Journal of Phonetics 49 (2015) 77–95

Contents lists available at ScienceDirect

Journal of Phonetics

journal homepage: www.elsevier.com/locate/phonetics

Research Article

Phonological status, not voice onset time, determines the acoustic realization of onset *f*0 as a secondary voicing cue in Spanish and English

Olga Dmitrieva [a,b,*], Fernando Llanos [b], Amanda A. Shultz [b], Alexander L. Francis [b]

[a] Stanford University, Stanford, CA 94305, USA
[b] Purdue University, West Lafayette, IN 47907-2038, USA

ARTICLE INFO

Article history:
Received 30 September 2013
Received in revised form
1 December 2014
Accepted 14 December 2014

Keywords:
Voicing
Onset *f*0
VOT
Secondary cues
English
Spanish

ABSTRACT

The covariation of onset *f*0 with voice onset time (VOT) was examined across and within phonological voicing categories in two languages, English and Spanish. The results showed a significant co-dependency between onset *f*0 and VOT across phonological voicing categories but not within categories, in both languages. Thus, English short lag and long lag VOT stops, which contrast phonologically, were found to differ significantly in onset *f*0. Similarly, Spanish short lag and lead VOT tokens are phonologically contrastive and also differed significantly in terms of onset *f*0. In contrast, English short lag and lead VOT stops, which are sub-phonemic variants of the same phonological category, did not differ in terms of onset *f*0. These results highlight the importance of phonological factor in determining the pattern of covariation between VOT and onset *f*0.

1. Introduction

Phonological features such as voicing are realized phonetically in terms of a constellation of coordinated articulatory gestures, and are manifested in the acoustic signal in terms of a variety of cues that contribute to the perception of the phonological feature in complex manner that is still poorly understood. Although there are many cases in which two acoustically distinct phenomena covary in the production and perception of a particular phonological feature, such covariation may result from the origin of the two cues in the same (or linked) articulatory gestures, or may have developed because the two cues contribute to the same perceptual response in a listener's auditory system. For example, both voice onset time (VOT), the time between the release of the consonant and onset of voicing, and onset *f*0, the fundamental frequency at the onset of the vowel following the stop, appear to covary cross-linguistically in the production of voicing (House & Fairbanks, 1953; Hombert, 1976; Lehiste & Peterson, 1961; Löfqvist, Baer, McGarr, & Story, 1989; Ohde, 1984). However, the factors responsible for this covariation are not entirely clear. Two different views on the nature of this relationship have been offered in the literature. A phonetic approach views the VOT–onset *f*0 correlation as automatic and physiologically determined (Hombert, Ohala, & Ewan, 1979; Löfqvist et al., 1989). According to this perspective the effect of voicing on both VOT and onset *f*0 is an automatic consequence of articulatory and/or aerodynamic settings involved in voicing production and is not directly controlled by the speaker. In contrast, a more phonological approach proposes that the connection between these two cues is intentional and phonologically-determined (Keating, 1984; Kingston & Diehl, 1994; Kingston, 2007). According to this perspective, the onset *f*0 cue serves to enhance the perception of voicing in [+voice] stops, thereby increasing the perceptual distinctiveness between [+voice] and [−voice] stops. In this paper we provide new evidence in support of a phonological influence on covariation between the onset *f*0 and VOT correlates of voicing in Spanish and English.

In support of the phonetic approach, Löfqvist et al. (1989) showed that higher levels of activity in the cricothyroid (CT) muscle, which controls the tension of the vocal folds, were detected in production of voiceless consonants by speakers of both Dutch and English (see also Hoole and Honda, 2011 for similar results in German). Greater tension is associated with higher rates of vocal fold

* Corresponding author at: Purdue University, West Lafayette, IN 47907-2038, USA. Tel.: +1 765 494 9330; fax: +1 765 496 1700.
E-mail address: odmitrie@purdue.edu (O. Dmitrieva).

0095-4470/$ - see front matter © 2014 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.wocn.2014.12.005

vibration and thus higher onset *f*0. While Löfqvist et al. (1989) argued that greater vocal fold tension in voiceless consonants may arise from the need to suppress vibration during the voiceless stop closure, Hoole and Honda (2011) suggest instead that vocal fold tensing during the production of voiceless consonants is aimed at a more precise control of voicing onset to prevent vibration from re-starting too soon after the voiceless consonant, leading to a crisper, sharper transition from voicelessness to modal phonation. The end result in either case is that both voicelessness and higher onset *f*0 may stem from the same articulatory gesture, namely tensing of the cricothyroid muscle. That is, a speaker aiming to produce an exemplar of a particular voicing category would implement it by means of an appropriate laryngeal setting. This setting then has a determinative effect on both the voicing of the stop, in particular in terms of its VOT value, and on the fundamental frequency of the following vowel. Consistent with this hypothesis, in the overwhelming majority of reports, voiceless stops are typically realized with *higher* onset *f*0.

However, evidence of a physiological basis underlying both voicelessness and high onset *f*0 values does not necessarily mean that the relationship between these two cues is purely physiological. It is possible that a connection which originally emerged due to physiological factors can become an intentional resource for increasing the perceptual distance between voiced and voiceless stops. A number of findings are consistent with this perspective. For example, onset *f*0 has been shown to covary with voicing even in cases where a *phonological* voicing distinction involves two types of stops both of which are *phonetically voiceless* (voiceless unaspirated and voiceless aspirated), such as word-initial stops in English (Ohde, 1984) and lenis vs. aspirated stops in Korean (Cho, Jun, & Ladefoged, 2002). These findings suggest that the onset *f*0 correlate might enjoy a certain degree of independence from its physiological precursors. According to this hypothesis, because it is a natural acoustic correlate of the phonetic voicing difference, onset *f*0 may be recruited to cue a phonologically related but phonetically different contrast between voiceless unaspirated and voiceless aspirated stops. In other words, onset *f*0 covariation becomes a property of phonological voicing rather than merely a byproduct of phonetic voicing.

In addition, *f*0 differences in a variety of languages have been shown to continue farther into the vowel than is thought to be necessary to control voicing during the consonant production. Hoole and Honda (2011) recently replicated and extended the findings of Löfqvist et al. (1989), showing that production of voiceless stops in German is associated with higher CT activity. However, they also found that there were significant differences in CT activity during the following vowel as well, for some participants in particular. Since the mechanics of voicing control in consonants do not require different CT activity during the following vowel, this articulation can be viewed as intentional and directed at increasing the acoustic difference between voiced and voiceless consonants. Further support for the intentional nature of the covariation between VOT and onset *f*0 comes from research which shows that this covariation may be minimized in tonal languages, where fundamental frequency is involved in cuing another important phonological distinction – lexical tones (Francis, Ciocca, Wong, & Chan, 2006; Gandour, 1974; Hombert, 1977). For example, Francis et al. (2006) showed that in Cantonese, short lag and long lag stops differed only minimally in terms of onset *f*0: the difference was considerably smaller in duration than that reported for non-tonal languages, such as English, and was not sufficient to influence perception of the relevant phonological contrast. Moreover, there is some evidence which suggests that onset *f*0 perturbation is not inevitable even if appropriate physiological conditions are met. Phonetic voicing differences that are not phonologically contrastive are not necessarily accompanied by onset *f*0 differences. For example, Kingston and Diehl (1994) reported that in Tamil, where stop voicing is allophonic, onset *f*0 does not correlate with voicing differences in stop consonants. This finding can be explained in a very straightforward manner: If onset *f*0 functions primarily as a cue to a *phonological* distinction, then it need not vary with VOT when that variation is simply phonetically conditioned (although, the phonological account does not necessarily preclude onset *f*0–VOT covariation in such cases).

The phonological (controlled) and the phonetic (automatic) view of onset *f*0 covariation with voicing are not irreconcilable. Recent research in this area has begun to support a hybrid approach: one which combines the ideas expressed by Löfqvist et al. (1989) as well as those of Kingston and Diehl (1994), among others, and gets us 'the best of both worlds'. Hoole and Honda (2011) propose that the CT activity patterns, which originate in the articulatory properties of voicing production, can be deliberately exaggerated by some speakers as part of an enhancement strategy aimed at increasing the perceptual distinctiveness of the voicing contrast. As a result, CT activity differences, as well as onset *f*0 differences, extend well into the vowel but only for some speakers. Chen (2011) examined voicing–*f*0 interactions in the tone-sandhi domain in Shanghai Chinese and found that the observed *f*0 patterns can be best explained by the interaction of phonetic and phonological factors. On the one hand, voicing-dependent *f*0 perturbation interacted with the larger pitch context (preceding lexical tone) suggesting a phonetic effect. At the same time, voicing-conditioned *f*0 differences were exaggerated in focus position, suggesting intentional manipulation by the speakers.

The present study builds upon this research by examining data particularly suitable for investigating the interaction between the phonetic and phonological factors in determining the patterns of voicing-onset *f*0 covariation. Specifically, we consider the case of a phonetically comparable voicing difference used contrastively in one language and non-contrastively (as phonetic variants of the same phoneme) in another. Examining such data allows for a more direct juxtaposition of phonetic and phonological effects on onset *f*0 and resulting findings will contribute to our understanding of the extent to which each one controls onset *f*0 patterns. The following sections will briefly review previous findings concerning onset *f*0 covariation with voicing across two major types of voicing contrast and introduce specific goals and hypotheses of the present study.

1.1. Voicing contrasts and onset *f*0

1.1.1. Across languages

It is generally accepted that VOT is the principal acoustic and perceptual correlate of voicing contrasts in syllable-initial position (Lisker, 1975, 1978; Raphael, 2005). Three types of VOT values are typically used by languages to distinguish voicing categories

(Cho & Ladefoged, 1999): lead VOT (laryngeal voicing begins during the stop closure, prior to release), short lag VOT (a very short or non-existent lag between the consonant release and the beginning of the following vowel), and long lag VOT (a relatively long period of aspiration-filled near-silence occurs between the stop release and the onset of vocalic voicing). Such types of stops are usually referred to as voiced, voiceless unaspirated, and voiceless aspirated, respectively. Languages can contrast all three stop series but often only two are selected. In 'voice' languages, lead VOT stops represent the [+voice] category and are contrasted with [−voice] short lag stops. In 'aspiration' languages, short lag stops represent the [+voice] category and are contrasted with [−voice] long lag stops. Thus, voice languages contrast phonetically voiced (lead) and phonetically voiceless (short lag) stops, while aspiration languages contrast two phonetically voiceless types of stops (short lag and long lag). Among the commonly referenced languages exhibiting a 'voice' contrast are Spanish, French, and Russian. Examples of languages with an 'aspiration' contrast include English (in initial position) and Cantonese. Based on the data available it is difficult to make definitive statements about how common particular types of voicing contrasts are. However, it appears that two-category contrasts may be found more frequently than three-category contrasts: In the UPSID database of 317 languages, about 50% of languages contrast two voicing categories, while only 25% contrast three (Maddieson, 1984).[1] Among the two-category languages, voice-type languages seem to dominate (Maddieson, 1984). However, it must be noted that many languages, including English, make use of one type of contrast in one phonetic context and another in others (see Section 1.1.2), and it is not always clear in large-scale language surveys how such discrepancies are resolved when determining the type of contrast said to be used in that language.

Both voice and aspiration languages have been examined with respect to the covariation between voicing and onset $f0$, although the data is much scarcer for voice languages. A significant covariation between phonological voicing and onset $f0$ has been reported for both aspiration and voice languages. For aspiration contrasts see multiple studies on English, including Ohde (1984), House and Fairbanks (1953), and Lehiste and Peterson (1961) among others[2]; also Lai, Huff, Sereno, and Jongman (2009) on Taiwanese, and Jeel (1975) and Reinholt Petersen (1983) on Danish. For work on voice languages see Hombert (1976) on French (two speakers), Caisse (1982) on French, Italian, Spanish, and Portuguese (a single speaker for each language) and Löfqvist et al. (1989) on Dutch (two speakers). Almost universally, and especially in the case of lead-short lag contrasts, a *higher* onset $f0$ was reported to co-occur with voiceless stops while a *lower* onset $f0$ co-occurred with voiced stops. This pattern is consistent (at least for voice languages) with the predictions of the vocal fold tension hypothesis. However, other findings support the interpretation that it is the *phonological* status of a segment rather than its VOT (or its underlying articulatory source) that plays a role in determining onset $f0$. For example, onset $f0$ is generally observed to be lower for [+voice] stops than for [−voice] ones, irrespective of whether that [+voice] category is realized with lead VOT (in voice languages) or short lag VOT (in aspiration languages) (Kingston & Diehl, 1994), although some violations of this tendency have been documented (see Chen, 2011 for review), particularly among tonal languages and languages with more than two contrasting stop series.

1.1.2. Within languages (across phonetic contexts)

Different types of voicing contrasts can also be employed by the same language in different phonetic contexts. Thus, English uses an aspiration contrast (short lag [+voice] vs. long lag [−voice]) in utterance-initial position, but in the intervocalic unstressed environment (*rabid-rapid*) English tends to exhibit a voice-type contrast (lead voicing [+voice] vs. short lag [−voice]). Despite these contextual differences, phonological stop consonant voicing in English shows a consistent pattern of onset $f0$ in both utterance-initial (Caisse, 1982; Lehiste & Peterson, 1961; Ohde, 1984) and intervocalic environments (House & Fairbanks, 1953; Hombert, 1976; Löfqvist et al., 1989; Ohde, 1984). In all reports, onset $f0$ is *higher* after [−voice] stops and *lower* after [+voice] stops, regardless of the precise phonetic realization of the [±voice] contrast. However, a trend that has not received much attention in the literature to date is that speakers of English produce a certain proportion of lead VOT [+voice] stops in utterance-initial position (Docherty, 1992). It is not known whether, in such stops, phonetic voicing takes precedence over phonological status in determining the onset $f0$ level.

1.2. The present study

Thus, research on onset $f0$ and voicing covariation provides evidence suggesting that both phonological and phonetic factors influence the relationship between VOT and onset $f0$. The phonetically-based view is supported by the fact that, in almost all reports, phonetically voiceless stops are realized with *higher* onset $f0$, as predicted by the vocal fold tension account (Löfqvist et al., 1989). In favor of the phonological approach is the fact that, in both voice and aspiration languages, phonologically voiced stops tend to exhibit lower onset $f0$ than do phonologically voiceless ones, although the production of the voicing contrast involves very different physiological and acoustic differences in aspiration languages as compared to voice languages (e.g., aspiration languages contrast two phonetically voiceless types of stops, while voice languages contrast phonetically voiced with phonetically voiceless ones).

Most studies of voicing and onset $f0$ have focused on cases in which phonetic differences along the VOT continuum correspond to phonological differences (contrastive voicing). However, cases in which phonetics and phonology are not in a one-to-one relationship present a better testing ground to contrast the phonetic and phonological hypotheses. Such cases include (i) those in which phonetically different stops correspond to the same phonological category (non-contrastive voicing or sub-phonemic variation) and

[1] The survey by Keating, Linker, and Huffman (1983), which focuses specifically on positional allophones of voiced and voiceless segments, suggests a more equal distribution; however this selection may not be as comprehensive as the UPSID survey due to its smaller size (51 languages).

[2] Many English studies used stimuli which actually involved a voice contrast (see section on voicing contrasts across contexts).

(ii) those in which phonetically identical stops are used for two distinct phonemic categories (across contexts in the same language or across languages).

A comparison between English (an aspiration language, at least in initial position) and Spanish (a voice language) with respect to phonetic and phonological voicing and onset *f*0 provides an opportunity to investigate both cases. In Spanish, utterance-initial [+voice] stops have lead VOT and [−voice] stops have short lag VOT. English utterance-initial [+voice] stops are often short lag VOT stops but can also have lead VOT (Docherty, 1992). English [−voice] initial stops are long lag VOT stops. Thus, the difference between lead voicing and short lag VOT in utterance-initial position is contrastive in Spanish but non-contrastive in English, as in (i), above. Furthermore, short-lag initial stops are [+voice] in English, but [−voice] in Spanish, as in (ii), above. Examination of onset *f*0 across the VOT types in English and Spanish can help determine the relative contributions of phonetic and phonological factors in defining the patterns of onset *f*0 covariation with voicing. Examination of short lag stops in both languages is particularly important in addressing this question. Specifically, the phonetic approach predicts higher onset *f*0 for short lag stops than for lead stops in both English and Spanish, while the phonological approach does not make such a prediction for English. Unlike the phonetic approach, the phonological account *does not require* English short lag stops to differ from English lead stops although it does not preclude this possibility. Additionally, according to the phonetic approach, lead VOT stops should have similar onset *f*0 properties across English and Spanish, and so too should short lag stops: Because they exhibit comparable VOT values, they should be realized with similar articulatory gestures, and therefore other acoustic properties derived from those gestures (e.g. onset *f*0) should also be similar. The phonological approach, on the other hand, makes no prediction regarding the similarity of onset *f*0 values in short lag stops in the two languages. On the contrary, it is possible that onset *f*0 values for short lag stops would differ across the two languages because they represent a [−voice] category in Spanish but a [+voice] one in English.

The phonetic predictions are less straightforward for the short lag–long lag contrast, since the physiological relationship between onset *f*0 and gestures related to longer VOT values is not well understood. The vocal fold-tension hypothesis predicts lower *f*0 after lead stops compared to plain voiceless and voiceless aspirated stops; however it predicts no difference between the latter two types. Given the empirical results of previous studies on English and languages with a similar type of voicing contrast, such as Danish[3] (Jeel, 1975; Lehiste & Peterson, 1961; Reinholt Petersen, 1983) we might expect a higher onset *f*0 after long lag stops than after short lag stops in English but this could be phonologically conditioned. Indeed, a phonological approach would specifically predict a difference in this direction since short lag stops represent a [+voice] category (=lower onset *f*0) while long lag stops represent a [−voice] category (=higher onset *f*0). The main predictions are summarized in Table 1.

The phonological approach can also be extended to predict gradient onset *f*0–VOT correlation patterns within each voicing category based on two assumptions. The first is that onset *f*0 variation is governed by considerations of phonological contrast enhancement, i.e. the goal of making members of contrasting categories more perceptually distinct. The second assumption is that perceptual cues to contrasts exist in a 'trading relation', i.e. when one cue is weakened or ambiguous, it will be compensated for by a stronger contribution from another cue (Repp, 1982). For example, there is evidence that secondary cues, such as onset *f*0, tend to contribute more to the voicing decisions when the primary cue, VOT, is ambiguous (Abramson & Lisker, 1965; Whalen, Abramson, Lisker, & Mody, 1990). Given that such trading relations between cues have been shown to exist in perception, it seems plausible that speakers may also compensate for relatively ambiguous primary cue values by emphasizing secondary cues in production, thus making potentially confusable stops more distinct from the contrasting ones.

Since low onset *f*0 is predicted to co-occur with lead VOT in the Spanish [+voice] category (see Table 1), both correlates can be expected to cue [+voice] category in Spanish and can therefore enter into a trading relation. Stops produced with a relatively short lead VOT (making them more similar to [−voice] stops) may be 'repaired' by emphasizing their low onset *f*0. If this enhancement strategy is implemented consistently across the range of VOT values within the [+voice] category, we would expect to see a negative correlation between VOT and onset *f*0 in Spanish [+voice] stops: as VOT increases (gets less negative, or closer to 0) onset *f*0 is expected to drop.

Similarly, if both high onset *f*0 and near-zero or slightly positive VOT are correlates of [−voice] Spanish stops, they can be used as cues for the [−voice] category. Smaller positive VOT makes [−voice] stops more similar to [+voice] ones, which may be compensated for by higher onset *f*0 values. Thus, a negative VOT–*f*0 correlation would be expected here as well: as VOT decreases, onset *f*0 is expected to rise.[4]

In English, the trading relation-based enhancement hypothesis would also predict a negative correlation between VOT and onset *f*0 within both [+voice] and [−voice] categories (provided the phonological predictions in part 3 of Table 1 are confirmed). Within the English [+voice] category, greater positive VOT values are ambiguous, making stops more similar to [−voice] ones. Thus a lower onset *f*0, characteristic of [+voice] stops, would be expected. Within the English [−voice] category, smaller positive VOT values are ambiguous, making stops more similar to [+voice] ones. Thus a higher onset *f*0, characteristic of [−voice] stops, would be expected.

Results of the present production study may also be relevant for theories of cue weighting and cue integration in perception of phonetic contrasts. A number of studies have demonstrated the importance of secondary cues, onset *f*0 in particular, in perceptual decisions, including identification of voicing category (Abramson & Lisker, 1985; Castleman & Diehl, 1996; Haggard, Ambler, & Callow, 1970; Oglesbee, 2008; Whalen, Abramson, Lisker, & Mody, 1993). However, the mechanisms underlying the integration of multiple cues in speech perception are currently under debate (Kingston & Diehl, 1995; Kingston, Diehl, Kirk, & Castleman, 2008).

[3] Aspiration contrast in initial position, with [+voice] stops realized with voicing lead elsewhere, normally in the intervocalic position (e.g. Danish).

[4] We were reminded by John Kingston that there is always much less VOT variation in short lag stops in comparison to lead or long lag stops. This smaller degree of VOT variability may, in turn, offer fewer possibilities for trading relations with *f*0 in short lag stops than in lead or long lag stops.

Table 1
Predictions of the phonetic and phonological accounts of onset *f*0 differences.

		Phonetic	Phonological
1.	Spanish	Lead<Short lag	Lead<Short lag
2.	*English*	*Lead<Short lag*	No prediction
3.	English	*Short lag=Long lag*	*Short lag<Long lag*

Note: Predictions not shared by the alternative approach are in italics. Cells representing critical cross-language comparisons are shaded.

According to one hypothesis, listeners learn to integrate multiple acoustic properties into a cue to a specific phonetic contrast because those properties covary in the ambient language (Holt, Lotto, & Kluender, 2001; Stilp, Rogers, & Kluender, 2010). Therefore, failure to integrate certain properties in the perception of a contrast receives a straightforward experience-based explanation if the properties in question do not in fact covary in the speakers' ambient language. In such cases, lack of experience with the relevant type of covariation would explain why listeners did not learn to treat these two cues as integral in perception. For example, a recent perceptual study (Llanos, Dmitrieva, Shultz, & Francis, 2013) demonstrated that onset *f*0 plays little role in voicing decisions for initial stops in Spanish (see also Oglesbee, 2008). Based on the specific pattern of responses, Llanos et al. (2013) argued against an experience-based explanation and in favor of an auditory enhancement account (for full details see Llanos et al., 2013; see also Kingston et al., 2008). However, rejection of the experience-based account in perception would be strengthened by demonstration that, despite the lack of perceptual integration, the two cues nevertheless covary in Spanish stop consonant production. In other words, listeners did not learn to integrate the two cues even though they covary in the ambient language. However, in order to make this argument, it is essential to determine whether or not those cues actually do covary in a given language. Studies of patterns of cue covariation in production can thus inform the development of theories of speech perception. The current study is the first large scale investigation of the covariation between voicing and onset *f*0 in Spanish initial stops, and the resulting data therefore also contribute to the assessment of experience-based explanations of cue integration in perception of these stops.

2. Methods

2.1. Participants

Twenty-four native speakers of Spanish (10 men, 14 women, mean age 29.4 years, ranging from 18 to 54 years of age) were recorded at the Centro de Ciencias Humanas y Sociales − Consejo Superior de Investigaciones Cientificas (CSIC-CCHS) in Madrid, Spain and 30 native speakers of American English (15 men, 15 women, mean age 25.0 years, ranging from 20 to 32 years of age) were recorded on the campus of Purdue University in West Lafayette, Indiana.[5] All participants were paid for their time.

All Spanish participants identified Spanish as their first language. Among the 24 participants, 20 were from Spain (10 of them from Madrid), 3 were from Latin America (Chile, Uruguay, and Mexico) and one chose not to report country of origin. All participants had some exposure to foreign languages, mostly in the classroom setting. A majority of participants had studied English; many reported having studied other languages, mostly those spoken elsewhere in Europe (e.g. French and German) and those spoken regionally in Spain.

Despite this experience this group of participants can be considered a reasonable approximation of monolingual speakers of Spanish for the purposes of the current study. All participants were born and grew up in a Spanish-speaking country; all had a Spanish-speaking country as their current primary country of residence. All participants were tested while immersed in a fully native Spanish-speaking environment. While 10 participants reported having spent time abroad in countries where languages other than Spanish were spoken, most visits were under 12 month in duration and the average temporal gap between the completion of the trip and the time of experiment was 2.4 years. Only 3 participants reported a stay of a considerable duration (between 12 and 24 months) in countries where languages with English-like VOT contrasts were spoken: English (2) and German (1).

All American participants indicated English as their first language and all but two were born in the United States. All participants were raised in the United States in a monolingual English-speaking environment. Every English-speaking participant had studied at least one currently spoken language other than English as part of the typical high school and college education in the United States, with Spanish dominating the list (23 participants). Most visits abroad were quite short in this group and no-one reported having resided in a country where languages with Spanish-like VOT contrasts were spoken. Based on the information available it is reasonable to conclude that foreign language experience in this group was mostly limited to formal classroom setting and did not result in proficiency beyond basic.

All participants reported having normal hearing and no history of speech or language disability.

2.2. Stimuli

Primary Spanish stimuli included four disyllabic, mostly monomorphemic, CVCV minimal pairs contrasting in the voicing of the initial, bilabial stops. In three of the pairs the voiced bilabial stop /b/ was represented orthographically as 'b': *bata-pata* (*robe*/*paw*), *beso-peso*

[5] These are the same participants, methods and procedures used by Shultz, Francis, and Llanos (2012).

(*kiss*/*weight*), *biso-piso* (*an encore; to give an encore, 1st p.sing./apartment; to step, 1st p.sing.*); in the remaining pair /b/ was spelled as 'v': *visa-pisa* (*ʋisa*/*to step on, 3rd p. sing.*). Three different front vowels were used across pairs: [i], [e], and [a]. With an exception of one item (*biso*), all stimuli were lexemes of high familiarity, as confirmed by a native speaker of Spanish (second author), and of comparable frequency: mean frequency of 33.6 words per million, ranging from 3 (*ʋisa*) to 91 (*peso*) (Almela, Cantos, Sánchez, Sarmiento, & Almela, 2005). The only exception is represented by the item *biso*, which corresponds to either a 1st person singular form of the verb *bisar* meaning 'to give an encore, to repeat' or a noun 'encore' and which was not listed in the Almela et al. (2005) frequency dictionary of Spanish. Because of the low frequency and familiarity of *biso*, which may affect cues to voicing (Goldrick & Rapp, 2007), a more familiar and frequent word (*ʋisa*) was also included.

Because preliminary examination of pilot data indicated a possibility that orthographic representation of /b/ stops may have an effect on phonetic properties of the consonants, and it was not possible to construct a complete, frequency- and familiarity-balanced, set of minimal pairs without including a 'v'-initial /b/ word in the first list, a second word-list was included for recording as well to permit comparison between 'b'-initial and 'v'-initial /b/ words. Three minimal pairs contrasting in the voicing of the initial bilabial stop were included in the second word-list: *vana-pana* (*ʋain*/*velvet*) *veto-peto* (*ʋeto*/*overalls*), *visa-pisa* (*ʋisa*/*to step on, 3rd p. sing.*). Across pairs, the same three front vowels that appeared for 'b'-initial words in list 1 were used in list 2 but, unlike in list 1, all /b/ stops in the second list were spelled as '*ʋ*'. Words were of high familiarity and comparable frequency (mean frequency of 3.4 words per million ranging from 2 to 5).

Sixteen distractor items were added to the first list and twelve distractor items (a subset of the first 16) were added to the second list. These words were all of the disyllabic (C)VCV structure (always CVCV orthographically) and had segments other than bilabial stops in initial position, including fricatives ([f] and [s] as in *fino* 'fine' and *sapo* 'toad' and interdental [θ] as in *cepa* 'rootstock, vine'[6]), velar and alveolar stops ([k], [d] as in *caso* 'event' and *dedo* 'finger'), sonorants ([m], [l], and [r] as in *mito* 'myth', *lodo* 'mud', and *raso* 'flat'), and vowels ([i] in words with an initial silent *h: hipo* 'hiccup' and *hilo* 'thread'). Distractor items were lexemes of high familiarity and comparable in frequency to target words (mean frequency of 56 words per million, ranging from 1.5 to 476). Most of the distractor items were minimal pairs for initial or medial consonants (e.g. *caso-raso, codo-lodo, foro-loro, seso-beso/peso*). Thus, list 1 consisted of 24 words (8 target words and 16 distractors) and list 2 consisted of 18 words (6 targets and 12 distractors). The target pair *visa-pisa* was included in both lists. All Spanish stimuli and distractor items had penultimate stress.

English stimuli consisted of four monomorphemic monosyllabic CVC minimal pairs, where members of the pair differed only in the voicing of the initial, bilabial stop consonants: *bat-pat, bet-pet, beat-Pete, bit-pit*. All target words had a comparable frequency (mean frequency of 36 words per million, ranging from 8 to 101) and high familiarity, estimated with the Washington University Speech and Hearing Lab Neighborhood Database (2013) (Washington University Speech & Hearing Lab). In addition to target words, eight distractor pairs were included in the word-list. Half of the distractor words were fricative-initial ([f] or [h] as in *fit* and *heap*); the remaining fillers had a non-bilabial stop as the initial segment ([d] or [k] as in *cat* and *deed*). All distractor items were minimal pairs for the initial consonant: e.g., *fig-dig, heap-keep, fat-cat*. Distractor items were comparable in frequency to target words (mean frequency of 131 words per million, ranging from 1 to 686) and equally high in familiarity. Full details can be found in Shultz et al. (2012).

2.3. Procedure

Participants were seated in front of the computer screen in a quiet room (US) or in a sound-attenuated booth (Spain). Stimuli were presented one at a time on the screen, black on white, in Times New Roman font, 72 or 48 points font size (Spain and US, respectively). Each word remained on the screen for 2 s and was followed by a 500 ms interval of blank screen. Stimuli were presented to US participants using a Dell Optiplex/Windows XP computer and E-Prime 1.2 interface (Schneider, Eschman, & Zuccolotto, 2002) and to Spanish participants using an ACER Pentium (R)/Windows XP computer and MATLAB and Statistics Toolbox Release (2001) graphical user interface written in-house. Participants were instructed to say each word aloud in a normal speaking voice as it appeared on the screen. In the recording of the first Spanish word-list, a set of 24 words (8 targets and 16 distractors) was presented to each participant five times (120 words in total, 40 targets), randomized for each of the five blocks. In the recording of the second Spanish word-list, a set of 18 words (6 targets and 12 distractors) was presented to each participant 5 times (90 words in total, 30 targets). All Spanish participants produced both lists. In the recording of the English word-list, a total of 24 words (8 targets and 16 distractors) was presented to each participant five times (120 words in total, 40 targets), randomized for each of the five blocks. Participants in both groups were given an opportunity to take a short break after each block.

On-screen presentation of the stimuli made it possible to control for the rate of speech and, to a great extent, intonation. Presentation of individual words ensured that both groups of participants pronounced the words with largely uniform (and similar) declarative statement intonation, realized with a falling pitch contour. Furthermore, because the words were produced in isolation, each constituted an intonation phrase, with a well-controlled prosodic boundary before and after each word. Finally, this elicitation method placed target words in absolute utterance-initial position, the most favorable context for eliciting the short lag allophone of English phonologically voiced stops.

Speech material was recorded in .wav format at 44.1 kHz sample rate, 16 bit quantization using a Marantz Professional solid state recorder (PMD 660) with a unidirectional hypercardioid microphone (Audio-Technica D1000HE) for American participants and using a

[6] Tokens pronounced with [θ] by speakers of central and northern dialects of Iberian Spanish are typically produced with [s] in other dialects of Spanish. The choice of realization is irrelevant for the present paper.

Alexis Multimix 16 USB recorder with a AKG C444L cardiod condenser microphone for Spanish participants. The recording session for each participant lasted 5–10 min.

2.4. Measurements

Measurements consisted of VOT and onset *f*0. Fundamental frequency was also measured at ten additional locations, evenly spaced every 10 ms after the initial onset *f*0 measurement point.[7] All measurements were performed with Praat 5.1 (Boersma & Weenink, 2009). VOT was measured from the beginning of the release burst of the stop consonant to the onset of voicing identified as an onset of periodic waveform and low-frequency voicing energy on the spectrogram (Francis, Ciocca, & Yu, 2003). Thus, for short lag and long lag tokens VOT encompassed the release burst and the aspiration period, if any, prior to the onset of the vowel. For the lead voicing tokens, VOT consisted of the prevoiced stop closure up to the beginning of the stop burst (Fig. 1).

Onset *f*0 was measured at the first point in time immediately following the end of the VOT portion at which the Praat default pitch tracking algorithm was able to detect periodicity. The average period between the observed onset of voicing and the first pitch measurement was 3 ms (sd 6 ms) for the Spanish group and 5 ms (sd 10 ms) for the English group.[8] In both languages, high vowels on average conditioned earlier pitch detection than non-high vowels (2 ms earlier on average). In English, pitch was also detected earlier after voiceless than after voiced stops (2 ms vs. 8 ms into the vowel), while the opposite was true for Spanish (4 ms vs. 2 ms into the vowel, respectively).

All resulting pitch values were visually examined for outliers potentially indicative of pitch doubling or pitch halving and other algorithm errors. Errors were corrected manually by taking the reciprocal of the waveform period (first identifiable period immediately after the VOT portion for onset *f*0 values). About 1% of all Spanish pitch measurements, 3% of English onset *f*0 measurements, and 6% of English non-onset pitch measurements were corrected in this manner.[9] To facilitate onset *f*0 comparison across genders, the *f*0 values for each participant were converted from Hz to semitones relative to each participant's mean onset *f*0 (cf. Shultz, et al. 2012). The formula used for this conversion was $12\ln(x/\text{individual mean onset } f0)/\ln 2$ (similar to the one found in Praat users' manual (Boersma & Weenink, 2009) but made relative to the individual mean instead of 100 Hz). The resulting values represent relative distance of each data point from the speaker's onset *f*0 mean on a logarithmic scale: positive values are instances of higher than average *f*0, negative values are lower than the average *f*0.

As a measure of reliability four participants were randomly selected from each group and VOT and onset *f*0 were re-measured for these participants by another experimenter. Measurement reliability was evaluated via correlation analysis applied to the series of measurements performed by the two experimenters. For the Spanish group, both VOT and onset *f*0 values were highly correlated between the two experimenters: $r=0.97$, $p<0.0001$ (VOT) and $r=0.94$, $p<0.0001$ (onset *f*0). The mean absolute difference between the values obtained by two experimenters was 2.5 ms for VOT and 1 Hz for onset *f*0. For the English group, likewise, a strong significant correlation was established for the VOT ($r=0.97$, $p<0.0001$) and onset *f*0 ($r=0.99$, $p<0.0001$) values reported by the two experimenters. The mean absolute difference between the two sets of VOT values was 2.2 ms and between the two sets of onset *f*0 values was 0.40 Hz.

2.5. Analysis

The duration of VOT for each VOT Type (lead, short lag, long lag) was examined in a Repeated Measures ANOVA in each language (with VOT Type as an independent factor). The duration of VOT for lead and short lag tokens was also examined across the two languages in a mixed-design Repeated Measures ANOVA with Language and VOT Type as independent variables.

Two types of statistical analysis were used in order to examine the connection between the voicing category/VOT type and the onset *f*0. A series of Repeated Measures (RM) ANOVAs was applied to the onset *f*0 data to evaluate the effect of VOT Type (lead, short lag, long lag) and Phonological Categories ([±voice]) on the *f*0 values across and within languages. In addition, within each phonological category, VOT and onset *f*0 data were submitted to correlation analyses to examine the hypothesis that more ambiguous VOT values are compensated for by more prototypical onset *f*0 values. All RM ANOVA analyses were checked for violations of the sphericity assumption and Greenhouse–Geisser correction of the degrees of freedom was applied when necessary. Corrected degrees of freedom are reported when the sphericity assumption was violated. All reported group means were calculated from means for each speaker and category (not from individual tokens).

3. Results

The following results are based on data obtained from the Spanish word-list 1 except for discussion of the orthographic influence, which compares measurements from word-list 1 and word-list 2.

[7] Because many of the syllables in question were closed and/or contained lax vowels, not all syllables were a full 100 ms in duration. In those cases, measurements were made at as many points as possible without extending into the coda consonant or beyond.

[8] This difference between languages was significant (RM ANOVA, $F(1,52)=6.525$, $p<0.05$). A certain amount of individual variability was present in both groups: for three English-speaking participants and three Spanish speaking participants pitch detection occurred later than on average for their respective language groups.

[9] The settings of the Praat autocorrelation algorithm were adjusted for two English-speaking participants to avoid a higher than average incidence of pitch tracking errors.

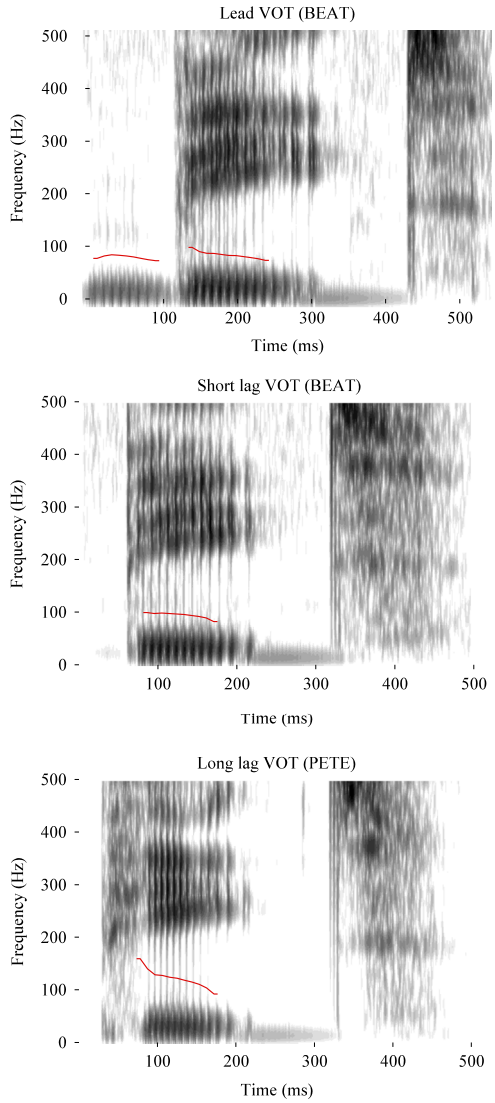

Fig. 1. Spectrograms and superimposed *f*0 trace for three sample stimuli. Top: Lead voicing VOT production of English *beat*; middle: short lag VOT production of English *beat*; bottom: long lag VOT production of English *Pete* (all by the same talker).

3.1. Prevoicing in English

In order to make the data analysis presented below clear, it is first necessary to discuss the results with respect to the proportion of prevoiced tokens among the [+voice] stops of the English-speaking participants.

Spanish [+voice] initial stops are reportedly produced exclusively with lead voicing VOT, and this expectation was confirmed in the present results. In contrast, the phonetic realization of English phonological voicing in stop consonants in initial position is reported to vary, both within and across talkers, between two distinct phonetic realizations: short lag VOT and lead voicing VOT, although there is little consensus as to the basis for this variation (see Shultz, 2011 for discussion). In the dataset reported here, approximately 31% of initial voiced stops produced by speakers of American English were prevoiced. Among the 30 US participants, only seven produced /b/-initial tokens exclusively with a short lag VOT and only one participant produced all /b/-initial tokens with lead voicing VOT. For the remaining 22 participants productions of the [+voice] category included both short lag VOT and lead voicing realizations. In this subgroup, 38% of all /b/ tokens showed lead voicing VOT. In most cases, within-participant productions were dominated by either prevoicing or short lag tokens. Only two participants' distributions were equally divided between short lag and lead voicing VOT (50% of each category). Fig. 2 demonstrates the percentages of lead vs. short lag tokens for each English speaker.

3.2. VOT results

In Spanish, [+voice] stops' VOT values centered around −94.7 ms (sd 31.5 ms) while [−voice] short lag stops had a mean VOT of 14 ms (sd 4.7 ms). The two distributions were significantly different from each other by Repeated Measures ANOVA: $F(1, 23) = 555.803$, $p < 0.001$; partial $\eta^2 = 0.960$.

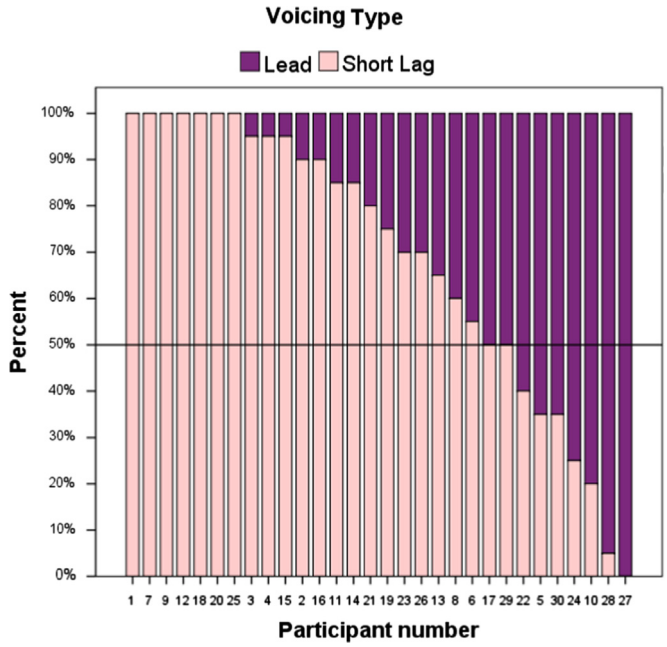

Fig. 2. The percentages of lead vs. short lag tokens for each English participant. Participants are listed according to the percent short lag tokens in the descending order.

In English, prevoiced [+voice] stops had an average VOT of −107.3 ms (sd 32 ms). English short lag [+voice] stops centered around 12.1 ms VOT (sd 5 ms). Long lag [−voice] stops in English had a mean VOT of 64.2 ms (sd 18.2 ms). All three distributions were significantly different from each other (one-way ANOVA with subject as a random factor: $F(2, 55)=793.238$, $p<0.001$; partial $\eta^2=0.966$).[10]

Across languages, the VOT duration of prevoiced and voiceless unaspirated tokens were examined in a mixed-design Repeated Measures ANOVA, with VOT type (within-subject) and Language (between-subject) as independent factors. There was no effect of Language on either short lag or lead tokens' VOT. There was also no Language–VOT interaction, suggesting that the VOT difference between lead VOT stops and short lag stops was of the same magnitude in both English as in Spanish.

3.3. Effect of orthography

As reported in Section 2, in the stimuli used for word-list 1, Spanish [+voice] stops were represented orthographically in two different ways: with a 'b' grapheme and with a 'v' grapheme. A standard assumption about the correspondence between spelling and pronunciation in Spanish is that both 'b' and 'v' are pronounced exactly alike as a voiced bilabial stop and for this reason no attempt was made to restrict stimuli to only one orthographic variant in word- list 1. However, to our knowledge, this hypothesis has not been empirically tested. Therefore, as a preliminary examination of this assumption, a second word-list with all *v*-initial /b/ targets was recorded and *b*-initial and *v*-initial stimuli in both word-lists were compared via Repeated Measures ANOVAs with respect to the VOT and onset *f*0 parameters. The results showed a significant effect of orthography on VOT, $F(1, 23),=5.146$, $p<0.05$; partial $\eta^2=0.183$, but not on onset *f*0. Voiced bilabial stops represented as 'v' had a significantly *shorter* prevoicing period (mean 'v' VOT: −91 ms) than stops represented as 'b' (mean 'b' VOT: −97 ms). This difference in VOT could be due to the fact that speakers were especially attentive to the spelling differences possibly because the two word-lists were recorded in separate blocks (see Section 4.3 for further discussion). While there was a significant difference between these two types of stops, these differences are overall quite small and they clearly can be viewed as sub-phonemic variants of the /b/-phoneme. Their acoustic parameters were well within the range of a Spanish [+voice] plosive: a strongly negative VOT and a lowered onset *f*0. Moreover, the difference in onset *f*0 between /p/ and /b/ cannot be a result of spelling-related differences, because there was no significant onset *f*0 difference between /b/ spelled as 'v' and /b/ spelled as 'b'. Thus, in the analysis presented below both orthographic variants of Spanish voiced stops included in word-list 1 are considered together under the [+voice] category, but further investigation of orthographic influences on Spanish consonant production is certainly warranted.

3.4. VOT type and onset *f*0

Two major VOT types are present in both Spanish and English. The first is lead voicing VOT, which is the sole manifestation of the [+voice] category in Spanish and which is also well attested as a phonetic variant of the [+voice] category in English. The second, short lag VOT, corresponds to the [−voice] category in Spanish and is generally considered to be the typical expression of the English

[10] This analysis was conducted as a between-groups ANOVA rather than a RM ANOVA in order to include the maximum number of participants. In an RM analysis, we would have had to exclude those English speakers who did not produce any tokens from one of the categories (8 in all).

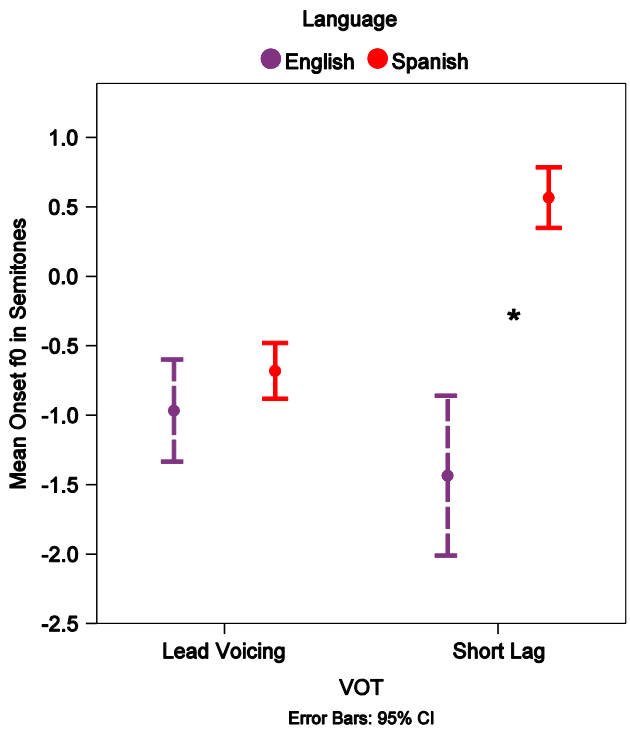

Fig. 3. Effect of language (dashed line: English; solid line: Spanish) and VOT (*x*-axis: lead voicing VOT and short lag VOT) on semitone-normalized onset *f*0 (*y*-axis).

Table 2
Means and standard deviations in semitones for onset *f*0 in Spanish and English lead and short lag stops.

	Lead VOT	Short Lag VOT
Spanish	−0.68 (sd 0.5)	0.56 (sd 0.5)
English	−0.96 (sd 0.8)	−1.4 (sd 1.3)

initial [+voice] category. Thus, the two VOT types are phonetic variants of the same phonological category in English, while in Spanish they correspond to the two opposing phonological classes. The semitone-normalized onset *f*0 values corresponding to these VOT types were submitted to a mixed-design Repeated Measures ANOVA, with VOT type (lead or short lag) as a within-subject factor and Language as a between-subject factor. In the English group, only data from those participants who produced both lead VOT and short lag stops were included in this analysis (22 participants).

Fig. 3 shows that lead VOT stops in both languages are very similar in terms of mean onset *f*0, while short lag stops differ considerably. The mean onset *f*0 of short lag stops in Spanish is much higher than the mean onset *f*0 of short lag stops in English. Both English lead and short lag stops exhibit lower than average onset *f*0 but are very similar to one another in magnitude with a large overlap of the confidence intervals. On the other hand, Spanish short lag stops exhibit a higher than average onset *f*0, setting them considerably apart from Spanish lead stops (as well as from both types of English [+voice] stops) that have lower than average onset *f*0 values.

The results of the omnibus mixed-design Repeated Measures ANOVA showed a significant effect of VOT type, $F(1, 44)=5.234$, $p<0.05$; partial $\eta^2=0.106$, and Language, $F(1, 44)=41.382$, $p<0.001$; partial $\eta^2=0.485$, on onset *f*0. Onset *f*0 of the stops with a short lag VOT was significantly *higher* than onset *f*0 of lead VOT stops. With respect to the language effect, the Spanish group demonstrated a significantly higher onset *f*0 overall (across lead and short lag VOT) than American participants. More importantly, there was also a significant interaction between VOT and Language: $F(1, 44)=25.373$, $p<0.001$, partial $\eta^2=0.366$. The interaction is due to the fact that the main VOT effect was driven by the Spanish group alone: Within the Spanish group, onset *f*0 was *higher* after short lag stops than after lead stops. Within the English-speaking group, on the other hand, onset *f*0 was slightly *lower* after short lag stops than after prevoiced ones (but not significantly so, as shown below). Means and standard deviations are reported in Table 2.

In order to evaluate the effect of the VOT type within each language, separate RM ANOVA analyses were applied to the onset *f*0 values in Spanish and English data. Within the English-speaking group, the difference in onset *f*0 across the two VOT types was not significant. However, within the Spanish-speaking group, the difference in onset *f*0 across the two VOT types was highly significant, $F(1, 23)=52.619$, $p<0.001$; partial $\eta^2=0.696$. Thus, the effect of the VOT type on onset *f*0 was highly significant in Spanish, in which the two portions of the VOT continuum correspond to different phonological categories, but was non-significant in English, in which the two portions of the continuum are subsumed within a single phonological category.

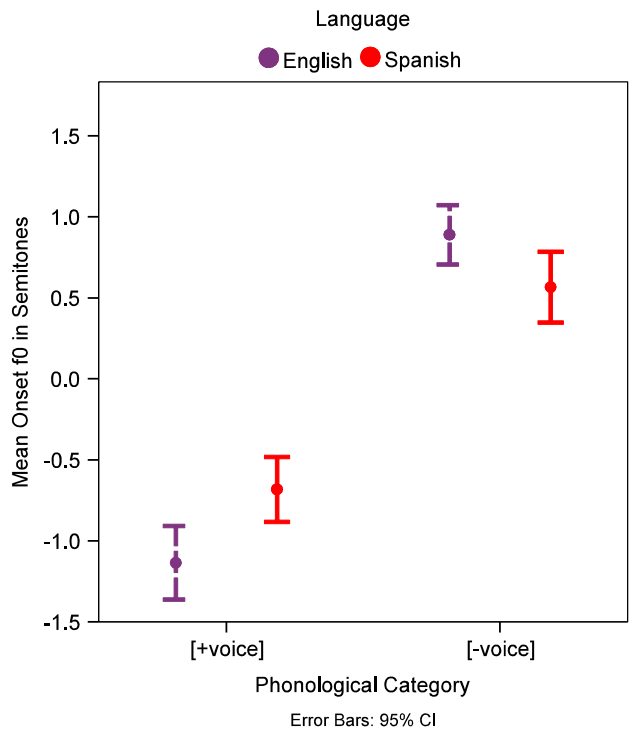

Fig. 4. Effect of language (dashed line: English; solid line: Spanish) and phonological category (*x*-axis: [+voice] and [−voice]) on semitone-normalized onset *f*0 (*y*-axis).

A separate independent-samples *t*-test was also applied to the onset *f*0 values within each VOT type to test for language-specific differences. The analysis showed that there was no significant difference in terms of onset *f*0 between lead VOT stops produced by Spanish and English participants. At the same time, the difference between Spanish and English short lag stops with respect to onset *f*0 was highly significant: $t(44)=-6.972$, $p<0.001$. Thus, Spanish short lag stops were significantly higher in onset *f*0 than English short lag stops.[11] Importantly, Spanish short lag stops in initial position represent a [−voice] category, while English short lags are [+voice].

3.5. Phonological voicing and onset *f*0

In the following analysis the VOT continuum is divided according to the phonological voicing categories of English and Spanish. Thus, both lead VOT and short lag VOT realizations of English /b/ are categorized as [+voice] and compared to English [−voice] long lag stops. In Spanish, as in the analysis above, lead VOT [+voice] stops are compared to short lag [−voice] stops. Collapsing lead and short lag tokens in English under the same category is justified by the results of the previous analysis, which showed that these two VOT types were not significantly different from each other in terms of onset *f*0.

Fig. 4 shows that both Spanish and English [+voice] stops have a lower than average onset *f*0, while both Spanish and English [−voice] stops have a higher than average onset *f*0. In both languages, [+voice] and [−voice] stops appear to be well separated in terms of onset *f*0. It can also be observed that English [+voice] and [−voice] onset *f*0 means are more different from each other than Spanish means.

A mixed-design Repeated Measures ANOVA was applied to the semitone-normalized onset *f*0 data from Spanish and American participants with Language (between-subject) and Phonological Category (within-subject) as independent factors. All participants in both groups were included in this analysis.

The results showed a significant effect of Phonological Category on onset *f*0, $F(1,52)=146.090$, $p<0.001$; partial $\eta^2=0.737$. Across the two languages, [−voice] stops exhibited a significantly *higher* onset *f*0 than [+voice] stops. There was no significant effect of Language ($p=0.204$).

There was also a significant Language by Phonological Category interaction, $F(1,52)=8.234$, $p<0.01$; partial $\eta^2=0.137$. The difference between [−voice] and [+voice] stops in terms of onset *f*0 was significantly greater in English than in Spanish.[12] Means and standard deviations are shown in Table 3.

In order to examine in more detail the effect of phonological voicing on onset *f*0 within each language, two additional RM ANOVAs were performed separately for each language group. These tests showed that the effect of Phonological Category on onset *f*0 was significant within each language: For Spanish, $F(1, 23)=52.619$, $p<0.001$; partial $\eta^2=0.696$; for English, $F(1, 29)=103.402$, $p<0.001$; partial $\eta^2=0.781$.

[11] The within VOT category *t*-tests produced statistically equivalent results whether all English-speaking participants were included in the analysis, or whether the analysis included only the sub-group of 22 participants who produced both short lag and lead VOT tokens.

[12] The main results of these analyses remain the same if only short lag tokens were included in English [+voice] category, discarding all lead VOT productions (including all from one participant who produced all /b/s with lead voicing, and one who produced all but one /b/ with lead voicing).

Table 3
Means and standard deviations in semitones for onset *f*0 in Spanish and English [+voice] and [−voice] stops.

	[+voice]	[−voice]
Spanish	−0.68 (sd 0.47)	0.57 (sd 0.52)
English	−1.14 (sd 0.61)	0.89 (sd 0.49)

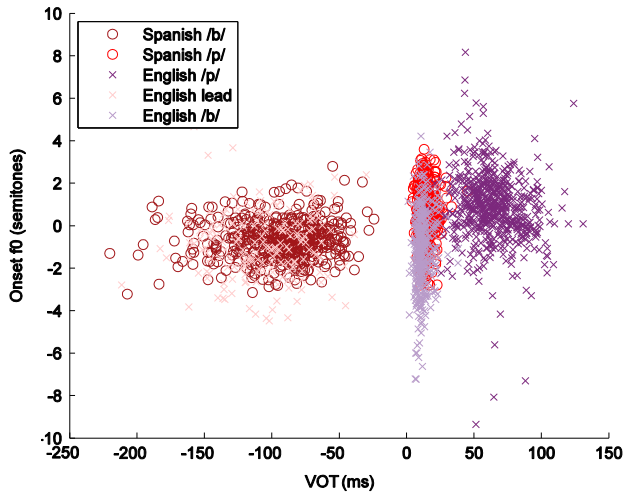

Fig. 5. Scatter plot of the VOT and corresponding semitone-normalized onset *f*0 for Spanish and English stops.

To test for language-specific effects on onset *f*0, a separate independent-samples *t*-test was performed on onset *f*0 values within each phonological voicing category with Language as an independent factor. The results showed a significant onset *f*0 difference between English and Spanish for both [+voice] and [−voice] stops. English [+voice] tokens were significantly *lower* in onset *f*0 than Spanish [+voice] tokens: $t(52)=3.003$, $p<0.01$ (this result was also upheld when only short lag stops were included in the English [+voice] category). English [−voice] token, on the other hand, were significantly *higher* in onset *f*0 than Spanish [−voice] tokens: $t(52)=2.345$, $p<0.05$.

3.6. Within-category VOT and onset *f*0 correlation

In the following analyses, onset *f*0 values are examined for correlation with VOT duration within each voicing category in each language (four correlation analyses). Fig. 5 shows the normalized onset *f*0 values for all Spanish and English participants plotted against corresponding VOT duration values.

A correlation between VOT and semitone-normalized onset *f*0 within each phonological category in each language group was examined via robust estimation of biweight midcorrelation coefficients, robust *r* (Wilcox, 2005). Within the English [+voice] category (including lead VOT and short lag stops), onset *f*0 was weakly positively correlated with VOT: robust $r=0.11$, $p<0.01$.[13] Within English [−voice] category (long lag VOT stops), there was a weak negative correlation between onset *f*0 and VOT: robust $r=-0.18$, $p<0.001$. A scatterplot of English VOT and onset *f*0 data including robust regression lines fitted to points within each phonological category is shown in Fig. 6. Within both Spanish voicing categories, onset *f*0 was uncorrelated with VOT (see Fig. 7).

3.7. The extent of *f*0 perturbation into the vowel

To examine the extent of *f*0 perturbation into the vowel beyond the onset *f*0 measurement point, ten *f*0 measurements taken every 10 ms after the onset *f*0 measurement point were submitted to a series of mixed-design Repeated Measures ANOVAs, with the variables of Language, Measurement Step, and Voicing as independent factors. Fig. 8 shows averaged normalized *f*0 contours (in semitones) for each language in each voicing condition. In both languages, the difference between voiced and voiceless contours becomes progressively smaller as the measurement point moves further into the vowel (Step increases). At the same time, the difference between Spanish and English contours becomes more and more pronounced. English contours in particular dip much lower than the speakers' average onset *f*0 (0 on the *y*-axis) towards the end of the vowel. This tendency is likely due, at least in part, to the pronounced presence of creaky voice in English productions, which often appeared towards the end of the vowel (recall that many vowels were close to or even shorter than 100 ms), significantly lowering English speakers' *f*0. Both the monosyllabic structure

[13] This result may appear to contradict the observation that mean onset *f*0 for English short lag stops is lower than for lead VOT stops (though not significantly so, see Table 2). This seeming contradiction is an artifact of using individual tokens in the correlation analysis while participants' means were used to calculate VOT type means in the RM ANOVA. Given the weak nature of the correlation and the non-significance of the mean differences, this discrepancy should be interpreted with caution if at all.

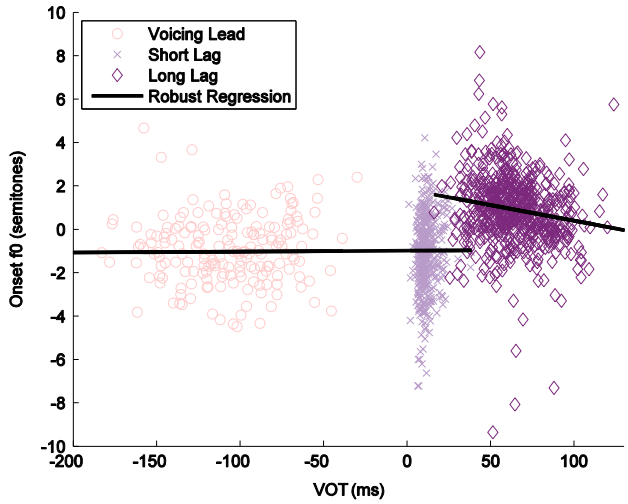

Fig. 6. Scatterplot of VOT and corresponding semitone normalized onset *f*0 for English [+voice] (lead and short lag stops) and [−voice] (long lag stops) categories, with robust regression lines fitted within each voicing category.

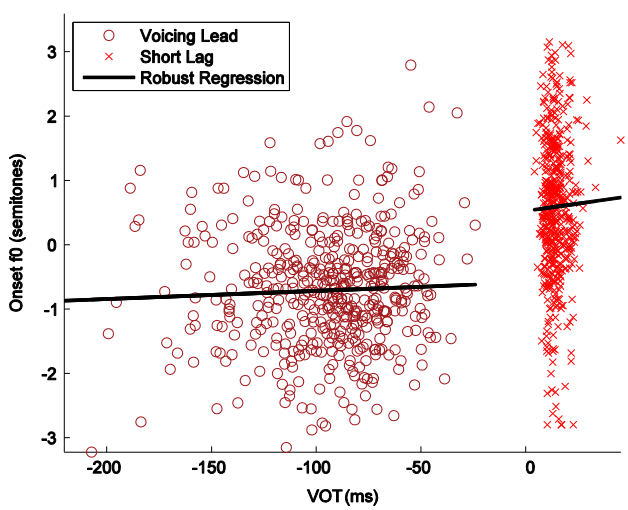

Fig. 7. Scatterplot of VOT and corresponding semitone normalized onset *f*0 for Spanish [+voice] (lead stops) and [−voice] (short lag stops) categories, with robust regression lines fitted within each voicing category.

of English stimuli and final [t], often pronounced with a simultaneous glottal constriction by English speakers, may have contributed to the creaky quality of the vowels.

The results of the omnibus mixed-design Repeated Measures ANOVA are presented in Table 4.

Of particular significance in this analysis are the interactions. The Voicing by Language interaction signifies that the effect of Voicing on *f*0 was not consistent across the two languages. Fig. 8 shows that the separation between the voiced and voiceless contours is more pronounced in English than in Spanish. The Step by Language interaction shows that the rate with which *f*0 changed across the measurement steps is not the same in Spanish and English. Fig. 8 demonstrates that *f*0 contours are considerably steeper in English than in Spanish data, especially after [−voice] stops. Finally, the Voicing by Step interaction indicates that the effect of Voicing on *f*0 was not constant across the measurement steps.

To further investigate the effect of voicing at different time-points within the vowel, separate Repeated Measures ANOVAs were conducted at each measurement step in each language. For English, this analysis established that the effect of Voicing was significant at each measurement point up to and including step 7. For the English group, because the initial onset *f*0 measurement point (step 0) was made, on average, 5 ms into the vowel, step 7 is located approximately 75 ms into the vowel.

For Spanish, it was found that the effect of Voicing on *f*0 was significant up to and including step 5 (approximately 53 ms into the vowel because Spanish onset *f*0 was measured on average 3 ms into the vowel). At steps 6, 7, and 8 (63 ms, 73 ms, and 83 ms) the effect of Voicing was not significant in Spanish. However, at steps 9 and 10 (93 ms and 103 ms) the effect of Voicing was significant again, albeit in the opposite direction, the pitch after voiced stops surpassing the pitch after voiceless stops, as shown by the crossover of the Spanish pitch contours in Fig. 8.

In order to address the issue of the apparently stronger effect of Voicing on *f*0 in English than in Spanish, independent samples *t*-test analyses of individual *f*0 ranges were conducted at the vowel step where in both English and Spanish Voicing ceased to have a

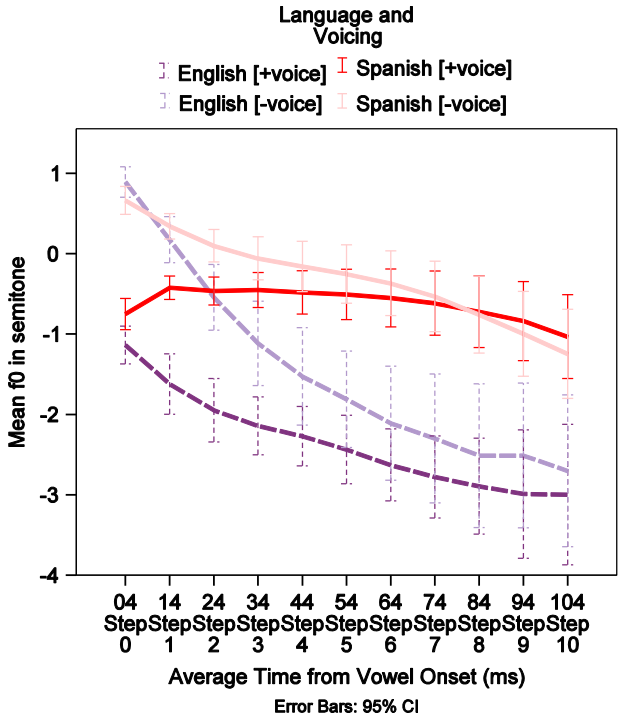

Fig. 8. Averaged and semitone-normalized *f*0 contours after voiced and voiceless stops in English and Spanish across the eleven measurement steps (with step 0 being the onset *f*0 measurement at approximately 4 ms after the vowel onset, and steps 1–10 in 10 ms intervals beyond that). Dashed lines corresponds to English data, solid lines correspond to Spanish data. Darker lines are for *f*0 contours after voiced stops. Note that the semitone scale is referenced to individual talkers' average onset *f*0 (i.e. "0"=average onset *f*0).

Table 4
Main effects and interactions in the omnibus mixed-design ANOVA with *f*0 values at 11 measurements steps as a dependent variable and Language (between-subject), Voicing, and Measurement Step (within-subject) as independent variables.

Factors	df	df error	*F*	*p*	Partial η^2
Language	1	47[a]	23.632	<0.001	0.335
Voicing	1	47	35.392	$p<0.001$	0.43
Measurement Step	1.59	74.717	36.044	$p<0.001$	0.434
Voicing by Language	1	47	7.66	<0.01	0.14
Step by Language	10	470	8.962	<0.001	0.16
Voicing by Step	2.795	131.37	24.916	<0.001	0.346

[a] This analysis was based on 20 Spanish and 29 English participants for whom *f*0 values were available at each measurement step. Five participants (4 Spanish and 1 English) were excluded because they did not provide a value at one or more of the last two or three measurement steps.

significant effect on *f*0 (step 8, around 83–85 ms into the vowel) and the distance between voiced and voiceless *f*0 contours was minimal. If English participants have a greater *f*0 range at this measurement step, it cannot be attributed to the enhancement of voicing-related differences in *f*0. Then, the greater separation between voiced and voiceless *f*0 contours in English may be at least partially explainable by crosslinguistic differences in f0 range.

The results showed a significant difference in *f*0 range across the two languages: $t(52)=5.939$, $p<0.001$. English participants showed a significantly higher *f*0 range (mean range: 10 semitones, sd 5.8) than Spanish participants (mean range: 3 semitones, sd 1).

4. Discussion

4.1. Lead and short lag stops

The first question to be addressed is whether onset *f*0 varied as a function of the VOT type in the two languages, independently of phonological status. The fact that two types of stops (lead and short lag) have different phonological status in Spanish and in English makes it possible to distinguish purely phonetic effects (of VOT) on onset *f*0 from phonological influences. A purely phonetic perspective would predict that a higher onset *f*0 should be associated with phonetically voiceless short lag stops in both languages, while a more phonological perspective would hold that this correspondence should be found in Spanish but may be absent in English where lead and short lag VOTs are not phonologically contrastive word-initially.

The results of the analysis of the VOT types that are present in both languages (lead voicing VOT and short lag VOT) and their patterning with onset *f*0 showed that lead stops and short lag stops were differentiated in terms of onset *f*0 only in Spanish and not in English. Crucially, only in Spanish do these two VOT types correspond to opposing phonological categories. In English, they are sub-phonemic variants of the same phoneme. While it has been shown numerous times that onset *f*0 in English varies predictably with VOT when VOT is a predictor of voicing status, the fact that, in these cases, VOT is itself governed by the *phonological* voicing status of the stop consonant means that phonetic and phonological factors are confounded. In the present study, the lack of any covariation between onset *f*0 and VOT differences within the [+voice] category (i.e. across lead voicing and short lag tokens) demonstrates that there is no predictable change in onset *f*0 as a result of non-phonologically governed phonetic variation in VOT.

Turning to the language-specific differences in the relationship between VOT and onset *f*0, it was observed that lead voicing stops in English did not differ in terms of onset *f*0 from lead voicing stops in Spanish. The two lead voicing distributions occupied approximately the same portion of the VOT continuum in both languages (between −25 and −220 ms VOT) and the two sets of onset *f*0 values overlapped considerably and were not significantly different.

In contrast, the behavior of the short lag tokens is dramatically dissimilar in the two languages. Spanish and English short lag stops are indistinguishable in terms of VOT duration, but are set apart quite impressively with respect to their onset *f*0 values. Spanish short lag stops are significantly higher in onset *f*0 than English short lag stops, as shown in Fig. 3. Thus, the onset *f*0 of initial voiceless unaspirated (short lag) stops across these languages appears to depend primarily on their phonological specification as [+voice] or [−voice]: In English, initial short lag stops are [+voice] and are associated with an onset *f*0 *lower* than in Spanish, in which short lag stops are [−voice] (see also Caisse, 1982). This result suggests that the phonological status of the consonant may carry more weight in determining the onset *f*0 patterns than do its phonetic properties, such as the presence or absence of laryngeal voicing (Keating, 1984; Kingston & Diehl, 1994; Kingston, 2007).

The crosslinguistic comparisons of onset *f*0 must be approached with some caution since differences in macro-prosody between languages may also be contributing to the observed *f*0 patterns. Efforts were made in this study to minimize language-specific prosodic effects on the recorded stimuli. All material was collected using the same procedures for Spanish and English. Words produced in isolation, with the pace controlled by one-by-one on-screen presentation, resulted in a uniform and similar falling intonation on each word across languages. While English stimuli were monosyllables and Spanish ones were disyllables, only initial, stressed syllables were analyzed in both cases. Certain prosodic differences are naturally expected in the realization of the H* L% declarative intonation in mono- vs. disyllables. For example, some data suggest that in English monosyllables the peak of the pitch accent is reached earlier than in disyllables (Xu & Xu, 2005). The necessity to reach the peak of *H tone earlier may have raised the overall onset *f*0 in the English monosyllables in comparison to the Spanish disyllables. However such a raising effect would only mitigate against the observed low onset *f*0 of English short lag stops, potentially reducing the observed crosslinguistic effect rather than contributing to it.

Finally, polysyllabic structure tends to have a 'compressing' effect on durational properties of syllables (Ladefoged & Johnson, 2011, p. 101). Thus, all else being equal, the English syllables, and perhaps their corresponding VOT values, may have been shorter if disyllables had been used. However, Umeda (1977) showed that consonant durations are less subject to word length effects than are vowels, suggesting that using disyllables instead of monosyllables might not have made much difference at all (see also Turk & Shattuck-Hufnagel, 2000). Moreover, as was shown in this study, sub-phonemic variation in VOT duration does not have a very pronounced effect on onset *f*0, making the possibility of cross-language differences appearing due to word length effects immaterial for the current *f*0 results.

The observation that the onset *f*0 of short lag consonants is so different across the two languages examined here suggests that the onset *f*0 property may be relatively malleable in the positive VOT range, particularly in the short lag range where it varies considerably depending on the type of contrast it is involved in (i.e. voice vs. aspiration contrast). This is also supported by the fact that distinct patterns of *f*0 perturbation, with aspirated stops either raising or lowering *f*0 compared to short lag stops, have been reported for contrasts located entirely within the positive VOT range (Francis et al., 2006; Xu & Xu, 2003; Kagaya & Hirose, 1974; see also reviews in Kingston & Diehl, 1994 and in Chen, 2011).[14]

The finding that, despite their differences in terms of VOT, English and Spanish [+voice] stops are similar in terms of onset *f*0 values may have implications for second language acquisition research. For example, Lotz, Abramson, Gerstman, Ingemann, and Nemser (1960) showed that speakers of Puerto-Rican Spanish tended to correctly identify naturally recorded English initial voiced stops as [+voice] (despite the fact that that English [+voice] stops are typically realized with short lag VOT, more similar to Spanish [−voice] stops). This pattern is consistent with the possibility that, when making voicing decisions in a second language, Spanish listeners may be giving greater weight to secondary cues, such as onset *f*0, in addition to the primary cues such as VOT (Llanos et al., 2013). As shown by the present results, English initial [+voice] (short lag) stops are quite different from Spanish [−voice] short lag stops in terms of onset *f*0 (and, therefore, possibly in terms of other secondary cues) and, in this respect are, in fact, more similar to Spanish [+voice] stops.

Thus, secondary cues, including onset *f*0, may be a guiding factor in allowing Spanish speakers to correctly identify English initial short lag stops as [+voice] despite their VOT values lying strongly within the range of Spanish [−voice] stops. In support of this hypothesis, a recent perceptual study by Llanos et al. (2013) showed that in the short lag VOT region Spanish listeners judged synthetic stops as predominantly [+voice] if onset *f*0 was low, even when no laryngeal voicing, obligatory in the production of native Spanish [+voice] tokens, was present.

[14] Note, however, that the effects reported by these studies are rather small and some of them may be subject to strong effects of inter-speaker variability (i.e. the data presented by Kagaya and Hirose (1975) is from a single speaker). Moreover, several of these studies concern tonal languages, which may also have significant consequences for onset *f*0 patterns.

4.2. Phonological voicing and onset *f*0

Both English and Spanish speakers made a clear distinction between their respective phonological voicing categories in terms of VOT and onset *f*0, with both languages demonstrating a significantly *higher* onset *f*0 for the [−voice] category than for [+voice] category despite the fact that the phonetic expression, in terms of VOT, of the corresponding phonological categories was quite different in the two languages.

A similar finding was reported by Hombert (1976) (also discussed by Hombert et al., 1979), who examined onset *f*0 patterns of English and French initial post-vocalic stops. Hombert et al. (1979) also observed that pitch perturbations caused by French and English voiceless stops were of the same magnitude. The present study, however, found a *greater* mean onset *f*0 difference between English voicing categories than between Spanish voicing categories. Thus, it appears that English speakers may further enhance the onset *f*0 difference between English voicing categories to a greater degree than do Spanish speakers. Furthermore, *f*0 measurements beyond vowel onset showed that English speakers maintained a voicing-based *f*0 difference farther into the vowel than Spanish speakers (85 ms vs. 53 ms). This result is also consistent with the hypothesis that English speakers enhance the *f*0 difference between initial voicing categories to a greater degree than Spanish speakers.

Alternatively, English speakers may simply have a greater *f*0 range for some unrelated reason such that they naturally produce particularly low *f*0 values in *f*0-lowering contexts, and/or particularly high ones in *f*0-raising contexts, independently of any enhancement intentions. To test this hypothesis, we compared *f*0 ranges across the two languages at approximately 83–85 ms into the vowel, where voicing effects on *f*0 disappear in both languages. Presumably, in this position any hypothetical effect of contrast-enhancement strategies is neutralized because the voicing-related *f*0 difference is no longer there to enhance. The results showed that English speakers did maintain a greater *f*0 range even in the absence of voicing-related *f*0 differences. A greater *f*0 range for English speakers may be attributable to the frequent presence of creaky voice, which may have lowered English speakers' *f*0 considerably with respect to their average *f*0 levels. This suggests that the difference between Spanish and English speakers in terms of the magnitude of the voicing-related effects on onset *f*0 could be due to cross-language differences in *f*0 range, and need not necessarily reflect a greater degree of enhancement of the onset *f*0 contribution to the voicing contrast in English as compared to Spanish.

Another noteworthy feature of English *f*0 measurements is that both ‘voiced’ and ‘voiceless’ *f*0 contours are consistently falling. Spanish, on the other hand, demonstrates a contrast between a rising contour for the ‘voiced’ category and a falling one for the ‘voiceless’ one. There is a lack of consensus concerning the expected shape of the *f*0 contour after English [+voice] stops that can be traced through numerous studies. For example, Umeda (1981) and Ohde (1984) report a falling trajectory for both voiced and voiceless contours, in agreement with the current results. In contrast, Lehiste and Peterson (1961) and Lea (1973), report a rising contour after voiced stops and a falling one after voiceless stops. It is possible that contour may be irrelevant: Silverman (1986) observed that the direction of the *f*0 trajectory after voiced vs. voiceless stops may change depending on the intonational context and concluded that the level but not the direction of *f*0 changes should covary consistently with voicing. As discussed by Ohde (1984) such variability in contours observed across experiments may also be related to a greater difficulty in obtaining accurate onset *f*0 measurements after English [+voice] stops. In the present study, we also observed that reliable onset *f*0 measurements in English could only be obtained significantly later after voiced stops (on average, 8 ms into the vowel) than after voiceless stops (on average, 2 ms into the vowel). If, however, the falling *f*0 contour observed for voiced stops in English is not an artifact of less reliable measurements, then it may be concluded that English *f*0 contours resemble greatly the consistently falling *f*0 contours that occur after both voiced and voiceless stops in aspirating languages, such as Cantonese (Francis et al., 2006). This resemblance suggests that, despite the prevalence of lead voicing among some English speakers, the English initial voicing contrasts may indeed belong in the ‘aspiration’ category and not among the true ‘voice’ contrasts, such as in Spanish.

Finally, within each phonological category in English and Spanish we saw little evidence for a consistent correlation between VOT and onset *f*0 values. We hypothesized that if trading relations exist between VOT and onset *f*0 in production, ambiguous VOT values may be compensated for by more prototypical onset *f*0 values, thus predicting a negative correlation between VOT and onset *f*0 within each phonological categories. Results showed that only within English [−voice] category (long lag stops) was there even a weak negative correlation between VOT and onset *f*0. The correlation was even weaker and in the positive direction within English [+voice] category (short lags and lead stops). No significant correlation was detected for either of Spanish categories. These results suggest that although onset *f*0 in both languages is a reliable correlate for the categorical difference between [+voice] and [−voice] stops, it does not differentiate less vs. more prototypical exemplars within each category.

4.3. Effect of orthography

An unexpected effect revealed a significant difference in the phonetics of Spanish initial voiced stops apparently connected to spelling differences. Although the pronunciations of initial “b” and “v” are typically assumed to be phonetically equivalent in modern Spanish, in the present experiment initial [+voice] stops spelled as “v” showed a significantly shorter lead VOT than did initial stops spelled as “b”. Among possible explanations for this effect is spelling pronunciation or ‘hypercorrection’. For example, an effect of orthography has been suggested to play a role in the phenomenon known as ‘incomplete neutralization’ – subtle but consistent phonetic traces of underlying representations, usually preserved in language's orthography, in the pronunciation of ‘neutralized’ phonemes (Fourakis & Iverson, 1984; Port & O'Dell, 1985; Jassem & Richter, 1989; Warner, Jongman, Sereno,& Kemps, 2004; Warner, Good, Jongman, & Sereno, 2006; Kharlamov, 2012). This difference may also be related to the efforts to promote a

historically-accurate fricative pronunciation of orthographic "v" by Spanish Real Academy through the beginning of 20th century (Martinez, 1986).

In light of this phonetic difference between the two orthographic variants, it is possible that a more detailed analysis would reveal other points of divergence between the two orthographic variants. An interesting further question to pursue is how pervasive this orthographic effect is in Spanish phonology and how much it depends on whether the elicitation task involved reading (cf. Damian & Bowers, 2003; Roelofs, 2006; Warner et al., 2006 and references therein). Ultimately, although they provide an interesting side note, the differences observed here are relatively small, and did not materially affect the central questions currently under investigation.

4.4. Implications for theories of speech perception

The present results may also have implications for theories of cue perception and integration. In particular, the findings presented here provide some support against experience-based explanations of cue integration between onset $f0$ and VOT in perceptual voicing categorization. Llanos et al. (2013) demonstrated that, within the native Spanish VOT range, onset $f0$ played a very modest perceptual role, affecting voicing decisions only in the positive VOT range. Moreover, the most ambiguous tokens – those with 0 ms VOT (the cross-over point in the VOT-based voicing judgments by Spanish speakers), which are predicted to be most strongly dependent on secondary cues to voicing, were not affected by onset $f0$ in voicing identification. This perceptual behavior could be explained by a lack of perceptual experience if the dependency between VOT and onset $f0$ was absent or very weak in Spanish. However, the current study showed a significant onset $f0$ difference between the two voicing categories in Spanish. Thus, as argued by Llanos et al. (2013), the observation that Spanish listeners did not rely on onset $f0$ to distinguish between voicing lead and short lag stops cannot be explained by a lack of experience with a covariation between onset $f0$ and VOT. The covariation is present in Spanish production, and yet Spanish listeners still do not seem to exploit it in perception. Building on the work of Kingston and coworkers (Kingston et al., 2008; Kingston & Diehl, 1994, 1995), Llanos et al. (2013) proposed instead that onset $f0$ is not used as a cue to voicing distinction in the lead-short lag range because prevoicing in lead stops constitutes a sufficiently salient cue and need not be reinforced by onset $f0$ differences. In the positive VOT range, prevoicing is absent, thus low frequency energy supplied by low onset $f0$ in short lag stops renders such stops more perceptually similar to truly voiced (=prevoiced) stops and more perceptually distinct from voiceless aspirated stops. The fact that onset $f0$ is used by listeners as a cue to voicing predominantly in the positive VOT range may also explain why, in the present study, trading relations between less prototypical VOT and more prototypical onset $f0$ were detected only in long lag [−voice] stops in English. If this is the range where onset $f0$ affects voicing judgments, then it is also the most plausible range in which to use onset $f0$ as an enhancing property as VOT values become less prototypically [−voiced].

5. Conclusions

The results of the present study showed that, in both Spanish and English, stops belonging to different phonological voicing categories were well-differentiated via the onset $f0$ parameter, with onset $f0$ being significantly higher for [−voice] stops than for [+voice] stops across both languages. However, the results also suggest that the connection between voicing and onset $f0$ is mediated by phonological as well as phonetic factors. As evidence for this claim, it was observed that a distinction between phonetically voiced (lead VOT) and voiceless (short lag VOT) stops did not necessarily result in an onset $f0$ difference, except in those cases in which a phonological boundary was involved: English short lag stops were not higher in onset $f0$ than English lead voicing stops, but Spanish short lag stops *were* higher in onset $f0$ than Spanish lead voicing stops. Thus, across languages, equivalent VOT types (short lag and lead voicing VOT) were differentiated via onset $f0$ *only* if they had a contrastive phonological status (in Spanish) but not if they were members of the same phonological category (in English).

While, there is, in all likelihood, a physiological basis for the VOT–onset $f0$ dependency (Hoole & Honda, 2011; Löfqvist et al., 1989), the present results suggest that onset $f0$ patterns can be shaped beyond this influence to serve the goals of the phonological system, in particular by making opposing phonological categories more perceptually distinct. The uncharacteristically low onset $f0$ of English initial short lag stops makes them more similar to lead stops and at the same time more acoustically distinct from the phonologically opposing long lag stops.

These results suggest that the cross-linguistic covariation observed between VOT and onset $f0$ is consistent with the manipulation of two cues that share a common articulatory basis but, more importantly, serve together to increase phonological distinctiveness, perhaps via a mechanism of auditory enhancement (Kingston & Diehl, 1995; Llanos et al., 2013). Although, these findings do not rule out the possibility that other patterns of covariation between primary and secondary acoustic cues may arise for other reasons, they do suggest that further research is necessary on a case-by-case basis until perhaps a larger pattern may emerge.

Acknowledgments

We are grateful to Prof. Juana Gil-Fernández for the use of her laboratory facilities at CSIC (Spain). We also thank Samantha Berger and Audrey Bengert for their assistance with data collection and Christie Wai Ling Law for assistance with reliability measurements. We also acknowledge John Kingston and two anonymous reviewers for helpful suggestions on a previous version of this article.

References

Abramson, A. S., & Lisker, L. (1965). Voice onset time in stop consonants: Acoustic analysis and synthesis. In *Proceedings of the 5th international congress of acoustics* (Vol. 51). A51, Liege.

Abramson, A. S., & Lisker, L. (1985). Relative power of cues: *F*0 shift versus voice timing. In V. Fromkin (Ed.), *Phonetic linguistics: Essays in honor of Peter Ladefoged* (pp. 25–33). New York: Academic.

Almela, R., Cantos, P., Sánchez, A., Sarmiento, R., & Almela, M. (2005). *Frecuencias del español. Diccionario y estudios léxicos y morfológicos*. Madrid: Universitas.

Boersma, P., & Weenink, D. (2009). *Praat: Doing phonetics by computer* (*Version 5.2*) [*Computer program*]. Amsterdam, The Netherlands: University of Amsterdam. Available online: ⟨http://www.praat.org⟩.

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

Castleman, W. A., & Diehl, R. L. (1996). Effects of fundamental frequency on medial and final [voice] judgments. *Journal of Phonetics*, *24*, 383–398.

Chen, Y. (2011). How does phonology guide phonetics in segment–*f*0 interaction?. *Journal of Phonetics*, *39*(4), 612–625.

Cho, T., Jun, S.-A., & Ladefoged, P. (2002). Acoustic and aerodynamic correlates of Korean stops and fricatives. *Journal of Phonetics*, *30*, 193–228.

Cho, T., & Ladefoged, P. (1999). Variation and universals in VOT: Evidence from 18 languages. *Journal of Phonetics*, *27*(2), 207–229.

Damian, M. F., & Bowers, J. S. (2003). Effects of orthography on speech production in a form-preparation paradigm. *Journal of Memory & Language*, *49*, 119–132.

Docherty, G. J. (1992). *The timing of voicing in British English obstruents* (pp. 29–32)Berlin: Walter de Gruyter29–32.

Fourakis, M., & Iverson, G. K. (1984). On the 'incomplete neutralization' of German final obstruents. *Phonetica*, *41*, 140–149.

Francis, A. L., Ciocca, V., & Yu, J. M. C. (2003). Accuracy and variability of acoustic measures of voicing onset. *Journal of the Acoustical Society of America,*, *113*(2), 1025–1032.

Francis, A. L., Ciocca, V., Wong, V. K. M., & Chan, J. K. L. (2006). Is fundamental frequency a cue to aspiration in initial stops?. *The Journal of the Acoustical Society of America*, *120*(5), 2884–2896.

Gandour, J. (1974). Consonant types and tone in Siamese. *Journal of Phonetics,*, *2*, 337–350.

Goldrick, M., & Rapp, B. (2007). Lexical and post-lexical phonological representations in spoken production. *Cognition*, *102*, 219–260.

Haggard, M., Ambler, S., & Callow, M. (1970). Pitch as a voicing cue. *The Journal of the Acoustical Society of America*, *47*, 613–617.

Holt, L. L., Lotto, A. J., & Kluender, K. R. (2001). Influence of fundamental frequency on stop-consonant voicing perception: A case of learned covariation or auditory enhancement?. *The Journal of the Acoustical Society of America*, *109*, 764–774.

Hombert, J. -M. (1976). The effect of aspiration on the fundamental frequency of the following vowel. In *Proceedings of the 2nd annual meeting of the BLS* (pp. 212–219).

Hombert, J.-M. (1977). Consonant types, vowel height, and tone in Yoruba. *Studies in African Linguistics,*, *8*(2), 173–190.

Hombert, J.-M., Ohala, J. J., & Ewan, W. G. (1979). Phonetic explanations for the development of tones. *Language*, *55*, 37–58.

Hoole, P., & Honda, K. (2011). Automaticity vs. feature-enhancement in the control of segmental *F*0. *Where do phonological features come from*, 131–174.

House, A. S., & Fairbanks, G. (1953). The influence of consonant environment upon the secondary acoustical characteristics of vowels. *The Journal of the Acoustical Society of America*, *25*, 105–113.

Jassem, W., & Richter, L. (1989). Neutralization of voicing in Polish obstruents. *Journal of Phonetics*, *17*, 317–325.

Jeel, V. (1975). *An investigation of the fundamental frequency of vowels after various Danish consonants, in particular stop consonants. Technical report No. 9*. Copenhagen: Institute of Phonetics, University of Copenhagen.

Kagaya, R., & Hirose, H. (1974). Fiberoptic, electromyographic and acoustic analyses of Hindi stop consonants. *Annual Bulletin of the Research Institute of Logopedics and Phoniatrics, University of Tokyo* no. 9 (pp. 27–46).

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

Keating, P., Linker, W., & Huffman, M. (1983). Patterns in allophone distribution for voiced and voiceless stops. *Journal of Phonetics*, *11*(3), 277–290.

Kharlamov, V. (2012). *Incomplete neutralization and task effects in experimentally-elicited speech: Evidence from the production and perception of word-final devoicing in Russian* (Ph.D. thesis). University of Ottawa.

Kingston, J. (2007). Segmental influences on F_0: Automatic or controlled. In C. Gussenhoven, & T. Riad (Eds.), *Tones and tunes*, Vol. 2 (pp. 171–210). Berlin: Mouton de Gruyter.

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

Kingston, J., & Diehl, R. (1995). Intermediate properties in the perception of distinctive feature values. In B. Connell, & A. Arvanti (Eds.), *Phonology and phonetic evidence: Papers in laboratory phonology IV* (pp. 7–27). Cambridge: Cambridge University Press.

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

Ladefoged, P., & Johnson, K. (2011). *A course in phonetics*, 6th edition.

Lai, Y., Huff, C., Sereno, J., & Jongman, A. (2009). The raising effect of aspirated prevocalic consonants on *F*0 in Taiwanese. In J. Brooke, G. Coppola, E. Görgülü, M. Mameni, E. Mileva, S. Morton, et al. (Eds.), *Proceedings of the 2nd international conference on East Asian Linguistics*, Simon Fraser University working papers in linguistics. Online document downloaded from ⟨http://www2.ku.edu/~kuppl/documents/Lai_EtAl.pdf⟩ (last checked 14.03.13).

Lea, W. A. (1973). Segmental and suprasegmental influences on fundamental frequency contours. *Consonant types and tones* (pp. 15–70).

Lehiste, I., & Peterson, G. E. (1961). Some basic considerations in the analysis of intonation. *The Journal of the Acoustical Society of America*, *33*, 419–425.

Lisker, L. (1975). Is it VOT or a first formant transition detector?. *Journal of the Acoustical Society of America*, *57*, 1547–1551.

Lisker, L. (1978). *In qualified defense of VOT*. *Language and Speech*, 21375–383.

Llanos, F., Dmitrieva, O, Shultz, A., & Francis, A. (2013). Auditory enhancement and second language experience in Spanish and English weighting of secondary voicing cues. *The Journal of the Acoustical Society of America*, *134*(3), 2213–2224.

Löfqvist, A., Baer, T., McGarr, N. S., & Story, R. S. (1989). The cricothyroid muscle in voicing control. *The Journal of the Acoustical Society of America*, *85*, 1314–1321.

Lotz, J., Abramson, A. S., Gerstman, L. J., Ingemann, F., & Nemser, W. J. (1960). The perception of English stops by speakers of English, Spanish, Hungarian and Thai. *Language and Speech*, *3*(2), 71–77.

Maddieson, I. (1984). *Patterns of sounds*. Cambridge: Cambridge University Press.

Martinez, C. F. (1986). Razones fonéticas del llamado betacismo. *Faventia*, *812*, 21–25.

MATLAB and Statistics Toolbox Release(2001). The MathWorks, Inc., Natick, MA, USA.

Oglesbee, E. (2008). *Multidimensional stop categorization in English, Spanish, Korean, Japanese, and Canadian French* (Ph.D. dissertation). Bloomington: Indiana University.

Ohde, R. (1984). Fundamental frequency as an acoustic correlate of stop consonant voicing. *The Journal of the Acoustical Society of America*, *75*, 224–240.

Port, R. F., & O'Dell, M. L. (1985). Neutralization of syllable-final voicing in German. *Journal of Phonetics*, *13*, 455–471.

Raphael, L. J. (2005). Acoustic cues to the perception of segmental phonemes. In D. B. Pisoni, & R. E. Remez (Eds.), *The handbook of speech perception* (pp. 182–206). Malden, MA: Blackwell.

Reinholt Petersen, N. (1983). *The effect of consonant type on fundamental frequency and larynx height in Danish. Technical report*. Copenhagen: Institute of Phonetics, University of Copenhagen.

Repp, B. H. (1982). Phonetic trading relations and context effects: New experimental evidence for a speech mode of perception. *Psychological Bulletin*, *92*(1), 81.

Roelofs, A. (2006). The influence of spelling on phonological encoding in word reading, object naming, and word generation. *Psychonomic Bulletin & Review*, *13*(1), 33–37.

Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-Prime user's guide*. Pittsburg, PA: Psychology Software Tools Inc.

Shultz, A. A. (2011). *Individual differences in cue weighting of stop consonant voicing in perception and production* (Master's thesis). West Lafayette, IN: Purdue University.

Shultz, A. A., Francis, A. L., & Llanos, F. (2012). Differential cue weighting in perception and production of consonant voicing. *The Journal of the Acoustical Society of America*, *132*(2), EL95–EL101.

Silverman, K. (1986). *F*0 segmental cues depend on intonation: The case of the rise after voiced stops. *Phonetica*, *43*(1-3), 76–91.

Stilp, C. E., Rogers, T. T., & Kluender, K. R. (2010). Rapid efficient coding of correlated complex acoustic properties. *Proceedings of the National Academy of Science*, *107*(50), 21914–21919.

Turk, A. E., & Shattuck-Hufnagel, S. (2000). Word-boundary-related duration patterns in English. *Journal of Phonetics*, *28*(4), 397–440.

Umeda, N. (1977). Consonant duration in American English. *Journal of the Acoustical Society of America*, *61*(3), 846–858.

Umeda, N. (1981). Influence of segmental factors on fundamental frequency in fluent speech. *The Journal of the Acoustical Society of America*, *70*(2), 350–355.

Warner, N., Good, E., Jongman, A., & Sereno, J. (2006). Orthographic versus morphological incomplete neutralization effects. *Journal of Phonetics*, *34*, 285–293.

Warner, N., Jongman, A., Sereno, J., & Kemps, R. (2004). Incomplete neutralization and other sub-phonemic durational differences in production and perception: Evidence from Dutch. *Journal of Phonetics*, *32*, 251–276.

Washington University in St. Louis Speech & Hearing Lab Neighborhood Database. Available from ⟨http://128.252.27.56/Neighborhood/SearchHome.asp⟩ (last accessed 02.08.13).

Whalen, D. H., Abramson, A. S., Lisker, L., & Mody, M. (1990). Gradient effects of fundamental frequency on stop consonant voicing judgments. *Phonetica*, *47*(1–2), 36–49.

Whalen, D. H., Abramson, A. S., Lisker, L., & Mody, M. (1993). *F*0 gives voicing information even with unambiguous voice onset times. *The Journal of the Acoustical Society of America*, *93*, 2152–2159.
Wilcox, R. R. (2005). *Introduction to robust estimation and hypothesis testing* (2nd ed.). San Diego: California Academic Press (Chapter 10).
Xu, C. X., & Xu, Y. (2003). Effects of consonant aspiration on Mandarin tones. *The Journal of the International Phonetic Association*, *33*, 165–181.
Xu, Y., & Xu, C. X. (2005). Phonetic realization of focus in English declarative intonation. *Journal of Phonetics*, *33*(2), 159–197.