Russian listeners would be, if anything, biased toward a voiceless response (due to a process of regressive voicing assimilation in consonant clusters, Bondarko, 1977), and the other in which Spanish listeners would be biased toward voiceless responses (*lapo/labo*) due to the lack of intervocalic spirantization of voiced stops (Hammond, 2001). The methods for generating each set were given as follows.

(a) *lapko/labko*. The stimuli for this experiment were a set of ten syllables of the form /*laCko* / (where C is ambiguous

between [p] and [b]) created according to the following procedure. The duration of the [a] vowel in the first syllable of each word was measured from the end of the transition from [l] to the last zero crossing of the vowel waveform before the stop gap of the following [p] or [b]. The stop gap was also measured. The token *lapko* was chosen as the base for manipulation because it was technically easier to reduce the amplitude of the [p] burst to make it more ambiguous, than it would have been to eliminate voicing in the closure and increase the amplitude of a weak [b] burst to make a *labko* token more ambiguous. By trial and error the stop gap was reduced by 13% and the burst amplitude was decreased by 97% so that the resulting token was perceived to be ambiguous between [p] and [b] by native speakers of English, Russian, and Spanish (not participants in the experiment). The duration of the first vowel [a] in the ambiguous token was then varied to range from 95 to 230 ms in ten steps (15 ms/step). The shortest step was chosen to be 95 ms because it was the shortest duration, which sounded natural to English, Spanish, and Russian listeners (not participants in the experiment) in the newly created token.

(b) *lapo/labo*. These stimuli were derived from the *lapko/labko* stimuli described above. To create these stimuli, the burst and stop gap associated with the consonant [k] in the *lapko/labko* stimuli were deleted immediately prior to manipulating the duration of the first vowel [a]. All other processes were identical. The beginning of the [k] was defined as the first zero crossing preceding the first upward-going excursion of the burst release. The end was defined as the first upward-going zero crossing at the onset of periodicity for the following [o]. The remaining gap between the end of the /p/ burst and the beginning of the /o/ vowel in the second syllable was reduced by 94% to make the Voice Onset Time (VOT) of the [p] consonant ambiguous for native speakers of English, Spanish, and Russian.

(c) *fápa/fapá*. The stimuli for this experiment were of a set of ten syllables of the form [fapa] created according to the following procedure. Duration, intensity, and *f*0 of the [a] in the first syllable were measured from the end of the noise portion of the fricative /f/ to the last zero crossing of the vowel waveform before the silent gap, and in the second syllable from the end of the release of the [p] closure to the last zero crossing of the vowel waveform. Using the PRAAT 4.1.21 manipulation editor (PSOLA resynthesis), the intensity and *f*0 of each vowel in both *fápa* and*fapá* were adjusted to the average value of the vowels in that position. Next, the first syllable in the manipulated *fápa* token and the second syllable in the manipulated *fapá* token (that is, the two "average" syllables derived from originally stressed syllables) were concatenated together into a new token of the form [fapa]. Finally, a ten step continuum was created by setting the duration of the first vowel [a] in this newly created ambiguous token so that it would range from 50 to 230 ms in ten (20 ms) steps. A shortest step of 50 ms in the newly created token was chosen because it was the minimal step, which sounded natural to four native speakers of English, Spanish, and Russian who were not participants in the experiment.

3. Procedure

On each trial listeners heard one stimulus with a simultaneous visual presentation of two possible answers, e.g., *fápa* and *fapá* (*lapko* and *labko*; *lapo* and *labo*). Russian participants saw these words written in Cyrillic. Responses were always in this order, left to right, corresponding to the left and right buttons on a response box. All participants were instructed that they would hear nonsense words produced by a native speaker of their language and that (in the *fápa/fapá* condition) an accent above the vowel in writing indicates lexical stress (this orthographic stress marking is conventional in all three languages, but more familiar to readers of Russian and Spanish than English). After participants pressed a button on the response pad corresponding to one of the two choices, there was a blink of the screen (250 ms) and another trial began. Each trial was self-paced with no limit on time to respond. In each block there were ten stimuli repeated ten times in random order. Average running time for each block was 10 min. Before each block began, all participants completed a practice session consisting of four trials using endpoint tokens from that block (steps 1 and 10). The structure of the practice session was identical to the experiment but responses were not recorded.

4. Analysis

(a) *Identification functions.* ID functions for the *lapko/labko* and *lapo/labo* series were calculated as the proportion of responses containing [b] sound (*labko* or *labo*) to all ten stimuli. The ID function for the *fápa/fapá* series was calculated as the proportion of second syllable stressed responses (*fapá*) to all ten stimuli. All ID functions were calculated as average group responses.

(b) *Logistic regression analysis.* Logistic regression functions were fitted to each subjects' identification data using statistical software (StatSoft, Inc, 2003). The logistic regression equation (Nissen, *et al.*, 2005) is

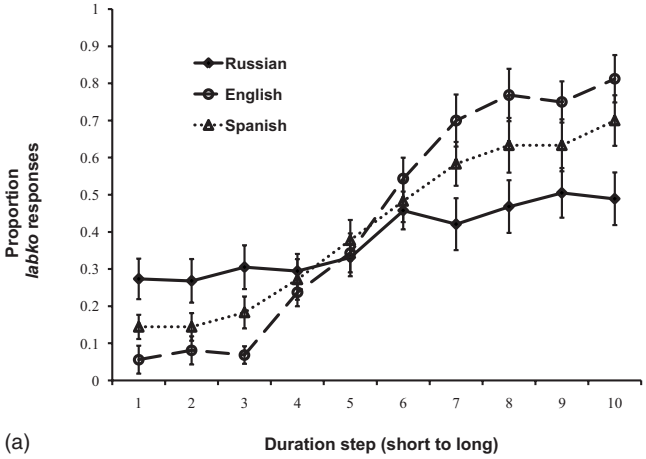
$$\log \frac{p}{1-p} = a + bx, \tag{1}$$

where p is the proportion correct, a is the regression intercept, b is the regression slope, and x is a stimulus level. The regression slope b demonstrates the steepness of the identification function slope, while the 50% crossover point of the identification function is derived by the following equation:

$$x = \frac{\log \frac{0.5}{1-0.5} - a}{b} = \frac{-a}{b}. \tag{2}$$

B. Results

ID functions for each stimulus set (*lapko/labko, lapo/labo*, and *fápa/fapá*) are shown in Figs. 2(a), 3(a), and 4(a),



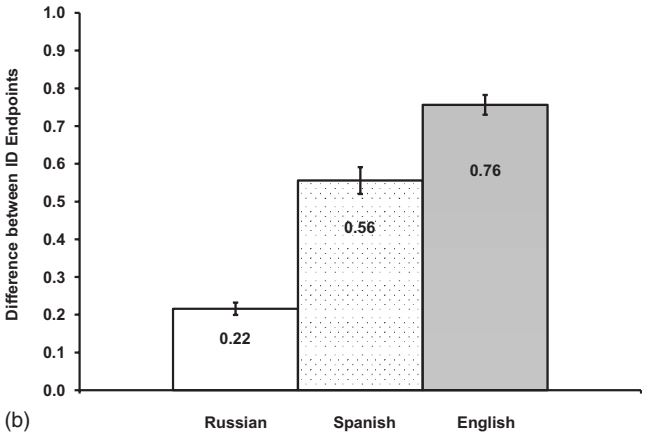
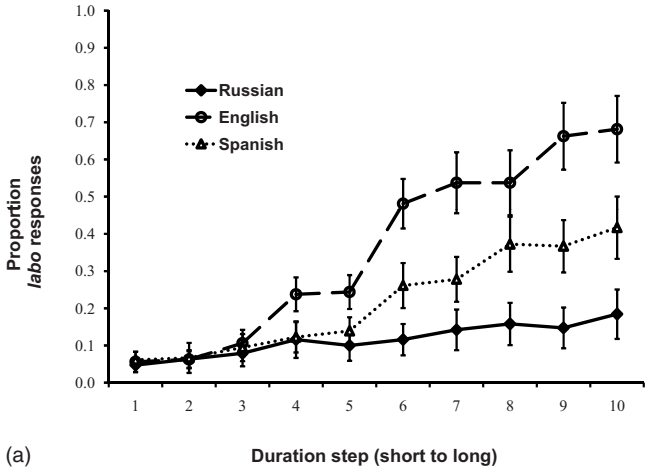


FIG. 2. Experiment 2: (a) Proportion of *labko* responses (voiced consonant) for each group of listeners at each step of duration of the first vowel (step 1=95 ms to step 10=230 ms). The error bars indicate standard error of the mean. (b) Difference between the endpoints (step 10 minus step 1) of the proportion of *labko* responses (voiced consonant) for each group of listeners. The error bars indicate standard error of the mean.



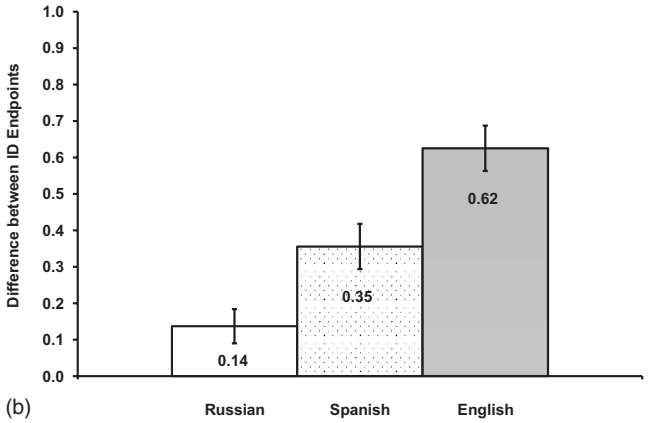


FIG. 3. Experiment 2: (a) Proportion of *labo* responses (voiced consonant) for each group of listeners at each step of duration of the first vowel (step 1=95 ms to step 10=230 ms). The error bars indicate standard error of the mean. (b) Difference between the endpoints (step 10 minus step 1) of the proportion of *labo* responses (voiced consonant) for each group of listeners. The error bars indicate standard error of the mean.

respectively. In order to compare the strength of each groups' reliance on duration, a measure of "duration reliance" was calculated as the difference between the endpoints of each continuum for each listener in each group (step 10 minus step 1), as shown in Figs. 2(b), 3(b), and 4(b), respectively. A mixed factorial ANOVA with one between group factor group (English, Spanish, Russian) and one within group contrast (*lapko/labko, lapo/labo*, and *fápa/fapá*) was run on the duration reliance measure for each stimulus set. Planned comparisons of selected means were conducted to identify significant differences between groups within contexts, and between contexts within groups.

Overall, there was a significant effect of group, $F(2,50)=14.91$, $p<0.001$, and contrast, $F(2,100)=50.48$, $p<0.001$, and a significant interaction between group and contrast, $F(4,100)=4.87$, $p=0.001$, suggesting that the three language groups used vowel duration differently as a cue to postvocalic consonant voicing as compared to lexical stress.

Planned comparison of means for the *lapko/labko* series [Fig. 2(b)] showed a significant difference between all pairs of groups: Spanish and English, $F(1,50)=6.67$, $p=0.01$; Russian and Spanish, $F(1,50)=10.12$, $p=0.003$, and Russian and English, $F(1,50)=32.40$, $p<0.001$. These results suggest that English listeners were more successful at using vowel duration for the identification of final consonant voicing than were Spanish and Russian listeners. In turn, Spanish listeners showed greater reliance on duration than did Russian listeners, which is exactly opposite the original prediction based on previously published descriptions of Spanish and Russian that suggested that the two groups should be equivalent in their use of duration as a cue to final consonant voicing.

Planned comparison of means for the *lapo/labo* series [Fig. 3(b)] demonstrated a significant difference between Spanish and English, $F(1,50)=5.89$, $p=0.02$; Russian and English, $F(1,50)=18.15$, $p<0.001$, but not Russian and Spanish, $F(1,50)=3.36$, $p=0.07$; listeners. These results again suggest that English listeners were more successful in using vowel duration for the identification of final consonant voicing than native Spanish and Russian listeners. The low proportion of voiced responses by Spanish listeners might be explained by the presence of a regular phonological process of spirantization of voiced stops in CVC position, meaning that Spanish listeners might be predisposed to hear all intervocalic stops (that are not spirantized) as voiceless. However, this does not explain the Russian listeners' performance, which was comparable to that of Spanish listeners. Thus, the importance of vowel duration is reduced before a C

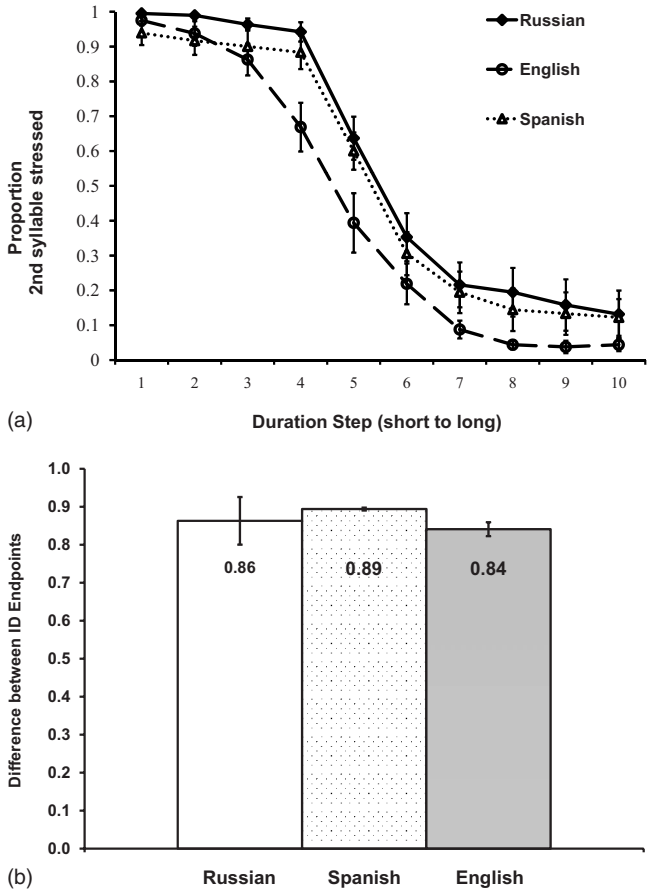


FIG. 4. Experiment 2: (a) Proportion of *fapá* (second syllable stressed) responses for each group of listeners at each step of duration of the first vowel (step 1=50 ms to step 10=230 ms). The error bars indicate standard error of the mean. (b) Difference between the endpoints (step 10 minus step 1) of the proportion of *fapá* (second syllable stressed) responses for each group of listeners. The error bars indicate standard error of the mean.

followed by a V, because in this context frication is presumably a stronger, more reliable cue to voicing than is preceding vowel duration.

Planned comparison of means for the *fápa/ fapá* series showed that there was no significant difference between any pair of groups: Russian and English, $F(1,50)=0.7$, $p=0.41$; Russian and Spanish, $F(1,50)=0.35$, $p=0.56$ and English and Spanish, $F(1,50)=1.93$, $p=0.17$. These findings suggest that all three groups were able to use vowel duration to the same extent as a cue to lexical stress differences.

Because of language-specific differences in extraneous phonological processes that might have affected perception of voicing in the two segmental contexts (*lapko/labko* and *lapo/labo*), planned comparisons of means were used to compare duration reliance in each context within each group. Results showed a significant difference between the reliance on duration in *lapko/labko* versus. *lapo/labo* contexts for English, $F(1,50)=6.19$, $p=0.02$ and Spanish, $F(1,50)=8.09$, $p=0.006$, but not Russian, $F(1,50)=2.53$, $p=0.12$, listeners. The lack of an observed difference for Russian listeners, who showed little evidence of using vowel duration as a cue to consonant voicing in either context, suggests that this property, although it is available in the acoustic patterns of Russian speech productions, is not weighted very heavily as a perceptual cue to voicing for Russian speakers. On the other hand, the observation that both Spanish and English listeners show less reliance on duration in the VCV as compared to the VCC context suggests that the Spanish pattern may not be solely due to the influence of a process of the spirantization of voiced stops in VCV position in Spanish, since no such process exists in English. It is possible that, in the case of English (and possibly Spanish speakers), the weight given to vowel duration may differ across phonetic contexts, such that in VCV context other acoustic cues, perhaps including the strength of the release burst, the presence of voicing during closure, onset *f*0 of the following vowel or F1 cutback (cf. Lisker, 1986) may contribute proportionally more to the identification of the consonant voicing than preceding vowel duration, which, in turn, receives greater weight when some of these other cues are unavailable due to the presence of a following consonant.

Because the measure of duration reliance is not so standard, further analyses using a more typical measure, logistic regression, were carried out. For each participant whose data could be summarized in this manner, a logistic curve was fit ted to their individual ID function, and the slope of the function and its 50% crossover point were computed (see Table II for group results and the number of people in each group who showed analyzable results). These data were analyzed separately for each context using a one-way ANOVA with three levels of one factor (language).

For the slope of the ID function, a significant effect of language was found only for the *lapko/labko* and *lapo/labo* series, $F(2,50)=9.89$, $p<0.001$ and $F(2,50)=10.67$, $p<0.001$, respectively. Post hoc comparisons were significant for English versus Russian in the *lapko/labko* series, while in the *lapo/labo* series English listeners differed significantly from both Russian and Spanish listeners. These results corroborate the findings of the duration reliance analysis above, suggesting that English listeners showed stronger categorization of consonant voicing on the basis of preceding vowel duration than did either Russian or Spanish listeners, though their difference from Spanish speakers disappeared in the *lapko/labko* context. In the *fápa/ fapá* series there was no significant difference between the groups, suggesting that all three used duration equally for identifying lexical stress.

When analyzing the 50% points, only three Russian participants showed ID curves that crossed the 50% line for either the *lapko/labko* or *lapo/labo* series. Thus, an ANOVA on the location of the 50% crossover point was run only on English and Spanish data for these two continua, but for the *fápa/ fapá* series all three language groups were included. For all three continua, no significant difference was found between English and Spanish listeners. However, in the *fápa/ fapá* series, there was a significant difference between English (step 4.4=118 ms) and Russian (step 5.6=142 ms) listeners. This difference may suggest that these listeners utilized their native language allophonic experience with vowel duration while listening to tokens. Notice that in Russian the duration of a stressed [a] vowel is 166 ms and the duration of the unstressed vowel is 95 ms. (Bolla, 1981), suggesting that a boundary midway between them should be around 130 ms, perhaps higher because other (palatalized) allophones of [a]

TABLE II. Steepness of the slope and location of the 50% crossover point of the identification functions for native speakers of English, Spanish, and Russian in experiment 2. The No. of listeners refers to the number of listeners exhibiting a quantifiable slope (col. 3) or identifiable 50% crossover point (col. 5), out of the total number of listeners in the specific language group.

		Slope		Location of 50% Crossover	
Series	Group	No. of listeners	Steepness (B1) (std)	No. of listeners	Step (std) (ms)
lapko/labko	English	16/16	0.71(0.35)	16/16	5.9 (2.29), 168
	Spanish	18/18	0.47 (0.4)	15/18	6.2 (1.25), 173
	Russian	17/19	0.24 (0.29)	5/19	5.7 (0.61), 165
lapo/labo	English	14/16	0.55 (0.40)	13/16	6.5 (2.76), 177
	Spanish	13/18	0.30 (0.22)	9/18	7.3 (0.93), 189
	Russian	7/19	0.08 (0.15)	3/19	6.4 (1.46), 176
fápa/fapá	English	16/16	1.5 (0.81)	16/16	4.4 (0.8), 118
	Spanish	18/18	1.45 (1.57)	17/18	5.3 (1.36), 136
	Russian	18/19	1.35 (0.71)	17/19	5.6 (1.09), 142

are all longer than 166 ms. Thus, it is possible that the pattern observed in the present study reflects a difference in the location of native boundaries between longer and shorter phonetic categories.

Finally, a correlation analysis was run between the duration reliance measure in native language contrasts (averaged over all three contexts in experiment 2) and the duration reliance measure calculated as the proportion of bit responses at step 9 minus step 1 along the duration continuum in experiment 1. For Spanish listeners, the correlation was significant, $r=0.57$, $p=0.01$; however, for Russian listeners, the correlation was not significant. This pattern of results suggests that there may be a connection between the use of vowel duration in the perception of non-native contrast and the use of duration in native contrasts for Spanish, but not for Russian, listeners.

C. Discussion

The results of the second experiment show an unexpected pattern of difference between Spanish and Russian listeners' use of vowel duration as a cue to native phonological contrasts. On the one hand, the results of the lexical stress experiment suggest that Russian, Spanish, and English speakers employ vowel duration to the same degree when identifying stress contrasts in their native languages despite the very different roles that duration plays in the stress systems of the three languages. Although these finding disagree with predictions based on traditional conceptualizations of the respective native language phonologies, they are consistent with previous findings that native Spanish speakers are able to use vowel duration as a cue to lexical stress (Diaz-Campos, 2000; Lecumberri, 2006; Manrique, and Signorini, 1983, also see Llisterri *et al.*, 2003 for opposite conclusions).

Similarly, the results of the consonant voicing experiments suggest that Spanish and Russian listeners relied less on vowel duration as an allophonic cue to following consonant voicing than did native English listeners, as might be predicted based on the results of previous production research (Chen, 1970). On the other hand, the finding that Spanish listeners rely more on vowel duration as a cue to consonant voicing than do Russian listeners is unexpected. In previous production studies, both Spanish and Russian listeners were shown to have almost identical ratios of vowel duration before voiceless consonants to vowel duration before voiced consonants (Spanish, 0.86; Russian, 0.82 or 0.85–0.90).[3] However, our results suggest that, perceptually, the two languages may differ in terms of the degree to which they rely on duration as a cue to following consonant voicing (little or none in Russian, more so in Spanish). Finally, our results suggest that English, and possibly Spanish listeners, may shift the weight they give to a particular cue (in this case vowel duration) to a particular contrast depending on the availability of other cues in a given phonetic context.

It must be noted that these results were derived from listeners with some experience listening to English, and it is possible that their behavior reflects transfer of experience with vowel duration in the L2 (English) back to their respective native languages as suggested by Flege's (1995) speech learning model, or perhaps activation of an "English mode" of listening (Grosjean, 1998). While this is certainly possible, the present findings agree clearly with the results of previous perceptual studies of monolingual Russian and Spanish listeners (Bondarko, 1998; Diaz-Campos, 2000; Manrique and Signorini, 1983). Moreover, differences between the two groups in the use vowel duration as a cue to syllable-final consonant voicing suggest that both groups were influenced by native phonological constraints during performance of the present task, and thus were not listening in a purely English-like manner. Thus, while we cannot rule out the possibility that these results reflect transfer back from the L2 to the L1, a simpler explanation is that Spanish listeners, like Russian listeners, do, in fact, have native language experience with using duration as a phonetic cue, at least at the prosodic level and, furthermore, that Spanish listeners may even have *more* experience using vowel duration as a cue to consonant voicing than do Russian listeners.

IV. GENERAL DISCUSSION

The results of experiment 1 provide further confirmation that non-native listeners weight the multiple acoustic cues to the English tense-lax vowel contrast differently than do native listeners. In agreement with previous studies, English listeners gave more weight to spectrum than to duration, but used both to some degree (Ainsworth, 1972; Hillenbrand *et al.*, 1995, 2000; Mermelstein, 1978), while native Spanish listeners gave no weight to spectrum and used duration alone (Bohn, 1995; Escudero, 2006; Escudero and Boersma, 2004). Russian listeners also showed little or no reliance on spectrum, using duration exclusively, in a manner comparable to Spanish listeners. This pattern of performance is consistent with a variety of more substantive explanations, including the transfer of native language experience with vowel duration as a perceptual cue, the influence of language-universal psychoacoustic biases (Bohn, 1995), and the application of residual L1 acquisition strategies (Escudero and Boersma, 2004).

With respect to the hypothesis of allophonic experience, the results of experiment 2 suggest that Spanish and Russian listeners rely to a similar degree on duration as a cue to lexical stress placement, consistent with the results of previous studies on monolingual speakers of both languages (Bolla, 1981; Diaz-Campos, 2000; Manrique and Signorini, 1983; Prieto, and van Santen, 1996). Thus, it is possible that both Spanish and Russian listeners are able to transfer their experience with vowel duration at an allophonic level to the perception of the English tense and lax vowel contrast.

However, the finding that Spanish listeners rely *more* on vowel duration as a cue to native consonant voicing than do Russian listeners (in experiment 2) but the two groups do not differ in their reliance on vowel duration in the perception of a non-native vowel contrast (in experiment 1) suggests that native allophonic experience does not explain all aspects of cross-language perception in this case.

Another possible explanation for Spanish listeners' use of duration in the non-native contrast in question, one that may also account for Russian listeners' performance in the present study, is the relative salience of the two acoustic properties (vowel duration and vowel quality). Acoustic dimensions of speech may differ in their relative salience (Burnham, 1986; Walley and Carrell, 1983) and the ease with which they can be learned (Francis *et al.* 2000; Francis and Nusbaum, 2002). In principle, unequal sensitivity to two dimensions could arise innately (Holt and Lotto, 2006) or, in the present case it could arise as a consequence of the smaller (as compared to English) vowel inventory of both Russian and Spanish, which might make speakers of these languages less sensitive to vowel quality differences (Hacquard *et al.*, 2007) and thus more likely to cue in on duration differences. This explanation is consistent with Bohn's (1995) *desensitization hypothesis*, in which it is proposed that duration will be employed regardless of the listener's degree of L1 experience with it when spectral properties are hard to distinguish.

Alternatively, Escudero and Boersma (2004) proposed that monolingual Spanish speakers start with no phonological categories distinguished according to duration, due to their limited experience with this feature in their native phonological and phonetic systems. Therefore, when they first encounter a phonological contrast that involves duration differences, they apply a first-language (L1) acquisition strategy based on distributional learning. Since they are able to detect two peaks in the distribution of [i] and [ɪ] tokens along the duration dimension, they learn to divide the duration continuum into two categories. This argument applies to some degree with respect to the Spanish listeners' performance in the present study, but requires some modification with respect to the Russian data.

The results of experiment 1 are consistent with the Escudero and Boersma (2004) hypothesis that Spanish listeners have learned the distribution of tense and lax vowels in English on the basis of duration alone using an L1 (distribution-based) acquisition mechanism, so long as we assume that Spanish listeners do not have sufficient phonetic experience with duration to have developed native duration-based categories. If we make this assumption, we must then interpret the results of experiment 2 as providing evidence for the transfer of experience with duration (in English) back to Spanish, a phenomenon that has been observed in other studies (Flege and Eefting, 1987) and is quite compatible with the speech learning model of L2 acquisition proposed by Flege (1995).

On the other hand, Russian listeners cannot be assumed to have developed duration-based category representations of English vowels based on L1 acquisition mechanisms, because they must already have native, duration-based categories (e.g., for stress, Bolla, 1981). In this case, the present data must be interpreted as Russian listeners' assimilating [i] to a longer (stressed) category, and [ɪ] to a shorter one (perhaps prestressed), in a pattern of two-category assimilation (Best, 1995). A similar explanation has been advanced by Ylinen *et al.* (2005) to explain monolingual Russian listeners' treatment of the Finnish vowel length contrast. This interpretation is also consistent with the observed lack of a significant correlation found between the results of experiments 1 and 2. If Russian listeners, like Spanish listeners, were using L2-based categories for the tense/lax contrast in experiment 1, we would expect some transfer of these newly learned contrasts back to their perception of contrasts that, in English, depend on duration (e.g., consonant voicing). The fact that Spanish, but not Russian, listeners show a pattern consistent with backward transfer of experience with duration, suggests that the Escudero and Boersma (2004) model does not apply in the case of Russian listeners, since they are able to apply established L1 categories to their perception of the L2 contrasts and do not need to invoke L1 learning mechanisms.

It seems plausible that all of the major factors discussed here, including native allophonic as well as phonological experience, individual differences in experience with specific dialects, and innate psychoacoustic proclivities, may contribute to some degree to the relative weighting of cues in cross-language speech perception. We have shown that native Spanish and Russian listeners have sufficient allophonic experience with duration in their native language to enable a

potential transfer to English. However, it is still possible that the overall dimension of duration is somehow more salient than spectrum for reasons unrelated to the listeners' native language. Our results cannot rule out a basic psychoacoustic explanation (Bohn, 1995), and are quite consistent with an explanation based on the application of residual L1 learning mechanisms (Escudero and Boersma, 2004). Future research to distinguish between these possibilities should employ a more controlled evaluation of general proficiency in the foreign language as well as the examination of truly monolingual listeners in comparison to those with more exposure to the foreign language.

ACKNOWLEDGMENTS

We are grateful to Robert Hammond, Margie Burns, and two anonymous reviewers for helpful suggestions and discussion related to earlier versions of this article. A preliminary version of the results of experiment 1 was presented at the 148th Meeting of the Acoustical Society of America, San Diego, CA, 15–19 November 2004, and some of the results of experiment 2 were presented at the 4th Joint Meeting of the Acoustical Society of America and the Acoustical Society of Japan, Honolulu, HI, 28 November–2 December 2006. Part of this research was supported by a grant from the College of Liberal Arts, Purdue University, to M. K., and by NIDCD Grant No. R03 DC006811 to A.L.F.

[1]Note that this pattern is the opposite of that found in the English [i]-[ɪ] contrast.

[2]The speaker was a former phonetics teacher and was fluent in English as well, and was specifically instructed to produce the stimuli in as "accent-free" a manner as possible (i.e., avoiding intervocalic spirantization). Russian listeners all reported that the stimuli sounded like Russian, while Spanish listeners reported that they sounded like Spanish, suggesting that the perceived linguistic background of the talker would not have affected listeners' responses in any language-dependent manner.

[3]Spanish values from Chen (1970), Russian values from Chen (1970) and calculated from values in Pye, (1986), Table 3.

Ainsworth, W. A. (**1972**). "Duration as a cue in the recognition of synthetic vowels," J. Acoust. Soc. Am. **51**, 648–651.

Beckman, M. E. (**1986**). *Stress and Non-Stress Accent*, Netherlands Phonetic Archives Vol. **7**, Foris Dordrecht.

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

Best, C. T., and Strange, W. (**1992**). "Effects of language-specific phonological and phonetic factors on cross-language perception of approximants," J. Phonetics **20**, 305–330.

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

Bolla, K. (**1981**). *A Conspectus of Russian Speech Sounds* (Akademiai Kiado, Budapes).

Bondarko, L. V. (**1977**). "*Zvukovoi stroi sovremennogo russkogo jazyka (Phonetic structure of modern Russian language)*," Moscow, Russia: Proveshenie.

Bondarko, L. V. (**1998**). "*Fonetika Sovremennogo Russkogo Jazyka (Phonetics of Modern Russian Language)*," University of St.-Petersburg, St.-Petersburg, Russia.

Burnham, D. K. (**1986**). "Developmental loss of speech perception: Exposure to and experience with a first language," Appl. Psycholinguist. **7**, 207–240.

Chen, M. (**1970**). "Vowel length variation as a function of the voicing of the consonant environment," Phonetica **22**, 129–159.

Crowther, C. S., and Mann, V. (**1992**). "Native language factors affecting use of vocalic cues to final consonant voicing in English," J. Acoust. Soc. Am. **92**, 711–722.

Diaz-Campos, M. (**2000**). "The phonetic manifestation of secondary stress in Spanish," in *Hispanic Linguistics at the Turn of the Millennium: Papers From the third Hispanis Linguistic Symposium*, edited by H. Campos, E. Herburger, A. Morales-Front, and T. J. Walsh (Cascadilla, Somerville, MA).

Escudero, P. (**2006**). "The phonological and phonetic development of new vowel contrasts in Spanish learners of English," in *English With a Latin Beat: Studies in Portugues/Spanish-English Interphonology, Studies in Bilingualism*, Vol. **31**, edited by B. O. Baptista, and M. A. Watkins (John Benjamins, Amsterdam), pp. 149–161.

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

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, MD), pp. 233–277.

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

Flege, J. E., and Eefting, W. (**1987**). "Production and perception of English stops by native Spanish speakers," J. Phonetics **15**, 67–83.

Fox, R. A., Flege, J. E., and Munro, M. J. (**1995**). "The perception of English and Spanish vowels by native English and Spanish listeners: A multidimensional scaling analysis," J. Acoust. Soc. Am. **97**, 2540–2551.

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

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

Fry, D. B. (**1955**). "Duration and intensity as physical correlates of linguistic stress," J. Acoust. Soc. Am. **27**, 765–768.

Fry, D. B. (**1958**). "Experiments in the perception of stress," Lang Speech **1**, 126–152.

Grosjean, F. (**1998**). "Studying bilinguals: Methodological and conceptual issues," Bilingualism: Lang. Cognit. **1**, 131–149.

Hacquard, V., Walter, M. A., and Marantz, A. (**2006**). "The effect of inventory on vowel perception in French and Spanish: An MEG study," Brain Lang **100**, 295–300.

Halle, P. A., Best, C., and Levitt, A. (**1999**). "Phonetic vs. phonological influences on French listeners' perception of American English approximants," J. Phonetics **27**, 281–306.

Hammond, R. M. (**2001**). *The Sounds of Spanish: Analysis and Application (With Special Reference to American English)* (Cascadilla, Somerville, MA).

Harris, J. (**1969**). *Spanish Phonology* (MIT, Cambridge, MA).

Hillenbrand, J., Clark, M. J., and Houde, R. A. (**2000**). "Some effects of duration on vowel recognition," J. Acoust. Soc. Am. **108**, 3013–3022.

Hillenbrand, J., Getty, L. A., Clark, M. J., and Wheeler, K. (**1995**). "Acoustic characteristics of American English vowels," J. Acoust. Soc. Am. **97**, 3099–3111.

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

Hualde, J. I. (**2005**). *The Sounds of Spanish* (Cambridge University Press, Cambridge).

Jamieson, D. G., and Moroson, D. E. (**1986**). "Training non-native speech contrasts in adults: Acquisition of the English /ð/-/θ/ contrast by francophones," Percept. Psychophys. **40**, 205–215.

Jenkins, J. J., and Yeni-Komshian, G. H. (**1995**). "Cross-language speech perception: Perspective and promise," in *Speech Perception and Linguistic Experience: Issues in Cross-Language Speech Research*, edited by W. Strange (York, Baltimore, MD), pp. 463–479.

Kewley-Port, D., and Watson, C. S. (**1994**). "Formant-frequency discrimination for isolated English vowels," J. Acoust. Soc. Am. **95**, 485–496.

Klatt, D. H. (**1976**). "Linguistic use of segmental duration in English: Acoustic and perceptual evidence," J. Acoust. Soc. Am. **59**, 1208–1221.

Lecumberri, L. G. (**2006**). "Perception of double stress by Spanish learners of English," in *English With a Latin Beat: Studies in Portugues/Spanish-English Interphonology*, edited by B. O. Baptista and M. A. Whatins (John Benjamins, Amsterdam).

Lisker, L. (**1986**). ""Voicing" in English: A catalogue of acoustic features signaling /b/ versus /p/ in trochees," Lang Speech **29**, 3–11.
Llisterri, J., Machuca, M., de la Mota, C., Riera, M., and Rios, A. (**2003**). "The perception of lexical stress in Spanish," in *Proceedings of the 15th International Congress of Phonetic Sciences*, edited by M. J. Sole, D. Recasens, and J. Romero, Barcelona, 3–9 August, casual productions [CD-ROM].
Manrique, A. M. B., and Signorini, A. (**1983**). "Segmental duration and rhythm in Spanish," J. Phonetics **11**, 117–128.
McAllister, R., Flege, J. E., and Piske, T. (**2002**). "The influence of L1 on the acquisition of Swedish quantity by native speakers of Spanish, English and Estonian," J. Phonetics **30**, 229–258.
McClaskey, C. L., Pisoni, D. B., and Carrell, T. D. (**1983**). "Transfer of training of a new linguistic contrast in voicing," Percept. Psychophys. **34**, 323–330.
Mendoz, E., Garballo, G., Cruz, A., Fresneda, M. D., Munoz, J., and Marrero, V. (**2003**). "Temporal variability in speech segments of Spanish: Context and speaker related differences," Speech Commun. **40**, 431–447.
Mermelstein, P. (**1978**). "On the relationship between vowel and consonant identification when cued by the same acoustic information," Percept. Psychophys., **23**, 331–336.
Nissen, S. L., Harris, R. W., Jennings, L., Eggett, D. L., and Buck, H. (**2005**). "Psychometrically equivalent trisyllabic words for speech reception threshold testing in Mandarin," Int. J. Audiol. **44**, 391–399.
Piske, T., MacKay, I. R. A., and Flege, J. E. (**2001**). "Factors affecting degree of foreign accent in an L2: A review," J. Phonetics **29**, 191–215.
Pisoni, D. B., Aslin, R. N., Perey, A. J., and Hennessy, B. L. (**1982**). "Some effects of laboratory training on identification and discrimination of voicing contrasts in stop consonants," J. Exp. Psychol. Hum. Percept. Perform. **8**, 297–314.
Prieto, P., and van Santen, J. (**1996**). "Secondary stress in Spanish: Some experimental evidence.," *Aspects of Romance Linguistics: Selected Papers From the Linguistic Symposium on Romance Languages*, Vol. **XXIV** pp. 337–356.
Pruitt, J. S., Jenkins, J. J., and Strange, W. (**2006**). "Training the perception of Hindi dental and retroflex stops by native speakers of English and Japanese," J. Acoust. Soc. Am. **119**, 1684–1696.
Pye, S. (**1986**). "Word-final devoicing of obstruents in Russian," Cambridge Papers in Phonetics and Experimental Linguistics **5**, 1–10.
Raphael, L. J. (**1972**). "Preceding vowel duration as a cue to the perception of the voicing characteristic of word-final consonants in American English," J. Acoust. Soc. Am. **51**, 1296–1303.
Schneider, W., Eschman, A., and Zuccolotto, A. (**2002**). *E-Prime User's Guide* (Psychology Software Tools Inc., Pittsburg, PA).
StatSoft, Inc. (**2003**). STATISTICA (*data analysis software system*), Version 6, www.statsoft.com.
Walley, A. G., and Carrell, T. D. (**1983**). "Onset spectra and formant transitions in the adult's and child's perception of place of articulation in stop consonants," J. Acoust. Soc. Am. **73**, 1011–1022.
Ylinen, S., Shestakova, A., Alku, P., and Huotilainen, M. (**2005**). "The perception of phonological quantity based on durational cues by native speakers, second-language users and non-speakers of Finnish," Lang Speech **48**, 313–338.
Zimmerman, S. A., and Sapon, S. M. (**1958**). "Note on vowel duration seen cross-linguistically," J. Acoust. Soc. Am. **30**, 152–153.