languages

MDPI

Article

The Influence of Heritage Language Experience on Perception and Imitation of Prevoicing

Emily J. Clare [1] and Jessamyn Schertz [1,2,*]

[1] Department of Language Studies, University of Toronto Mississauga, Mississauga, ON L5L 1C6, Canada
[2] Department of Linguistics, University of Toronto, Toronto, ON M5S 3G3, Canada
* Correspondence: jessamyn.schertz@utoronto.ca

Abstract: This work tests the effect of heritage language background on imitation and discrimination of prevoicing in word-initial stops. English speakers with heritage languages of Spanish (where prevoicing is obligatorily present) or Cantonese (where prevoicing is obligatorily absent), as well as monolingual English speakers, imitated and discriminated pairs of stimuli differing minimally in prevoicing, both in English (participants' dominant language) and Hindi (a foreign language), and they also completed a baseline word reading task. Heritage speakers of Spanish were expected to show the highest performance on both imitation and discrimination, given the contrastive status of prevoicing in Spanish. Spanish speakers did indeed show more faithful imitation, but only for Hindi, not English, sounds, suggesting that imitation performance can differ based on language mode. On the other hand, there were no group differences in imitation of prevoicing in English or in discrimination in either language. Imitation was well above chance in all groups, with substantial within-group variability. This variability was predicted by individual discrimination accuracy, and, for Cantonese speakers only, greater prevoicing in baseline productions corresponded with more faithful imitation. Overall, despite an expectation for differences, given previous evidence for the influence of heritage languages on production and perception of English voiced stops, our results point to a lack of cross-language influence on perception and imitation of English prevoicing.

Keywords: phonetic imitation; heritage languages; perception; prevoicing; cross-language influence; voice onset time; bilingualism

Citation: Clare, Emily J., and Jessamyn Schertz. 2022. The Influence of Heritage Language Experience on Perception and Imitation of Prevoicing. *Languages* 7: 302. https://doi.org/10.3390/languages7040302

Academic Editors: Olga Dmitrieva, Chiara Celata, Esther de Leeuw and Natalia Kartushina

Received: 18 July 2022
Accepted: 15 November 2022
Published: 27 November 2022

1. Introduction

Speech perception and production are characterized by extreme sensitivity to fine-grained acoustic differences. At the same time, this sensitivity is shaped and constrained by language-specific knowledge. This work investigates English speakers' sensitivity to prevoicing in phonologically voiced stops /b d ɡ/. Specifically, we test whether the status of prevoicing in a heritage language (Spanish or Cantonese) influences ability to imitate and discriminate prevoicing in English (participants' dominant language) and a foreign language (Hindi), tapping into the broader question of whether and to what extent early exposure to a contrastive feature in a heritage language enhances sensitivity to that dimension more generally. We also test individual predictors of imitation: whether successful imitation of prevoicing is predicted by individual participants' discrimination ability and/or by their use of prevoicing in baseline productions.

Prevoicing is in free variation in English phonologically voiced stops /b d ɡ/: words beginning with these sounds can be produced either with or without prevoicing (i.e., [b] or [p]). Given that the phonologically voiceless counterparts /p t k/ are always aspirated word-initially in English, aspiration is considered the primary cue to the laryngeal contrast, while the presence vs. absence of prevoicing in /b d ɡ/ is not critical to the contrast and therefore considered to be less important. While English /b d ɡ/ are described in most textbook

descriptions as being canonically produced without prevoicing (e.g., Catford 2001; Ladefoged and Johnson 2014), more recent reports show that prevoicing is prevalent in many dialects, with substantial variability both within and across dialects (e.g., Walker 2020; see Herd 2020 for a review).

Languages differ in the role that prevoicing plays in cuing phonological contrast (Lisker and Abramson 1964; summarized in Table 1). In "true voicing" languages such as Spanish, among many others, prevoicing is consistently present in phonologically voiced stops /b d ɡ/, while phonologically voiceless stops /p t k/ are unaspirated. In contrast to English, then, the critical acoustic difference defining the laryngeal contrast in true voicing languages is the presence/absence of prevoicing, and this is assumed to be the primary cue for listeners in distinguishing between the categories. On the other hand, there are languages such as Cantonese where prevoicing is never present. In these languages, as in English, aspiration is the primary cue to the laryngeal contrast; however, unlike English, Cantonese unaspirated stops are obligatorily voiceless (see Cho and Ladefoged 1999, for a phonetic analysis of these and 15 other languages).

Table 1. Phonetic implementation of the word-initial stop laryngeal contrast (using the example of labial stops) across the three languages relevant to this work.

Language	Phonological Voiced Series	Phonological Voiceless Series	Prevoicing in Voiced Series	Primary Cue to Contrast
English	[b] or [p]	[p^h]	variable	aspiration
Spanish	[b]	[p]	always present	prevoicing
Cantonese	[p]	[p^h]	always absent	aspiration

These cross-linguistic differences in production lead to different predictions about listeners' sensitivity to prevoicing. In languages like Spanish where prevoicing is a critical cue to the phonological contrast, listeners must pay attention to it and be able to manipulate it. In languages where it is absent or optional, like Cantonese or English, there is no need to pay attention to or be able to produce it in order to understand or produce lexical distinctions. At the same time, in English, some speakers do produce it systematically (e.g., Lisker and Abramson 1964), suggesting that for these speakers, it is indeed important to the contrast. Furthermore, even for those English speakers who do not consistently produce prevoicing, it may be useful to pay attention to it, given that it does show variation across speakers and dialects and therefore may be a sociolinguistic and/or stylistic marker (see Drager 2010 for discussion and examples). Nevertheless, it is expected that the contrastive status of prevoicing in true voicing languages would lead to greater sensitivity and imitation ability overall.

Support for this idea comes from work comparing cross-language perception of voice onset time (VOT) continua ranging from stops with negative VOT (i.e., with prevoicing) to long-lag VOT (i.e., with aspiration). A sharp change in identification of voiced vs. voiceless stops, and greater discrimination accuracy, around the zero-VOT point for true voicing languages, but at a higher value of VOT for English speakers, indicates greater sensitivity to the presence of prevoicing in true voicing languages (discrimination: Williams 1974; identification: Benkí 2005; Caramazza et al. 1974; Schertz et al. 2020). Support for the importance of prevoicing in true voicing languages also comes from van Alphen and Smits (2004), who showed that prevoicing was the best of several acoustic dimensions in predicting Dutch listeners' identification of naturally produced stimuli.

Results from work comparing imitation of VOT continua across languages is also consistent with the idea that true voicing language speakers are better able to manipulate prevoicing than English speakers. For example, Flege and Eefting (1988) tested English and Spanish speakers' imitation of a VOT continuum and found that English speakers did not, for the most part, produce prevoiced stops, whereas Spanish speakers did. Furthermore, Olmstead et al. (2013) found that Spanish, but not English, speakers showed gradient imitation of prevoicing durational differences along a VOT continuum (see also Podlipský

and Šimáčková 2015 for Czech). Nevertheless, in both Flege and Eefting (1988) and Olmstead et al. (2013), English speakers also did show, on average, negative VOT values when imitating the negative VOT range of the continuum, albeit with much smaller negative values than the Spanish speakers. A plausible interpretation of this, supported by individual results in Flege and Eefting (1988), is that a small subset of English-speaking participants did show imitation of prevoicing.

Given these findings, along with the variable use of prevoicing documented in English speakers' productions, and the fact that it is prevalent in many dialects of English, we expect there to be a wide range of sensitivity to, and imitation of, prevoicing in English listeners. However, this has not, as far as we know, been directly tested, nor have potential predictors of variability been explored. In the current work, we explore one potential predictor of sensitivity to prevoicing: heritage language background.

We use the term "heritage language" to refer a language that is learned in the home as a child but that is not the dominant language of the community (see Nagy 2015 for discussion of different definitions of the term). A well-studied but still open question is the extent of interaction vs. independence of languages in heritage speakers, including whether the phonological system of the heritage language influences the phonology of the dominant language.[1] Chang (2021) provides a thorough review of work examining the phonetic and phonological systems of heritage speakers. He notes that heritage speakers' production of the dominant language is often not perceptibly different from that of monolingual speakers. At the same time, a growing body of work shows that this is not always the case, and that productions can often differ systematically, albeit subtly, from those of monolinguals.

The stop laryngeal contrast has been particularly well-studied in this domain, and much of the work converges to show both language-specificity and the presence of cross-language influence. Most studies have found that speakers maintain language-specific categories; for example, Spanish-English heritage speakers produce Spanish /b/ differently than English /b/, reflecting differences between monolingual productions in the two languages (e.g., Antoniou et al. 2010 for Greek-English; Balukas and Koops 2015 for Spanish-English; Kang et al. 2016 for Tagalog-English; Newlin-Łukowicz 2014 for Polish-English). At the same time, many of these same studies also find evidence of cross-language influence, the extent of which can vary considerably depending on the participants, the sound, and the linguistic context.

Phonologically voiced stops, the target of the current work, appear to be particularly susceptible to cross-language influence in production, as compared to their voiceless counterparts, as shown by several studies looking at heritage speakers of true voicing languages living in English-dominant communities, similar to the population in the current study. Kang et al. (2016) found that Tagalog-English heritage speakers produce voiceless stops /p t k/ in each language with similar VOT values as monolinguals, but that their voiced stops /b d ɡ/ exhibit bidirectional cross-language interference: heritage speakers produced more prevoicing than monolingual English speakers in English stops, and less prevoicing than monolingual Tagalog speakers in Tagalog stops. Similarly, Newlin-Łukowicz (2014) reported more frequent prevoicing of English stops by Polish-English heritage speakers than by monolinguals, despite showing similar VOT values for voiceless stops. Antoniou et al. (2010) found that Greek-English heritage speakers' stops were almost indistinguishable from those of monolinguals in both languages, except for English /b/, where they showed more prevoicing than monolinguals (see also Schwartz 2022 for more discussion and examples). In sum, although differences can be subtle, there is strong evidence for the influence of the phonetic system of a heritage language on production of the dominant language, and this influence appears to be particularly robust in the case of voiced stops.

Evidence for both language-specificity and interaction has also been found in perception, mostly by comparing category boundaries along a VOT continuum, which are lower in true voicing languages as compared to English. Some studies have shown that even early bilinguals show similar perceptual patterns in their two languages, intermediate to monolinguals of each language (e.g., Caramazza et al. 1974; Williams 1977a). Other work has

found language-specific perceptual patterns (Casillas and Simonet 2018; Dmitrieva 2019; Gonzales and Lotto 2013; Schertz et al. 2020), indicating different perceptual strategies for the different languages, but these differences tend to be quite small and/or only present in extremely balanced bilinguals. Furthermore, Antoniou et al. (2012) showed that perceptual strategies in Greek-English bilinguals were similar across language modes and biased toward their dominant language (English). Overall, while it is difficult to compare the two modalities directly, the combined findings of previous work suggest that cross-language interaction may have a larger impact on perception than on production.

Most studies above examined cross-language influence by comparing category boundaries in different language modes. A distinct but related question is whether *sensitivity* to a dimension might be affected by the status of that dimension in a heritage language. Some studies have demonstrated long-lasting impacts of early exposure to a heritage language: even after not using it for the majority of their lives, those with early exposure to a language have been shown to out-perform late learners in discriminating contrasts from that language. For example, Tees and Werker (1984) found that students with early exposure to Hindi in the first 1–2 years of their lives, but little subsequent exposure, were very successful in discriminating Hindi stop contrasts that are very difficult for English speakers, even after training (see also Oh et al. 2010, but cf. Pallier et al. 2003).

Turning to the question directly relevant to the current work, whether sensitivity due to L1 phonological patterns might influence sensitivity in other languages as well, Chang (2016) tested whether early exposure to a heritage language (Korean) would influence perceptual sensitivity in a dominant language (English). He found that heritage Korean speakers out-performed monolingual English speakers in discrimination of the place of articulation of English unreleased final stops. In English, final stops are often released, such that there are cues to the place of articulation in the stop burst. However, in Korean, final stops are obligatorily unreleased, such that the place of articulation must be fully identified based on coarticulatory cues in the vowel. The higher performance by Korean heritage speakers was attributed to enhanced sensitivity to these coarticulatory cues due to their early exposure to Korean. Chang's work provides evidence that heritage language experience can influence the extent of sensitivity to dimensions that may be less important in the dominant language (see also Chang 2018).

However, it appears that the extension of L1 sensitivity to other languages may be selective. In Korean, unlike in English, f0 is a primary cue to the laryngeal (i.e., voicing) contrast, and Korean speakers show spontaneous imitation of f0 differences in stops (Kwon 2019). However, when doing the same task in English, Korean-English bilinguals performed similarly to monolingual English speakers (Kwon 2021), failing to replicate the f0 imitation shown in the Korean task. This suggests that participants were not drawing on their L1 phonology when completing the English task, but rather using language-specific strategies. Taken together, past work therefore provides somewhat mixed predictions for whether heritage language experience would influence imitation strategies in the dominant language. Based on the findings of Chang (2016) and the large body of work showing cross-language influence in phonologically voiced stops, there is good reason to believe that Spanish heritage speakers should show greater sensitivity to, and imitation of, prevoicing in English stops. On the other hand, given findings of influence from the dominant language on perception of the heritage language (Antoniou et al. 2012), we might expect heritage speakers' performance to be equivalent to that of monolinguals. Finally, given that language-specific imitative strategies been found in previous work (e.g., Kwon 2021), it is also possible that performance would differ depending on the language being tested.

In the current work, three groups of English speakers (heritage speakers of Spanish or Cantonese and monolinguals), all born and residing in the United States or Canada, and all of whom learned English in early childhood, completed imitation and discrimination tasks on pairs of stimuli differing in the presence/absence of prevoicing, as well as a word reading task to measure baseline levels of prevoicing. Stimuli were presented from two languages: English (participants' dominant language), and Hindi (a foreign language unfa-

miliar to all participants)[2], with two types of manipulations of prevoicing absence (natural vs. artificial, described below), and at two places of articulation (labial and coronal).

The goal of this study was to examine the upper limit of participants' capabilities, and as such, the instructions, procedure, and materials used in the tasks were chosen to maximize imitation and perceptual sensitivity. Instructions were very explicit: participants were directly asked to pay attention to and imitate differences that may be very subtle, as explicit instructions to imitate have been shown to elicit greater imitation than more implicit tasks (Dufour and Nguyen 2013). Stimuli were always presented in pairs differing minimally in the presence vs. absence of prevoicing (or, for filler trials, aspiration); this direct juxtaposition not only was expected to draw attention to the differences, but was also expected to enhance imitation, based on previous work showing hyperarticulation of contrastive phonetic features in the presence of minimal pairs (e.g., Baese-Berk and Goldrick 2009). Finally, a relatively small set of stimuli were used, and all items were stop-initial, further drawing attention to the target contrast. Together, these design choices were shaped by our overall goal of testing the extent to which speakers can perceive and imitate differences in prevoicing under optimal, highly controlled conditions, with the understanding that this may not reflect day-to-day processing.

Our primary research question was how imitation and perceptual sensitivity of a freely varying phonetic dimension differ based on language background, but here we also lay out our motivations for the other manipulations included in the design. We included two stimulus languages (English vs. Hindi) given findings that imitative strategies can differ based on language mode (Kwon 2021). We included pairs of stimuli with a truly minimal difference in prevoicing (artificial prevoicing absence condition) to see how well prevoicing alone is perceived and imitated, in comparison to pairs that had been naturally produced with and without prevoicing (natural prevoicing absence condition). We expected the natural condition to be easier to discriminate and imitate on account of the multiple naturally occurring acoustic differences. We included two places of articulation mainly for purposes of generalizability; our study is not designed to test place of articulation effects in a systematic way. Nevertheless, based on previous findings showing that listeners may be more sensitive to the presence of voicing at more posterior places of articulation (Kharlamov 2022), we might expect that prevoicing in coronals is more perceptually salient—and more easily imitated—than prevoicing in labials.

In terms of our primary question of interest, namely how early exposure to a true voicing language affects imitation and discrimination of prevoicing, we hypothesized that Spanish speakers would outperform the other two groups, based on (1) the contrastive status of prevoicing in Spanish, vs. Cantonese or English, and (2) the demonstrated susceptibility of phonologically voiced stops to cross-language influence. We also explored two potential predictors of individual variability in imitation: discrimination accuracy and baseline prevoicing levels. We expected that individual imitative performance would be related to individual discrimination ability, and that those individuals who produced prevoicing in English baseline productions might be better imitators.

2. Materials and Methods

2.1. Participants

68 participants' data are included in the current study. An additional 25 participated, but their data was not used due to poor recording quality (n = 18) or because they did not meet our language background or residence criteria (n = 7). Participants were recruited through Psychology courses at the University of Toronto (n = 7) and through the online experiment platform Prolific (n = 61), and they received course credit or monetary compensation for their time.

Participants fell into three language groups: Spanish/English (n = 20), Cantonese/English (n = 25), and Monolingual English (n = 23) (we refer to these groups as Spanish, Cantonese, and Monolingual, respectively). All participants were between the ages of 18 and 37 and had lived in the United States or Canada since birth (see Table 2 for a break-

down of demographic information). Participants in the bilingual groups learned Spanish or Cantonese as a first language in the home; some learned English simultaneously in the home, and all learned it, at the latest, at the onset of schooling. Self-reported proficiency in speaking and understanding on a scale from 1 (low) to 7 (high), and percentage of time that each language is/was used with family, friends, and at school are shown in Figure 1. While considering themselves more proficient in English, most also maintain a relatively high level of proficiency in both speaking and understanding their heritage language. In terms of use, English is clearly dominant with friends and at school, while there is a wide range of variability in use with family. Participants in the Monolingual group learned only English in the home, and self-reported as not proficient in any other language.

Table 2. Summary of demographic information. Age and English Age of Acquisition (AoA) give means and ranges in parentheses. For English AoA, 0 refers to "from birth." Gender was self-reported with terms chosen by the participant; in the table below, F encompasses {female/woman /she/her}, M {male/man}, NB {nonbinary/gender fluid/they}.

LangGroup	N	Age	Gender	Current Residence	English AoA
Spanish/English	20	22 (18–30)	14F, 4M, 2NB	US (19), Canada (1)	4 (0–8)
Cantonese/English	25	25 (19–37)	20F, 5M	US (7), Canada (18)	3 (0–6)
Monolingual English	23	25 (18–30)	9F, 13M, 1NB	US (18), Canada (5)	0

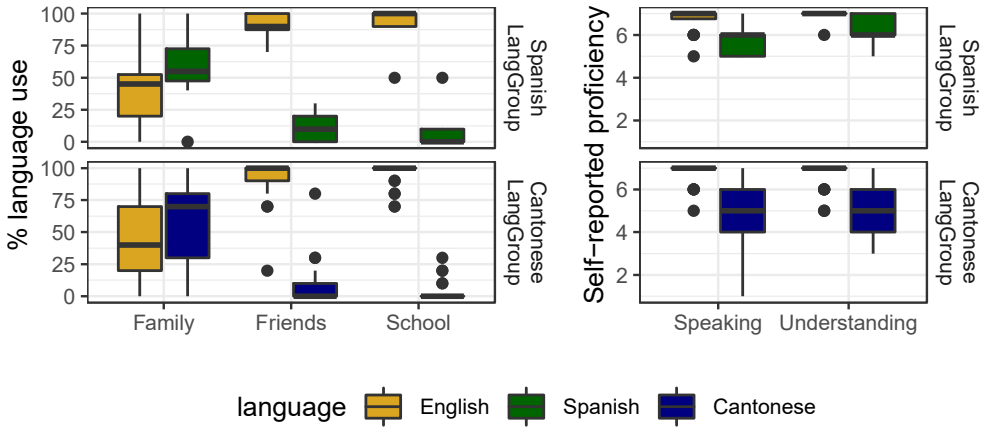

Figure 1. Distribution of self-reported proficiency, on a scale from 1 to 7, for Spanish/English (**top**) and Cantonese/English (**bottom**) speakers in speaking and understanding English and Spanish or Cantonese (**left**), and percentage use of each language with family, friends, and at school (**right**).

2.2. Materials

Lists of 18 (Spanish and Cantonese) or 22 (English) words beginning with phonologically voiced and voiceless stops at all places of articulation (labial, coronal, velar) before the vowels /a/, /i/, and /u/ was prepared for a baseline reading task for each language (Table A1 in Appendix A). Words were presented in native orthography, and for Cantonese, participants chose in advance whether words would be presented in traditional ($n = 16$) or simplified ($n = 9$) characters. For Cantonese and Spanish, the English translation of the word was also provided; this was because it was expected that some Cantonese participants would have limited reading proficiency, and that seeing the English translation might help them recognize the character.

The stimuli for the main task consisted of minimal pairs of English and Hindi words beginning with stops at two places of articulation, differing in prevoicing or aspiration. Stimuli were based on natural productions of speakers of each language. For English, a native speaker from the midwestern United States recorded the words *pie*, *buy*, *tie*, and *die*. *Buy* and *die* were produced both with and without prevoicing, such that there were three baseline tokens for each place of articulation (e.g., *pie* [p^{h}aj]; *buy*, [baj] and [paj]). Baseline

stimuli for Hindi were the recordings of the words [pʰɑl] *knife blade*, [pɑl] *nature*, [bɑl] *hair*, [t̪ʰɑl] *platter*, [t̪ɑl] *beat (n.)*, and [d̪ɑl] *lentil* associated with the *Illustrations of the IPA* article on Hindi by Ohala (1994). Acoustic characteristics of the baseline stimuli are provided in Appendix B, and the stimuli themselves are available in supplemental materials.

Along with these naturally produced stimuli, "artificial" voiceless tokens were created by splicing out prevoicing from the naturally produced prevoiced stimuli. For example, the English production of *buy* with prevoicing [baj] was manipulated such that the prevoicing was removed, creating a phonetically voiceless token that differed minimally from the prevoiced token. We indicate the artificial voicelessness as [b̥], in contrast to the naturally produced voiceless unaspirated token [p]. The same manipulation was done with aspirated tokens to create artificial unaspirated stimuli [p⁼].

The main imitation/discrimination task, described below, made use of "trial sets" focusing on pairs of stimuli differing in prevoicing. As shown in Table 3, there were 8 pairs of target stimuli, representing the fully crossed set of 2 languages (English, Hindi), 2 places of articulation (labial, coronal), and 2 types of lack of prevoicing (natural vs. artificial). An additional 8 pairs of stimuli, differing in presence/absence of aspiration along the same parameters, were used as filler trial sets (Table A2 in Appendix A).

Table 3. The 8 pairs of target stimuli used in the main task, with each pair differing in the presence/absence of prevoicing. Hindi coronal stops are all dental; the dental diacritic is omitted for readability. [b̥] and [d̥] refer to tokens where prevoicing has been artificially removed from naturally prevoiced tokens. An additional 8 pairs of stimuli differing in presence/absence of aspiration, not shown here, were also presented as fillers.

		English	**Hindi**
Place	**Manipulation**	**Prevoicing Present ~ Absent**	**Prevoicing Present ~ Absent**
labial	natural	[baj] ~ [paj]	[bɑl] ~ [pɑl]
	artificial	[baj] ~ [b̥aj]	[bɑl]~ [b̥ɑl]
coronal	natural	[daj] ~ [taj]	[dɑl] ~ [tɑl]
	artificial	[daj] ~ [d̥aj]	[dɑl] ~ [d̥ɑl]

2.3. Procedure

The experiment was run using the online platform Gorilla (Anwyl-Irvine et al. 2020). Participants gave informed consent and then completed a headphone check (Woods et al. 2017). Prior to the main task, they performed a baseline word-reading task, where they read aloud words in isolation as they appeared on the screen (2 repetitions of the wordlist given in Table A1 of Appendix A). All participants completed the baseline word-reading in English, and the bilingual groups also completed it in Spanish or Cantonese. This was followed by the main task, encompassing both imitation and discrimination, which consisted of two blocks, the first of which included only English stimuli, and the second of which included only Hindi stimuli. Each of these blocks consisted of a series of eight "trial sets," each corresponding to a pair of syllables differing in presence/absence of prevoicing or aspiration.

A sample trial set is shown in Table 4. First, participants heard the pair of stimuli, repeated twice in sequence. Second, participants heard the pair, also repeated twice in sequence, and were asked to imitate what they had heard immediately after each stimulus. Finally, participants completed four ABX discrimination trials presented in a randomized order. In each of these trials, they were asked to decide whether a third stimulus matched the first or second word they had heard. For each trial set, it was randomly chosen which member of the stimulus pair would be the A stimulus (presented first throughout the duration of the trial set) versus the B stimulus (presented second). Participants were explicitly told that they were hearing words of English or Hindi, depending on the block.

Table 4. Summary of a sample trial set from the English block.

Phase	A	B	X
LISTEN: you will listen to two words of English. Try to pay attention to the differences.	[baj]	[paj]	–
	[baj]	[paj]	–
IMITATE: Now you will be asked to repeat the words. After you hear each word, please repeat it out loud, trying to imitate it as closely as possible.	[baj]	[paj]	–
	[baj]	[paj]	–
DECIDE: Please decide whether the third word sounds more like the first or second word you heard.	[baj]	[paj]	[baj]
	[baj]	[paj]	[paj]
	[baj]	[paj]	[baj]
	[baj]	[paj]	[paj]

The English block was always presented first, followed by the Hindi block. The first trial set presented to all participants in the English block (and thus the first trial set of the experiment) was a filler trial set where stimuli differed in aspiration. Apart from this, the order of trial sets within each block was randomized. Within the 8 trial sets of each language block, half were target trial sets which differed minimally in prevoicing, and the other half were filler trial sets which differed minimally in aspiration. In total there were 8 target trial sets per participant for analysis (2 languages * 2 VoicingAbsent types * 2 places of articulation).

After the main task, participants completed a demographic questionnaire and the Bilingual Language Profile (Birdsong et al. 2012). The experiment took about 40 min in total.

2.4. Data Processing, Phonetic Annotation and Measurements

Only phonologically voiced tokens (for the baseline wordlist task) and imitation of pairs differing in prevoicing (for the imitation/discrimination task) are analyzed in the current work. Phonologically voiceless stops and pairs differing in aspiration were considered fillers; an overview of performance on these filler items is provided at the beginning of the Results section, but no analysis is included.

For the main imitation task, out of a possible 2176 imitations of target tokens, 12 were missing, and several tokens were omitted because of noise (n = 11) or because they were produced with a different manner of articulation (fricative or approximant; n = 4), leaving 2149 tokens for analysis. In some cases, the presence of higher-frequency formants provided evidence of prenasalization. This happened consistently in two monolingual speakers and 3 Spanish speakers, and sporadically in others. These were considered prevoiced for the purposes of this analysis.

For the baseline wordlist task, 286 Cantonese, 1345 English, and 328 Spanish tokens beginning with phonologically voiced stops were analyzed (due to technical errors with recording and/or participants skipping words, 136 Cantonese, 151 English, and 32 Spanish tokens were omitted).

Each token was manually annotated for prevoicing, as evidenced by the presence of a voice bar in the spectrogram and/or periodicity in the waveform during stop closure. When present, the beginning of prevoicing was considered the beginning of periodicity in the waveform, and ended just before the stop burst, or the end of periodicity, whichever came first. Figure 2 shows examples of tokens annotated as having prevoicing present (a and b) and absent (c). The primary metric used in our analysis is binary presence/absence of prevoicing, with any amount of prevoicing counting as present, even if it ended before the stop burst, as in Figure 2a. We also took measurements of prevoicing duration

and intensity. Quantitative analysis of these measures is outside the scope of the current work; however, we performed exploratory analyses examining the phonetic properties of prevoicing, and found strikingly little difference in the nature of prevoicing across groups. Furthermore, statistical analysis of our primary question using prevoicing duration, rather than presence/absence of prevoicing, as the response variable of interest, resulted in the same conclusions as in the analysis reported here.

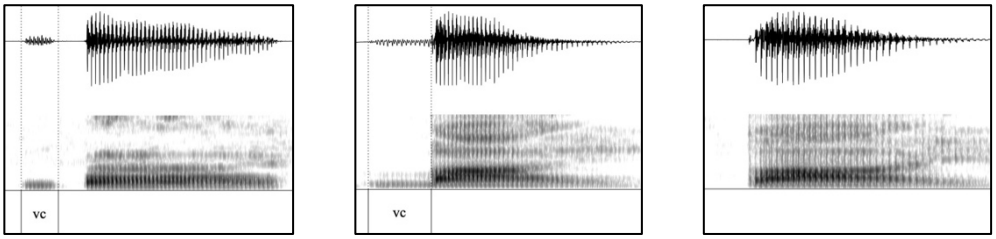

Figure 2. Examples of tokens produced with and without prevoicing (labeled as 'vc' when present): (**a**) Imitation of English [baj] by a monolingual English speaker with voicing present; (**b**) Imitation of English [baj] by a Cantonese-English speaker with voicing present; (**c**) Imitation of English [b̥aj] by the same Cantonese-English speaker with voicing absent.

2.5. Statistical Analysis

First, we calculated by-participant imitation and discrimination scores for each combination of our factors of interest (Stimulus Language, VoicingAbsent Type, Place of Articulation). For imitation, we calculated the proportion of tokens produced with prevoicing when imitating tokens that had prevoicing, and the proportion of tokens produced with prevoicing when imitating tokens without prevoicing. The difference between these two was taken as the imitation score, which ranged from 1 (completely faithful imitation) to −1 (completely non-faithful imitation; i.e., always producing the opposite prevoicing value as the stimulus). The discrimination score was calculated as each participants' average percentage correct (ranging from 0 to 100).[3]

We used linear mixed-effects regression (lme4 package in R; Bates et al. 2015) to analyze how participants' imitation and discrimination of presence/absence of prevoicing differed based on our factors of interest. We created two models, one for imitation and one for discrimination. The response variable for each was the imitation or discrimination score, and predictor variables were, with reference levels in italics: Language Group (*Spanish/English*, Cantonese/English, Monolingual English), Stimulus Language (*English*, Hindi), and VoicingAbsent Type (*Artificial*, Natural). Two-way interactions between Stimulus Language and each of the other predictors were included as we expected that the effect of the predictors might differ based on the stimulus language. All predictor variables were simple-coded (e.g., −0.5, 0.5 for two-level factors) with the reference levels given in italics. Random by-participant intercepts were included. P-values were computed using the lmerTest package (Kuznetsova et al. 2017). When appropriate, we performed follow-up Chi-squared tests with non-adjusted *p*-values, using the *phia* package (De Rosario-Martínez 2015).[4]

We also used linear regression models to test individual-level predictors of imitation: specifically, whether participants' imitative performance was related to their discrimination ability and/or their baseline use of prevoicing in English, using the same indices described above.

3. Results

To ensure that participants understood the task and were completing it as instructed, we first examined performance on the filler trials differing in aspiration, which were expected to be easy for all participants to discriminate and imitate. Discrimination was above 75% for all participants (mean 95%), and all participants showed robust imitation of the

aspiration contrast (mean difference 75 ms between unaspirated and aspirated stops); furthermore, there were no clear differences between language groups in performance on the filler trials. The following analysis focuses on our question of interest: imitation and discrimination of trials differing in prevoicing.

3.1. Baseline Wordlist

Figure 3 shows the percentage of phonologically voiced tokens produced with prevoicing in the baseline wordlist task. The boxplots show the distribution of the percentage of tokens produced with prevoicing by each speaker. For example, for the Monolingual group, since the median is near zero, this means that about half of speakers produced no tokens with prevoicing, while the other half produced at least some tokens (ranging from zero to about 80%) with prevoicing. Cantonese speakers produced no prevoicing at all in their Cantonese productions, and most Spanish speakers produced the majority of Spanish words with prevoicing. The percentage of English tokens produced with prevoicing by the two heritage groups are intermediate between the Monolingual group and the productions in their other language. Despite the fact that all participants are early learners of English, there are systematic group differences in English prevoicing levels based on language background, indicating cross-language influence: Cantonese speakers produce very few tokens as prevoiced (but more than in Cantonese), and Spanish participants produce more prevoicing than the other two language groups in English (but less than they do in Spanish). This is consistent with previous work showing more prevoicing in the English stops of Spanish-English bilinguals than in monolingual speakers (Williams 1977a).

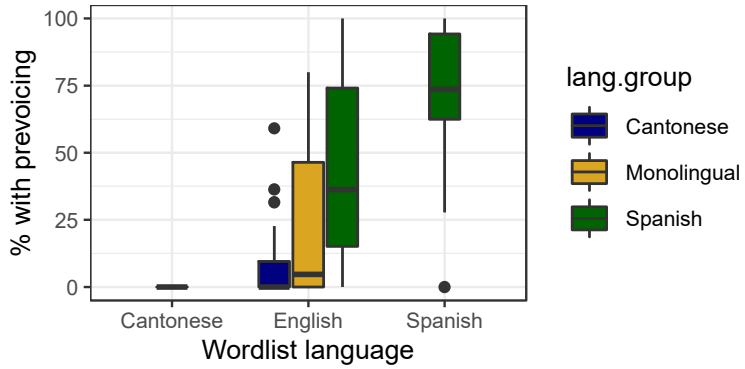

Figure 3. Percentage of phonologically voiced tokens produced with prevoicing in the baseline word reading task. Boxplots show distributions of by-participant means.

3.2. Imitation

Results for imitation of stimuli differing in presence/absence of prevoicing are shown in Figure 4, and statistical results are shown in Table 5. Recall that the response variable, and the y-axis of the figures, is the individual imitation score, with positive values indicating greater-than-chance imitation, and 1 indicating completely faithful imitation. In the statistical results, the beta-coefficients represent the difference in imitation scores between the two levels of the relevant factor. Breakdowns showing the percentage of voiced productions for stimuli with voicing present vs. absent separately are shown in Appendix C.

Our primary question was whether imitation differed across language groups, with the hypothesis that Spanish speakers should show more imitation than the other two groups. We first note that all groups showed clear imitation, with the vast majority of individual speakers showing at least some imitation (i.e., positive scores). While there was no significant main effect of Language Group for either the Spanish-Monolingual or Spanish-Cantonese comparisons[5], there was a significant interaction between Stimulus Language and both of these comparisons, indicating that the effect of Language Group differs depending on whether participants were imitating English or Hindi stimuli. Follow-up tests confirm the pattern shown in the graph: when imitating Hindi, the Spanish group showed

higher imitation scores than both the Monolingual group ($\chi^2 = 4.39$, $p = 0.036$) and the Cantonese group ($\chi^2 = 6.77$, $p = 0.009$)[6]. However, the Spanish group did not differ from either of the other two groups in imitation of English ($p > 0.1$ for both).

Table 5. Statistical results for the imitation task. Effects in bold are significant ($p < 0.05$).

	Estimate	SE	t	*p*
(Intercept)	**0.306**	**0.030**	**10.158**	**<0.001**
Stimulus Language (Hindi vs. *English*)	**0.199**	**0.032**	**6.248**	**<0.001**
Language Group (Monolingual vs. *Spanish*)	−0.075	0.076	−0.994	0.324
Language Group (Cantonese vs. *Spanish*)	−0.106	0.074	−1.423	0.160
VoicingAbsent Type (natural vs. *artificial*)	**0.137**	**0.032**	**4.309**	**<0.001**
Place of Articulation (coronal vs. *labial*)	**0.089**	**0.032**	**2.802**	**0.005**
StimLang * LangGroup (Mono vs. *Spanish*)	**−0.207**	**0.080**	**−2.597**	**0.010**
StimLang * LangGroup (Cantonese vs. *Spanish*)	**−0.225**	**0.079**	**−2.869**	**0.004**
StimLang * VoicingAbsent Type	**0.213**	**0.063**	**3.358**	**0.001**
StimLang * Place	**0.132**	**0.063**	**2.083**	**0.038**

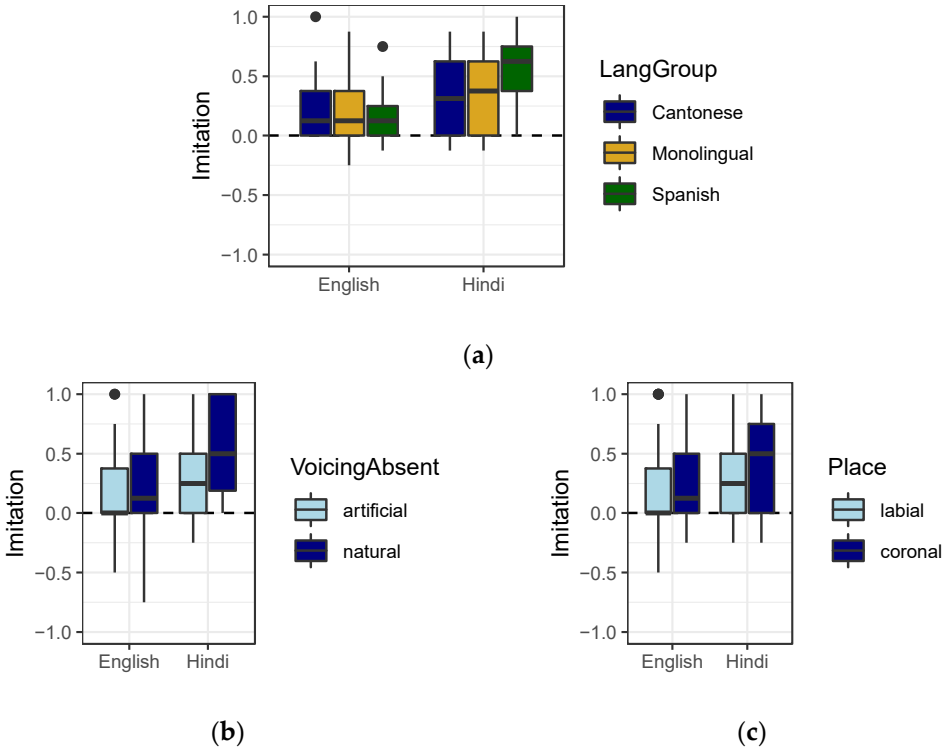

Figure 4. Distribution of by-participant imitation scores across (**a**) Language Groups, (**b**) artificial vs. natural VoicingAbsent conditions, and (**c**) Places of Articulation. Imitation scores represent the difference in proportion of prevoiced tokens when imitating stimuli where prevoicing is present vs. absent; values of 1 indicate completely faithful imitation.

The three other factors we tested, Stimulus Language, VoicingAbsent Type, and Place of Articulation, all showed significant effects. The main effect of Stimulus Language, in which participants showed more imitation of Hindi than English, held across all other factors, but to different extents, as shown by the interaction with the other two factors: namely, the benefit was much larger for artificial than natural VoicingAbsent types, and larger for coronal than labial stops. For VoicingAbsent Type, pairs with naturally voiceless tokens were imitated significantly more faithfully than those with artificially removed voicing, as shown by the significant main effect. However, follow-up tests based on the interaction with Stimulus Language indicate that this effect was only significant when listening

to Hindi ($\chi^2 = 29.45$, $p < 0.001$), and that there was no significant difference in VoicingAbsent Type when listening to English ($p > 0.1$). There was also greater imitation of coronal stops than labial stops overall, but follow-up tests based on the interaction with Stimulus Language showed that the Place of Articulation effect was significant in Hindi ($\chi^2 = 11.96$, $p < 0.001$), but not in English ($p > 0.1$).

To sum up, participants in all language groups showed significant imitation of pairs of stimuli differing in the presence/absence of prevoicing. Spanish speakers showed greater imitation than the other Language Groups, but only when imitating Hindi, but not English, stimuli. In addition, across Language Groups, imitation of Hindi stimuli was higher overall, and prevoicing imitation in the Hindi block was modulated by artificial vs. natural voicelessness as well as the place of articulation of the stimulus, while there was no evidence of a role for these factors in imitation in the English block.

3.3. Discrimination

Results for discrimination of pairs of stimuli differing in presence/absence of prevoicing are shown in Figure 5, and statistical results are shown in Table 6. The response variable, and the y-axis of the figures, is by-participant discrimination scores. In the statistical results, the beta-coefficients represent the difference in scores between the two levels of the relevant comparison.

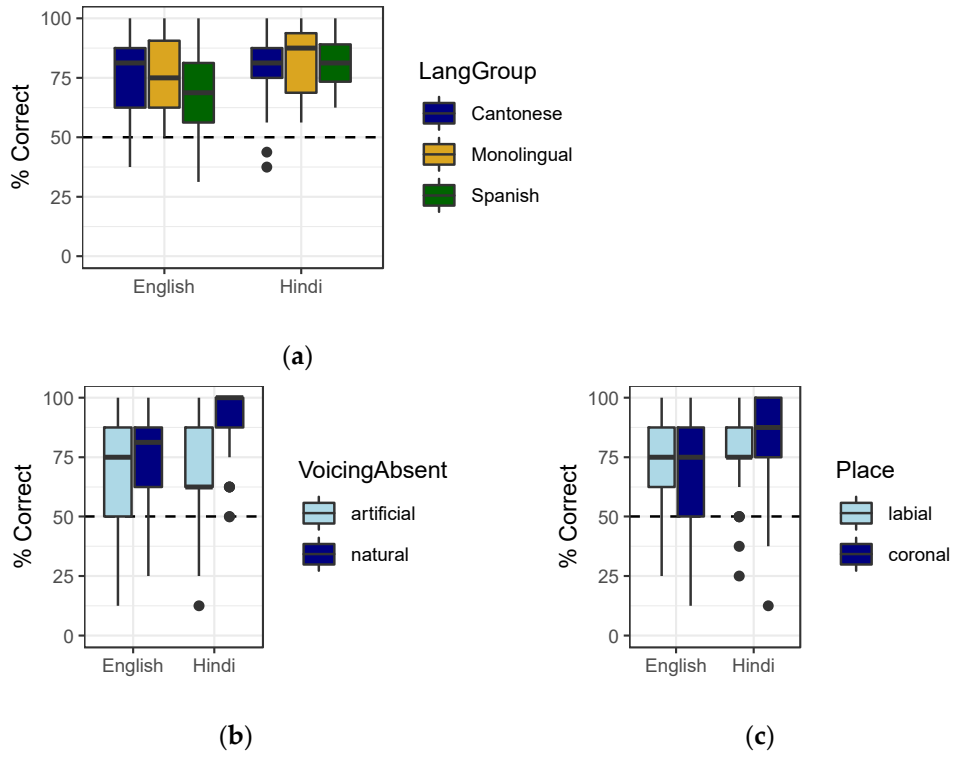

Figure 5. Discrimination accuracy for pairs of stimuli differing in prevoicing, broken down by (**a**) Language Groups, (**b**) artificial vs. natural VoicingAbsent conditions, and (**c**) Places of Articulation. Boxplots show distributions of by-participant means.

All groups showed above-chance discrimination (i.e., greater than 50% accuracy), with the vast majority of speakers showing above-chance discrimination. Our primary question was whether discrimination differed across language groups, with the hypothesis that Spanish speakers would show greater sensitivity than the other two groups. However, there was no main effect of Language Group for either comparison[7], and while there was an interaction of Stimulus Language and the Spanish-Cantonese comparison, follow-up tests showed that the difference between Spanish and Cantonese groups was not significant for either Stimulus Language (English: $\chi^2 = 2.15$, $p = 0.143$; Hindi: $\chi^2 = 0.63$, $p = 0.436$).

Turning to the effects of the other three factors: as with imitation, discrimination was higher for Hindi than for English stimuli, as shown by the significant main effect of Stimulus Language, but follow-up tests based on the interactions show that this effect was only significant for artificial, and not natural, VoicingAbsent Types (artificial: $\chi^2 = 0.23$, $p = 0.630$; $\chi^2 = 2.15$, $p < 0.001$), and only for labial, but not coronal, stops (labial: $\chi^2 = 1.36$, $p = 0.243$; $\chi^2 = 21.71$, $p < 0.001$). The significant main effect of VoicingAbsent Type indicates that pairs with naturally voiceless tokens were better discriminated than those with artificially removed prevoicing overall, and follow-up tests showed that this effect held in both Hindi (mean accuracy 91% for natural vs. 69% for artificial; $\chi^2 = 67.81$, $p < 0.001$) and, to a lesser extent, in English (mean accuracy 76% for natural vs. 67% for artificial; $\chi^2 = 11.31$, $p < 0.001$). For Place of Articulation, although there was no significant main effect, follow-up tests based on the interaction with Stimulus Language indicate that labials were better discriminated than coronals in English (mean accuracy 76% for labial vs. 68% for coronals; $\chi^2 = 7.92$, $p = 0.005$), while there was no significant difference in Hindi (mean accuracy 81% for labial vs. 79% for coronals; $p < 0.1$).

Table 6. Statistical results for the discrimination task. Effects in bold are significant ($p < 0.05$).

	Estimate	SE	t	*p*
(Intercept)	**0.762**	**0.015**	**50.454**	**<0.001**
Stimulus Language (Hindi vs. *English*)	**0.078**	**0.019**	**4.111**	**<0.001**
Language Group (Monolingual vs. *Spanish*)	0.035	0.038	0.912	0.365
Language Group (Cantonese vs. *Spanish*)	0.015	0.037	0.395	0.694
VoicingAbsent Type (natural vs. *artificial*)	**0.155**	**0.019**	**8.201**	**<0.001**
Place of Articulation (coronal vs. *labial*)	−0.028	0.019	−1.504	0.133
StimLang * LangGroup (Mono vs. *Spanish*)	−0.069	0.048	−1.448	0.148
StimLang * LangGroup (Cantonese vs. *Spanish*)	**−0.099**	**0.047**	**−2.121**	**0.034**
StimLang * VoicingAbsent Type	**0.131**	**0.038**	**3.445**	**0.001**
StimLang * Place	**0.094**	**0.038**	**2.475**	**0.014**

Overall, discrimination performance was well above chance, with higher performance for Hindi than English in certain conditions but with no indication of differences based on the language background of participants. Naturally produced voiceless tokens were easier to discriminate from naturally produced prevoiced tokens as compared to those where voicing had been artificially removed, and this effect was even more pronounced in Hindi than English. There was also a small difference in place of articulation for English, but not Hindi.

3.4. Individual Predictors of Imitation: Discrimination and Baseline Voicing

Our final set of analyses explored individual predictors of imitation: whether participants' faithfulness of imitation was related to individual discrimination accuracy and baseline use of prevoicing in English. For each participant, we calculated three scores: (1) imitation, (2) discrimination, and (3) English baseline prevoicing. The imitation and discrimination indices were calculated as above, but averaged across all factors (Stimulus Language, Place and VoicingAbsent Type), such that there was a single datapoint per participant. The baseline voicing score was the percentage of phonologically voiced tokens produced with prevoicing in the English wordlist reading task.

The relationship between individual imitation and discrimination scores, broken down by Language Group, is shown in the left panel of Figure 6. As shown in the graph, there is a positive relationship between discrimination and imitation, and this holds for all language groups (the higher line for the Spanish group reflects the overall higher degree of imitation for the Spanish group, as discussed above). We tested the strength of the relationship between imitation and discrimination, and whether it differed across Language Group, via a linear regression model with a response variable of individual imitation and predictor variables of individual discrimination accuracy, Language Group (centered), and

their interaction. Results showed a significant effect of discrimination accuracy on imitation ($\beta = 0.009$, SE = 0.002, t = 4.316, $p < 0.001$), but no other significant main effects or interactions (all $p > 0.1$). This indicates that participants who are better at perceiving the difference tend to show more imitation, and that this relationship does not appear to differ across the three language groups. The overall correlation coefficient between individual imitation and discrimination (not considering language group) is r = 0.463.

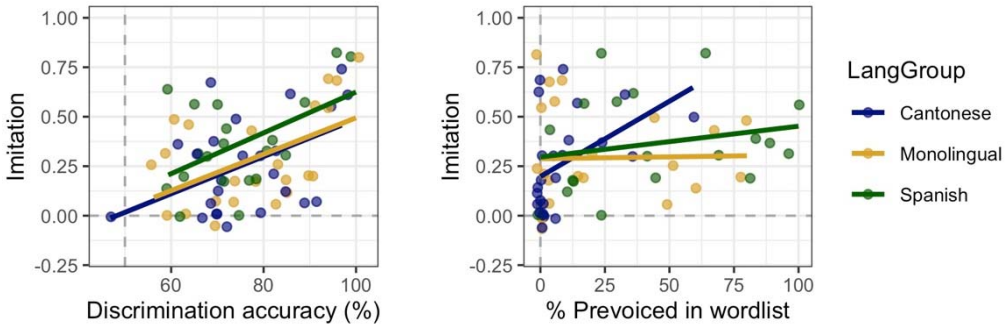

Figure 6. Mean by-participant imitation values, as predicted by individual discrimination accuracy (**left**) and baseline prevoicing (percentage of tokens produced with prevoicing in the English wordlist, **right**), with the best-fit regression line for each language group. Points are jittered for readability.

The relationship between imitation and baseline prevoicing, broken down by Language Group, is shown in the right panel of Figure 6. Because the different language groups varied systematically in their baseline prevoicing scores (e.g., Cantonese speakers had much lower baseline values than Spanish speakers), we did not include both baseline prevoicing and Language Group as predictors in the same model, given their high degree of collinearity. Instead, we tested the degree to which baseline prevoicing predicted imitation for each language group separately, using three separate models, each with the response variable of individual imitation and a predictor variable of baseline prevoicing. The effect of baseline prevoicing was significant for the Cantonese group ($\beta = 0.008$, SE = 0.003, t = 2.421, $p = 0.024$), but not for the two other groups (both $p > 0.1$). This reflects the patterns shown in the graph: there does not appear to be a relationship between baseline prevoicing for the Spanish or Monolingual groups, but there does for the Cantonese group. This appears to be driven by the fact that many of the Cantonese participants who showed little to no prevoicing in the English baseline reading task also did not show imitation. However, this relationship is not particularly strong, and leaves much of the variance unexplained; for example, there were several participants who showed no prevoicing but did show substantial imitation.

4. Discussion

4.1. Summary of Results

The primary focus of this study was to investigate imitation and discrimination of word-initial prevoicing, a phonetic dimension in free variation in English, and whether this sensitivity differs based on heritage language background. We had hypothesized that Spanish heritage speakers would perform better than Cantonese heritage speakers and English monolinguals, given the contrastive status of prevoicing in Spanish. We found partial support for this hypothesis: Spanish speakers did show more imitation than the other two groups, but crucially, only when imitating stimuli from a foreign language (Hindi). No group differences were found in imitation of English, and no differences were found in discrimination accuracy. Overall, participants from all language groups showed sensitivity to, and imitation of, minimal differences in prevoicing, albeit with substantial individual variability.

We found clear parallels between perception and production, including group-level regularities (e.g., imitation and discrimination were both highest for naturally produced Hindi contrasts), and an individual relationship between discrimination accuracy and imitation that held across all language groups. This relationship was robust, but not strongly predictive ($r = 0.463$), suggesting that there are important predictors of imitation other than discrimination acuity. On the other hand, there was little evidence for a relationship between baseline voicing and imitation, with the exception of a correlation for one group (Cantonese speakers), discussed below.

Naturally produced distinctions between prevoiced and nonprevoiced stimuli were easier to discriminate than pairs in which the prevoicing had been artificially removed from one of the tokens. This is expected since there are multiple acoustic differences in any naturally produced pair of utterances, whereas the stimuli in the artificial condition were identical except for the removed prevoicing. For Hindi, but not for English, the natural pair also showed more imitation. This is likely because the natural pair [b]-[p] in Hindi represents two different phonemes, and while prevoicing is the primary cue thought to differentiate this contrast, other, secondary cues, such as f0 at vowel onset, also differentiate the contrast in production (e.g., Schertz and Khan 2020). On the other hand, in English, secondary cues such as f0 do not appear to vary based on the presence vs. absence of prevoicing in unaspirated stops (Dmitrieva et al. 2015). This discrepancy is also most likely the reason for the overall advantage for both imitation and discrimination of Hindi vs. English stimuli, which was driven almost entirely by the natural VoicingAbsent Type.

In terms of the effect of place of articulation, we found that numerically, performance was better on coronal stops in Hindi but better on labial stops in English (these differences were only significant in imitation for Hindi and only in discrimination for English). Since we only had one baseline set for each place of articulation, we cannot interpret this finding generally; we think it likely that the difference is based on the properties of the specific tokens used for each language rather than a general discrepancy in perception/imitation of place of articulation differences across languages. Future work should take on a more systematic investigation of potential effects of place of articulation on imitation, given that some work has found that perception of the laryngeal contrast is subtly modulated by this factor (Kharlamov 2022).

As noted above, our tasks were designed to examine participants' imitation and discrimination capabilities under ideal circumstances; i.e., using instructions, presentation, and stimuli expected to maximize both the perceptual salience of the target feature and participants' tendency to imitate. However, given the explicit nature of our paradigm, the results may not reflect use of these cues in everyday speech processing. Future work could complement these findings by testing how sensitivity differs when the contrast is less salient, as well as the extent to which imitation of prevoicing is reduced in more implicit tasks (e.g., Dufour and Nguyen 2013).

4.2. Influence of Language Background and Language Mode on Imitation and Discrimination

We hypothesized that Spanish heritage speakers would be more sensitive to prevoicing than the other two groups because of the contrastive status of prevoicing in Spanish. This was expected to carry over to English based on previous findings of cross-language influence from heritage languages in production of voiced stops, bolstered by the fact that we might expect even more cross-language influence in tasks tapping into perception, as did the imitation and discrimination tasks in our current study. The fact that the Spanish group showed more faithful imitation of differences than the other two groups when imitating Hindi stimuli provides support for the idea that a contrastive dimension in L1 increases ability to make use of this cue more generally. However, this increased imitative performance did not hold in English, participants' dominant language; Spanish speakers' imitation of English was comparable to the other two groups (and less faithful than their imitation of Hindi). This is in line with the findings of Antoniou et al. (2012) that the dominant language may have a prevailing influence on perception in bilinguals. Although

Spanish speakers may have enhanced ability to imitate prevoicing, they may have learned from their experience with English that it is not an important dimension, and therefore do not draw on this ability, particularly in an English context.

An open question is why Spanish speakers did not show better *discrimination* of prevoicing than the other groups, at least in Hindi, where the imitation benefit was found. While we do not have a compelling explanation for this, discrimination performance was relatively high in general, possibly due to the high perceptual salience of the contrast in our paradigm, so this lack of difference could be due to a ceiling effect. In other words, it is possible that language-background-based differences may indeed be present, but were obscured because of the high performance by all groups. This possibility could be tested in future work, using a modified discrimination paradigm in which the target contrast is less salient, or in which the task is more implicit, drawing less attention to the target contrast.

Our primary finding was the surprising lack of group differences in English, particularly given that there were strong reasons based on empirical work to expect cross-language influence in this particular domain (e.g., Chang 2016; Kang et al. 2016). Contrary to our predictions, all groups performed similarly in English: Spanish speakers did not show the expected advantage. Our results were more in line with those of Kwon (2019, 2021), where Korean speakers showed convergence to f0 differences in stops in Korean, but not in English. Although the methods differ substantially between the two studies, both demonstrate language-specificity in phonetic imitation or convergence, adding to the body of work showing language-specific strategies in speech perception tasks more generally (e.g., Casillas and Simonet 2018; Dmitrieva 2019; Gonzales and Lotto 2013; Schertz et al. 2020).

In contrast to Chang (2016), we did not find evidence for a benefit of early exposure on sensitivity to an acoustic cue with greater importance in the heritage language. While there are many differences between the two studies that could potentially be the cause of this difference, we think that the most plausible explanation lies in nature of the specific cues that were the focus of the two studies. In Chang's work, the cue in question was coarticulatory information to word-final stop place of articulation, which is always still useful in English, and sometimes necessary (in cases where there is no stop release burst). On the other hand, the cue in the current work, presence/absence of prevoicing, is never necessary to distinguish the English laryngeal contrast, since presence/absence of aspiration can always be relied on. Relatedly, producing prevoicing is never necessary in English, whereas coarticulatory cues are always produced. Therefore, the extent to which the "heritage language benefit" (i.e., enhanced sensitivity) is present may depend on the role of this cue in the dominant language. The findings of Kwon (2021) showing that convergence to f0 did not occur in Korean speakers' performance on an English task (whereas it had been shown to occur in Korean tasks) are also consistent with this interpretation, since f0, like prevoicing, is only a very weak cue to the laryngeal contrast in English (e.g., Schertz et al. 2020).

As with any case study, caution must be used in generalizing, and as with any null result, there are multiple possible reasons for a lack of group differences. For example, our data showed a large amount of individual variability, so it is possible that the expected language background-based effect could have been masked by the noise in the data, or lack of power more generally. However, descriptive statistics (means and distributions) do not show any sign of systematic differences across groups, so we do not think the lack of differences is likely to be attributable to low power—or if it is, the effect size must be very small. Another possible reason could be because of participants' dominance in English (and less in their heritage language); however, given that our participants' language background and dominance were quite similar to those in previous work which did show evidence of cross-language influence (e.g., Chang 2016; Kang et al. 2016), and that there were clear language-based differences in baseline productions in the expected direction, this explanation does not seem particularly likely. Therefore, although there is always a possibility that the expected effect exists and we simply failed to find it, the expectations for difference were strongly grounded in previous empirical work and widely-held assump-

tions about the importance of prevoicing in true voicing languages, and the fact that the expected difference was not found points to the need to check these assumptions.

4.3. Sensitivity to Prevoicing: An Understudied Domain

Although stop voicing is probably the most well-studied phonological contrast, there is a surprising dearth of studies directly testing sensitivity to the presence vs. absence of prevoicing. Instead, most evidence that exists needs to be extrapolated from work looking at identification or discrimination of VOT continua. Therefore, our results both contribute to the empirical research in this domain and raise questions for future work.

First, we found a wide range of variability of both discrimination and imitation of English prevoicing. As expected, given that it is necessary to perceive a distinction in order to imitate it, individual performance across the two tasks was correlated. However, we also expected that imitation might be correlated with baseline voicing levels. We did find some evidence for this, in that for Cantonese speakers, there was a correlation between baseline voicing and imitation. However, this was only weakly predictive—there were several Cantonese speakers who showed no prevoicing in baseline but showed faithful imitation—and the correlation did not hold for the other two groups, despite the wide range of variability in those groups. Almost half of the monolingual speakers showed no prevoicing in baseline, so if this was a robust effect, we would have expected to see it in that group as well. Determining the sources of variability in sensitivity to prevoicing in English is therefore an important topic for future work.

Second, our hypothesis for higher performance by Spanish speakers was based on the assumption that prevoicing is the primary cue to the laryngeal contrast in Spanish, given that it is consistently present in the production of Spanish voiced stops. However, direct tests of sensitivity to prevoicing in true voicing languages are surprisingly rare, and a close look at the results that exist show that the primacy of prevoicing as a cue might not be as overwhelming as often assumed. Williams (1977b) examined monolingual Spanish listeners' perception of initial stops differing minimally in the presence/absence of prevoicing (where prevoicing was either naturally present or artificially removed, as in our artificial VoicingAbsent condition) via a forced-choice identification task, asking listeners whether each token was phonologically voiced (/b/), voiceless (/p/), or 'other.' There were significantly more voiced responses for tokens where prevoicing was present (94%) than when it was (artificially) absent (52%), leading to the conclusion that prevoicing was a "sufficient" cue to voiced stop perception. However, the fact that 52% of tokens without prevoicing were still perceived as voiced suggests that the presence vs. absence of prevoicing alone does not categorically account for listeners' judgments. As discussed by Williams (1977b), other cues therefore also likely play a role (see also van Alphen and Smits 2004). Similarly, if the presence/absence of prevoicing is an unambiguous cue to category membership, it would be expected that it would be easy to discriminate stimuli with even a small amount of prevoicing from those without. Results from monolingual Spanish listeners' discrimination of a VOT continuum (Williams 1974) showed that peak accuracy occurred around the 0 VOT point, as expected; however, discrimination accuracy was only around 60%. Therefore, the assumption of high sensitivity to prevoicing for true voicing language listeners may be too strong, and there is a need for more direct tests of the relative sensitivity to prevoicing in monolingual speakers from L1s which differ in the contrastive use of prevoicing.

4.4. Conclusions

This study found some support for the hypothesis that the contrastive status of prevoicing in a heritage language increases imitative performance in other languages: Spanish heritage speakers showed more faithful imitation of minimal differences in prevoicing in Hindi than Cantonese heritage speakers or monolingual English speakers. Crucially, however, there were no group-based differences in performance in English. This shows that imitative performance can differ based on language mode, consistent with language-

mode-based differences found in other perceptual tasks in previous work (e.g., Gonzales and Lotto 2013). The group-based differences were small, and sensitivity to and imitation of prevoicing appears to be highly variable, both among monolingual English speakers and among speakers with heritage languages varying in the status of prevoicing. Despite the strong expectation that heritage Spanish would facilitate sensitivity, language background accounted for surprisingly little of the variability in imitation and discrimination. Both the group results and the patterns of individual variability point to the need for more direct tests of factors conditioning variability in perception and imitation of prevoicing specifically, and of features in free variation more generally.

Author Contributions: Conceptualization, E.J.C. and J.S.; methodology and implementation, E.J.C. and J.S.; analysis, E.J.C. and J.S; writing—original draft preparation, J.S.; writing—review and editing, E.J.C. and J.S.; funding acquisition, J.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Natural Sciences and Engineering Research Council of Canada.

Institutional Review Board Statement: This research was approved by the Research Ethics Board of the University of Toronto.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: Data, stimuli, and code for analysis are publicly available at https://osf.io/c37rg/?view_only=6390be6fc264424c94b6a8aa40b19a7d (accessed on 14 November 2022).

Acknowledgments: We would like to thank Crystal Chow for help with the creation of experiment materials.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Table A1. Stimuli for baseline word reading task. For Cantonese, only simplified characters are shown here, but participants had a choice to see simplified or traditional characters.

English		Spanish		Cantonese	
/bɑt/	bot	/base/	*base* 'base'	/ba1hɐn4/	疤痕 'scar'
/bitʃ/	beach	/bisi/	*bici* 'bike'	/bin3fa3/	变化 'change'
/bum/	boom	/buska/	*busca* 'search'	/bun1ʊk1/	搬屋 'to move house'
/dɑt/	dot	/dato/	*dato* 'date'	/da2tsi6/	打字 'to type'
/dip/	deep	/digo/	*digo* 'I say'	/din6si6/	电视 'TV'
/dun/	dune	/dutʃa/	*ducha* 'shower'	/dʊŋ1tin1/	冬天 'winter'
/gɑt/	got	/gafas/	*gafas* 'glasses'	/ga1tɪŋ4/	家庭 'family'
/gis/	geese	/gia/	*guía* 'guide'	/gin6hɔŋ1/	健康 'health'
/gun/	goon	/gusto/	*gusto* 'taste'	/gu1tsɛ1/	姑妈 'father's older sister'
/pɑt/	pot	/pasa/	*pasa* 'happens'	/pa4san1/	爬山 'to climb a mountain'
/pitʃ/	peach	/piso/	*piso* 'floor'	/pin1fʊk1/	蝙蝠 'bat (animal)'
/pul/	pool	/puso/	*puso* 'I put'	/pun4gwɐt1/	盆骨 'pelvis'
/tɑt/	tot	/tako/	*taco* 'taco'	/ta1mun4/	他们 'they/them'
/tiθ/	teeth	/tigɾe/	*tigre* 'tiger'	/tin1hei3/	天气 'weather'
/tun/	tune	/tuve/	*tuve* 'I had'	/tʊŋ4hɔk6/	同学 'classmate'
/kɑp/	cop	/kasa/	*casa* 'house'	/ka1tʊŋ1/	卡通 'cartoon'
/kip/	keep	/kinse/	*quince* 'fifteen'	/kit3hiu2/	揭晓 'to disclose/reveal'
/kul/	cool	/kuvo/	*cubo* 'cube'	/ku1ŋa4/	箍牙 'to get braces'
/baj/	buy				
/daj/	die				
/paj/	pie				
/taj/	tie				

Table A2. The 8 pairs of filler stimuli used in the main task, with each pair differing in the presence/absence of aspiration. Hindi coronal stops are all dental; the dental diacritic is omitted for readability. [p$^{=}$] and [t$^{=}$] refer to tokens where aspiration has been artificially removed from naturally aspirated tokens. Note that four of the individual items (English [paj, taj] and Hindi [pɑl, tɑl]) are also used in target pairs (see Table 3).

		English	Hindi
Place	**Manipulation**	**Aspiration Present ~ Absent**	**Aspiration Present ~ Absent**
labial	natural	[p^{h}aj] ~ [paj]	[p^{h}ɑl] ~ [pɑl]
	artificial	[p^{h}aj] ~ [p$^{=}$aj]	[p^{h}ɑl] ~ [p$^{=}$ɑl]
coronal	natural	[t^{h}aj] ~ [taj]	[t^{h}ɑl] ~ [tɑl]
	artificial	[t^{h}aj] ~ [t$^{=}$aj]	[t^{h}ɑl] ~ [t$^{=}$ɑl]

Appendix B

Table A3. Acoustic characteristics (VOT and voicing intensity) of baseline stimuli. All stimuli were normalized to mean 70 dB intensity. For sounds with prevoicing, includes the duration, maximum intensity, and difference between maximum intensity of voicing and maximum intensity of vowels. For all syllables, includes positive voice onset time (duration from beginning of burst to onset of voicing in the following vowel.

	English				Hindi		
Word	**VOT (ms)**	**Prevoicing Duration (ms)**	**Peak Intensity of Voicing (dB)**	**Word**	**VOT (ms)**	**Prevoicing Duration (ms)**	**Peak Intensity of Voicing (dB)**
[baj]	13	122	67.81	[bɑl]	5	106	70.66
[daj]	13	125	68.42	[dɑl]	8	122	70.04
[paj]	15	0	-	[pɑl]	7	0	-
[taj]	13	0	-	[dɑl]	11	0	-
[p^{h}aj]	71	0	-	[p^{h}ɑl]	131	0	-
[t^{h}aj]	47	0	-	[t^{h}ɑj]	127	0	-

Appendix C

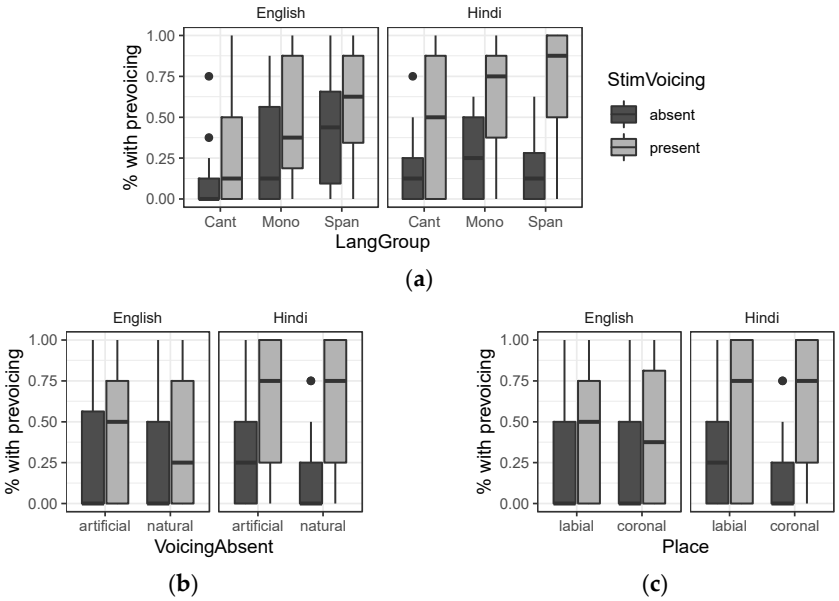

Figure A1. Distribution of by-participant mean percentage of voicing in stimuli where prevoicing was absent (dark grey) or present (light grey), broken down by Stimulus Language (left vs. right panels) and (**a**) Language Group, (**b**) VoicingAbsent Type, and (**c**) Place of Articulation.

Notes

[1] While not always the case, the language of the wider community is often the dominant language of heritage speakers, and this is the case in the participants in our study. For simplicity, in this work we use "dominant language" to refer to the language of the wider community, and "monolinguals" to refer to non-heritage native speakers of this wider community language.

[2] We chose Hindi for the foreign language because the stop laryngeal contrast includes both prevoicing and aspiration, resulting in a four-way phonemic distinction between [b], [b^h], [p], and [p^h] (Hussain 2018). As detailed in the Methods section, our stimulus manipulations required natural baseline stimuli including both prevoicing and aspiration.

[3] We chose to use percent accuracy for its transparency and for consistency with comparable work (Nielsen and Scarborough 2015), and because we the ABX task we used eliminates much of the potential response bias present in AX discrimination tasks, reducing the need for a measure like d-prime that corrects for this bias.

[4] We chose to use by-participant indices (i.e., an aggregate measure) as our response variable, instead of the raw data, for two reasons. First, it allowed for a more direct comparison between participants' perception and production, and because the models using raw data with the appropriate random effects structures failed to converge, indicating that more data is likely needed to support such models.

[5] The difference between Cantonese and Monolingual groups, not tested directly in this model, was also non-significant: (estimate = 0.03, SE = 0.071, t = 0.43, $p = 0.904$), as calculated by using the *emmeans* package in R (Lenth 2022).

[6] One possibility was that the better imitative performance by Spanish speakers in Hindi was driven by naturally produced trials; in order to test this, we compared an additional model that included the three-way interaction between Language Group, VoicingAbsent Type, and StimulusLanguage. None of the interactions involving Language Group and VoicingAbsent type were significant, and this model was not significantly different than the one used in our analysis, indicating that the effect was not driven by the natural stimuli.

[7] The difference between Cantonese and Monolingual groups, not tested directly in this model, was also non-significant: (estimate = 0.02, SE = 0.036, t = 0.555, $p = 0.844$), as calculated by using the *emmeans* package in R (Lenth 2022).

References

Antoniou, Mark, Catherine T. Best, Michael D. Tyler, and Christian Kroos. 2010. Language context elicits native-like stop voicing in early bilinguals' productions in both L1 and L2. *Journal of Phonetics* 38: 640–53. [CrossRef]

Antoniou, Mark, Michael D. Tyler, and Catherine T. Best. 2012. Two ways to listen: Do L2-dominant bilinguals perceive stop voicing according to language mode? *Journal of Phonetics* 40: 582–94. [CrossRef] [PubMed]

Anwyl-Irvine, Alexander L., Jessica Massonnié, Adam Flitton, Natasha Kirkham, and Jo K. Evershed. 2020. Gorilla in our midst: An online behavioral experiment builder. *Behavior Research Methods* 52: 388–407. [CrossRef] [PubMed]

Baese-Berk, Melissa, and Matthew Goldrick. 2009. Mechanisms of interaction in speech production. *Language and Cognitive Processes* 24: 527–54. [CrossRef]

Balukas, Colleen, and Christian Koops. 2015. Spanish-English bilingual voice onset time in spontaneous code-switching. *International Journal of Bilingualism* 19: 423–43. [CrossRef]

Bates, Douglas, Martin Maechler, Ben Bolker, and Steve Walker. 2015. Lme4: Linear Mixed-Effects Models Using Eigen and S4. Available online: https://cran.r-project.org/web/packages/lme4/index.html (accessed on 14 November 2022).

Benkí, José R. 2005. Perception of VOT and first formant onset by Spanish and English speakers. In *Proceedings of the 4th International Symposium on Bilingualism*. Edited by James Cohen, Kara McAlister, Kellie Rolstad and Jeff MacSwan. Somerville: Cascadilla Press, pp. 240–48.

Birdsong, David, Libby M. Gertken, and Mark Amengual. 2012. *Bilingual Language Profile: An Easy-to-Use Instrument to Assess Bilingualism*. Austin: COERLL, University of Texas at Austin. Available online: https://sites.la.utexas.edu/bilingual/ (accessed on 14 November 2022).

Caramazza, A., G. Yeni-Komshian, and E. B. Zurif. 1974. Bilingual switching: The phonological level. *Canadian Journal of Psychology/Revue Canadienne de Psychologie* 28: 310–18. [CrossRef]

Casillas, Joseph V., and Miquel Simonet. 2018. Perceptual categorization and bilingual language modes: Assessing the double phonemic boundary in early and late bilinguals. *Journal of Phonetics* 71: 51–64. [CrossRef]

Catford, John Cunnison. 2001. *A Practical Introduction to Phonetics*, 2nd ed. Oxford: Oxford University Press.

Chang, Charles B. 2016. Bilingual perceptual benefits of experience with a heritage language. *Bilingualism: Language and Cognition* 19: 791–809. [CrossRef]

Chang, Charles B. 2018. Perceptual attention as the locus of transfer to nonnative speech perception. *Journal of Phonetics* 68: 85–102. [CrossRef]

Chang, Charles B. 2021. Phonetics and phonology of heritage languages. In *The Cambridge Handbook of Heritage Languages and Linguistics*. Edited by Silvina Montrul and Maria Polinsky. Cambridge: Cambridge University Press, pp. 581–612. [CrossRef]

Cho, Taehong, and Peter Ladefoged. 1999. Variation and universals in VOT: Evidence from 18 languages. *Journal of phonetics* 27: 207–29. [CrossRef]

De Rosario-Martínez, Helios. 2015. "Package 'Phia'". Available online: https://CRAN.R-project.org/package=phia (accessed on 14 November 2022).
Dmitrieva, Olga. 2019. Transferring perceptual cue-weighting from second language into first language: Cues to voicing in Russian speakers of English. *Journal of Phonetics* 73: 128–43. [CrossRef]
Dmitrieva, Olga, Fernando Llanos, Amanda A. Shultz, and Alexander L. Francis. 2015. Phonological status, not voice onset time, determines the acoustic realization of onset f0 as a secondary voicing cue in Spanish and English. *Journal of Phonetics* 49: 77–95. [CrossRef]
Drager, Katie. 2010. Sociophonetic variation in speech perception. *Language and Linguistics Compass* 4: 473–80. [CrossRef]
Dufour, Sophie, and Noël Nguyen. 2013. How much imitation is there in a shadowing task? *Frontiers in Psychology* 4: 346. [CrossRef] [PubMed]
Flege, James Emil, and Wieke Eefting. 1988. Imitation of a VOT continuum by native speakers of English and Spanish: Evidence for phonetic category formation. *The Journal of the Acoustical Society of America* 83: 729–40. [CrossRef]
Gonzales, Kalim, and Andrew J. Lotto. 2013. A *Bafri*, un *Pafri*. *Psychological Science* 24. [CrossRef] [PubMed]
Herd, Wendy. 2020. Sociophonetic voice onset time variation in Mississippi English. *The Journal of the Acoustical Society of America* 147: 596. [CrossRef]
Hussain, Qandeel. 2018. A typological study of Voice Onset Time (VOT) in Indo-Iranian languages. *Journal of Phonetics* 71: 284–305. [CrossRef]
Kang, Yoonjung, Sneha George, and Rachel Soo. 2016. Cross-language Influence in the Stop Voicing Contrast in Heritage Tagalog. *Heritage Language Journal*. [CrossRef]
Kharlamov, Viktor. 2022. Phonetic effects in the perception of VOT in a prevoicing language. *Brain Sciences* 12: 427. [CrossRef]
Kuznetsova, Alexandra, Per B. Brockhoff, and Rune H. B. Christensen. 2017. lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software* 82: 1–26. [CrossRef]
Kwon, Harim. 2019. The role of native phonology in spontaneous imitation: Evidence from Seoul Korean. *Laboratory Phonology: Journal of the Association for Laboratory Phonology* 10: 10. [CrossRef]
Kwon, Harim. 2021. A non-contrastive cue in spontaneous imitation: Comparing mono- and bilingual imitators. *Journal of Phonetics* 88: 101083. [CrossRef]
Ladefoged, Peter, and Keith Johnson. 2014. *A Course in Phonetics*, 7th ed. Belmont: Wadsworth Publishing. ISBN 1285463404.
Lenth, Russell V. 2022. Emmeans: Estimated Marginal Means, aka Least-Squares Means, Version 1.8.1-1. Available online: https://CRAN.R-project.org/package=emmeans (accessed on 14 November 2022).
Lisker, Leigh, and Arthur S. Abramson. 1964. A Cross-Language Study of Voicing in Initial Stops: Acoustical Measurements. *WORD* 20: 384–422. [CrossRef]
Nagy, Naomi. 2015. A sociolinguistic view of null subjects and VOT in Toronto heritage languages. *Lingua. International Review of General Linguistics. Revue Internationale de Linguistique Generale* 164: 309–27. [CrossRef]
Newlin-Łukowicz, Luiza. 2014. From interference to transfer in language contact: Variation in voice onset time. *Language Variation and Change* 26: 359–85. [CrossRef]
Nielsen, Kuniko Y., and Rebecca Scarborough. 2015. Perceptual asymmetry between greater and lesser vowel nasality and VOT. Paper presented at ICPhS, Glasgow, UK, August 10–14.
Oh, Janet S., Terry Kit-Fong Au, and Sun-Ah Jun. 2010. Early childhood language memory in the speech perception of international adoptees. *Journal of Child Language* 37: 1123–32. [CrossRef]
Ohala, Manjari. 1994. Hindi. *Journal of the International Phonetic Association* 24: 35–38. [CrossRef]
Olmstead, Annie J., Navin Viswanathan, M. Pilar Aivar, and Sarath Manuel. 2013. Comparison of native and non-native phone imitation by English and Spanish speakers. *Frontiers in Psychology* 4: 475. [CrossRef]
Pallier, Christophe, Stanislas Dehaene, J.-B. Poline, Denis LeBihan, A.-M. Argenti, Emmanuel Dupoux, and Jacques Mehler. 2003. Brain imaging of language plasticity in adopted adults: Can a second language replace the first? *Cerebral Cortex* 13: 155–61. [CrossRef]
Podlipský, Václav Jonás, and Sárka Šimáčková. 2015. Phonetic imitation is not conditioned by preservation of phonological contrast but by perceptual salience. Paper presented at ICPhS, Glasgow, UK, August 10–14.
Schertz, Jessamyn, and Sarah Khan. 2020. Acoustic cues in production and perception of the four-way stop laryngeal contrast in Hindi and Urdu. *Journal of Phonetics* 81: 100979. [CrossRef]
Schertz, Jessamyn, Kathy Carbonell, and Andrew J. Lotto. 2020. Language Specificity in Phonetic Cue Weighting: Monolingual and Bilingual Perception of the Stop Voicing Contrast in English and Spanish. *Phonetica* 77: 186–208. [CrossRef] [PubMed]
Schwartz, Geoffrey. 2022. Asymmetrical cross-language phonetic interaction. *Linguistic Approaches to Bilingualism* 12: 103–32. [CrossRef]
Tees, Richard C., and Janet F. Werker. 1984. Perceptual flexibility: Maintenance or recovery of the ability to discriminate non-native speech sounds. *Canadian Journal of Psychology* 38: 579–90. [CrossRef]
van Alphen, Petra M., and Roel Smits. 2004. Acoustical and perceptual analysis of the voicing distinction in Dutch initial plosives: The role of prevoicing. *Journal of Phonetics* 32: 455–91. [CrossRef]
Walker, Abby. 2020. Voiced stops in the command performance of Southern US English. *The Journal of the Acoustical Society of America* 147: 606. [CrossRef] [PubMed]

Williams, Lee. 1974. Speech Perception and Production as a Function of Exposure to a Second Language. Ph.D. thesis, Harvard University, Cambridge, MA, USA.
Williams, Lee. 1977a. The perception of stop consonant voicing by Spanish-English bilinguals. *Perception & Psychophysics* 21: 289–97. [CrossRef]
Williams, Lee. 1977b. The voicing contrast in Spanish. *Journal of Phonetics* 5: 169–84. [CrossRef]
Woods, Kevin J. P., Max H. Siegel, James Traer, and Josh H. McDermott. 2017. Headphone screening to facilitate web-based auditory experiments. *Attention, Perception & Psychophysics* 79: 2064–72. [CrossRef]