Research Article

VOICE ONSET TIME IN MULTILINGUAL SPEAKERS: ITALIAN HERITAGE SPEAKERS IN GERMANY WITH L3 ENGLISH

*Miriam Geiss**

University of Konstanz

Sonja Gumbsheimer

University of Konstanz

Anika Lloyd-Smith

University of Konstanz

Svenja Schmid

University of Konstanz

Tanja Kupisch

University of Konstanz UiT the Arctic University of Norway

Abstract

This study brings together two previously largely independent fields of multilingual language acquisition: heritage language and third language (L3) acquisition. We investigate the production of fortis and lenis stops in semi-naturalistic speech in the three languages of 20 heritage speakers (HSs) of Italian with German as a majority language and English as L3. The study aims to identify the extent to which the HSs produce distinct values across all three languages, or whether crosslinguistic influence (CLI) occurs. To this end, we compare the HSs' voice onset time (VOT) values with those of L2 English speakers from Italy and Germany. The language triad exhibits overlapping and distinct VOT realizations, making VOT a potentially vulnerable category. Results indicate CLI from German into Italian, although a systemic difference is maintained. When speaking English, the

HSs show an advantage over the Italian L2 control group, with less prevoicing and longer fortis stops, indicating a specific bilingual advantage.

INTRODUCTION

In the assessment of crosslinguistic influence (CLI)[1] in populations of multilingual speakers, most studies to date have concentrated on the effects of CLI in one language only. Depending on the researcher's interest, this is either the heritage language (HL), the majority language (ML), or a foreign language. Recently, the call has been made to shift the focus from studies in which the target language is investigated in isolation from the other languages in a speaker's repertoire, and toward studies that investigate the acquisition of the phenomenon of interest in *all* the speakers' languages (Rothman et al., 2019). This is because, quite logically, a phenomenon cannot be transferred into another language if it has not been (fully) acquired. A further point has been to include more diverse learner populations (e.g., Rothman et al., 2019). For example, third language (L3) acquisition research has to date mainly been concerned with L3 acquisition in consecutive language learners who grew up monolingually, and for whom the L3 is the second foreign language. Less frequently studied is the population of heritage speakers (HSs), who grow up with two languages in early childhood, and for whom the L3 is the first foreign language. Yet, HS L3 acquirers provide an interesting case, because they have two early-acquired languages to draw from, unlike consecutive learners, who grew up monolingually.

In response to these gaps in research, we investigate patterns of phonetic-phonological CLI in the three languages of 20 HSs of Italian with German as ML and English as the third chronological language, and compare them with speakers who have acquired only one language during early childhood. The main goals are to find out how these speakers produce voice onset time (VOT) in their three languages, and to shed light on how the early-acquired languages of HSs interact with the acquisition of an L3. To this end, patterns of CLI are assessed in the production of fortis and lenis stops in all three languages and by comparing the HSs to monolingual[2] and L2 control groups in each language. In contrast to previous VOT studies, which have focused exclusively on fortis stops, and often use word reading lists or picture naming tasks (e.g., Gabriel et al., 2018; Llama & López-Morelos, 2016, 2020), we examine the production of both fortis *and* lenis stops. Our study is based on semi-naturalistic speech, which is deemed as ecologically more valid.

Although several L3 models have been proposed to account for morphosyntactic transfer (see, e.g., Puig-Mayenco et al., 2020, for an overview), we still know little about the processes that drive CLI in the phonological domain (see, e.g., Cabrelli & Pichan, 2021; Kopečková, 2016). This is even truer for HSs, who have so far only seldom been the focus of L3 phonology research. Having acquired two languages in early childhood—before any assumed critical period—means that HSs have two native languages to draw from, which may inform our understanding of L3 processes. Yet despite exposure to the HL and the ML from early childhood, monolingual-like phonological acquisition cannot be taken for granted in the two languages of early bilinguals. This is because phonological CLI may occur (i) bidirectionally in early bilinguals (e.g., Kehoe, 2015; Kupisch, 2019)

and (ii) regressively in L3 learners (Cabrelli Amaro, 2013). We also know that the accents of HSs are frequently perceived to sound different from those of monolingual speakers of the HL (e.g., Kupisch et al., 2014; Lloyd-Smith et al., 2020), and the same has even been shown for the ML in certain populations (Kupisch et al., 2020). Thus, the importance of investigating the phonologies of all three languages seems paramount to understanding and explaining patterns of CLI into the L3.

The paper is structured as follows. The background section provides an overview of VOT patterns in Italian, German, and English, and discusses previous research on VOT in multilingual constellations. The method and results sections present the analyses from the VOT studies in the early-acquired languages and in L3 English, respectively. We end with a discussion of results, and a brief conclusion.

BACKGROUND

VOT IN ITALIAN, GERMAN, AND ENGLISH

VOT is considered to be the most salient cue that differentiates the language-specific realizations of lenis (/b, d, g/) and fortis (/p, t, k/) stops. It refers to the interval between the release of the stop and the beginning of vocal cord vibrations (Lisker & Abramson, 1964). The phonological categories of fortis and lenis can be realized as different phonetic categories, that is, different types of VOT. According to Lisker and Abramson, there exist three types of VOT: (i) voicing lead or prevoicing (voicing starts before the release; < 0 ms), (ii) short-lag VOT (voicing begins with the release or shortly after it; 0–35 ms), and (iii) long-lag VOT (voicing starts late after the release; > 35 ms). The three different patterns are displayed in Figure 1, which summarizes characteristics of the stop consonants and their VOT patterns in the three languages investigated in this study. The values used in Figure 1 are only approximations and are compromised by the methodology and by the data type.

Italian is considered to be a voicing language, where prevoicing with negative VOTs characterizes lenis stops, and fortis stops display short-lag (VOT values up to 30 ms) (see Bortolini et al., 1995; Kupisch & Lleó, 2017). German, by contrast, is considered to be an aspirating language. Phonologically voiced stops are said to be produced with short-lag, whereas phonologically voiceless stops are produced with aspiration and a longer VOT (long-lag) (Fischer-Jørgensen, 1976; Haag, 1979; Neuhauser, 2011; Stock, 1971). English is classified as an aspirating language, which is generally said to display the same VOT patterns as German (see, e.g., Lisker & Abramson, 1967; Keating et al., 1981,
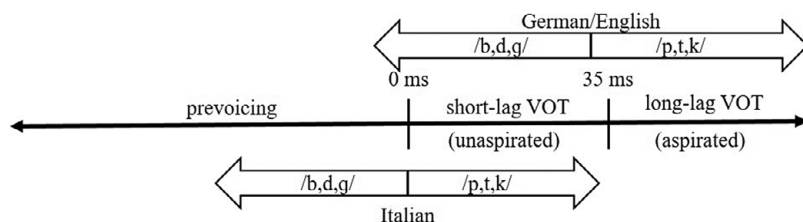

FIGURE 1.   Comparison of stop categories in Italian, German, and English.

for VOT in English stops). Thus, German and English fortis stops have longer VOTs than Italian fortis stops.[3] However, the distinction between the languages is somewhat less clear with regard to lenis stops, because some studies have also reported instances of prevoicing for English (Docherty et al., 2011; Lisker & Abramson, 1964) and for German (e.g., Hamann & Seinhorst, 2016; Stock, 1971; Stoehr et al., 2017), suggesting that common assumptions about German and English VOT patterns need to be treated with caution. If it is correct that German and English also display prevoicing in some contexts, then this leads to more (partial) overlap between the patterns, which may in turn induce more CLI (see Kehoe, 2015, for discussion).

Findings on VOT values reported in the literature differ due to several factors, such as place of articulation (PoA; Ladefoged & Maddieson, 1996), position of the stop in the syllable (Lisker & Abramson, 1964), type of data (e.g., read speech vs. naturalistic speech), vocalic contexts (Lein et al., 2016), and speech rate (Miller et al., 1986). Therefore, we consider it problematic to take values from the literature as a point of comparison and provide control data from monolingual speakers who did the same experiment as the HSs. These control data will be important for the first half of our study, which examines HL acquisition. The varieties of German relevant in this study are Southern German varieties, which are known to have lower VOT values for all stop consonants compared with Northern Standard German (see Braun, 1996, for an overview of VOT patterns in German varieties).

VOT IN EARLY BILINGUAL DEVELOPMENT

VOT in early bilingual children and early bilingual adults is relatively well-studied in language combinations that display different VOT patterns, because predictions for language (non-) separation and CLI are straightforward. For example, as outlined above, the VOT patterns of the Romance and the Germanic language families (often) differ in that the former are voicing languages and the latter aspirating languages, which means that CLI can be verified by means of VOT production. In the following review, we make reference to studies that involve German and Italian whenever possible but we also include language pairs that have comparable VOT patterns.

In monolingual language development, the contrast between short-lag and long-lag VOT is acquired relatively early, around 2;0–2;6 (Davis, 1995; Kehoe et al., 2004; Macken & Barton, 1979). By contrast, the distinction between prevoicing and short-lag VOT is acquired comparatively late, after age 4, due to more complex motor activities needed to coordinate the laryngeal closure and the vocal fold vibrations for prevoicing (see Allen, 1985, for French; Bortolini et al., 1995, for Italian; Macken & Barton, 1980, for Spanish). Stoehr et al. (2018) showed that monolingual Dutch children do not prevoice lenis plosives consistently up until the age of 6. Differences in the acquisition process are consistent with degrees of markedness (see, e.g., Davis, 1995; Kehoe et al., 2004).

Studies on early bilingual development have shown that bilingual children distinguish fortis and lenis stops in their two languages from early on, but there may be delays due to CLI. For example, Kehoe et al. (2004) studied four simultaneous German-Spanish bilinguals (aged 2;0–3;0), who all grew up in Germany. In German, two of the children behaved in a target-like manner[4] and produced fortis stops with long-lag VOT, while the other two produced short-lag VOT, which can be interpreted as a delay in the acquisition

of long-lag VOT, possibly due to CLI from Spanish. In Spanish, none of the four children produced lenis stops with prevoicing, which indicates CLI, or general difficulties in the acquisition of prevoicing, which are also found with monolinguals (see Deuchar & Clark, 1996, for a similar case). In Fabiano-Smith and Bunta's (2012) study of Spanish-English simultaneous bilingual children in the United States (aged 3;0–4;0), the production of /p/ and /k/ in Spanish did not differ from Spanish monolinguals, but English productions of /k/ were comparably short. Again, two interpretations are possible: CLI from Spanish, or a delay in the acquisition of long-lag VOT, which is comparatively marked and, therefore, susceptible to delays independently of bilingualism. Stoehr et al. (2018) studied simultaneous Dutch-German bilingual children (ages 3;7–5;11) in the Netherlands and found bi-directional influence. The children produced lenis stops similarly in German and in Dutch, and differently from monolinguals in both languages. In their production of fortis stops, by contrast, the bilinguals showed a clear separation between Dutch and German, resulting in target-like short-lag VOT in Dutch and long-lag VOT in German. As the examples show, it is often difficult to tease apart CLI from late acquisition due to markedness, especially in the acquisition of prevoicing, which is also late acquired in monolinguals. One consistent finding, however, is that, if the speakers' languages have different VOT patterns, speakers will form separate categories, that is, their productions reflect language-specific patterns that approximate those of monolinguals in each of the two languages. This means that, in early bilingual children, no evidence of "fused systems," in early bilingual terminology, or "hybrid values," in second language acquisition terminology, has been provided. However, the studies on bilingual children leave open whether the VOT patterns will eventually be acquired in a target-like manner.

In addition to CLI, the heterogeneous nature of existing findings may be explained by diverse types of methodologies (see the HL Study section), varying conditions for multilingualism, intra-linguistic factors, or sociolinguistic variables. For example, the situation of French-English bilinguals in Canada is different from that of Italian bilinguals in Germany, because there are far more opportunities for using both languages in the former setting. Early bilinguals in the latter setting are likely to be more strongly dominant in the ML and, as a result, CLI has often been shown to occur uni-directionally from the ML to the HL, although there are some noticeable exceptions that have shown VOT values in the ML that differ from the monolingual baseline (e.g., Kupisch & Lleó, 2017; Mayr & Siddika, 2018). A further methodological aspect, related to linguistic factors, is the type of stops studied, with evidence suggesting that, when compared with monolinguals, differences are more likely in the production of lenis stops than in the production of fortis stops (Sundara et al., 2006; although see Fowler et al., 2008, for an exception). Nevertheless, studies have shown that HSs are able to develop different phonetic categories for the stops in their two languages, but these categories are not necessarily monolingual-like (e.g., Flege, 1991; Flege & Eefting, 1987). Finally, Hrycyna et al. (2011) and Nagy and Kochetov (2013) stress the importance of the HSs' attitudes and relations toward their HL. Among three groups of HSs (Ukrainian, Russian, Italian), only the Italian HSs were resilient to influence from English. A possible explanation for this difference is that the Italian community in Toronto receives a lot of institutional support, while the Russian HSs do not seem to feel a strong cultural need to maintain their HL.

Table 1 summarizes existing studies with early bilinguals during adulthood, indicating the sounds that have been studied, whether a difference was found between the languages

TABLE 1.   Overview of studies with adult early bilinguals

	Languages	Stops investigated	Significant difference between languages?	Significant difference to baseline?
Sundara et al. (2006)	Canadian French-English	/t/ /d/	Yes yes	French: no English: yes (only for lenis)
Fowler et al. (2008)	Canadian French-English	/p, t, k/	Yes	French: yes English: yes
Nagy and Kochetov (2013)	English-Italian English-Russian English-Ukrainian	/p, t, k/	n.a.	Italian: no Russian: yes Ukrainian: yes
Mayr and Siddika (2018)	English-Sylheti	/t̪ ʈ k, p t k gʱ bʱ, b d̪ ḍ g, b d g/	Yes Yes	English: yes Sylheti: no
Lein et al. (2016)	German-French	/k/	Yes	German: yes French: no
Kupisch and Lleó (2017)	German-Italian	/k/	Yes	German: yes Italian: yes
Flege and Eefting (1987)	English-Spanish	/p, t, k/	Yes	English: yes Spanish: yes
Flege (1991)	English-Spanish	/t/	Yes	n.a.

Note: The latter of the two languages indicates the HL, except for the studies conducted in Canada (because neither French nor English is a HL in this context).

and, finally, whether the bilinguals showed a difference to the (monolingual) baseline. Note that, if no comparison was made with monolinguals but across generations, we considered the first generation as baseline. All studies provide evidence in favor of language separation, but they differ in terms of whether or not there was a difference to the baseline.

L3 PHONOLOGY IN HSs

Studies examining L3 phonology in HSs have rendered quite mixed results, but several central trends may be identified. First, some studies on VOT acquisition have suggested dominance in the ML to be a driving factor, meaning that CLI from the HL tends to be negligible. For example, Llama and López-Morelos (2016) found that English-dominant Spanish HSs produced L3 Canadian French fortis stops in line with English, even though transferring from Spanish would have been more facilitative. Llama and López-Morelos (2020) confirmed this in a later study in which they investigated fortis stops in adolescent HSs of Spanish with English as ML and L3 French in a Canadian immersion context. In L3 French, the bilinguals transferred negatively from English, and were in line with English monolingual controls. The authors also examined the speakers' background languages, and found identical-to-target values in the ML English, and close-to-target values in the HL Spanish for /p/ and /k/, while the values for /t/ were slightly longer. Statistical analyses showed that, while they had created separate categories for their HL and their ML, their L3 production patterned with the ML. In the same vein, Gabriel et al. (2016) found no difference from German monolinguals in the perception and production

of L3 French fortis stops in HSs of Mandarin, who theoretically could have transferred shorter values from their HL. However, some evidence for the (co-)occurrence of CLI from the HL also exists, e.g., in HSs with a high degree of metalinguistic awareness (Gabriel & Rusca Ruths, 2015; Özaslan & Gabriel, 2019) or a high proficiency in the HL (Lloyd-Smith et al., 2017).

A second observation is that HSs may have a bilingual advantage in L3 phonology as compared with monolingual peers. In two studies by Dittmers et al. (2018) and Gabriel et al. (2018), German-dominant HSs of Turkish and Russian were shown to produce shorter, more target-like values for the fortis stops /p, t, k/ in L3 French when compared with German monolinguals, because fortis stops in Turkish and Russian are produced with short-lag VOT, whereas in German they are produced with long-lag. Advantages for HSs acquiring L3s have also been found for other phonological phenomena, including the production of rhotic sounds in L3 Spanish (Kopečková, 2016), speech rhythm in L3 French (Gabriel & Rusca Ruths, 2015), and word-final voiced obstruents in L3 French and English (Özaslan & Gabriel, 2019). Although these studies all used small samples and, therefore, do not allow for generalization, what they have in common is that they suggest that HSs can benefit from specific properties of their HL if there is overlap with the target property in the L3. However, these studies do not allow us to comment on whether there are any across-the-board or language general advantages for HSs acquiring L3 phonology.

Third, it is possible that HSs will form hybrid VOT values, or converged phonological systems. This was the case for two VOT studies by Wrembel (2014, 2015) that examined L3 learners of German and French from several different language backgrounds. In particular, two groups of L1 Polish-L2 German and L1 German-L2 English speakers produced VOT in L3 French with a slight overshoot, while L1 Polish-L2 English speakers produced VOT in L3 German with a slight undershoot, which in both cases was argued to reflect hybrid values from the background languages. Merged values across the three languages of child-aged early bilingual speakers of Pomeranian and Brazilian Portuguese acquiring English in the United States were also found by Tessmann Bandeira and Zimmer (2012).

One additional possibility is that phonological CLI occurs from the typologically closest language. Cabrelli and Pichan (2021) found evidence for transfer from the typologically closest language in the production of voiced intervocalic stops in L3 Brazilian Portuguese and in L3 Italian, which are realized as [–continuant] in English, Brazilian Portuguese, and Italian, but as [+continuant] in Spanish. Their results showed that the majority of participants produced Spanish-like [+continuant] stops, regardless of whether Spanish was acquired as an L1, as an L2, or as a HL. These results were interpreted by the authors as evidence for the Typological Primacy Model (Rothman, 2011, 2015).

In summary, the above research leaves open the question of how CLI will obtain in the three languages of the early bilinguals in this study. We therefore pose the following research questions (RQs):

RQ1 Do HSs differentiate between the ML (German) and HL (Italian) with regard to VOT values?
RQ2 Do they differ from monolinguals in Italian and German?

The answer to these questions will be crucial to the L3 study, because the two background languages serve as potential transfer sources. If there is CLI, the two transfer sources may not correspond to the patterns we find in German and Italian monolinguals. For the L3 acquisition study, we then ask:

RQ3 Do L3 VOT patterns in English differ from those of their two first languages (Italian or German)?

RQ4 Does the acquisition of two first languages aid the acquisition of an L3, that is, do HSs behave differently compared with L2 learners?

METHOD

Our study examines VOT production in three different languages: German, Italian, and English, acquired across four different contexts (L1, HL, L2, and L3). Accordingly, we divide the discussion of results into two sections, discussing first the acquisition of VOT in the early-acquired languages, followed by the discussion of English as a foreign language. To this end, we first address RQ1 and RQ2 by comparing the German-Italian bilingual HSs to the respective monolingual control groups; next, for the L3 study, we focus on VOT in L3 English, comparing HSs to L1 German and L1 Italian controls in English, as well as to L1 English controls (RQ3 and RQ4).

PARTICIPANTS

A total of 20 German-Italian HSs, 20 Italian monolinguals, and 20 German monolinguals participated in the HL study (see Table 2). All bilinguals grew up in South Germany and acquired Italian as an HL from birth. Seven bilinguals have one German- and one Italian-speaking parent (exposure to German from age 0), while 13 have two Italian-speaking parents (exposure to German between 2 and 6 years; $M = 2.7$). The HSs were exposed to different varieties of Italian. The Italian and German monolingual controls were exposed to the same regional varieties as the HSs. Proficiency in all three languages (Italian, German, and English) was measured using a Yes/No vocabulary task, which consisted of 50 real words (full verbs) and 25 pseudowords taken from the placement test for the DIALANG (Alderson, 2005, p. 80), and adapted for use in a self-directed experiment in Presentation® (see Lloyd-Smith et al., 2021, for details on the test and its scoring). The total score was 75 for this task. The results showed significantly higher scores for the ML German ($M = 70.75$, range = 64–74, $SD = 3.13$) than for the HL Italian, which also displayed a much larger range ($M = 57.85$, range 39–68, $SD = 8.18$, $F(1,38) = 43.36$, $p < .001$). In German, the HSs did not differ significantly from the monolinguals

TABLE 2. Participant profiles

	HL study			L3 study		
	HS	L1 German	L1 Italian	L1 English	L1 German	L1 Italian
N (female: male ratio)	20 (15: 5)	20 (17: 3)	20 (13: 7)	20 (16: 4)	20 (18: 2)	20 (15: 5)
Mean age (range)	23 (17–29)	26 (19–33)	25 (19–39)	37 (17–68)	25 (19–33)	26 (19–37)

($M = 71.2$, range $= 65$–75, $SD = 2.67$, $F(1,32) = 0.24$, $p = .63$), but they differed significantly from the Italian monolinguals in the Italian test ($M = 70.05$, range 59–74, $SD = 3.02$, $F(1,38) = 39.13$, $p < .001$). The larger range in the Italian test showed that some HSs were more balanced with almost equal proficiencies, while others were fairly unbalanced (as a group) with Italian as their weaker language. No participant scored higher in Italian than in German.

In the L3 study, the HSs were tested in English. English was the first foreign language for all speakers, and was first learned at school between 6 and 11 years of age.[5] Their current contact with English was limited to holidays, contact with (social) media, and through contact at university. None studied English as a subject, and none had spent more than 2 weeks in an English-speaking country. We compared the HSs with three control groups, including 20 L1 native English speakers (10 with Australian and New Zeeland English, five with American English, four with British English, and one with South African English; for VOT in varieties of English, see, footnote 4), with 20 L1 German-L2 English speakers, and with 20 L1 Italian-L2 English speakers (see Table 2). The reason for including the L2 control groups was to identify the relative influence of either German or Italian on the L3. English proficiency was evaluated for all groups using the English version of the Yes/No vocabulary test, which showed that all non-native groups were matched for proficiency. Out of a total of 75 points, the HSs attained a mean of 63.75 points in English (range: 44–74, $SD = 7.35$), the L1 English controls a mean of 73.6 points (range 66–75, SD = 2.23),[6] the L1 German controls a mean of 66.8 points (range 58–74, $SD = 4.72$), and the L1 Italian controls a mean of 67.65 points (range: 55–75, $SD = 4.46$). The HSs differed significantly from the English monolinguals ($F(1,38) = 32.84$, $p < .001$). However, we did observe a difference neither between the HSs and the L1 German ($F(1,38) = 2.44$, $p = .13$), nor between the HSs and the L1 Italian ($F(1,38) = 4.11$, $p = .05$).

MATERIALS AND PROCEDURE

The stops of interest were the fortis stops /p/, /k/ and the lenis ones /b/, /g/. The coronal stops /t/ and /d/ were not included because they have different PoAs in the three languages with potential effects on VOT duration (Lisker & Abramson, 1964). We selected stop-initial words (mostly nouns) that could be portrayed in simple pictures, controlling for the following vowel (/a/ or /i/), word length (mono- or disyllabic), and position in the syllable (initial position in stressed syllable). This resulted in a total of 32 target words; see Online Supplementary Material 1 for a full list of stimuli.

All participants were recruited in an academic context and tested at the University of Konstanz. They signed informed consent before taking part in the study.[7] We tested the bilingual participants in all three languages in three different sessions of approximately 45 min (in which they also completed the vocabulary test and a background questionnaire). To avoid language influence, the sessions were scheduled several days apart and were led by a native speaker of the target language. The experimental design was meant to elicit the target stops in semi-spontaneous speech.[8] The VOT data were elicited by means of a picture-cued storytelling task, where participants were asked to tell a story that contained the things or actions they saw on different PowerPoint slides. Before the experiment, the participants had to name the things and actions they saw on the slides

to ensure that they recognize the target items. In cases where the participants did not recognize the items, the experimenter provided them.

RECORDINGS AND MEASUREMENTS

The data were recorded with an Olympus Linear PCM Recorder LS-11 with uncompressed 24 bit / 96 kHz recording capability. Phonetically trained coders analyzed VOTs taking into account waveforms and spectrograms in Praat (Boersma & Weenink, 2015). In the analysis, all words, target words, or other words produced by the participants that fulfilled the above-mentioned criteria were included. We measured positive VOT as the period between the release of the closure (peak of the first visible burst) and the onset of voicing (peak of the first periodic wave) (Lisker & Abramson, 1964). In the case of lenis stops, we coded devoicing[9] for positive VOTs and prevoicing for negative VOTs (clear periodic waveform during closure) as a categorical variable.[10] We did not consider lenis stops with a preceding nasal because of coarticulation effects. Figure 2 shows measurements of short-lag, long-lag, and prevoiced VOT. All reported VOTs were cross-checked by at least one additional coder.[11] A total of 1.4% of all data points were excluded from the analysis due to hesitations, stutters, or distorted noise. Because Miller et al. (1986) show an effect of speaking rate on VOT, we also measured the participants' speech rate by counting the number of syllables per 30 s in a fluent part of the recording. A correlation test (Pearson's r), however, revealed no correlation between VOT and speech rate within the three languages ($r_{ge} = -.02$, $r_{it} = -.07$, $r_{en} = -.04$). Therefore, we did not include speech rate in further statistical analyses.

STATISTICAL ANALYSIS

The statistical analyses were based on mixed-effects regression models in R, using the package lme4 (Bates et al., 2015) and lmerTest (Kuznetsova et al., 2017) to obtain p-values. For the fortis stops /p/ and /k/, we defined linear mixed-effects regression models with VOT as dependent variable. In the analysis of the lenis stops /b/ and /g/, we followed the approach taken in Stoehr et al. (2018) and converted VOT into a categorical dependent variable with two levels: "prevoicing" for negative VOT and "devoicing" for positive VOT, which was entered in a logistic mixed-effects regression model.[12] We used different independent variables in the models: "Language Background" (HL study: HSs vs. monolinguals; L3 study: L1-E vs. HSs, L1-G, L1-I) was the independent variable of interest in the between-group analyses that compared the monolinguals and HSs. The variable "Language" (German vs. Italian vs. English) was used to compare the HSs' three languages using a within-group design, and to analyze the monolinguals of each language. The stop itself (PoA), the vowel following the stop (/a/ vs. /i/), the context preceding lenis stops (voiceless vs. voiced), and word length (number of syllables) were included as four additional independent variables to address potential variance in the data. "Participant" and "word" were added as random effects. For an overview of all model specifications, including interaction terms, fixed effects, random effects, and random slopes, see Supplementary Materials 2 and 3; the complete presentation of their effects, as well as the effect size (R^2) of each model can be found in Supplementary Material 4.
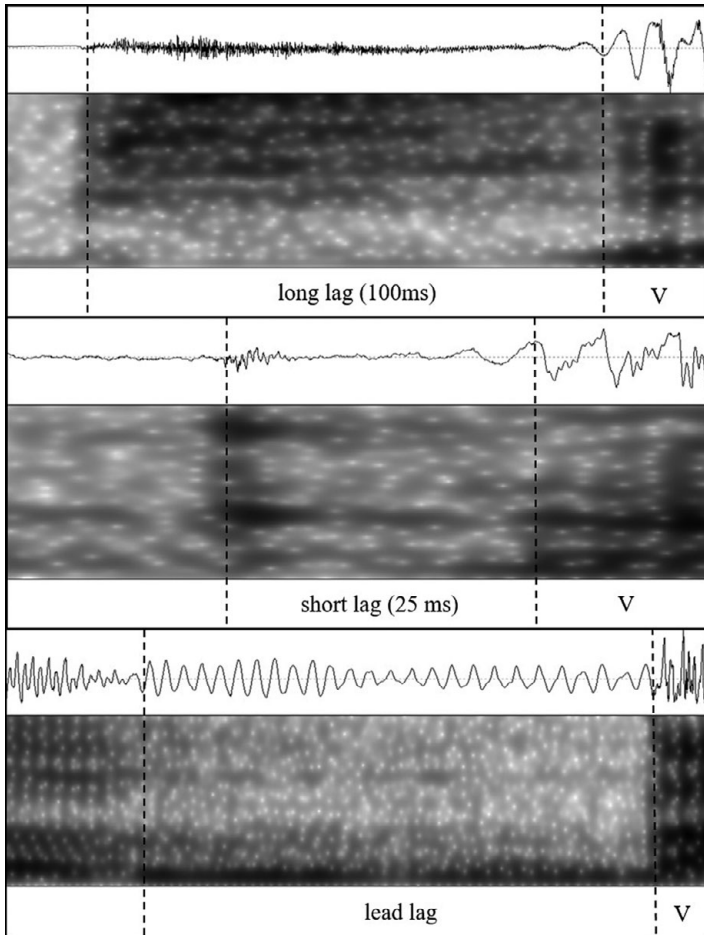
FIGURE 2.   Examples of measurements of short-lag, long-lag, and prevoicing VOT using Praat.

RESULTS

In this section, we first present the results for VOT in Italian and German, comparing HSs in their two languages and with L1 speakers of German and Italian, followed by the results for the L3 study. Each section begins with the descriptive statistics,[13] and then presents the statistical effects of Language Background and Language on VOT.

HL STUDY

For fortis stops, the results are summarized in Table 3 and Figure 3, showing mean VOTs, standard deviations (*SD*s), and total number (*N*) of fortis stops in each language for monolinguals and HSs. German monolinguals produced the longest VOTs and Italian monolinguals the shortest VOTs on average. The HSs' VOT values fell in between the two monolingual groups, and they produced higher VOTs in German than in Italian.

TABLE 3.   VOT of fortis stops by language in ms (mean value in ms, *SD*, and total number, *N*)

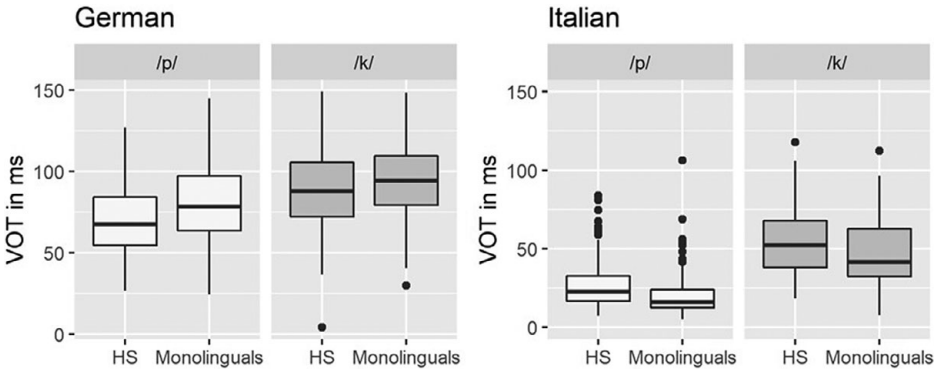|		German		Italian	
		HS	Monolinguals	HS	Monolinguals
/p/	*M* [ms]	**70.14**	**80.07**	**26.05**	**19.81**
	SD	22.61	23.84	14.13	11.99
	N	154	172	235	213
/k/	*M* [ms]	**88.86**	**94.69**	**54.81**	**48.04**
	SD	23.78	23.06	21.25	20.02
	N	200	232	169	188
Group total (*N*)		354	404	404	401

FIGURE 3.   VOT of fortis stops by language in ms.

The results for lenis stops (Table 4; Figure 4) showed that Italian monolinguals produced the highest percentage of prevoiced stops. German monolinguals produced the lowest percentage with only slightly lower percentages than the HS. The percentage of prevoicing in Italian was slightly lower for the HSs compared with the Italian monolingual controls. There was some interspeaker variation: one monolingual and two HSs prevoiced less than 50% of the time in Italian, and two monolinguals and five HSs prevoiced more than 25% of the time in German.

We ran four sets of mixed effects analyses (Table 5). The first analysis confirmed that German and Italian monolinguals produced VOT differently. As anticipated for fortis stops, the German monolinguals had significantly longer VOTs than the Italian monolinguals ($\beta = -44.81$, $SE = 9.00$, $t = -4.98$, $p < .001$). An interaction between language and vowel ($\beta = 14.94$, $SE = 4.91$, $t = 3.04$, $p < .01$) revealed, as expected, that Italians produced longer VOTs if the stop preceded the vowel /i/ than if it preceded /a/ ($\beta = 12.94$, $SE = 3.35$, $t = 3.86$, $p < .01$). We did not observe this effect for German ($\beta = -3.06$, $SE = 3.88$, $t = -0.79$, $p = .44$). An interaction between language and stop ($\beta = -10.46$, $SE = 5.13$, $t = -2.04$, $p < .05$) indicated that both groups produced shorter VOTs for /p/, as expected (German: $\beta = -15.88$, $SE = 4.90$, $t = -3.88$, $p < .01$; Italian: $\beta = -27.16$,

TABLE 4. Mean percentage (%), *SD*, and total number of prevoiced stops by language and language background

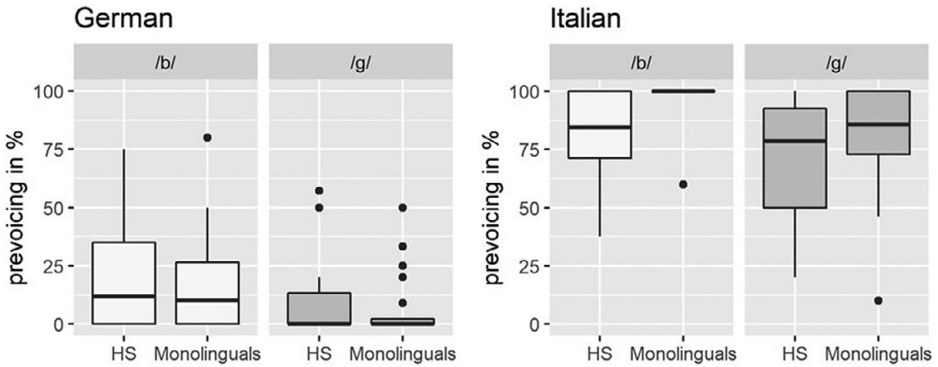|		German		Italian	
		HS	Monolinguals	HS	Monolinguals
/b/	*M* [%]	**20.15**	**17.32**	**79.82**	**98.08**
	SD	40.26	37.94	40.31	13.80
	N	134	179	114	104
/g/	*M* [%]	**10.17**	**6.03**	**70.45**	**76.43**
	SD	30.35	23.91	45.75	42.58
	N	118	116	176	157
Group total (*N*)		252	295	290	261
Range prevoiced		0–66.67%	0–50%	30.77–100%	26.67–100%

FIGURE 4. Percentage of prevoiced stops in German and Italian by HSs and monolinguals.

TABLE 5. Summarized statistical effects of language and language background on VOT

	German monolinguals vs. Italian monolinguals	RQ1: HS German vs. Italian	RQ2: HS vs. monolinguals (German and Italian)
fortis	Italian: shorter VOT ***	Italian: shorter VOT **	German: n.s. Italian: HS with longer VOTs*
lenis	German: lower percentage of prevoicing ***	German: lower percentage of prevoicing ***	German: n.s. Italian: lower percentage of prevoicing**

***p < .001.
**p < .01.
*p < .05.
n.s. p > .05.

SE = 3.46, *t* = −7.85, *p* < .001). For lenis stops, monolingual Germans produced fewer prevoiced stops than Italians (*β* = 7.62, *SE* = 1.22, *z* = 6.25, *p* < .001).

The following analysis concerns RQ1, that is, whether the HSs produced language specific VOT patterns. The results showed that the HSs' VOTs for fortis stops were

significantly longer in German than in Italian ($\beta = -26.70$, $SE = 7.51$, $t = -3.56$, $p < .01$). HSs produced a higher percentage of lenis stops with prevoicing in Italian than in German ($\beta = 3.67$, $SE = 0.61$, $z = 6.06$, $p < .001$).

The next analysis tests whether HSs behave like monolinguals in German and Italian, respectively (RQ2). For German, we did not observe a difference between HSs and the monolingual controls either for fortis stops ($\beta = 10.16$, $SE = 6.29$, $t = 1.61$, $p = .11$) or for lenis stops ($\beta = -0.22$, $SE = 0.54$, $z = -0.41$, $p = .68$). For Italian fortis stops, the HSs and the Italian monolinguals differed significantly, as HSs produced overall longer, that is, more German-like, VOTs ($\beta = -11.31$, $SE = 4.77$, $t = -2.37$, $p < .05$). For lenis stops, the HSs produced a lower percentage of prevoicing than the Italian monolinguals ($\beta = 3.12$, $SE = 1.01$, $z = 3.11$, $p < .01$). An interaction of language background and PoA ($\beta = -2.59$, $SE = 0.93$, $z = -2.80$, $p < .01$) indicated that monolinguals prevoiced /g/ less often than /b/ ($\beta = -3.30$, $SE = 0.90$, $z = -3.67$, $p < .001$), while there was no detectable difference for HSs ($\beta = -0.43$, $SE = 0.33$, $z = -1.31$, $p = .19$).

L3 STUDY

As Table 6 and Figure 5 illustrate, HSs produced slightly higher VOTs for fortis stops in their L3 English than the L1 English control group. These two groups fell between the L1 German speakers, who produced the longest, and the L1 Italians, who produced the shortest VOTs on average. Because both English and German are described as languages with long-lag VOT, the difference between the respective L1 speakers is somewhat unexpected. On the other hand, we are not aware of any previous study that has compared these two languages based on the same methodology.

Table 7 and Figure 6 show the results for lenis stops. English monolinguals produced the lowest percentage of lenis stops with prevoicing. L1 Italians produced by far the highest percentage of prevoiced stops, thus differing significantly from the other three groups. L1 Germans had the same amount of prevoicing as the HSs.

The results of the statistical analysis, including mixed effect regression models,[14] are summarized in Table 8. The first analysis compared monolingual VOT production in the three languages. English monolinguals produced significantly longer VOTs for fortis

TABLE 6.   VOT (ms) of English fortis stops by language background (mean value, *SD*, and total number)

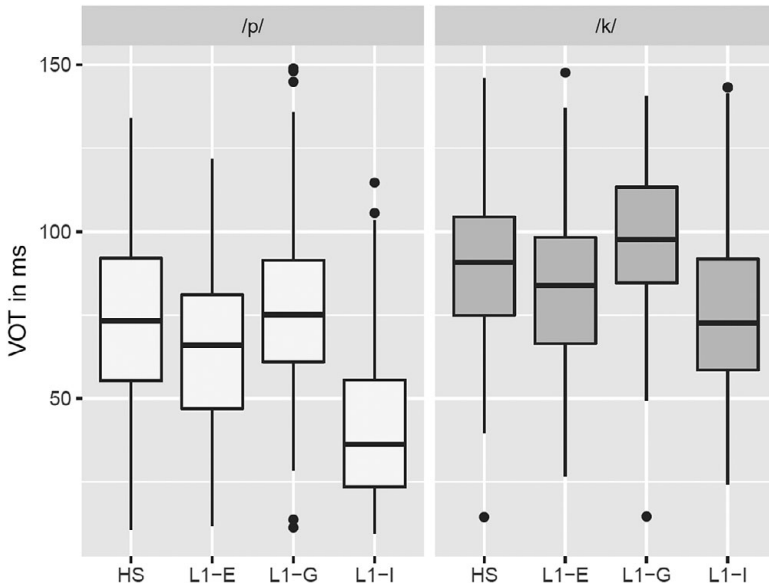|		English			
		HS	L1 English	L1 German	L1 Italian
/p/	*M* [ms]	73.62	65.60	78.02	41.08
	SD	25.79	24.41	24.30	24.61
	N	238	203	215	213
/k/	*M* [ms]	90.32	83.99	97.82	76.17
	SD	22.40	23.09	21.26	23.31
	N	236	196	199	206
Group total (*N*)		474	399	414	419

FIGURE 5. Mean VOT values (ms) of English fortis stops by language background.

TABLE 7. Mean percentage (%), *SD*, and total number of prevoiced stops in English by language background.

		English			
		HS	L1 English	L1 German	L1 Italian
/b/	*M* [%]	**16.51**	**6.98**	**15.00**	**70.52**
	SD	37.21	25.55	35.81	45.73
	N	212	172	180	173
/g/	*M* [%]	**10.38**	**9.80**	**13.21**	**51.75**
	SD	30.59	29.83	33.97	50.15
	N	183	153	174	143
Group total (*N*)		395	325	354	316
Range prevoiced		0–55.17%	0–40%	0–53.3%	4.4–100%

stops than Italian monolinguals ($\beta = -26.37$, $SE = 8.78$, $t = -3.00$, $p < .01$), as anticipated based on the literature. However, the results of the comparison between English and German monolinguals did not mirror those of the literature, because English monolinguals produced significantly shorter VOTs than German monolinguals ($\beta = 20.26$, $SE = 6.28$, $t = 3.22$, $p < .01$). An interaction between language and PoA further showed that VOT in /p/ was produced shorter than in /k/ in all three languages, as expected (English: $\beta = -18.69$, $SE = 3.33$, $t = -5.62$, $p < .001$; Italian: $\beta = -27.67$, $SE = 3.00$, $t = -9.23$, $p < .001$; German: $\beta = -15.99$, $SE = 3.91$, $t = -4.09$, $p < .001$). Monolingual English participants produced fewer prevoiced /b, g/ than Italian monolinguals ($\beta = 9.03$,
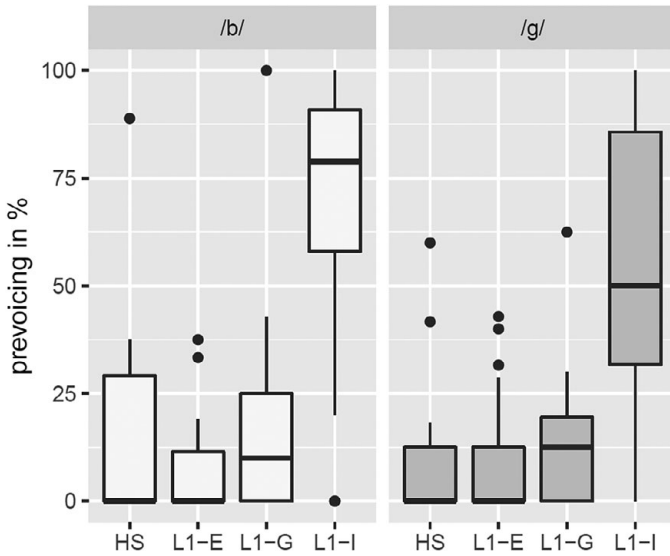
FIGURE 6.   Percentage of prevoicing of English lenis stops by language background.

TABLE 8.   Summarized statistical effects of language and language background on VOT

	Monolingual English vs. Italian and German	RQ3: HS English vs. Italian and German	RQ4: L1-E vs. HS, L1 German and L1 Italian (in English)
fortis	Italian: shorter VOT** German: longer VOT**	Italian: Shorter VOT*** German: n.s.	HS: n.s. L1 German: longer VOT* L1 Italian: shorter VOT*
lenis	Italian: Higher percentage of prevoicing*** German: n.s.	Italian: Higher percentage of prevoicing*** German: n.s.	HS: n.s. L1 German: higher percentage of prevoicing* L1 Italian: higher percentage of prevoicing***

***$p < .001$.
**$p < .01$.
*$p < .05$.
n.s. = $p > .05$.

$SE = 1.23$, $z = 7.30$, $p < .001$), while the English and German monolinguals did not differ significantly in this respect ($\beta = 1.41$, $SE = 0.75$, $z = 1.88$, $p = .06$).

The next analyses tested whether HSs produced language specific VOTs, that is, we compared their VOTs in English with Italian and German (RQ3). We observed that they produced significantly longer VOTs for fortis stops in English than in Italian ($\beta = -27.89$, $SE = 7.34$, $t = -3.80$, $p < .001$), while their English VOT productions did not differ significantly from their German productions ($\beta = 1.49$, $SE = 5.32$, $t = 0.28$, $p = .78$). An

interaction between language and PoA showed that HSs produced VOTs in /p/ shorter than in /k/ in all three languages (English: $\beta = -18.92$, $SE = 3.38$, $t = -5.61$, $p < .001$; Italian: $\beta = -28.79$, $SE = 2.49$, $t = -11.55$, $p < .001$; German: $\beta = -21.82$, $SE = 3.40$, $t = -6.41$, $p < .001$). In English, HSs produced a significantly lower percentage of lenis stops with prevoicing than in Italian ($\beta = 4.04$, $SE = 0.40$, $z = 10.16$, $p < .001$), while their percentage of prevoicing did not differ in English and German ($\beta = 0.52$, $SE = 0.42$, $z = 1.24$, $p = .21$).

The last set of analyses considered RQ4, that is, whether the acquisition of two first languages aids the acquisition of an L3. To answer this question, we examined the VOTs produced when speaking English, comparing whether HSs performed differently from L2 learners of English (L1 German and L1 Italian). The results showed that the VOTs of HSs did not differ from those of English monolinguals either for fortis stops ($\beta = 6.20$, $SE = 5.38$, $t = 1.15$, $p = .25$) or for lenis stops ($\beta = 0.68$, $SE = 0.60$, $z = 1.14$, $p = .26$). In the between-group comparison with the L2 learners, there was a significant difference between VOT duration of English monolinguals and L1 Germans for fortis stops ($\beta = 14.33$, $SE = 5.49$, $t = 2.61$, $p < .05$) and for the percentage of prevoicing in lenis stops ($\beta = 1.22$, $SE = 0.60$, $z = 2.03$, $p < .05$). The L1 Italians also differed from English monolinguals in producing significantly shorter VOTs in fortis stops ($\beta = -12.26$, $SE = 5.49$, $t = -2.24$, $p < .05$) and a higher percentage of lenis stops with prevoicing ($\beta = 4.29$, $SE = 0.61$, $z = 7.06$, $p < .001$).

DISCUSSION

This study examined the VOT patterns of HSs of Italian in their two L1s, Italian and German, as well as in their L3 English, in comparison to monolinguals and, in the L3 Study, also to L2 learners of English with either L1 German or Italian. In the following, we summarize our findings and interpret them in the light of CLI and a potential bilingual advantage.

VOT IN THE EARLY-ACQUIRED LANGUAGES

RQ1 was concerned with whether HSs differentiate between their HL (Italian) and their ML (German) in their production of VOTs. For fortis stops, which display short-lag in Italian and long-lag in German, we found significantly higher VOTs in German than in Italian. For lenis stops, which are mostly prevoiced in Italian and sometimes prevoiced also in German, we found that the proportion of prevoicing was significantly higher in Italian than in German. These results speak in favor of separate VOT patterns, which was expected given previous work testing both languages of bilingual speakers.

It is noteworthy that, in German, the monolinguals and bilinguals produced a considerable number of prevoiced stops, although in most of the relevant literature, German is characterized as having short-lag VOT for lenis stops (e.g., Kehoe et al., 2004). However, the finding is consistent with that of Braun (1996), indicating shorter VOTs and prevoicing in South German varieties (see Stoehr et al., 2017, for another case of prevoicing in German). Crucially, the monolinguals and bilinguals in our study did not differ in this respect. As mentioned above, the HSs produced more prevoiced stops in Italian than in German, which speaks in favor of separate VOT patterns.

RQ2 was concerned with whether the HSs performed like monolinguals in Italian and German. For German, we found that the HSs were not different from monolingual speakers for both the production of fortis stops (produced with long-lag VOT) and lenis stops (produced with short-lag VOT). This is consistent with most of the literature on HSs, showing no differences between bilinguals in their ML and monolingual baselines (e.g., Lein et al., 2016), although some studies also found influence into the ML (e.g., Mayr & Siddika, 2018, for lenis stops in English; Kupisch & Lleó, 2017, and Dittmers et al., 2018, for fortis stops in German). Future studies could investigate the effects of fundamental frequency and the first formant frequency at vowel onset, since some studies have shown that these acoustic measurements also play a role in the production of stops (see, e.g., Schwartz et al., 2019, on VOT in Polish). Including these measurements could provide valuable insights into the nature of (lenis) stops in general and for bilingual language acquisition in particular. The findings might reveal similarities between monolinguals and bilinguals that are currently missed out in VOT studies in the area of language acquisition.

In Italian, the HSs produced significantly higher VOTs than monolingual speakers, which we interpret as CLI from German, despite maintaining systemic differences between the languages. As for lenis stops, the HSs prevoiced significantly less compared with monolingual Italian controls. One possible explanation for this finding is CLI from German, where lenis stops are more likely to be produced with short-lag VOT (although, as we have shown, prevoicing is not entirely excluded). Another possible explanation is that prevoiced stops are more *marked* and later acquired than lenis stops with short-lag VOT, and that by the time prevoicing is typically acquired our HSs were massively exposed to German. We do not see these two explanations as being mutually exclusive. Notice also that there was a high inter-speaker variability in the production of lenis stops, but this was true for both mono- and bilinguals, as mentioned above. This suggests that prevoicing is not only challenging in bilingual acquisition but also in monolingual acquisition. Moreover, prevoicing is an area of variation; it is natural that bilinguals are inclined to exploit an option that is present in both languages but less marked (Kupisch, 2019). Given the significant main effect of language background and the smaller variability of prevoicing found in Italian monolinguals,[15] we are more inclined to interpret our findings in the light of CLI. Another argument suggesting that CLI from the ML can overpower markedness is that CLI was found both with long-lag stops (the least marked category) and with prevoiced stops (the most marked category).

VOT IN L3 ENGLISH

We turn now to the last two RQs, which pertained to VOT in the L3 English study. RQ3 aimed at ascertaining whether the HSs produced different VOT values in L3 English than in Italian and/or German. For fortis and lenis stops, the production of stops did not differ from those in German. No evidence of CLI from Italian was found.

RQ4 was concerned with whether the HSs would have an advantage over their monolingual peers, based on their knowledge of two language systems. Comparing HSs with English monolinguals, we found no significant difference for the production of fortis and lenis stops. In comparison, the monolingual Germans display longer VOT values for fortis stops (although their values are still in the long-lag range) and a higher percentage of prevoicing for lenis stops. The L1 Italian control group produced fortis

VOT values that were significantly shorter than target, and used significantly more prevoicing for the lenis stops. These results indicate the HSs were by no means disadvantaged by the shorter VOT values in Italian and, from a statistical perspective, did not perform differently from the L1 German peers ($\beta = -8.13$, $SE = 5.40$, $t = -1.51$, $p = .14$).

In summary, the HSs produced clearly differentiated values in Italian and German, which is argued to be evidence for separate VOT patterns, although with some CLI attested from the ML to the HL. In L3 English, the HSs VOT productions did not differ from those of L1 English and the HSs outperformed the L1 Italian control group. In theory, these results pattern both with studies that have shown phonological CLI from the typologically closest language (e.g., Cabrelli & Pichan, 2021), and also with studies that argue for CLI from the dominant language (e.g., Gabriel et al., 2016; Llama & López-Morelos, 2016, 2020; Lloyd-Smith et al., 2017). However, it is debatable to what extent typological proximity (in the sense of genealogical relatedness) plays a role when languages have a different phonological make-up. For example, while English and German have similarities on the suprasegmental level, there are many differences in their phoneme inventories. In this respect, it could be interesting for future studies to compare languages pairing within one family, specifically languages that have a more similar phonological make-up (e.g., Italian and Spanish) and languages that are more different in their phonological make-up (e.g., Italian and French). To test the impact of dominance further, more work is needed on language combinations that are typologically entirely unrelated (e.g., Spanish and Basque) to exclude potential effects of typological similarity.

A BILINGUAL ADVANTAGE?

The results for all speaker groups and languages are summarized in Figures 7 and 8. As these figures show, the VOT values obtained for fortis stops differed across the three languages, with longer values attested for German than for English, and significantly shorter values obtained for Italian.

Figure 7 illustrates that, while the L1 Italians differed from the English monolinguals, the HSs did not, producing longer VOT and less prevoicing than the L1 Italians (see Figure 8), likely due to facilitative CLI from German (although this was non-facilitative when speaking Italian). Interestingly, the HSs also had an advantage over German monolinguals when speaking English, because their fortis stops were shorter, likely due to CLI from Italian (but possibly also because their VOTs in German were slightly shorter-than-target to begin with). Therefore, while it is tempting to interpret this result as evidence for a bilingual advantage, our data rather suggest that the HSs transferred their VOT values from German, which led to an advantage when speaking English. This result is reminiscent of that obtained by Dittmers et al. (2018) and Gabriel et al. (2018) who found that HSs of Turkish and Russian converged more closely to target for VOT in L3 French than their German monolingual peers, due to shorter VOTs transferred from their HLs. It is also true that, being a cross-sectional study with speakers at the later stages of L3 acquisition, our data does not allow us to say whether the facilitative effect of knowing German was present from the early stages of L3 learning.

Future studies that approach L3 phonological acquisition from a longitudinal perspective (see, e.g., Kopečková, 2016) will be promising in delivering insights into how L3
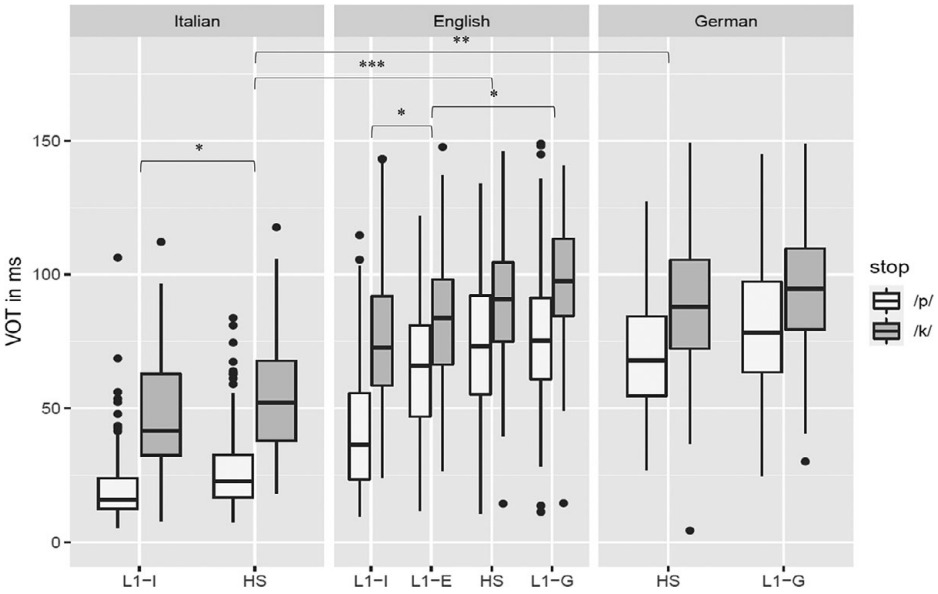
FIGURE 7.   VOT (ms) of fortis stops in Italian, English, and German by HSs and monolinguals.
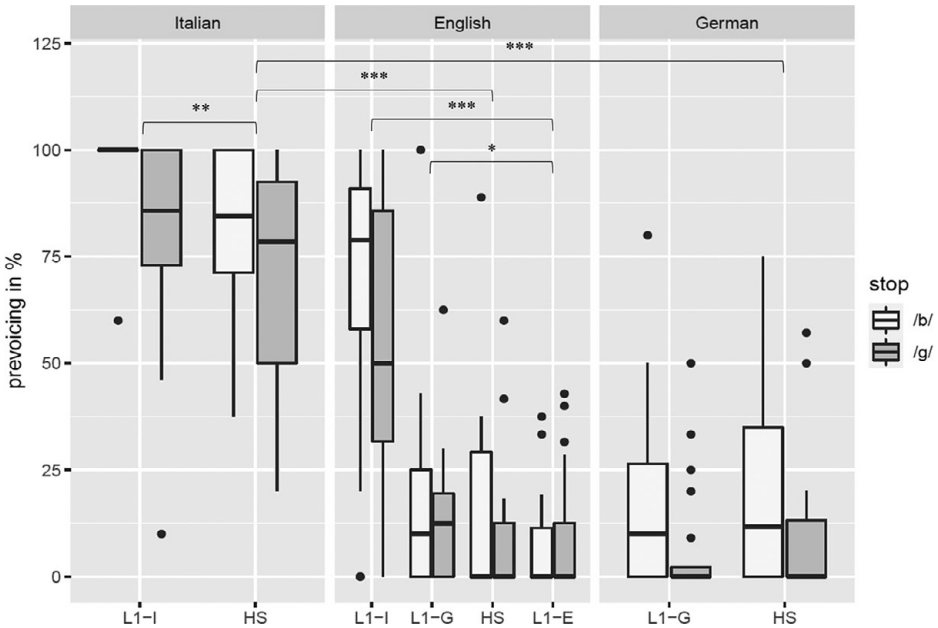*** $p < .001$, ** $p < .01$, * $p < .05$.

FIGURE 8.   Percentage of prevoiced stops in Italian, English, and German by HSs and monolinguals.
*** $p < .001$, ** $p < .01$, * $p < .05$.

phonology develops. Nonetheless, our results provide further evidence for the idea put forward by Kopečková (2016), namely that HSs acquiring an L3 can benefit from specific properties of their HL if there is overlap between the patterns. This leaves open the question of whether general bilingual advantage would obtain when HSs learn properties that cannot be transferred from any of their languages, as would be the case when learning a language that is typologically unrelated to the previously learned languages, or an artificial language.

CONCLUSION

We set out to explore whether heritage bilinguals show evidence of two separate VOT patterns in their two languages, German (the ML) and Italian (the HL), and whether there is CLI into L3 English. We found evidence for two separate VOT patterns: In Italian, the HSs produced fortis stops with short-lag VOT and lenis stops predominantly with prevoicing. However, compared with monolingual Italians, the percentage of prevoicing was significantly lower, and the VOTs for fortis stops was longer, suggesting CLI from German. In German, the HSs produced lenis stops with or without prevoicing and fortis stops with long-lag VOT, not differing from monolinguals. Our results thus confirmed the existence of separate VOT patterns for German and Italian, thereby providing a solid basis from which to interpret CLI into English. In English, the HSs produced fortis and lenis stops with no difference from English monolinguals. They had an advantage over Italian monolinguals whose VOT productions were significantly different from those of English monolinguals, and performed not different from L1 German controls. This can be taken as evidence for a facilitative role of the background languages in the acquisition of a foreign language.

SUPPLEMENTARY MATERIALS

To view supplementary material for this article, please visit http://doi.org/10.1017/S0272263121000280.

NOTES

[1]We use the term *CLI* to indicate (bidirectional) influence from any language in a speaker's repertoire.

[2]By monolingual, we mean people who grew up speaking only one language at home before age 6. The participants in this study are college-educated. Given the German education system, students learn at least one foreign language and are, therefore, not functionally monolingual.

[3]For VOT in regional variations of English, see, e.g., Lisker and Abramson (1964) for American English, Docherty (1992) for British English, and Antoniou et al. (2010) for Australian English.

[4]When referring to other scholar's work, expressions such as "target-like" or "identical-to-target" refer to their interpretations, that is, absence of statistical difference is interpreted as identical-to-target.

[5]Our results did not indicate any relation between the bilinguals' AoO in English and their proficiency in English as measured by the receptive vocabulary task, which is why we did not consider their amount of exposure at school to be a significant factor.

[6]The reported range has a relatively low limit, but this is the effect of one outlier. The other participants scored 70 or higher.

[7]In the case of minors, parental consent was obtained.

[8]Spontaneous speech could be more revealing than more controlled language samples because, in free speech, speakers have less control over their productions, which might facilitate access to procedural knowledge, which is precisely what we are interested in, because the two source languages of our participants were acquired naturalistically.

[9]Although we measured the positive VOT of lenis stops, we will not report on those measurements here, since our focus is on prevoicing for lenis stops. Additionally, the low number of devoiced stops in Italian monolinguals (for /b/ $N = 2$ and /g/ $N = 37$) did not allow for a statistical between-group analysis, but we report them here: If there is devoicing in German and Italian, /b/ was produced with VOTs of 13 ms in both languages, and /g/ with 32 ms (German) and 25 ms (Italian). Devoicing in the L3 mirrors devoicing in the HL, being within the range of short-lag VOT.

[10]Lenis stops were coded as a categorical variable because there are no clear values or ranges of values for negative VOT associated with /b d g/. Additionally, the measurements of negative VOT reported in the literature (see, e.g., Lisker & Abramson, 1964) show that the values for the three stops overlap. Because these VOT values vary considerably, they do not allow for firm conclusions about potential CLI.

[11]Problematic cases were discussed jointly by all authors.

[12]Stoehr et al.'s (2018) approach follows from the characteristics of Dutch. In Dutch, the presence versus absence of prevoicing is more important than the actual duration of prevoicing. Our rationale for treating prevoicing as a categorical variable is outlined in note 10.

[13]In the presentation of our descriptive statistics, we follow Stoehr et al. (2018).

[14]In the first and second model, German and Italian were compared with English, respectively, because English is of main interest here (for a comparison of German and Italian, see the HL Study section).

[15]The reported range in Table 4 suggested a big variability, but this was the effect of one outlier, which is visible in Figure 4.

REFERENCES

Alderson, J. C. (2005). *Diagnosing foreign language proficiency: The interface between learning and assessment.* Continuum.

Allen, G. D. (1985). How the young French child avoids the pre-voicing problem for word-initial voiced stops. *Journal of Child Language*, *12*, 37–46.

Antoniou, M., Best, C. T., Tyler, M. D., & Kroos, C. (2010). Language context elicits native-like stop voicing in early bilinguals' productions in both L1 and L2. *Journal of Phonetics*, *38*, 640–653.

Bates, D., Maechler, M., Bolker, B., & Walkers, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*, 1–48.

Boersma, P., & Weenink, D. (2015). Praat: Doing phonetics by computer (version 6.0.56) [computer software]. http://www.praat.org/.

Bortolini, U., Zmarich, C., Fior, R., & Bonifacio, S. (1995). Word-initial voicing in the productions of stops in normal and preterm Italian infants. *International Journal of Pediatric Otorhinolaryngology*, *31*, 191–206.

Braun, A. (1996). Zur regionalen distribution von VOT im Deutschen. *Untersuchungen zu Stimme und Sprache*, *96*, 19–32.

Cabrelli Amaro, J. (2013). Methodological issues in L3 phonological acquisition research. *Studies in Hispanic and Lusophone Linguistics*, *6*, 101–117.

Cabrelli, J., & Pichan, C. (2021). Initial phonological transfer in L3 Brazilian Portuguese and Italian. *Linguistic Approaches to Bilingualism*, *11*, 131–167. https://doi.org/10.1075/lab.18048.cab.

Davis, K. (1995). Phonetic and phonological contrasts in the acquisition of voicing: Voice onset time production in Hindi and English. *Journal of Child Language*, *22*, 275–305.

Deuchar, M., & Clark, A. (1996). Early bilingual acquisition of the voicing contrast in English and Spanish. *Journal of Phonetics*, *24*, 351–365.

Dittmers, T., Gabriel, C., Krause, M., & Topal, S. (2018). The production of voiceless stops in multilingual learners of English, French, and Russian: Positive transfer from the heritage languages. In C. Belz, S. Mooshammer, S. Fuchs, S. Jannedy, O. Rasskazova, & M. Zygis (Eds.), *Proceedings of the 13th conference on phonetics and phonology in the German-speaking countries*, (pp. 41–44). ZAS.

Docherty, G., Watt, D., Llamas, C., Hall, D., & Nycz, J. (2011). *Variation in voice onset time along the Scottish-English border* (pp. 591–594). Proceedings of the XVIIth international congress of phonetic science, Hong Kong.

Docherty, G. J. (1992). *The timing of voicing in British English obstruents*, vol. *9*. Walter de Gruyter.

Fabiano-Smith, L., & Bunta, F. (2012). Voice onset time of voiceless bilabial and velar stops in 3-year-old bilingual children and their age-matched monolingual peers. *Clinical Linguistics & Phonetics*, *26*, 148–163.

Fischer-Jørgensen, E. (1976). Some data on north German stops and affricates. *Annual Report of the Institute of Phonetics, University of Copenhagen*, *10*, 149–200.

Flege, J. E. (1991). Age of learning affects the authenticity of voice onset time (VOT) in stop consonants produced in a second language. *The Journal of the Acoustical Society of America*, *89*, 395–411.

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

Fowler, C. A., Sramko, V., Ostry, D. J., Rowland, S. A., & Hallé, P. (2008). Cross language phonetic influences on the speech of French–English bilinguals. *Journal of Phonetics*, *36*, 649–663.

Gabriel, C., Krause, M., & Dittmers, T. (2018). VOT production in multilingual learners of French as a foreign language: Cross-linguistic influence from the heritage languages Russian and Turkish. *Revue française de linguistique appliquée*, *23*, 59–72.

Gabriel, C., Kupisch, T., & Seoudy, J. (2016). VOT in French as a foreign language: A production and perception study with mono-and multilingual learners (German/mandarin-Chinese). *SHS Web of Conferences,* 27.

Gabriel, C., & Rusca Ruths, E. (2015). Der Sprachrhythmus bei deutsch-türkischen L3-Spanischlernern: Positiver transfer aus der Herkunftssprache? In S. Witzigmann, & J. Rymarczyka (Ed.), *Mehrsprachigkeit als chance. Herausforderungen und Potentiale individueller und gesellschaftlicher Mehrsprachigkeit (inquiries in language learning—Forschungen zu Psycholinguistik und Fremdsprachendidaktik)*, (pp. 185–204). Lang.

Haag, W. K. (1979). An articulatory experiment on voice onset time in German stop consonants. *Phonetica: International Journal of Speech Science*, *36*, 169–181.

Hamann, S., & Seinhorst, K. (2016). *Prevoicing in standard German plosives: Implications for phonological representations.* Thirteenth Old World Conference in Phonology, Budapest.

Hrycyna, M., Lapinskaya, N., Kochetov, A., & Nagy, N. (2011). VOT drift in 3 generations of heritage language speakers in Toronto. *Canadian Acoustics*, *39*, 166–167.

Keating, P., Mikoś, M., & Ganong, W. (1981). A cross-language study of voice onset time in the perception of initial stop voicing. *Journal of the Acoustical Society of America*, *70*, 1261–1271.

Kehoe, M. (2015). Cross-linguistic interaction: A retrospective and prospective view. In E. Babatsouli & D. Ingram (Eds.), *Proceedings of the international symposium on monolingual and bilingual speech 2015* (pp. 141–167). Institute of Monolingual and Bilingual Speech.

Kehoe, M., Lleó, C., & Rakow, M. (2004). Voice onset time in bilingual German-Spanish children. *Bilingualism: Language and Cognition*, *7*(1), 71–88.

Kopečková, R. (2016). The bilingual advantage in L3 learning: A developmental study of rhotic sounds. *International Journal of Multilingualism*, *13*, 410–425.

Kupisch, T. (2019). Recent developments in early bilingualism. *Bilingualism: Language and Cognition*, *21*, 653–655.

Kupisch, T., Barton, D., Klaschik, E., Lein, T., Stangen, I., & van de Weijer, J. (2014). Foreign accent in adult simultaneous bilinguals. *The Heritage Language Journal*, *11*, 123–150.

Kupisch, T., & Lleó, C. (2017). Voice onset time in German-Italian simultaneous bilinguals: Evidence on cross-language influence and markedness. In M. S. Yavaş, M. Kehoe, & W. Cardoso (Ed.), *Romance-Germanic bilingual phonology*, (pp. 79–98). Equinox.

Kupisch, T., Lloyd-Smith, A., & Stangen, I. (2020). Perceived global accent in Turkish heritage speakers in Germany: Exposure and use are more important than AoO. In F. Bayram (Ed.), *Turkish as a Heritage Language. Studies in Bilingualism (SiBiL)*, (pp. 207–228). John Benjamins.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*, 1–26.

Ladefoged, P., & Maddieson, I. (1996). *The sounds of the world's languages* (Vol. *1012*). Blackwell.

Lein, T., Kupisch, T., & van de Weijer, J. (2016). Voice onset time and global foreign accent in German–French simultaneous bilinguals during adulthood. *International Journal of Bilingualism*, *20*, 732–749.

Lisker, L., & Abramson, A. S. (1964). A cross-language study of voicing in initial stops: Acoustical measurements. *Word*, *20*, 384–422.

Lisker, L., & Abramson, A. S. (1967). Some effects of context on voice onset time in English stops. *Language and Speech*, *10*, 1–28.

Llama, R., & López-Morelos, L. P. (2016). VOT production by Spanish heritage speakers in a trilingual context. *International Journal of Multilingualism*, *13*, 444–458.

Llama, R., & López-Morelos, L. P. (2020). On heritage accents: Insights from VOT production by trilingual heritage speakers of Spanish. In E. Babatsouli, & M. J. Ball (Ed.), *An anthology of bilingual child phonology*. Multilingual Matters.

Lloyd-Smith, A., Einfeldt, M., & Kupisch, T. (2020). Italian-German bilinguals: The effects of HL use on accent in the early-acquired languages. *The International Journal of Multilingualism*, *24*, 289–304.

Lloyd-Smith, A., Gyllstad, H., & Kupisch, T. (2017). Transfer into L3 English: Global accent in German-dominant heritage speakers of Turkish. *Linguistic Approaches to Bilingualism*, *7*, 131–162.

Lloyd-Smith, A., Gyllstad, H., Kupisch, T., & Quaglia, S. (2021). Heritage language proficiency does not predict syntactic CLI into L3 English. *International Journal of Bilingual Education and Bilingualism*, *24*, 435–451. https://doi.org/10.1080/13670050.2018.1472208.

Macken, M. A., & Barton, D. (1979). The acquisition of the voicing contrast in English: A study of voice onset time in word-initial stop consonants. *Journal of Child Language*, *7*, 41–74.

Macken, M. A., & Barton, D. (1980). The acquisition of the voicing contrast in Spanish: A phonetic and phonological study of word-initial stop consonants. *Journal of Child Language*, *7*, 433–458.

Mayr, R., & Siddika, A. (2018). Inter-generational transmission in a minority language setting: Stop consonant production by Bangladeshi heritage children and adults. *International Journal of Bilingualism*, *22*, 255–284.

Miller, J. L., Green, K. P., & Reeves, A. (1986). Speaking rate and segments: A look at the relation between speech production and speech perception for the voicing contrast. *Phonetica*, *43*, 106–115.

Nagy, N., & Kochetov, A. (2013). Voice onset time across the generations: A cross-linguistic study of contact-induced change. In P. Siemund, I. Gogolin, M. Schulz, & J. Davydova (Ed.), *Multilingualism and language contact in urban areas: Acquisition—Development—Teaching—Communication*, (pp. 19–38). John Benjamins.

Neuhauser, S. (2011). Foreign accent imitation and variation of VOT and voicing in plosives. In *Proceedings of the XVIIth International Congress of Phonetic Science*, (pp. 1462–1465). Hong Kong.

Özaslan, M., & Gabriel, C. (2019). Final obstruent devoicing in English and French as foreign languages: Comparing monolingual German and bilingual Turkish-German learners. In C. Gabriel, J. Grünke, & S. Thiele (Ed.), *Romanische Sprachen in ihrer Vielfalt: Brückenschläge zwischen linguistischer Theoriebildung und Fremdsprachenunterricht (Romanische Sprachen und ihre Didaktik)*, (pp. 209–177): Stuttgart.

Puig-Mayenco, E., González Alonso, J., & Rothman, J. (2020). A systematic review of transfer studies in third language acquisition. *Second Language Research*, *36*, 31–64.

Rothman, J. (2011). L3 syntactic transfer selectivity and typological determinacy: The typological primacy model. *Second Language Research*, *27*, 107–127.

Rothman, J. (2015). Linguistic and cognitive motivations for the typological primacy model (TPM) of third language (L3) transfer. Timing of acquisition and proficiency considered. *Bilingualism: Language and Cognition*, *18*, 179–190.

Rothman, J., González Alonso, J., & Puig-Mayenco, E. (2019). *Third language acquisition and linguistic transfer (Cambridge studies in linguistics)*. Cambridge University Press.

Schwartz, G., Wojtkowiak, E., & Brzoza, B. (2019). *Beyond VOT in the Polish laryngeal contrast.* Proceedings of the 19th ICPhS in Melbourne.

Stock, D. (1971). *Untersuchungen zur Stimmhaftigkeit hochdeutscher Phonemrealisationen* (Vol. 28). Buske.

Stoehr, A., Benders, T., Van Hell, J. G., & Fikkert, P. (2017). Second language attainment and first language attrition: The case of VOT in immersed Dutch–German late bilinguals. *Second Language Research*, *33*, 483–518.

Stoehr, A., Benders, T., Van Hell, J. G., & Fikkert, P. (2018). Heritage language exposure impacts voice onset time of Dutch–German simultaneous bilingual preschoolers. *Bilingualism: Language and Cognition*, *21*, 598–617.

Sundara, M., Polka, L., & Baum, S. (2006). Production of coronal stops by simultaneous bilingual adults. *Bilingualism: Language and Cognition*, *9*, 97–114.

Tessmann Bandeira, M., & Zimmer, M. C. (2012). The dynamics of interlinguistic transfer of VOT patterns in multilingual children. *Linguagem & Ensino*, *Pelotas*, *15*, 341–364.

Wrembel, M. (2014). VOT patterns in the acquisition of third language phonology. *Concordia Working Papers in Applied Linguistics*, *5*, 751–771.

Wrembel, M. (2015). Cross-linguistic influence in second vs. third language acquisition of phonology. In U. Gut, R. Fuchs, & E.-M. Wunder (Ed.), *Universal or diverse paths to English phonology*, (pp. 41–70). Mouton De Gruyter.