Auditory enhancement and second language experience in Spanish and English weighting of secondary voicing cues

Fernando Llanos
School of Languages and Cultures, Purdue University, West Lafayette, Indiana 47907-2038

Olga Dmitrieva
Department of Linguistics, Stanford University, Stanford, California 94305

Amanda Shultz
Linguistics Program, Purdue University, West Lafayette, Indiana 47907-2038

Alexander L. Francis[a)]
Department of Speech, Language, and Hearing Sciences, Purdue University, West Lafayette, Indiana 47907-2038

(Received 12 March 2013; revised 5 July 2013; accepted 18 July 2013)

The role of secondary cues in voicing categorization was investigated in three listener groups: Monolingual English ($n = 20$) and Spanish speakers ($n = 20$), and Spanish speakers with significant English experience ($n = 16$). Results showed that, in all three groups, participants used onset *f*0 in making voicing decisions only in the positive voice onset time (VOT) range (short lag and long lag tokens), while there was no effect of onset *f*0 on voicing categorization within the negative VOT range (voicing lead tokens) for any of the participant groups. These results support an auditory enhancement view of perceptual cue weighting: Onset *f*0 serves as a secondary cue to voicing only in the positive VOT range where it is not overshadowed by the presence of pre-voicing. Moreover, results showed that Spanish learners of English gave a significantly greater weight to onset *f*0 in their voicing decisions than did listeners in either of the other two groups. This result supports the view that learners may overweight secondary cues to distinguish between non-native categories that are assimilated to the same native category on the basis of a primary cue.
 [http://dx.doi.org/10.1121/1.4817845]

PACS number(s): 43.71.An, 43.71.Hw [JMH] Pages: 2213–2224

I. INTRODUCTION

Phonetic contrasts are typically realized in terms of multiple acoustic cues, although not all cues contribute in equal measure to the perceptual identification of the contrast. Cues that are perceptually dominant with respect to a particular phonetic contrast are considered primary. Less perceptually prominent acoustic cues to the contrast are referred to as secondary cues. Despite their relative perceptual inconspicuousness, secondary cues have been shown to influence category judgments although their effect is not as consistent as that of the primary cues and may be subject to additional conditions.

The present study addresses the role that secondary acoustic cues play in the perception of voicing across languages with different phonetic realizations of the voicing feature. In particular, the study examines the relative contribution of voice onset time (VOT) and onset *f*0 to the perception of stop voicing in English and Spanish. Differences in the phonetic implementation of voicing in these two languages make it possible to test the predictions of two contrasting theories of the basis for perceptual contribution of the secondary cue (onset *f*0) to stop consonant voicing: A distributional theory in which secondary cue weighting derives from perceptual experience with correlations between primary and secondary cues (Holt *et al.*, 2001) and an auditory enhancement theory in which secondary cue weights derive from their ability to contribute to the perception of a higher-level, integrated perceptual cue (Kingston and Diehl, 1995).

In addition, the present study examines the effects of second language experience on the perceptual use of onset *f*0. Previous studies suggest that second language learners may overweight secondary cues to distinguish between non-native categories that are assimilated to the same native category in terms of the primary cue (Bohn, 1995; Escudero and Boersma, 2004; Escudero *et al.*, 2009; Kondaurova and Francis, 2008, 2010). Given that English voicing categories assimilate to the same Spanish category in terms of VOT, the present study will provide new data relevant to evaluating the effect of second language exposure on the weighting of secondary cues.

A. Voicing

With respect to voicing, Lisker (1986) identified 16 acoustic correlates that differentiate voiced and voiceless stops in English. The most commonly studied of these correlates include the relative timing of the burst release and onset of voicing (VOT; Abramson and Lisker, 1970); the fundamental frequency at the onset of voicing (onset *f*0; Haggard

a)Author to whom correspondence should be addressed. Electronic mail: francisa@purdue.edu

et al., 1970; Ohde, 1984); the frequency of the onset of the first formant ($F1$ frequency); the time between the onset of voicing and the onset of the first formant ($F1$ cutback) (Liberman *et al.*, 1958; Stevens and Klatt, 1974); the duration of the oral closure (Keating, 1984); and the relative amplitude of aspiration noise between the burst release and the onset of voicing (Repp, 1979). Of these, VOT has been consistently and repeatedly shown to be the dominant cue to stop consonant voicing in English (Abramson and Lisker, 1985; Lisker, 1978).

Cross-linguistically, measured VOT values range from strongly negative (voicing onset greatly precedes the burst) to strongly positive (voicing onset lags behind the burst to a significant degree) (Cho and Ladefoged, 1999; Keating, 1984; Lisker and Abramson, 1964). The VOT continuum can be divided into three ranges corresponding to distinct phonetic categories across a variety of languages: Voicing lead (voicing begins typically between 200 and 40 ms prior to the burst release; also referred to as negative VOT or prevoicing), short lag (voicing starts shortly after the burst release, between 0 and 20 ms), and long lag (voicing starts well after the burst release, between 40 and 100 ms; also characterized by aspiration noise during the VOT lag). Interestingly, the presence of a short lag category appears to be ubiquitous: All of the languages surveyed by Cho and Ladefoged (1999), Keating (1984), and Lisker and Abramson (1964) include a voicing category characterized by a short lag VOT in their phonological inventory. Thus, it seems that every two-category language contrasts a short lag category with either a voicing lead category or a long lag category. However, while lead stops are typically heard as [+voice] and long lag stops are typically heard as [−voice], short lag stops may be perceived as either [+voice] or [−voice], depending on the category they are contrasted with. For example, in utterance initial position, English listeners typically treat short lag stops as [+voice] in contrast to long lag stops (as [−voice]), whereas Spanish listeners treat short lag stops as [−voice] in contrast to voicing lead stops ([+voice]).[1]

Another acoustic cue to stop consonant voicing that has been studied extensively is onset *f*0, which tends to positively correlate with VOT across voicing categories in a great variety of languages (Kingston and Diehl, 1994). This covariation has been attributed to the physiological properties of voicing production (Löfqvist *et al.*, 1989): An increase in the longitudinal tension of the vocal folds during voicing suppression in voiceless stops leads to the increase in the onset *f*0 on the following vowel. This pattern of a higher onset *f*0 following [−voice] stops as compared to [+voice] ones has been observed in both lead and long lag languages (lead: French, Spanish, and Dutch, see Caisse, 1982; Hombert, 1978; Löfqvist *et al.*, 1989; and lag: Danish, English, and Taiwanese, see House and Fairbanks, 1953; Jeel 1975; Lai *et al.*, 2009; Lehiste and Peterson, 1961; Petersen, 1983), suggesting that the covariation between VOT and onset *f*0 is conditioned by the presence of phonological voicing specification and not by the physiological conditions of voicing production (Keating, 1984; Kingston and Diehl, 1994). This conclusion is further supported by data from languages with allophonic consonant voicing. For example, in Tamil, where voicing of consonants is predictable from their phonetic environment, the presence or absence of voicing in terms of VOT has no relationship to onset *f*0 (Kohler, 1982, 1984).

While VOT is commonly accepted as the primary cue to voicing categorization (Abramson and Lisker, 1985; Lisker, 1978), onset *f*0 has also been shown to contribute to voicing decisions. For example, Idemaru and Holt (2011) showed that onset *f*0 differences significantly affected listeners' judgments of the voicing status of ambiguous VOT tokens, and Whalen *et al.* (1993) showed that the onset *f*0 differences enhanced perceived typicality even of unambiguous VOT tokens. The observation that the perceptual contribution of a secondary cue (such as onset *f*0) in the phonetic decision varies as a function of the perceptual contribution of the primary one (such as VOT) suggests that acoustic cues to a particular phonetic contrast are not evaluated individually. Rather, the way cues are perceived—the relative weight they are given in a phonetic decision—depends on the contribution of other cues in the signal (Repp, 1982; McMurray and Jongman, 2011). Although the primacy of primary acoustic cues may well derive from simple auditory biases or nonlinearities (Holt and Lotto, 2006; Holt *et al.*, 2004; Stevens, 1972), the relative weight given to secondary cues has been attributed to both associative learning of cue distributions in linguistic input, and to the integration of low-level cues into higher level ones (see discussion by Francis *et al.*, 2008).

B. Theories of perceptual interaction

The ability of onset *f*0 to supplement or alter the perception of voicing in conjunction with VOT has been attributed both to associative learning of distributional properties in the input (Holt *et al.*, 2001; Stilp *et al.*, 2010) and also to the enhancement of the auditory representation of one cue by the presence of another (Kingston, *et al.*, 2008; Kingston and Diehl, 1995). According to the associative learning hypothesis, listeners learn to give more weight to secondary cues that contribute more reliably to a phonetic percept, i.e., that co-vary more strongly with the primary cue. On the other hand, according to the theory of auditory enhancement, the weight given to a secondary cue is determined by the degree to which that cue enhances the same pattern of response in the auditory system that is engendered by the primary cue.

In support of the associative learning theory, Holt *et al.* (2001) showed that Japanese quail trained on stimuli with co-varying VOT and onset *f*0 were able to learn the pattern of covariation to which they were exposed, and were subsequently able to generalize the learned pattern to categorize novel stimuli. Such learning occurred whether the correlation between onset *f*0 and VOT was positive (with high onset *f*0 values corresponding to longer VOT values, as found in English) or negative (with high onset *f*0 values corresponding to shorter VOT values, a pattern opposite that of English). Thus, the pattern of covariation was learned independently of the acoustic properties of the cues that co-varied.

In support of the auditory interaction of VOT and onset *f*0, Kingston *et al.* (2008) showed that the perceptual

integration of multiple correlates to English voicing (onset *f*0, *F*1 onset, and closure duration) is not determined by their covariation in the input but by the mutual enhancement of a more perceptually fundamental auditory cue. While onset *f*0, *F*1 onset, and closure duration are positively correlated across English voicing categories, only onset *f*0 and *F*1 onset contribute to the enhancement/inhibition of the perception of voicing continuation (identified by the presence of low frequency energy in the vicinity of the consonant burst release). Kingston *et al*. (2008) established that onset *f*0 and *F*1 onset are perceptually integral (in the sense of Garner, 1974) with voicing continuation. In contrast, the acoustic cue of closure duration is not perceptually integral with voicing continuation, despite the fact that the two are correlated in production. These results suggest that the weight given to a secondary cue may be determined mainly by the degree of enhancement that it provides to the perception of the primary cue within the context of an integrated, multi-cue percept.

C. The role of cue weighting

Despite their differences, both auditory and associative learning theories make similar predictions about the weight given to onset *f*0 in long lag languages such as English. Associative theories predict that listeners exposed to the positive correlation between longer VOT values and higher onset *f*0 values that already exist in the ambient language will learn to weight onset *f*0 accordingly. Similarly, auditory theories predict that these two cues, by virtue of their mutually enhancing nature, will be perceived as integral and thus onset *f*0 will also contribute to the voicing decision. Both theories predict that listeners will incorporate onset *f*0 into their voicing decisions in long lag languages. On the other hand, the two types of theories may make different predictions about the relative weight that will be given to onset *f*0 in the perception of voicing in lead languages.

Associative learning theories predict that languages will weight onset *f*0 according to the degree to which it co-varies with VOT in the speech to which listeners have been exposed. Since VOT and onset *f*0 co-vary relatively well in lead languages just as they do in long lag languages, this theory predicts that listeners from both languages will give onset *f*0 similar weight.

In contrast, auditory enhancement theories predict that onset *f*0 will be given little weight in lead languages, because, in these languages, the contribution of onset *f*0 differences to the enhancement of the voicing continuation property is relatively small compared to that provided by the presence of voicing immediately prior to the burst. In lead languages, [+voice] stops are typically characterized by a period of pre-voicing which, in itself, contributes a great deal of low frequency energy in the immediate vicinity of the burst. Thus, its presence may drastically reduce the relevance of any concomitant lowering of onset *f*0 to the perception of [+voice] stops. This contrasts with the circumstances in long lag languages in which both [+voice] and [−voice] stops are most commonly characterized by relatively little low frequency energy in the immediate vicinity of the burst (aside from contextually determined variants such as those discussed in footnote 1). In cases in which the perception of voicing continuation is not dominated by the presence of pre-voicing, relatively small differences in onset *f*0 are sufficient to enhance the perception of the voicing continuation property, and thus the perception of most [+voice] vs [−voice] stop contrasts, meaning that listeners would be expected to give considerably more weight to onset *f*0 in long lag languages than in lead languages.[2]

Moreover, if the contribution of onset *f*0 to the perception of voicing is only through its role in enhancing the perception of low frequency energy around the burst, then auditory enhancement theories also predict that listeners will only be affected by onset *f*0 differences when those differences have a clearly enhancing effect. Differences in onset *f*0 should only affect voicing decisions when other cues, such as the presence of pre-voicing, do not dominate the phonetic decision. Thus, it might be expected that onset *f*0 differences will matter primarily in the positive range of VOT values, and will have little or no effect on voicing decisions in the negative VOT range for both lead and long lag languages. In the present paper, we address these questions by examining stop consonant voicing perception by native speakers of Spanish, a lead language, and English, a long-lag language.

D. Effects of second language experience

It is also possible that experience with a second language may affect weighting of secondary cues, whether or not those cues play a significant role in a listeners' native language. For example, many studies have shown that Spanish learners of English tend to overweight the duration cue to the English tense/lax vowel contrast ([i] as in "bit" vs [i] as in "beat") (Bohn, 1995; Escudero and Boersma, 2004; Escudero *et al*., 2009; Kondaurova and Francis, 2010), even though this cue does not play a significant role in the native Spanish vowel system (Kondaurova and Francis, 2008).

One possible explanation for such overweighting of a secondary cue might be to compensate for the difficulties faced in distinguishing between non-native categories with primary cue values within the range of a single native category [in the terminology of Best *et al*. (1988) a *single category contrast*, at least along the primary dimension]. According to this argument, listeners who find themselves unable to rely on a familiar cue (i.e., the presence/absence of pre-voicing that serves as a primary cue to the Spanish voicing contrast), may increase their dependence on secondary cues, in this case including (but perhaps not limited to) onset *f*0.[3]

Although there are cases in which overweighting of secondary cues may be detrimental (e.g., Iverson *et al*., 2003), under some circumstances it can be successful and the case of stop consonant voicing, like the English tense/lax vowel contrast, may represent an optimal context in which listeners might benefit from overweighting secondary cues. As in the case of the English vowels [i] and [i], the English [+voice] and [−voice] categories are assimilated to a single phonological category in Spanish in terms of the primary cue (in this case VOT, not vowel formant frequencies). Thus, if Spanish listeners make little use of onset *f*0 as a cue to

voicing in their native language (as predicted by the auditory enhancement theory), the case exactly parallels that of Spanish listeners learning the English tense/lax vowel contrast and permits a conceptual replication of that research using a new contrast. If, on the other hand, Spanish listeners do use onset *f*0 as a voicing cue, as predicted by associative learning theory, an investigation of the weight given to onset *f*0 by Spanish learners of English would provide new data on whether the phenomenon of overweighting is constrained to previously unattended cues (such as duration in the tense/lax vowel contrast) or also to cues that are already relevant in the native language. To address this question, we examine stop consonant voicing perception in a third group of listeners: Native speakers of Spanish with significant English experience.

II. METHODS

A. Subjects

Twenty native speakers of American English (E-US; 12 women, 8 men; mean age 21 yrs) and 16 native speakers of Spanish (S-US; 7 women, 9 men; mean age 28 yrs; mean years of English immersion 4.6 yrs) were recruited on the campus of Purdue University in West Lafayette, IN. Twenty native speakers of Spanish (S-SP; 9 men, 11 women; mean age 28.2 yrs) were recruited at the Centro de Ciencias Humanas y Sociales—Consejo Superior de Investigaciones Cientificas in Madrid, Spain.

The 20 native English-speaking participants had an average of 3.4 yrs of experience with a currently spoken language other than English, beginning this exposure on average at age 15. Of these, 13 had studied Spanish (3.35 yrs, starting at age 14.8), 5 studied French (5.1 yrs, from age 13.4), 2 studied German (4.5 yrs, from age 14.5), 2 studied Japanese (9 months, from age 18.5), and 1 studied Chinese (1 semester, from age 17). Note that totals add up to more than 20 because some had studied more than 1 language. Three English speakers had lived in a non-English environment for a period greater than 1 month, one in Belgium (4 months, at age 21 yrs) one in Spain (3 months, age 3), and one in Korea (13 months, age 4). Thus, although we have no data on individual degrees of second language fluency for these participants, it is safe to say that their second language competence, as a group, likely approximates that of a typical American college student who grew up speaking only English at home.

Among the 16 participants in the S-US group, four were from Spain (including 1 from the Basque Country), 11 were from Latin America (9 from Colombia, 2 from Venezuela), and 1 chose not to report a country of origin. Given that these participants were recruited on the campus of Purdue University it is reasonable to assume that all had a considerable experience speaking and listening in English. All S-US participants reported having studied at least one foreign language. English dominated the list (14 participants), which also included French (7), Italian (2), German (1), and Russian (1). One participant was a Spanish-Basque bilingual. The average duration of stay in a country where languages other than Spanish were spoken was 4.5 yrs. The average duration of stay in an English-speaking country (predominantly the USA, in one case the USA and Canada) was 4.6 yrs, ranging from 6 months to 12 yrs.

Among the 20 participants in the S-SP group, 14 reported being from Spain and 5 from Latin America (1 chose not to report country of origin). Those from Spain listed birthplaces of Madrid (6), Alicante (1), Badajoz (1), Cordoba (1), and Barcelona (1), or did not provide a city (4). Of the participants from Latin America, two were from Venezuela, two from Chile, and one from Mexico (no cities specified). Although 17 participants reported having studied English in a Spanish speaking environment, only 6 participants reported having lived in an English-speaking country, and of those 6, 4 were there for a year or less. One participant had been in the US for two years, and another for four. However, none reported having been a resident in an English-speaking country during the year prior to the experiment (average of 2.2 yrs since overseas residence).

Rosner *et al.* (2000) have shown that Castillian Spanish differs significantly from some Latin American dialects in terms of the production of VOT. Specifically relevant for the present paper, their measure of Castillian /b/ and /p/ VOT values differed significantly from those found in Guatemalan Spanish as published by Williams (1977a): For /b/, −91.5 ms (Castillian), vs −120.3 (Guatemalan); for /p/, 13.1 ms (Castillian) vs 9.8 ms (Guatemalan). Further research is necessary to determine whether there are correspondingly significant dialectal differences in *perceptual* VOT boundary locations, but the magnitude of the reported differences between production means across Spanish dialects is quite small when compared to differences between any Spanish dialect and English and thus no attempt was made to distinguish between listeners on the basis of native dialect.

All interactions with Spanish-speaking participants were conducted in Spanish, including recruitment posters, scheduling emails, and all written and spoken instructions. Participants were also engaged in a brief (approximately 5 min) conversation in Spanish by a native (Castillian) Spanish speaker prior to beginning the experiment. English speakers were similarly recruited, engaged, instructed, and tested in English, interacting only with native English speakers during the experiment. Participants were paid at the rate of $10/€8 per hour for about half an hour of participation. All participants reported having no history of speech or hearing disorder.

B. Stimuli

The stimuli, similar to those used by Shultz *et al.* (2012), were created using the Klatt speech synthesizer (Klatt, 1980) implemented in Praat 5.2 (Boersma and Weenink, 2009) with 16 bit precision at a 44.1 kHz sampling rate. Tokens ranged from a Spanish [+voice] /ba/ to an English [−voice] /pa/ varying orthogonally in VOT (from −60 to 60 ms in equal steps of 10 ms) and onset *f*0 (ranging from 90 to 150 Hz in equal steps of 20 Hz, with the *f*0 contour subsequently changing from this starting value to 120 Hz over the first 50 ms of voicing). The vowel was a low, central/back vowel, as in the Spanish word *papa*.

In order to maintain the same burst properties across all tokens, a sound file consisting of a single burst was generated by setting to zero the amplitude of all non-burst parameters in a Klatt template corresponding to a token of 0 ms of VOT and 120 Hz of onset *f*0 (see below). Then, 52 separate parameter files (corresponding to each combination of onset *f*0 and VOT) were created with the amplitude parameter set to zero throughout the duration of the burst. Separate sound files were created for each of these burst-less syllables generated from each of these parameter files and then a copy of the burst sound file was added to each of these 52 burst-less sound files to create the final stimuli. The duration of the burst was set at 4 ms with the amplitude rising from 0 to 25 dB over the first millisecond and falling to 0 dB during the last millisecond. A fricative formant (300 Hz, 100 Hz bandwidth) was used to enhance bilabial quality.

The five [a]-vowel formants began 1 ms after the end of the burst. Formant transitions for *F*1–*F*3 lasted 35 ms out the total vowel duration (315 ms). *F*1 began at 220 Hz and rose to 710 Hz. *F*2 began at 900 Hz, rising to 1240 Hz, and *F*3 rose from 2000 to 2500 Hz. *F*4 and *F*5 were held constant at 3600 and 4500 Hz, respectively. Formant bandwidths were constant at the following values: *F*1: 50 Hz; *F*2: 70 Hz; *F*3: 110 Hz; *F*4: 170 Hz; *F*5: 250 Hz.

For short lag and long lag tokens, the *f*0 parameter began at a value of either 90, 110, 130, or 150 Hz, and converged to 120 Hz over the next 50 ms. It subsequently fell to 95 Hz at 40 ms before the end of the vowel, and from there to 50 Hz at the end of the vowel. For these tokens, initial voicing amplitude was 60 dB and remained at that level for 20 ms, subsequently falling to 50 dB over the remaining vowel duration.

For voicing lead tokens, the *f*0 parameter was held constant at 120 Hz from the beginning of voicing until the end of the burst. Immediately after the burst, at the onset of the vowel, the *f*0 parameter was set to the corresponding onset *f*0 value (i.e., 90, 110, 130, or 150 Hz). The *f*0 frequency contour during the vocalic portion was shaped in the same way as for short lag and long lag tokens. The amplitude of voicing was set at 45 dB during the pre-voicing period of voicing lead tokens, with a subsequent increase to 60 dB during the burst. After the burst, the intensity of voicing was held constant at 60 dB over the next 20 ms and then fell linearly to 50 dB at the end of the vowel. Aspiration amplitude was linearly interpolated from 20 to 25 dB as a function of VOT duration (from 0 to 60 ms). Aspiration reached its maximum amplitude during the first millisecond immediately after the burst and fell to 0 dB over the last millisecond before voicing began. See Fig. 1 for spectrograms, waveforms, and *f*0 contours from representative stimuli.

C. Procedure

Experimental procedures were similar to the perceptual task used by Shultz *et al.* (2012). Stimuli were presented to participants at a comfortable listening level using a MATLAB 7.10 interface (MathWorks, 2010). For the E-US and S-US participants, Sennheiser HD 280 pro headphones were used

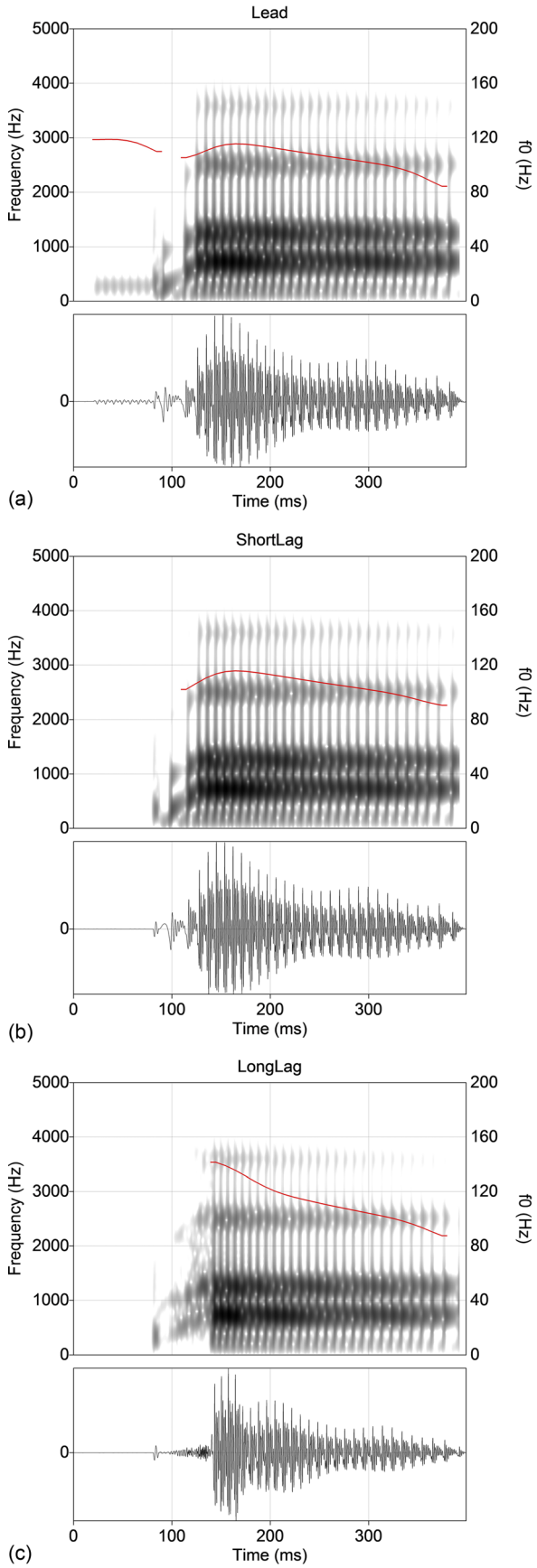

FIG. 1. (Color online) Spectrograms, waveforms, and superimposed f0 trace (right *Y* axis) for three sample stimuli: (a) A token with the most negative VOT (longest lead voicing) and lowest onset *f*0; (b) a token with an intermediate (short lag) VOT and moderate onset *f*0; (c) a token with the longest positive VOT (long lag) and highest onset *f*0.

with a Soundblaster Live! Sound card on a Dell Optiplex/Windows XP computer. For the S-SP group, AKG K240 headphones were used with an ACER Pentium (R)/Windows XP computer with the onboard sound card.

On each trial, listeners were presented with a single token and were asked to choose which of two syllables (/pa/ or /ba/) they heard. The choice was made by using the computer mouse to click on one of two on-screen buttons labeled "BA" and "PA." The left–right order of the response buttons was counterbalanced across participants. After each trial the mouse pointer was automatically re-centered between the two on-screen response buttons to avoid response bias. Participants were not limited in response time, but were instructed to respond as quickly and accurately as possible. After each response, there was a 400 ms pause before the beginning of the next trial. Participants completed a total of 11 blocks of 52 tokens each (572 trials). The first block was treated as familiarization and was not analyzed.

D. Analysis

Statistical analyses were used to address three theoretical questions: (1) Were participants using both onset *f*0 and VOT in their voicing categorization (onset *f*0 users vs nonusers); (2) to what extent did each of these cues contribute to listeners' voicing decisions (cue weighting); and (3) was the contribution of onset *f*0 uniform across the VOT range, or was it constrained to just one range (i.e., positive or negative values). To answer the first question, participants' responses were analyzed using logistic regression with Wald tests applied to the logistic model fitted to each individual participant's response pattern. This analysis was used to determine whether or not each of the two dimensions, VOT and onset *f*0, contributed significantly to each participant's model of voicing categorization. Since all participants were *a priori* expected to depend primarily on VOT, this analysis was useful for establishing the number of "onset *f*0 users" in each group. A χ^2 test applied to the number of onset *f*0-users in each group was used to determine whether there was a significant difference in the number of such users across the groups. These logistic models computed for each individual listener were also used to calculate VOT category boundaries as a way of determining whether or not listeners in the S-US group exhibited an influence of English exposure on voicing categorization.

To answer the second (cue weighting) question, logistic function fitting was used to obtain standardized β-coefficients for each subject. These are indicative of the relative contribution of VOT and onset *f*0 to the /ba/2011/pa/ categorization (Morrison and Kondaurova, 2009). Separate repeated measures analyses of variance (ANOVAs) applied to the individual β-coefficients were then used to test for a significant difference in the perceptual weighting of VOT and onset *f*0 across groups.

To answer the third question (uniformity of *f*0 weighting across the VOT range), a new battery of Wald tests was applied to each onset *f*0 user's logistic model. For this analysis, separate models were generated for the positive and negative VOT ranges for each onset *f*0 user to determine the contribution of onset *f*0 in each range separately.

III. RESULTS

The results of the Wald tests (95% confidence, $\alpha = 0.05$) applied to the logistic models of individual participants' responses showed that, while all subjects relied on VOT, not everyone used onset *f*0 in their voicing decisions. Specifically, only 10 out of 20 listeners in the E-US group, 10 out of 16 listeners in the S-US group, and 9 out of 20 listeners in the S-SP group demonstrated a significant use of onset *f*0 in voicing categorization. However, a χ^2 test of homogeneity did not reveal any significant differences between the number of onset *f*0 users in the three populations of listeners.

The VOT boundary was calculated for each listener in both S-US and S-SP groups and compared to the VOT boundaries obtained for listeners in the E-US group. The calculation was made using the logistic curves modeling each subject's performance at the intermediate onset *f*0 level of 120 Hz. The VOT boundary was established by identifying the VOT value at the 50% point in the identification curve (i.e., median level, where the participants' responses to the categorization task were at chance). Figure 2 shows the logistic curves used for VOT boundary calculation for listeners in the three groups with individual 50% points marked along the *X* axis.

To test for between-group differences in the VOT boundary location, the obtained values were submitted to a one-way between-group ANOVA. Results showed a main effect of group membership on VOT boundary, $F(2,53) = 19.46$, $p < 0.001$; means: S-SP = 2.8 ms; E-US = 22.4 ms; S-US = 17.2 ms. *Post hoc* pairwise comparisons (Tukey HSD) of mean group differences showed a significant difference between S-SP and E-US and between S-SP and S-US with the 95% confidence intervals spanning [11.76 to 27.33] and [6.02 to 22.53], respectively. This means that the VOT boundary of the monolingual Spanish group was significantly different from the VOT boundaries of both the monolingual English group and the group of Spanish listeners immersed in an English-speaking environment, but there was no difference between the latter two groups.

To identify between-group differences in the perceptual weights associated with VOT and onset *f*0, the individual β-coefficients for VOT and onset *f*0 for the subjects in all three groups were submitted to two separate one-way between-group ANOVAs. For each ANOVA, only participants who showed significant use of the cue being tested were included in the analysis: Thus, all listeners (56 total) were included in the VOT analysis, but only the 10 E-US, 10 S-US, and 9 S-SP listeners (29 total) who were identified as onset *f*0 users according to the previously described Wald tests were included in the onset *f*0 analysis. Results showed no significant effect of group membership on weighting of VOT, $F(2,53 = 2.00$, $p = 0.068$, but there was a significant effect of group membership on weighting of onset *f*0, $F(2,26) = 6.39$, $p = 0.005$. *Post hoc* pairwise comparisons (Tukey HSD) of the onset *f*0 group means showed a significant difference between S-US and E-US, and between S-US and S-SP with two 95% confidence intervals spanning [−0.70 to −0.09] and

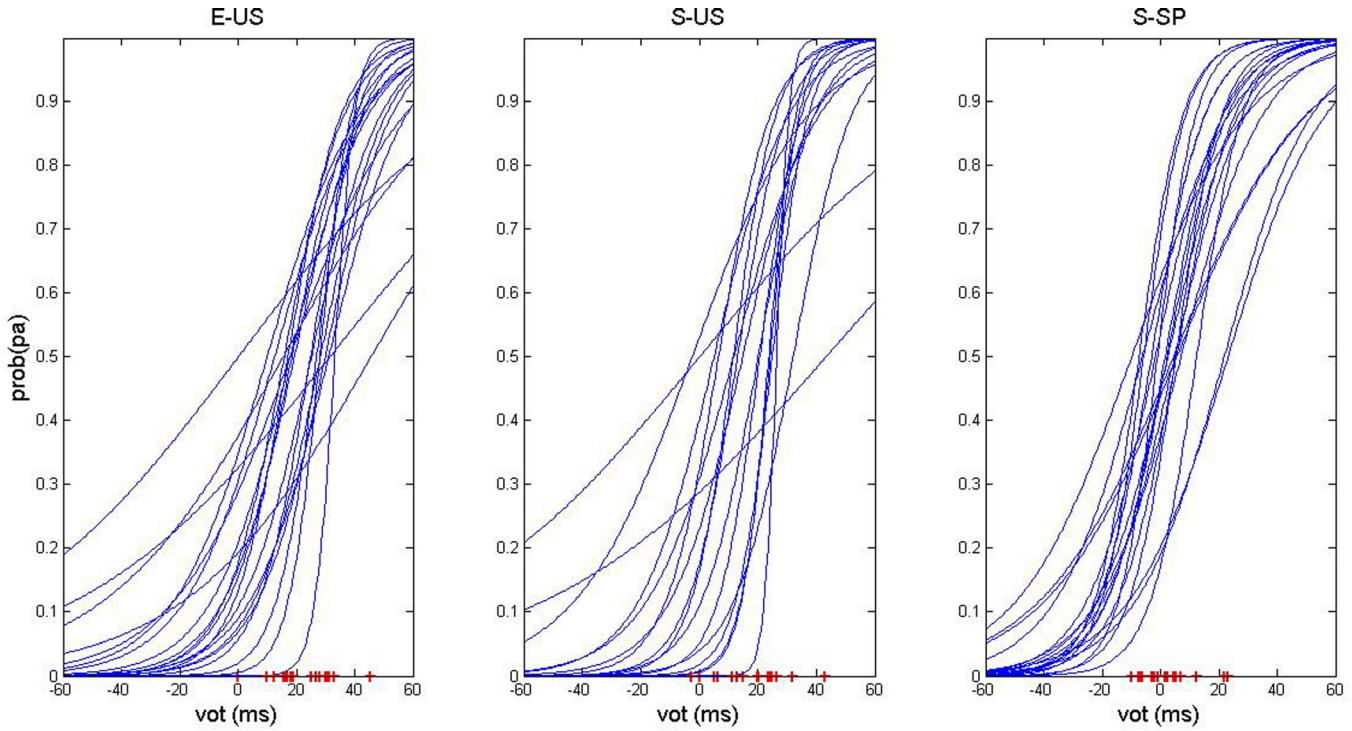

FIG. 2. (Color online) Logistical curves fitted to each individual participants' identification responses. For each listener, the 50% point is also marked below as thicker hash marks crossing the *X* axis. Panels display results for participants in different groups. Left: English speakers tested in the US (E-US); middle: Spanish speakers tested in the US (S-US); right: Spanish speakers tested in Spain (S-SP).

[0.04 to 0.67], respectively (means: S-US = 0.89; S-SP = 0.53; E-US = 0.48). This shows that Spanish speakers with significant English experience assigned a significantly greater weight to onset *f*0 than did the two remaining participant groups, while there was no difference between the monolingual English and the monolingual Spanish groups.

Between-group differences in onset *f*0 weight are displayed in Fig. 3. The first row shows the averaged

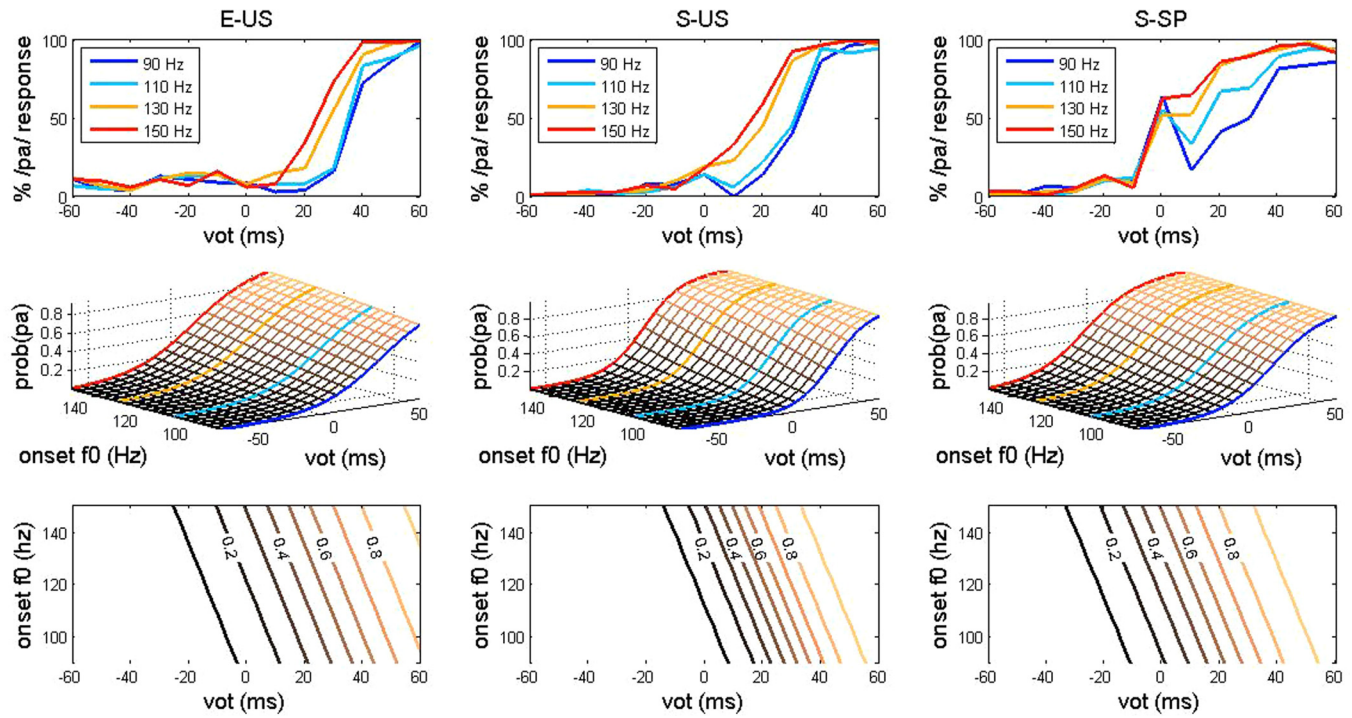

FIG. 3. (Color online) Top row: Mean identification curves for each group of listeners. Middle row: Three-dimensional logistic surface fitted to each group's model (*X* axis: Probability of [pa] response; *Y* axis: Onset *f*0 frequency in hertz; *Z* axis: VOT in milliseconds). Bottom row: Probability contour plots for each group's logistic model. Each contour line delineates the set of points in the VOT (*X* axis) by onset *f*0 (*Y* axis) plane that exhibit the same proportion (i.e., probability) of /pa/ responses, from 0 to 1 in 0.2 steps. In all rows, panels display results for different groups. Left: English speakers tested in the US (E-US); middle: Spanish speakers tested in the US (S-US); right: Spanish speakers tested in Spain (S-SP).

identification curves for each *f*0 step for each group, while the second and third rows show results of logistic modeling. In the second row, onset *f*0 weight is represented in terms of the slope of the decision surface (steeper slope meaning greater weight for onset *f*0), while in the third row it is reflected in the relative distance between the probability contour lines (probability of /pa/ response). Each contour line delineates the set of points in the VOT-onset *f*0 space that exhibit the same proportion of /pa/ responses. The distance between two contour lines represents the number of VOT units required to increase the proportion of /pa/ responses for a given level of onset *f*0. As a consequence, smaller distances reflect greater onset *f*0 weight per VOT unit. A greater weight assigned to onset *f*0 by the S-US group is visually most apparent in the third row display, where the distance between the probability contour lines is appreciatively smaller for the S-US group than for the E-US and S-SP groups.

Moreover, all three types of display indicate that the perceptual effect of onset *f*0 seems to be restricted to the moderately positive VOT range for all groups of participants. In particular, in the top row, identification curves for the lower and higher *f*0 values are most clearly separated in the positive VOT range. Similarly, the slopes of the surfaces in the middle row are all steepest toward the right side of the VOT continuum, and, finally, in the bottom row, the contour lines are closest between about 10 and 50 ms of VOT.

In order to test the hypothesis that onset *f*0 effect was operative only in the positive VOT range, a series of Wald tests with a 95% confidence ($\alpha = 0.05$) was applied to each participant's logistic model to determine the contribution of onset *f*0 in both the negative and positive VOT range separately. As above, only participants who relied on onset *f*0 in voicing categorization (onset *f*0 users) were included in the analysis. The result showed that none of the participants in any group were significant users of onset *f*0 in the negative VOT range. On the other hand, the results of tests on the positive ranges showed a significant contribution of onset *f*0 for all participants' models in all three groups, except for one in the E-US group. These results confirm that all effects of onset *f*0 are confined to the positive VOT range.

IV. DISCUSSION

A. Associative learning and auditory enhancement

The results demonstrate that monolingual native speakers of Spanish (S-SP) do not differ significantly from native speakers of English (E-US) in terms of the degree to which they use onset *f*0 in voicing decisions. These two groups showed similar overall weighting of onset *f*0 in voicing identification, consistent with the prediction of the associative learning theory that the combined use of multiple cues will result from the presence of a positive correlation between cues in the ambient language [cf. Shultz *et al.* (2012) and Kingston and Diehl (1994) for evidence of such correlations in the production of lead and long lag languages]. However, a closer look at the patterns of responses reveals that the cue weightings computed over the entire range of VOT (including both positive and negative values) may be misleading, since the use of onset *f*0 in voicing categorization was not constant across different ranges of the VOT continuum. By splitting the analyses across the positive and negative VOT ranges, additional details may be observed.

First, none of the groups tested showed an effect of onset *f*0 in the negative VOT range. For the E-US group, this lack of onset *f*0 effect could be explained by the fact that the negative VOT is not included in the English phonological system as an independent category in utterance-initial position. English listeners' lack of experience with onset *f*0 covariation with VOT in the negative VOT range in utterance-initial position could thus prevent them from successfully using it in perception. However, the same explanation would predict a lack of an onset *f*0 effect in the long lag VOT range for monolingual Spanish listeners because these tokens likewise do not exist as an independent category in this position for these listeners. However, this expectation is not supported by the observations from the S-SP group. Thus, an explanation based solely on native language category experience is not viable. On the other hand, this response pattern is consistent with the auditory enhancement theory, which predicts that, in the negative VOT range, the presence of the pre-voicing cue may completely overpower the comparatively subtle contribution from onset *f*0 to the perception of low frequency energy proximal to the burst release, rendering the *f*0 cue irrelevant for perception of voicing irrespective of listeners' experience with specific distributions of cues.

Second, as shown clearly in the top row of Fig. 3, the S-SP group did not show an effect of differences in onset *f*0 at the most linguistically ambiguous VOT value (0 ms). If the role of secondary cues is to facilitate the disambiguation of tokens with ambiguous primary cue values, one would expect the greatest effect of onset *f*0 precisely at the most ambiguous VOT values for each language group: Around 0 ms for S-SP, and around 20 ms for E-US and S-US listeners (using the English boundary). In fact, for the E-US group of listeners, there is a noticeable onset *f*0 effect identifiable on both sides of the categorical boundary (i.e., across the range from 20 to 50 ms VOT). This result is consistent with the hypothesis that onset *f*0 plays a relatively modest role in Spanish voicing decisions in comparison to English. Moreover, the fact that S-SP listeners *do* show an onset *f*0 effect in the positive VOT range, well within their [−voice] category, suggests that their weighting of onset *f*0 results mainly from auditory enhancement, in that even speakers of languages without a long lag VOT category still seem to make use of onset *f*0 for voicing category decisions in the upper reaches of the VOT continuum.

Another explanation for the relative lack of onset *f*0 effect for Spanish participants in their native VOT range may be found among other perceptually salient statistical properties of the input in lead languages. Cross-linguistic research (Keating 1984; see also discussion by Holt *et al.*, 2004) has shown that the temporal separation between the prototypical VOT values corresponding to voiced and voiceless categories in lead languages is significantly greater than in long lag languages. More specifically, lead languages exhibit a separation of approximately 40 ms between the

right-most edge of the [+voice] category VOT distribution (at approximately −40 ms) and the left-most edge of the [−voice] category distribution (at 0 ms) along the VOT continuum. In contrast, long lag languages exhibit a smaller gap (∼20 ms) between the two categories: From about +20 ms to about +40 ms VOT (Keating, 1984). A greater separation along the VOT continuum between the voicing categories in lead languages as compared to long lag languages could make VOT-based decisions more reliable in lead languages, therefore making it less necessary for listeners to refer to onset *f*0 when identifying voicing contrasts. However, this explanation still fails to account for the observation of onset *f*0 weighting in the positive VOT range by lead language listeners, and is thus less parsimonious than one based in auditory enhancement.

B. Effects of second language experience

Results showed that listeners in all three groups relied on VOT to the same degree in voicing categorization. In terms of the VOT boundary, the results indicated an effect of *L*2 exposure in the S-US group. While there was a significant difference between the S-SP and E-US groups, with both groups adhering to the VOT boundary characteristic of their respective native languages (0 ms vs 20 ms, see Lisker and Abramson, 1970; Williams, 1977a,b), the S-US group demonstrated an English-like VOT boundary that was significantly different from that of the S-SP group but not the E-US group. This suggests that S-US listeners were somehow influenced to hear the stimuli in an English-like manner, despite the fact that the experimental procedure was conducted entirely in Spanish by a native Spanish speaker. It must be noted that the magnitude of the difference between the S-US and S-SP boundaries (14.4 ms) is quite large compared to the shift (2.1 ms) induced by changing the language context for bilingual listeners, even those who are not very fluent in the second language (Flege and Eefting, 1987a). Moreover, although all interactions with Spanish participants were carried out only in Spanish, and every Spanish participant was engaged in a brief conversation in Spanish by a native speaker of Spanish prior to commencing the experiment, such methods may still not be sufficient to completely determine the linguistic mode in which listeners operate (see Caramazza *et al*., 1973; Williams, 1977b; Elman *et al*., 1977; Flege and Eefting, 1987a; Hazan and Boulakia, 1993; Garcia-Sierra *et al*., 2009; Antoniou *et al*., 2012 for a discussion). Thus, the most likely explanation for these findings is that these listeners had become quite fluent in English and were responding in a fundamentally English mode despite the use of Spanish in the immediate testing context. Factors that may have contributed to this influence include participants' extensive and regular exposure to English as current students at an English-speaking university in a mostly monolingual English-speaking environment as well as the inclusion of many highly English-like VOT values in the experimental stimuli.

Interestingly, the English-like VOT boundary location adopted by the S-US group was also accompanied by a difference in overall weighting of onset *f*0, suggesting that this may also be an effect of exposure to English. Although there was no significant difference between the groups in terms of the *number* of onset *f*0 users, those who used onset *f*0 in the S-US group relied on onset *f*0 for voicing identification significantly more than did onset *f*0 users in either the S-SP or E-US group. While the shift in the VOT boundary for the S-US group can be explained as a shift to conform to the boundaries exhibited by native speakers of English, this increased weighting of onset *f*0 conforms to neither the native English nor the native Spanish pattern. Thus, in terms of onset *f*0 weighting, the S-US listeners who were immersed in an English-speaking environment at the time of testing are clearly different from the S-SP listeners, but not in a way that can be explained simply in terms of having attained a greater similarity to the pattern exhibited by the E-US group. A tendency to overweight secondary cues to contrast in the non-native environment observed in the current study is similar to that reported by earlier studies, including studies of English vowel perception by Spanish listeners (Best *et al*., 1988). More importantly, it seems that listeners made special use of a secondary cue that is not relevant in the native contrast, similar to the pattern discussed by Kondaurova and Francis (2008, 2010).

A variety of explanations may be identified for how *L*2 exposure might also cause non-native listeners to rely more heavily even than native listeners on a particular secondary acoustic cue. One possibility is that non-native listeners' over-reliance on onset *f*0 might result from compensation for a reduction in the weight given to VOT (as might occur due to the perceived unreliability of the native category boundary), but this cannot be the case here as all three groups showed comparable weighting of VOT.

Alternatively, the over-use of onset *f*0 might be caused by the greater cognitive load imposed by processing speech in a non-native environment. Previous research has shown that, under increased cognitive load, listeners tend to assign greater weight to secondary cues to a given phonetic contrast, including onset *f*0 in voicing contrasts (Gordon *et al*., 1993; Mattys and Wiget, 2011). This explanation suggests that Spanish learners of English may be giving more weight to multiple secondary cues in addition to onset *f*0.

On the other hand, it is possible that the effects observed here are unrelated to English exposure, and instead derive from differences in the countries of origin of the participants in the S-US and S-SP groups. That is, the group effect observed in weighting of onset *f*0 may reflect Spanish dialectal differences that have not previously been identified. Specifically, 10/16 listeners in the S-US group were from Latin America, while only 6/20 in the S-SP group were from Latin America (see Sec. II A for more details on the listeners' countries of origin). As Rosner *et al*. (2000) show, there are dialectal differences in the realization of VOT boundaries in Spanish (both between Latin America and Spain, and within each region as well) and, therefore, it is possible that there are also differences in onset *f*0 and the correlation between the two properties across dialects. Confirmation of this hypothesis has to await verification until further data are available, comparing the use of different cues to voicing in different Spanish dialects.

Finally, there is also a way in which the observed pattern of responses of the S-US group appears to be intermediate between that of the monolingual English and the monolingual Spanish participants. For stimuli with rising onset *f*0 contours (the 90 and 110 Hz curves), the S-SP group appears to show an increase in /ba/ responses (a decrease in /pa/ responses) as VOT increases from 0 to 10 ms. This pattern of responses may result from a combination of two factors: The general tendency for identification to become ambiguous in the vicinity of a category boundary and the influence of onset *f*0 in the positive VOT range. As VOT increases toward the Spanish VOT boundary around 0 ms, responses approach 50% because onset *f*0 is not informative but VOT is increasingly ambiguous. As soon as VOT is positive, the onset *f*0 effect manifests itself in the clear and systematic separation of the onset *f*0 curves. In contrast, for the E-US group, the onset *f*0 effect appears at higher VOT values, because the VOT boundary is located well into the positive range (24.3 ms). Thus, for the E-US group, the 90 and 110 Hz onset *f*0 curves simply lag behind the other two curves with respect to their eventual increase toward the category boundary. Interestingly, even though the S-US group exhibits an English-like VOT boundary, they still show a small amount of the S-SP-like increase in /pa/ responses approaching the Spanish VOT boundary, and a corresponding "dip" in the 90 and 110 Hz onset *f*0 curves at 10 ms of VOT. It is possible that these listeners may be showing a broader category boundary effect, or perhaps some vestige of their native category boundary, that leads them to treat the 0 ms tokens as more ambiguous than, e.g., the −10 ms tokens (though not to the same degree as do the S-SP listeners, perhaps due to group differences in English experience). Ultimately, this pattern of results is consistent with the idea that some aspects of the phonetic categorization of advanced second language learners and bilinguals may be intermediate between those of the corresponding monolingual groups (Flege and Eefting, 1987b). In this case, the S-US listeners seem to have adopted the location of the English VOT category boundary (i.e., the 50% point), but still show some effects of the Spanish boundary in terms of a greater degree of uncertainty regarding tokens with 0 ms VOT.

C. Implications for theories of phonological voicing

Findings reported here also have implications for theories of phonological voicing. A detailed examination of the experimental results determined a great similarity between the monolingual E-US and S-SP groups in terms of onset *f*0 weighting in the positive and negative ranges of the VOT continuum. However, these ranges have different linguistic meanings for these two groups: Low onset *f*0 weight in the negative VOT range means little for English speakers since in the initial position it is not a separate phonemic category in their native inventory. On the other hand, for Spanish speakers, a lack of onset *f*0 effect in this range is significant since negative VOT in Spanish corresponds to one of their native voicing categories. The argument is reversed for the positive VOT range. High onset *f*0 weight in this VOT range is relatively unimportant for Spanish listeners but it is meaningful for English listeners who make a major voicing distinction in this part of the VOT continuum. Thus, even though both types of phonemic contrast are usually addressed according to the same category distinction of [± voice], they are organized very differently in terms of the way in which specific phonetic cues relate to phonological categories. In lead languages, pre-voicing itself, that is, a considerable amount of low frequency energy provided by the vocal fold vibration during stop closure, is a dominant, and, it appears, largely sufficient cue to the voicing distinction. In long lag languages, secondary cues have a greater chance to influence the perception of voicing continuity; thus the contrast tends to be based more heavily on multiple cues.

V. CONCLUSION

The results presented here showed that both Spanish and English-speaking listeners used onset *f*0 in their voicing decisions, but only within the positive range of the VOT continuum. Thus, the resulting impression was that across the whole VOT continuum both monolingual English and Spanish listeners were comparable in their use of onset *f*0. However, since different areas of the VOT continuum are linguistically significant for Spanish and English, in effect, only English-speaking participants gave weight to the onset *f*0 parameter in the voicing decisions of their native language. Spanish listeners, on the other hand, did not use onset *f*0 within the larger portion of the VOT range encompassing their prototypical native voicing categories. These findings are in agreement with the prediction of the auditory enhancement theory stating that listeners will only integrate two acoustic cues when one cue provides an enhancing effect to the other (phonetically relevant) cue. The results also provide support for a view of phonological organization of voicing across language which approaches lead-based contrasts and lag-based contrasts as fundamentally different and relying on different types of phonetic cues. Expanding on this perspective, the present results further suggest that, while lead-based languages may rely mainly on pre-voicing as a cue, lag-based languages may make more extensive use of multiple cues.

Although the predictions of the associative learning theory were not borne out by the results of the current study, it must be noted that these predictions were based on the results of a relatively small number of available studies of the relevant production patterns, and many of these studies included only two or three subjects. It is possible that the results of an acoustic study currently under way will provide a richer, more detailed picture of the precise patterns of covariance between onset *f*0 and VOT/phonological voicing in lead languages.

Finally, the present study suggests the need for a more detailed investigation of perception and production of voicing by Spanish individuals with and without significant exposure to English. In the present case, listeners in the S-US group show some expected patterns of results. For example, they seem to have acquired the English VOT boundary relatively effectively, while still retaining some influence of the

Spanish 0 ms boundary (contributing to the greater ambiguity of the 0 ms stimulus for this group as for the S-SP group). However, based on the present results, further research is needed to determine whether, or to what degree, second language learners may be adopting cue weighting strategies that reflect properties unique to the learning contexts (i.e., increasing the weight given to secondary cues more generally), rather than simply representing a stage of weighting intermediate between that of the first and target language.

ACKNOWLEDGMENTS

We are grateful to Professor Juana Gil-Fernández for use of her laboratory facilities at CSIC (Spain), and to Professor Alejandro Cuza for comments on an earlier draft of this manuscript. We also thank Audrey Bengert, Samantha Berger, and A. Danielle Yu for their assistance in running subjects.

[1]It is important to note that these characterizations are based primarily on perceptual findings. In production the situation is much more complex. For example, in English, phonologically voiceless stops may be realized with short-lag VOT at the beginning of unstressed syllables, while syllable-initial phonologically voiced stops may be realized with pre-voicing when preceded by a word ending in a vowel. Moreover, there can be considerable variability both within and across talkers in the degree to which English speakers exhibit pre-voicing of phonologically voiced stops (Shultz *et al*., 2012; Zlatin, 1974). Still, English listeners have been consistently shown to identify stop consonants with a short VOT lag as voiced, and those with a long VOT lag as voiceless (Abramson and Lisker, 1970, 1985; Lisker, 1978; Holt *et al*., 2004; Zlatin, 1974).

[2]Future research is needed to address the very interesting question of what interactions might occur in languages like Thai that subdivide the VOT continuum into three voicing categories (pre-voiced, short-lag, and long-lag). One possibility is that an auditory enhancement theory might predict that onset *f*0 would have a greater effect on the short-lag/long-lag contrast than on the pre-voiced/short-lag one, while the predictions of a learned covariation theory would depend on the correlations observed in production between VOT and onset *f*0 across the three categories.

[3]We are grateful to an anonymous reviewer for pointing out that Abramson and Lisker (1970, 1973) already discussed a possible psychoacoustic basis for Spanish listeners' better-than-expected voice timing discrimination in the extreme lag end of the VOT continuum.

Abramson, A. S., and Lisker, L. (**1970**). "Discriminability along the voicing continuum: Cross-language tests," in *Proceedings of the Sixth International Congress of Phonetic Sciences*, Academia, Prague, pp. 569–573.

Abramson, A. S., and Lisker, L. (**1973**). "Voice timing perception in Spanish word-initial stops," J. Phonetics **1**, 1–8.

Abramson, A. S., and Lisker L. (**1985**). "Relative power of cues: F0 shift versus voice timing," in *Phonetic Linguistics: Essays in Honor of Peter Ladefoged*, edited by V. Fromkin (Academic Press, Inc., Orlando, FL), pp. 25–33.

Antoniou, M., Tyler, M. D., and Best, C. T. (**2012**). "Two ways to listen: Do L2-dominant bilinguals perceive stop voicing according to language mode?," J. Phonetics **40**(4), 582–594.

Best, C. T., McRoberts, G. W., and Sithole, N. M. (**1988**). "Examination of perceptual reorganization for nonnative speech contrasts: Zulu click discrimination by English-speaking adults and infants," J. Exp. Psychol. Hum. Percept. Perform. **4**, 45–60.

Boersma, P., and Weenink, D. (**2009**). "Praat: Doing phonetics by computer" (Version 5.2), Computer program, University of Amsterdam, Amsterdam, The Netherlands. Available online: http://www.praat.org (Last viewed July 9, 2013).

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 Press, Baltimore, MD), pp. 279–304.

Caisse, M. (**1982**). "Cross-linguistic differences in fundamental frequency perturbation induced by voiceless unaspirated stops. Berkeley," M.A. thesis, University of California-Berkeley.

Caramazza, A., Yeni-Komshian, G., Zurif, E., and Carbone, E. (**1973**). "The acquisition of a new phonological contrast: The case of stop consonants in French-English bilinguals," J. Acoust. Soc. Am. **54**, 421–428.

Cho, T., and Ladefoged, P. (**1999**). "Variation and universals in VOT: Evidence from 18 languages," J. Phonetics **27**, 207–229.

Elman, J., Diehi, R., and Buchwald, S. (**1977**). "Perceptual switching in bilinguals," J. Acoust. Soc. Am. **62**, 971–974.

Escudero, P., Benders, T., and Lipski, S. (**2009**). "Native, non-native and L2 perceptual cues weighting for Dutch vowels: The case of Dutch, German, and Spanish listeners," J. Phonetics **37**, 452–466.

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., and Eefting, W. (**1987a**). "Cross-language switching in stop consonant perception and production by Dutch speakers of English," Speech Commun. **6**, 185–202.

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

Francis, A., Kaganovich, N., and Driscoll-Huber, C. (**2008**). "Cue-specific effects of categorization training on the relative weighting of acoustic cues to consonant voicing in English," J. Acoust. Soc. Am. **124**, 1234–1251.

Garcia-Sierra, A., Diehl, R. L., and Champlin, C. (**2009**). "Testing the double phonemic boundary in bilinguals," Speech Commun. **51**, 369–378.

Garner, W. R. (**1974**). *The Processing of Information and Structure* (Erlbaum, Hillsdale, NJ), 194 pp.

Gordon, P. C., Eberhardt, J. L., and Rueckl, J. G. (**1993**). "Attentional modulation of the phonetic significance of acoustic cues," Cogn. Psychol. **25**, 1–42.

Haggard, M., Ambler, S., and Callow, M. (**1970**). "Pitch as a voicing cue," J. Acoust. Soc. Am. **47**, 613–617.

Hazan, V. L., and Boulakia, G. (**1993**). "Perception and production of a voicing contrast by French-English bilinguals," Lang Speech **36**, 17–38.

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**(5), 3059–3071.

Holt, L. L., Lotto, A., and Diehl, R. (**2004**). "Auditory discontinuities interact with categorization: Implications for speech perception," J. Acoust. Soc. Am. **116**, 1763–1773.

Holt, L. L., Lotto, A. J., and Kluender, K. R. (**2001**). "Influence of fundamental frequency on stop-consonant voicing perception: A case of learned covariation or auditory enhancement?," J. Acoust. Soc. Am. **109**, 764–774.

Hombert, J. M. (**1978**). "Consonant types, vowel quality, and tone," in *Tone: A Linguistic Survey*, edited by V. Fromkin (Academic, New York), pp. 77–111.

House, A. S., and Fairbanks, G. (**1953**). "The influence of consonantal environment upon the secondary acoustical characteristics of vowels," J. Acoust. Soc. Am. **25**, 105–113.

Idemaru, K., and Holt, L. L. (**2011**). "Word recognition reflects dimension based statistical learning," J. Exp. Psychol. Hum. Percept. Perform. **37**(6), 1939–1956.

Iverson, P., Kuhl, P. K., Akahane-Yamada, R., Diesch, E., Tohkura, Y., Kettermann, A., and Siebert, C. (**2003**). "A perceptual interference account of acquisition difficulties for non-native phonemes," Cognition **87**, B47–B57.

Jeel, V. (**1975**). "An investigation of the fundamental frequency of vowels after various Danish consonants, in particular stop consonants," Copenhagen: Annual Report of the Institute of Phonetics, University of Copenhagen, Vol. 9, pp. 191–211.

Keating, P. (**1984**). "Phonetic and phonological representations of stop consonant voicing," Language **60**, 286–319.

Kingston, J., and Diehl, R. (**1994**). "Phonetic knowledge," Language **70**, 419–454.

Kingston, J., and Diehl, R. (**1995**). ''Intermediate properties in the perception of distinctive feature values," in *Phonology and Phonetic Evidence: Papers in Laboratory Phonology IV*, edited by B. Connell and A. Arvanti (Cambridge University Press, Cambridge, England), pp. 7–27.

Kingston, J., Diehl, R. L., Kirk, C. J., and Castleman, W. A. (**2008**). "On the internal perceptual structure of distinctive features: The [voice] contrast," J. Phonetics **36**, 28–54.

Klatt, D. H. (**1980**). "Software for a cascade/parallel formant synthesizer," J. Acoust. Soc. Am. **67**, 971–995.
Kohler, K. J. (**1982**). "F0 in the production of lenis and fortis plosives," Phonetica **39**, 199–218.
Kohler, K. J. (**1984**). "Phonetic explanation in phonology. The feature fortis/lenis," Phonetica **41**, 150–174.
Kondaurova, M. V., and Francis, A. L. (**2008**). "The relationship between native allophonic experience with vowel duration and perception of the English tense/lax vowel contrast by Spanish and Russian listeners," J. Acoust. Soc. Am. **124**(6), 3959–3971.
Kondaurova, M. V., and Francis, A. L. (**2010**). "The role of selective attention in the acquisition of English tense and lax vowels by native Spanish listeners: comparison of three training methods," J. Phonetics **38**(4), 569–587.
Lai, Y., Huff, C., Sereno, J., and Jongman, A. (**2009**). "The raising effect of aspirated prevocalic consonants on F0 in Taiwanese," in *Proceedings of the 2nd International Conference on East Asian Linguistics*, edited by J. Brooke, G. Coppola, E. Görgülü, M. Mameni, E. Mileva, S. Morton, and A. Rimrott, Simon Fraser University Working Papers in Linguistics. Online document downloaded from http://www2.ku.edu/~kuppl/documents/Lai_EtAl.pdf (Last viewed March 14, 2013).
Lehiste, I., and Peterson, G. (**1961**). "Some basic considerations in the analysis of intonation," J. Acoust. Soc. Am. **33**, 419–425.
Liberman, A. M., Delattre, P. C., and Cooper. F. S. (**1958**). "Some cues to the distinction between voiced and voiceless stops in initial position," Lang Speech **1**, 153–167.
Lisker, L. (**1978**). "In qualified defense of VOT," Lang Speech **21**(4), 375–383.
Lisker, L. (**1986**). "'Voicing' in English: A catalogue of acoustic features signaling /b/ versus /p/ in trochees," Lang Speech **29**, 3–11.
Lisker, L., and Abramson, A. S. (**1964**). "A cross-linguistic study of voicing in initial stops: Acoustical measurements," Word **20**, 384–422.
Lisker, L., and Abramson, A. S. (**1970**). "The voicing dimension: Some experiments in comparative phonetics," in *Proceedings of the Sixth International Congress of Phonetic Sciences*, Academia, Prague, pp. 563–567.
Löfqvist, A., Baer, T., McGarr, N. S., and Story, S. R. (**1989**). "The cricothyroid muscle in voicing control," J. Acoust. Soc. Am. **85**, 1314–1321.
Mattys, S. L., and Wiget, L. (**2011**). "Effect of cognitive load on speech recognition," J. Mem. Lang. **65**, 145–160.
McMurray, B. and Jongman, A. (**2011**). "What information is necessary for speech categorization? Harnessing variability in the speech signal by integrating cues computed relative to expectations," Psychol. Rev. **118**(2), 219–246.
Morrison, G. S., and Kondaurova, M. V. (**2009**). "Analysis of categorical response data: Use logistic regression rather than endpoint-difference scores or discriminant analysis," J. Acoust. Soc. Am. **126**(5), 2159–2161.
Ohde, R. N. (**1984**). "Fundamental frequency as an acoustic correlate of stop consonant voicing," J. Acoust. Soc. Am. **75**(1), 224–230.
Petersen, N. R. (**1983**). "The effect of consonant type on fundamental frequency and larynx height in Danish," Annual Report Inst. Phon., University of Copenhagen, Vol. 17, pp. 55–86.
Repp, B. H. (**1979**). "Relative amplitude of aspiration noise as a voicing cue for syllable-initial stop consonants," Lang Speech **22**, 173–189.
Repp, B. H. (**1982**). "Phonetic trading relations and context effects: New experimental evidence for a speech mode of perception," Psychol. Bull. **92**, 81–110.
Rosner, B. S., López-Bascuas, L. E., García-Albea, J. E., and Fahey, R. P. (**2000**). "Voice-onset times for Castilian Spanish initial stops," J. Phonetics **28**, 217–224.
Shultz, A. A., Francis, A. L., and Llanos, F. (**2012**). "Differential cue weighting in perception and production of consonant voicing," J. Acoust. Soc. Am. **132**(2), EL95–EL101.
Stevens, K. N. (**1972**). "The quantal nature of speech: Evidence from articulatory-acoustic data," in *Human Communication: A Unified View*, edited by E. E. J. David and P. B. Denes (McGraw-Hill, New York), pp. 51–66.
Stevens, K. N., and Klatt, D. H. (**1974**). "Role of formant transitions in the voiced-voiceless distinction for stops," J. Acoust. Soc. Am. **55**, 653–659.
Stilp, C. E., Rogers, T. T., and Kluender, K. R. (**2010**). "Rapid efficient coding of correlated complex acoustic properties," Proc. Natl. Acad. Sci. U.S.A. **107**(50), 21914–21919.
Whalen, D. H., Abramson, A. S., Lisker, L., and Mody, M. (**1993**). "F0 gives voicing information even with unambiguous voice onset times," J. Acoust. Soc. Am. **93**, 2152–2159.
Williams, L. (**1977a**). "The voicing contrast in Spanish," J. Phonetics **5**, 169–184.
Williams, L. (**1977b**). "The perception of stop consonant voicing by Spanish English bilinguals," Percept. Psychophys. **21**(4), 289–297.
Zlatin, M. A. (**1974**). "Voicing contrast: perceptual and productive voice onset time characteristics of adults," J. Acoust. Soc. Am. **56**(3), 981–994.