Article

Second language attainment and first language attrition: The case of VOT in immersed Dutch–German late bilinguals

Antje Stoehr
Radboud University Nijmegen, The Netherlands

Titia Benders
Macquarie University, Australia

Janet G van Hell
Pennsylvania State University, USA

Paula Fikkert
Radboud University Nijmegen, The Netherlands

Abstract
Speech of late bilinguals has frequently been described in terms of cross-linguistic influence (CLI) from the native language (L1) to the second language (L2), but CLI from the L2 to the L1 has received relatively little attention. This article addresses L2 attainment and L1 attrition in voicing systems through measures of voice onset time (VOT) in two groups of Dutch–German late bilinguals in the Netherlands. One group comprises native speakers of Dutch and the other group comprises native speakers of German, and the two groups further differ in their degree of L2 immersion. The L1-German–L2-Dutch bilinguals ($N = 23$) are exposed to their L2 at home and outside the home, and the L1-Dutch–L2-German bilinguals ($N = 18$) are only exposed to their L2 at home. We tested L2 attainment by comparing the bilinguals' L2 to the other bilinguals' L1, and L1 attrition by comparing the bilinguals' L1 to Dutch monolinguals ($N = 29$) and German monolinguals ($N = 27$). Our findings indicate that complete L2 immersion may be advantageous in L2 acquisition, but at the same time it may cause L1 phonetic attrition. We discuss how the results match the predictions made by Flege's Speech Learning Model and explore how

Corresponding author:
Antje Stoehr, Centre for Language Studies, Radboud University Nijmegen, PO Box 9103, 6500 HD Nijmegen, The Netherlands.
Email: a.stohr@let.ru.nl

far bilinguals' success in acquiring L2 VOT and maintaining L1 VOT depends on the immersion context, articulatory constraints and the risk of sounding foreign accented.

I Introduction

Adults speaking a second language (L2) are likely to be identified as non-native speakers due to properties of their first language (L1) in their L2 speech (Brennan et al., 1975; Ferguson and Garnica, 1975; Flege, 1980, 1981; Scovel, 1969). Immersion in an L2 environment may cause the L2 to play a dominant role in everyday life, and may reduce the use of the L1 and contact to other native speakers. While L2 immersion can be beneficial to approach a native accent in the L2, the associated reduced L1 use may cause linguistic abilities in the L1 to deteriorate, a phenomenon known as L1 attrition (Freed, 1982; Schmid, 2004). When L1 attrition affects the domains of phonology or phonetics, it can surface as foreign-accented L1 speech (Bergmann et al., 2016; De Leeuw et al., 2010; Hopp and Schmid, 2013). The present study combines investigations of L2 attainment and L1 attrition in the speech of two groups of late bilinguals who differ in their degree of L2 immersion to assess potential bidirectional L1–L2 influences in their phonetic systems.

Bidirectional L1–L2 influences in a bilingual's speech can be explained by the Speech Learning Model (SLM; Flege, 1995). The SLM postulates that bilinguals have a common L1–L2 phonetic space and that these phonetic systems remain to some degree flexible in adulthood. If an L2 sound is not perceived as sufficiently different from an L1 sound, it may be classified as this phonetically similar L1 sound, a process known as 'equivalence classification'. As a result of equivalence classification in perception, also the speaker's production of that L2 sound may be different from native speakers' productions.

New L2 categories can be established provided they are perceived as sufficiently different from existing L1 sounds. Nevertheless, new L2 categories in a bilingual's L1–L2 phonetic space may still deviate from those of monolingual native speakers, for example to maintain contrasts with the bilingual's L1 categories. Hence, the speech of an L2 speaker who acquired new L2 categories may still deviate from native speech.

The SLM's assumption that phonetic systems remain flexible over the lifespan also implies that L1 categories can change under the influence of L2 acquisition, which can lead to a foreign accent in the L1. For this reason, the SLM has previously been used to interpret phonetic L1 attrition (Bergmann et al., 2016; Chang, 2012; Mayr et al., 2012). In order to understand how phonetic categories are organized in a speaker who accommodates two languages, it is important to characterize phonetic properties in both L2 and L1 speech (Chang, 2012; De Leeuw et al., 2012, 2013; Flege and Eefting, 1987a, 1987b; Mayr et al., 2012; Mennen, 2004; Sancier and Fowler, 1997).

Bilinguals' linguistic skills in the L2 are typically established by comparing their speech against monolingual native speech (Abrahamsson and Hyltenstam, 2009;

Bongaerts et al., 1997). If the goal is to determine to what extent bilinguals have been able to adapt to the phonetic environment in which they actually acquire the L2, a comparison against monolingual native speakers may be unsuitable (for similar thoughts on heritage language acquisition, see Rothman, 2007). For example, consider an L2 learner who acquires the L2 in the home country where he or she is exposed to other non-native speakers (e.g. non-native instructors or fellow L2 speakers in the home country) or to a native speaker with attrited L1 speech (e.g. an immigrant from the L2 country). In this case, comparing L2 speakers with monolingual native speakers implies that L2 speakers are evaluated against a type of speech to which they are barely exposed.

The monolingual reference point is also problematic because bilinguals are affected by cross-linguistic competition between their two languages (Cook, 2007; Hopp and Schmid, 2013; Kroll et al., 2006; Kupisch et al., 2013; Rothman and Treffers-Daller, 2014; Schmid et al., 2014). In addition, bilinguals presumably have to accommodate more phonetic categories than monolinguals. For example, consider a native speaker of Dutch who acquired German as L2 and a monolingual native speaker of German. The L2 speaker's phonetic system comprises L1-Dutch and presumably L2-German sounds, while the monolingual's phonetic system only comprises L1-German sounds. The mere process of becoming bilingual, with more phonetic categories to accommodate, may make the monolingual state impossible to attain. If we aim to test to what extent L2 speakers approach the speech of their linguistic environment, both the characteristics of the language to which they are exposed and the fact that they are bilingual need to be acknowledged. These two considerations make it important to compare bilinguals to native speakers who have been exposed to a comparable linguistic environment and who are bilinguals themselves (Cook, 2007; Hopp and Schmid, 2013; Kroll et al., 2006; Kupisch et al., 2013; Rothman and Treffers-Daller, 2014; Schmid et al., 2014).

A bilingual's daily linguistic environment is largely determined by the country of residence and may influence the linguistic skills in both L1 and L2. Bilinguals immersed in the L2 country are likely to be exposed to more speakers of their L2 compared to L2 speakers who live in their home country. The number of speakers who provide linguistic input has recently been identified as an important factor in the early stages of monolinguals' phonotactic learning (Seidl et al., 2014) and heritage speakers' lexical development (Gollan et al., 2015). Furthermore, quality and quantity of native language input play a crucial role in maintaining a native-like L1 accent after immigration to an L2 country (De Leeuw et al., 2010; Mayr et al., 2012). Input quality, quantity and diversity as captured through the country of residence are possibly also crucial factors in L2 acquisition.

The present study specifically focuses on the production of voice onset time (VOT) in two groups of late bilingual adults who live in binational households either in their home country or the L2 country, and who are L2 speakers and potentially L1 attriters. VOT is an acoustic cue that can contribute to a perceived foreign accent in both L2 speakers and L1 attriters (Flege, 1984; Flege and Eefting, 1987b; Major, 1987; Riney and Takagi, 1999; Sancier and Fowler, 1997; Schoonmaker-Gates, 2015). The present research enriches the existing literature on VOT in L2 attainment and L1 attrition in three important ways. First, it implements the methodological considerations on L2 attainment outlined above by evaluating L2 speech against the speech of native speakers who are

bilinguals themselves and whose speech is characteristic to the L2 speakers' linguistic environment. Second, it brings together investigations of L2 attainment and L1 attrition in the same speakers. Third, the present experiments cover VOT production in voiceless and voiced plosives to allow insight into the speakers' voicing contrasts. By addressing these three considerations, the present study allows assessing the possible restructuring of bilinguals' voicing systems.

VOT is the most important acoustic cue to distinguish voiced and voiceless plosives, and describes the time interval between a plosive's burst release and the onset of voicing (Abramson and Lisker, 1973; Lisker and Abramson, 1964). The VOT continuum can be divided into three phonetic categories: prevoicing (negative VOT), short lag (short positive VOT) and aspiration (long positive VOT). Dutch contrasts prevoiced 'voiced' and short lag 'voiceless' plosives (e.g. Lisker and Abramson, 1964). German contrasts short lag 'voiced' and aspirated 'voiceless' plosives (e.g. Jessen, 1998). Thus, depending on the language, short lag plosives can be phonologically classified as 'voiceless' (in Dutch) or 'voiced' (in German). Although voiced plosives do not require prevoicing in German, adult native speakers sometimes prevoice initial singleton plosives (Fischer-Jørgensen, 1976; Hamann and Seinhorst, 2016; Jessen, 1998; Kohler, 1977; Stock, 1971).

In production, prevoicing, short lag and aspiration differ in the required velopharyngeal activity, which is reflected in children's acquisition order (Allen, 1985; Bortolini et al., 1995; Kager et al., 2007; Kewley-Port and Preston, 1974; Khattab, 2000; Macken and Barton, 1980a, 1980b; MacLeod, 2016; Stoehr et al., 2017): across different languages, children produce the least complex short lag VOT in their early babbles. Around their second birthday, children acquiring an aspiration language produce aspiration, for which the glottis must remain open throughout consonantal closure. Substantially later, possibly in the early school years, children speaking a prevoicing language attain adult-like prevoicing, for which the glottis must be closed considerably before consonantal release and, additionally, vocal fold vibration must be initiated and sustained (Kewley-Port and Preston, 1974).

Within each phonetic category, small VOT differences can arise depending on the consonantal place of articulation (e.g. Lisker and Abramson, 1964) and, in the case of voiceless aspirated plosives, word length (Flege et al., 1998; Yu et al., 2015). In addition, male speakers produce optional prevoicing more frequently than female speakers (Ryalls et al., 1997), which can be ascribed to sex differences in vocal tract morphology (Fitch and Giedd, 1999).

I   Previous research into VOT in L2 acquisition

When bilinguals speak two languages that implement the voicing contrast differently, as is the case for the participants in the present study, a potential influence from L1 to L2 can be measured in their VOT. For voiceless plosives, three different acquisition patterns have been observed in late bilinguals whose L1 is a prevoicing language (Arabic, Dutch, French or Spanish) and who learn an aspiration L2 (English or German): (1) native-like acquisition (Schmid et al., 2014; Simon, 2009; Simon and Leuschner, 2010, the phonetically trained participants); (2) differential acquisition (Flege, 1987, 1991; Flege and Eefting, 1987a, 1987b; Simon and Leuschner, 2010, the phonetically

untrained participants); and (3) complete L1-to-L2 transfer (Flege, 1987, the least experienced participants; Flege and Port, 1981).

The native-like VOT acquisition pattern has been observed in highly advanced L1-immersed native speakers of Belgian Dutch with L2-English (and some participants with L3-German). The late bilinguals produced VOT in English (and German) voiceless plosives similar to monolingual native speakers (Simon, 2009; Simon and Leuschner, 2010). Similarly, native speakers of Dutch in the Netherlands reached comparable VOT durations in English as English native speakers who were also immersed in a Dutch environment (Schmid et al., 2014). These studies demonstrate that native-like aspiration of voiceless plosives can be acquired without L2 immersion.

The differential VOT acquisition pattern occurs when bilinguals produce VOT differently in their L2 than in their L1, but still deviate from native speakers' VOT in the L2. This pattern has been observed in bilinguals with L1-Spanish who learned L2-English as adults: their VOT was longer in English than in Spanish, but their English VOT was nevertheless shorter than that of monolingual English speakers (Flege, 1991). The same pattern emerged in bilinguals with L1-Spanish who learned L2-English during childhood, and occurred irrespective of whether they were immersed in an English environment or not (Flege and Eefting, 1987a). Similar results come from Dutch native speakers in the Netherlands with L2-English and L3-German who were not formally instructed in L2 and L3 phonetics. The speakers produced distinct VOT values for Dutch short lag voiceless plosives versus English and German aspirated voiceless plosives. Yet, their aspirated VOT productions in English and German still appeared shorter than the VOT of English and German monolinguals, although no direct statistical comparison was administered (Simon and Leuschner, 2010). L2 speakers with some level of L2 proficiency can thus differentiate L1 and L2 plosives in VOT, but do not necessarily reach native-like VOT.

The complete L1-to-L2 VOT transfer pattern has been observed in L1-Arabic speakers with L2-English in the USA (Flege and Port, 1981). Their VOT for English voiceless plosives was similar to Arabic and was therefore shorter than the VOT of English monolinguals. Although the L2 speakers were immersed in the L2 country for several years, they did not show evidence for phonetic differentiation between L1 and L2 VOT. L2 immersion thus does not always lead to the acquisition of new – be it native-like or differential – L2 VOT for voiceless plosives.

In sum, most studies on L2 VOT dealt with the acquisition of voiceless plosives. For long lag voiceless plosives, native-like acquisition, differential acquisition, and complete L1-to-L2 transfer have been observed, as was described above. For the acquisition of short lag voiceless plosives, native-like acquisition has never been reported, but it has only been addressed in one study, on English L2 speakers of French (Flege, 1987).

Studies on late bilinguals' production of voiced plosives reveal two acquisition patterns: native-like acquisition and L1-to-L2 transfer. The native-like acquisition pattern has been observed for L2 short lag voiced plosives in only one sample of Dutch native speakers with L2-English even though they were not immersed in the L2-speaking country (Schmid et al., 2014). The L1-to-L2 transfer pattern of L1 prevoicing to L2 short lag has also been observed, even in advanced and phonetically trained L2 speakers (Simon, 2009; Simon and Leuschner, 2010). Similarly, bilinguals who acquired their L2 during childhood tend to produce voiced plosives with prevoicing in both languages, especially

when their dominant language requires prevoicing (Flege and Eefting, 1987a; Hazan and Boulakia, 1993; MacLeod and Stoel-Gammon, 2009; Sundara et al., 2006).

No data are yet available on the opposite scenario: late bilinguals' acquisition of L2 prevoiced voiced plosives when their L1 does not require prevoicing. The present study fills this gap in the literature by contributing data on the production of voiced plosives in Dutch by native speakers of German.

In sum, native-like attainment and even VOT differentiation between L1 and L2 do not seem to require immersion, and do not automatically result from immersion. Two studies suggest that VOT differentiation may instead be related to language experience. This relationship was observed for the acquisition of voiceless plosives in bilinguals whose L1 was a prevoicing language (Spanish) learning an aspiration L2 (English), as well as in bilinguals with an aspiration L1 (English) learning a prevoicing L2 (French) (Flege, 1987; Flege and Eefting, 1987a). The more advanced L2 speakers in these two studies produced different VOT in their L2 than in their L1, but still showed differential VOT acquisition. Only the less experienced L2 speakers displayed full L1-to-L2 transfer and thus did not produce language-specific VOT. These studies suggest that language experience contributes to differentiating VOT between L2 and L1, but it may not necessarily be a sufficient predictor for native-like VOT acquisition in the L2.

2 Previous research into VOT in phonetic attrition

In some L2 speakers, the reverse of L1-to-L2 influence can be observed, namely an influence from L2 to L1. Bilinguals whose L2 has become the dominant language, for example through L2 immersion, are generally more prone to L1 attrition than L1-dominant bilinguals (Schmid and Köpke, 2007). The present study also investigates speech production in L2-immersed bilinguals, who may be affected by L1 attrition.

Research on L1 VOT in phonetic attrition is sparse, but there is broad evidence for L1 phonetic attrition at the segmental level (Bergmann et al., 2016; Chang, 2012; De Leeuw et al., 2013; Flege, 1987; Flege and Hillenbrand, 1984; Major, 1992; Mayr et al., 2012; Sancier and Fowler, 1997; Ulbrich and Ordin, 2014; Ventureyra et al., 2004) and the suprasegmental level (De Leeuw et al., 2012; Mennen, 2004). L1 attrition affecting the segmental or suprasegmental level may surface as a global foreign accent (Bergmann et al., 2016; De Leeuw et al., 2010; Hopp and Schmid, 2013). Most of these studies on L1 phonetic attrition reported changes in the realization of L1 speech sounds or prosody under the influence of long term L2 use (for short term L2 use, see Chang, 2012), and thus represent a context of language use that is similar to that of the participants in the present study.

Phonetic attrition can surface as a drift of the L1 VOT values towards the L2 VOT values. Four studies have observed phonetic attrition surfacing as durational changes in VOT in highly proficient L2 speakers (Flege, 1987; Major, 1992; Mayr et al., 2012; Sancier and Fowler, 1997). The bilinguals in these studies spoke Dutch, French or Portuguese, which have voiceless short lag plosives, in addition to English, which has voiceless aspirated plosives, like German. Native speakers of English produced shorter VOT in English voiceless plosives when they frequently used French or Portuguese (Flege, 1987; Major, 1992). This was irrespective of whether they were immersed in the

L2 or L1 context. Similarly, L1 speakers of French or Portuguese who were immersed in L2-English produced voiceless plosives with longer VOT in L1-French and L1-Portuguese than the respective monolinguals (Flege, 1987; Sancier and Fowler, 1997). Further support for L1 phonetic attrition of VOT comes from a case study of a monozygotic twin who emigrated from the Netherlands to the United Kingdom 30 years before testing (Mayr et al., 2012). Her VOT production was evaluated against the speech of the other twin who lived in the Netherlands throughout her life. The emigrated twin exhibited longer – and therefore more English-like – VOT in voiceless plosives than the Netherlands-based twin. By contrast, the emigrated twin's L1-Dutch voiced plosives remained prevoiced and were thus not affected by L1 phonetic attrition. These four studies suggest that changes to the L1 VOT may be limited to bilinguals with high L2 proficiency, but appear to occur independently of the immersion context (Flege, 1987).

A more nuanced view on the role of the immersion context on durational changes to L1 VOT and target-like L2 VOT production is provided by longitudinal data of one Portuguese–English late bilingual (Sancier and Fowler, 1997). The speaker produced longer – and thus more English-like – VOT in L1-Portuguese and L2-English after several months of L2 immersion in the USA. In turn, the speaker produced shorter – and thus more Portuguese-like – VOT after subsequent L1 immersion in Brazil. These durational VOT changes were perceived by native listeners of Brazilian Portuguese who rated the speech as more accented right after the informant's stay in the USA than after a stay in Brazil. This study suggests that changes to L1 VOT do not necessarily reflect an irreversible loss of native-like L1 VOT.

Although L1 attrition surfacing as durational VOT changes has been observed in highly proficient L2 speakers (Flege, 1987; Major, 1992; Mayr et al., 2012; Sancier and Fowler, 1997), high L2 proficiency does not automatically lead to attrition of L1 VOT. Dutch L1 speakers who acquired native-like aspiration in L2-English maintained short lag VOT in Dutch voiceless plosives (Simon, 2009; Simon and Leuschner, 2010). These speakers lived in their L1 country, which suggests that it may be easier to maintain native-like L1 VOT with frequent native L1 input.

The observed cases of L1 VOT drift in voiceless plosives are in line with the Speech Learning Model's (SLM) assumed flexibility of L1 phonetic categories (Flege, 1995), and showed that L2 VOT can influence L1 VOT. This influence is not limited to an L2 immersion context, but rather seems related to frequency of language use. In addition, frequent L1 exposure through L1 immersion may help to prevent L1 attrition in highly proficient L2 speakers.

Only the case study of Mayr et al. (2012) included investigations of VOT in voiced plosives, but found no evidence for phonetic attrition of L1 prevoicing. The present study follows up on this finding to address whether voiced plosives are indeed resistant to durational changes of L1 VOT, while voiceless plosives are frequently affected.

3   The current study

This study investigates VOT in the L1 and L2 speech of Dutch–German binational couples living in the Netherlands. Each couple consists of one partner with L1-Dutch and L2-German and one partner with L1-German and L2-Dutch. Within each couple, interactions in both

languages are common as the two partners have at least one child that they raise bilingually. The L1-Dutch speakers are frequently exposed to German and to non-native Dutch at home through their German partner and their bilingual child or children. Similarly, the L1-German speakers are frequently exposed to Dutch and non-native German at home. The exposure to German in both groups of bilinguals is limited to the family context. Exposure to Dutch occurs, on the other hand, in a variety of contexts and through multiple speakers.

In addition to a difference in immersion, the two groups face a different acquisition task: to produce target L2 VOT, the L1-Dutch speakers need to suppress Dutch prevoicing and learn to produce German aspiration. The L1-German speakers need to suppress German aspiration and learn to produce Dutch prevoicing.

This study combines investigations of VOT in L2 acquisition and L1 attrition in both voiceless and voiced plosives in the same speakers. Addressing the speakers' two languages and both voicing categories is essential to draw conclusions about the structure of bilinguals' phonetic space and voicing systems. The use of bilingual couples as participants allows addressing L2 attainment by comparing one group of bilinguals' L2 to the other group of bilinguals' L1, which offers two crucial advantages. First, a comparison between the L2 of one group of bilinguals and the L1 of the other group of bilinguals accounts for the characteristics of the speech to which the L2 speakers are daily exposed in their immediate social environment. Second, the L1 speech of bilinguals rather than monolinguals represents target speech that L2 speakers can in fact approach, as both groups' phonologies encompass a similar number of phonemes.

The three questions we are specifically asking regarding both groups of bilinguals are whether both acquisition contexts allow to: (1) produce VOT differently in L1 and L2; (2) realize VOT in the L2 similarly to native speakers who are bilingual themselves; and (3) maintain L1 VOT that is similar to a monolingual control group consisting of speakers representative of the linguistic environment in which the participants acquired and used their L1 before they became bilingual.

Regarding the L1-Dutch speakers, we hypothesize that they produce longer than monolingual-like VOT in L1 voiceless plosives, but maintain native-like prevoicing in L1 voiced plosives (compare Mayr et al., 2012). In L2-German, we expect the L1-Dutch speakers to produce voiceless plosives with longer VOT than in Dutch, but shorter VOT than the L1-German speakers. We further expect transfer of L1 prevoicing to L2 voiced plosives.

Regarding the L1-German speakers, we hypothesize to find shorter than monolingual-like VOT in L1 voiceless plosives, and possibly prevoiced voiced plosives to maintain a clear voicing contrast. If the L1-German speakers are indeed capable of producing prevoicing in L1-German and L2-Dutch, which has never been addressed in previous research, we expect them to be able to suppress aspiration and produce L2-Dutch voiceless plosives with target-like short lag VOT.

II Method

1 Participants

Ninety-seven speakers divided over four groups participated in this study: bilinguals with L1-Dutch and L2-German ($N = 18$, 5 female), henceforth the L1D–L2G speakers;

bilinguals with L1-German and L2-Dutch (*N* = 23, 19 female), henceforth the L1G–L2D speakers; Dutch monolinguals (*N* = 29; 26 female); and German monolinguals (*N* = 27, 26 female). All participants were parents of preschoolers. Table 1 provides detailed information on the participants.

Sixteen of the L1D–L2G speakers have had formal instruction to German in high school; the other two learned German only as adults when they met their German partner. The average age of first exposure to German of the L1D–L2G speakers was 13 years (range 1–28, *SD* = 6).[1] Regular exposure to German commenced for all L1D–L2G speakers when they met their German spouse in early adulthood. Further exposure to German now comes from their bilingual child or children. Twelve L1D–L2G speakers reported frequent use of Dutch and German. Six reported frequent use of Dutch and occasional use of German.

The L1G–L2D speakers learned Dutch at an average age of 23 years (range 8–33, *SD* = 6), when they moved to the Netherlands. One participant learned Dutch at school before she was regularly exposed to Dutch through her partner. Twenty-two of the participants in this group reported frequent use of German and Dutch. One participant reported frequent use of German and occasional use of Dutch.

Although not all participants reported knowledge of an additional language besides Dutch and German, schooling in the Netherlands and Germany requires all students to study English. Language teachers in these countries are, traditionally, non-native speakers of English.

The majority of the bilingual participants were 17 Dutch–German binational couples, contributing one partner to the L1D–L2G group and the other partner to the L1G–L2D group. One additional participant in the L1D–L2G group and six participants in the L1G–L2D group participated without their partners. The bilinguals were tested in different provinces across the Netherlands.

Of the Dutch monolinguals, two reported some knowledge of German, and three reported speaking English sporadically. All Dutch monolinguals were tested in or around Nijmegen in the Central Eastern Netherlands. Four of the monolingual German participants had some knowledge of Dutch, but none of them reported regular use of a language different from German. The German monolinguals were tested in Central Western Germany (*N* = 27) and Northern Germany (*N* = 2). Like the bilinguals, all monolinguals had studied English in high school.

2   Materials and procedure

The target plosives were voiceless /p/, /t/ and /k/ and voiced /b/ and /d/. As /g/ is not a native phoneme of Dutch, it was not included in this study for either language. For each language and plosive, six target words were selected that were picturable, plosive-vowel-initial nouns, such as the Dutch word *kast* ('cupboard'). The complete set of target words can be found in Tables 7 and 8 in Appendix 1. Twenty-three of the 30 Dutch target words[2] and 10 of the 30 German target words were monosyllabic. The remainder of the target words were disyllabic and carried stress on the initial syllable.

Testing took place in a quiet room in the participants' homes, after the participants signed informed consent for their family to participate in the study. When both

Table 1. Participant overview.

Participant	L1	Gender	Frequent German	Frequent Dutch	Age of acquisition of L2 (years)	Dutch at work	L2 active	L2 passive	Additional L2*
L1-G-01	German	F	✓	✓	20	✓	4	4	
L1-G-02	German	M	✓	✓	13	✗	5	5	
L1-G-03	German	M	✓	✓	?	✓	4	4	
L1-G-04	German	F	✓	?	31	✓	4	5	
L1-G-06	German	F	✓	✓	23	✗	3	4	
L1-G-07	German	F	✓	✓	20	✓	4	4	
L1-G-10	German	F	✓	✓	24	✓	5	5	
L1-G-12	German	F	✓	✓	20	✓	3	4	
L1-G-13	German	F	✓	✓	25	✓	4	4	
L1-G-15	German	F	✓	✓	20	✓	5	5	
L1-G-16	German	F	✓	✓	8	✓	5	5	FR
L1-G-17	German	F	✓	✓	25	✗	3	4	
L1-G-18	German	F	✓	✗	27	✗	4	5	
L1-G-19	German	F	✓	✓	23	✓	5	5	
L1-G-21	German	F	✓	✓	25	✓	4	4	
L1-G-23	German	F	✓	✓	33	✓	4	5	
L1-G-24	German	F	✓	✓	30	✓	4	5	FR
L1-G-26	German	F	✓	✓	25	✓	4	4	DAN, POR, NOR
L1-G-27	German	F	✓	✓	20	✓	4	4	
L1-G-29	German	M	✓	✓	16	✓	5	5	
L1-G-31	German	M	✓	✓	23	✓	4	4	
L1-G-32	German	F	✓	✓	19	✓	3	4	
L1-G-33	German	F	✓	✓	33	✗	4	4	
L1-D-02	Dutch	F	✓	✓	13	✓	4	4	
L1-D-03	Dutch	F	✓	✓	?	✓	4	4	
L1-D-06	Dutch	M	✓	✓	12	✓	3	4	
L1-D-07	Dutch	M	✓	✓	14	✓	3	3	
L1-D-10	Dutch	M	✗	✓	14	✓	4	4	
L1-D-11	Dutch	F	✓	✓	28	✓	4	4	
L1-D-12	Dutch	M	✗	✓	14	✓	2	2	
L1-D-16	Dutch	M	✓	✓	12	✓	3	4	ITA, DAN
L1-D-18	Dutch	M	✗	✓	13	✓	3	4	
L1-D-19	Dutch	M	✓	✓	13	✓	3	3	
L1-D-21	Dutch	M	✓	✓	1	✓	4	4	
L1-D-24	Dutch	M	✓	✓	12	✓	4	4	
L1-D-26	Dutch	M	✗	✓	12	✓	4	4	
L1-D-27	Dutch	M	✓	✓	6	✓	3	3	
L1-D-29	Dutch	F	✓	✓	13	✓	3	3	
L1-D-31	Dutch	F	✗	✓	25	✓	3	3	
L1-D-32	Dutch	M	✓	✓	13	✓	4	4	
L1-D-33	Dutch	M	✗	✓	14	✓	2	3	

Notes. * All speakers had instruction in English during high school. Codes: ✓ = yes, ✗ = no, ? = no information provided. Additional L2: DAN = Danish, FR = French, ITA = Italian, NOR = Norwegian, POR = Portuguese. L2 active: 5 = native fluency, 4 = very fluent, 3 = quite fluent, 2 = somewhat fluent, 1 = limited fluency, 0 = virtually no fluency. L2 passive: 5 = native understanding, 4 = excellent understanding, 3 = good understanding, 2 = some understanding, 1 = limited understanding, 0 = almost no understanding.

participants from a couple completed the task during the same testing session, the other participant left the room during the recordings. The participants were shown pictures of the target words and they were asked to name them at a comfortable pace without a determiner. The participants then filled out a language background questionnaire, while their children completed three tasks for a different study (Stoehr et al., 2017). Finally, the participants named the pictures in their other language. The language order was counterbalanced across participants. The picture naming took approximately three minutes per language. At the end of the session, the participants and their child were compensated with €10 or a book.

3   Recordings and VOT measurements

Recordings were made with an Olympus Linear PCM Recorder LS-10 with uncompressed 24 bit / 96 kHz recording capability. VOT measurements were performed in Praat (Boersma and Weenink, 2014) taking into account waveforms and spectrograms viewed at zero to 5,000 Hz. The burst onset was measured as the onset of abrupt energy release. The onset of voicing was defined as the first periodic component of the waveform and was measured at the preceding zero-crossing (Francis et al., 2003). Intercoder reliability based on 25% of the data indicated 99% agreement. Measurements of voiceless plosives were considered in agreement when they differed less than 10 ms (Fabiano-Smith and Bunta, 2012). Coding of voiced plosives was considered in agreement when both coders rated VOT as either prevoiced or short lag. Only tokens that allowed unambiguous measurements without coarticulation or speech overlap entered the analyses. Figure 1 shows examples of VOT measurements of prevoicing, short lag, and aspiration, respectively.

III   Results

In this section, we first provide an overview of the descriptive statistics of voiceless plosives (Table 2 and Figure 2) and voiced plosives (Tables 3 and 4, Figure 3). We then present the statistical models (Table 5) before we turn to the statistical effects of *Language* and *Language Background* on VOT, which are summarized in Table 6.

Table 2 provides the means and standard deviations of VOT per voiceless plosive over participants by language and language background. Both groups of bilinguals produced overall longer VOT in German than in Dutch. In each language, the bilinguals produced L1 VOT intermediate to the monolinguals' L1 VOT and the L2 VOT of the other group of bilinguals. In Dutch, the L1D–L2G speakers produced minimally longer VOT than the monolinguals, and shorter VOT than the L1G–L2D speakers. In German, the L1G–L2D speakers produced VOT that was intermediate to the monolinguals' overall longer VOT and the L1D–L2G speakers overall shorter VOT. Figure 2 visualizes these findings by consonantal place of articulation.

VOT of voiced plosives was bimodally distributed in 47 of the 70 participants in Dutch and in 51 of the 68 participants in German. VOT of voiced plosives was therefore treated categorically as either prevoiced (negative VOT) or short lag (short positive
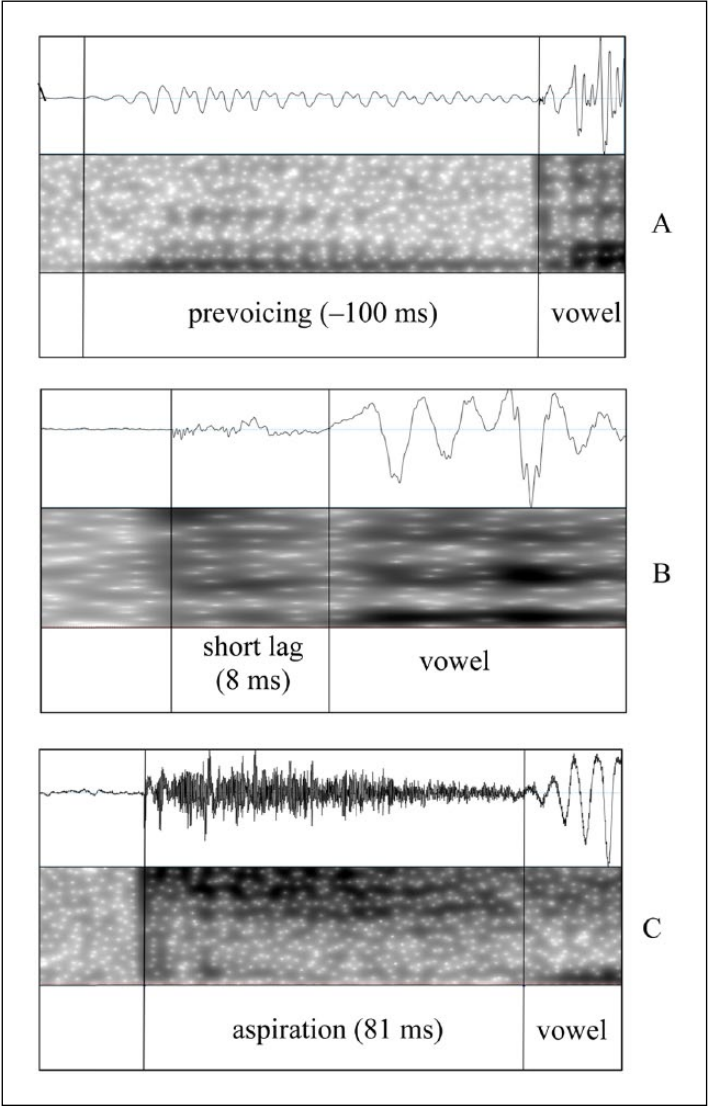
Figure 1. Acoustic landmarks from top to bottom: A. prevoicing, B. short lag, C. aspiration.

VOT). Table 3 shows the mean percentages and standard deviations of the voiced plosives produced with prevoicing (and inversely related short lag VOT) over participants together with the total number of analysable prevoiced and short lag tokens per voiced plosive by language and language background. Both groups of bilinguals produced overall more prevoiced tokens in Dutch than in German, although this difference is more pronounced in the L1G–L2D speakers. In Dutch, the L1D–L2G speakers produced the

Table 2. Voice onset time (VOT) in ms by place of articulation over participants.

		Dutch			German		
		LIG–L2D	LID–L2G	MonoD	LID–L2G	LIG–L2D	MonoG
/p/	M	21	10	8	23	38	45
	SD	15	6	5	19	17	18
	Tokens	147	105	173	109	140	159
/t/	M	31	23	21	48	59	69
	SD	13	9	10	20	19	17
	Tokens	141	111	179	108	140	169
/k/	M	43	31	28	44	58	72
	SD	16	13	10	18	18	20
	Tokens	139	110	171	112	140	165
	Overall M	32	21	19	38	52	62

Notes. LIG-L2D = bilinguals with German as first language and Dutch as second language; LID-L2G = bilinguals with Dutch as first language and German as second language; MonoD = Dutch monolinguals; MonoG = German monolinguals.

Figure 2. Voice onset time (VOT) of voiceless plosives by language background over participants.
Notes. LIG-L2D = bilinguals with German as first language and Dutch as second language; LID-L2G = bilinguals with Dutch as first language and German as second language; MonoD = Dutch monolinguals; MonoG = German monolinguals.

highest percentage of voiced plosives with prevoicing, closely followed by the monolingual Dutch speakers. This small between-group difference may be ascribed to the larger number of males in the L1D–L2G group, who typically produce more prevoicing than females (Ryalls et al., 1997). The L1G–L2D speakers produced a lower percentage of

Table 3. Mean percentage of prevoiced plosives by place of articulation over participants.

		Dutch			German		
		LIG–L2D	LID–L2G	MonoD	LID–L2G	LIG–L2D	MonoG
/b/	M % prevoiced	66	91	87	87	38	26
	SD	35	24	22	20	37	34
	Tokens	95/143	96/106	149/172	93/107	53/140	42/158
/d/	M % prevoiced	64	82	79	64	26	22
	SD	33	22	24	30	31	24
	Tokens	86/139	93/110	133/165	66/103	37/145	38/177
	Overall M	65	87	83	76	32	24

Notes. LIG-L2D = bilinguals with German as first language and Dutch as second language; LID-L2G = bilinguals with Dutch as first language and German as second language; MonoD = Dutch monolinguals; MonoG = German monolinguals.
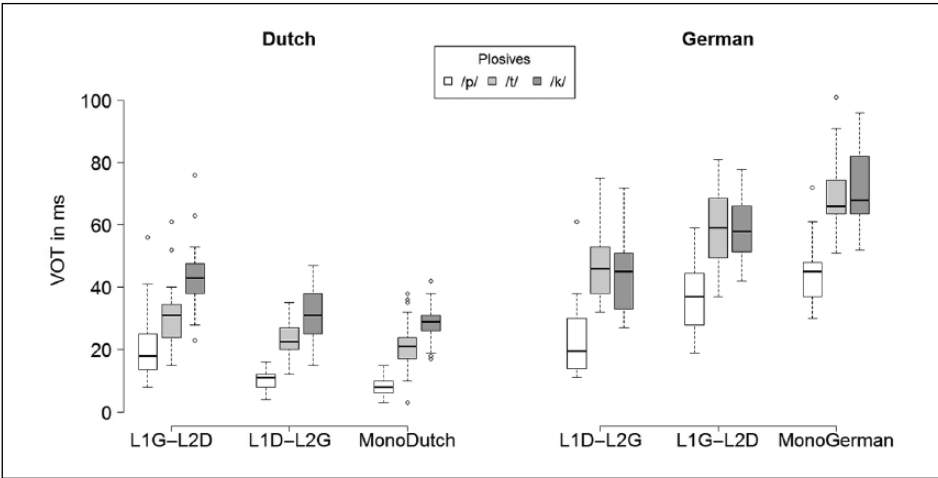

Figure 3. Percentage of voiced plosives produced with prevoicing by language background over participants.
Notes. LIG-L2D = bilinguals with German as first language and Dutch as second language; LID-L2G = bilinguals with Dutch as first language and German as second language; MonoD = Dutch monolinguals; MonoG = German monolinguals.
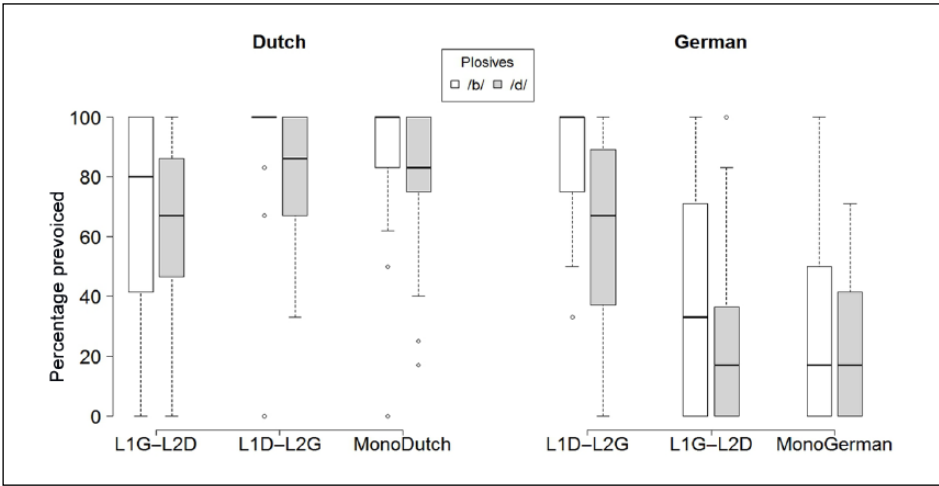
prevoiced plosives in Dutch than the two groups of Dutch native speakers. In German, the monolinguals produced the lowest percentage of prevoiced plosives, followed by the L1G–L2D speakers. The L1D–L2G speakers produced the highest percentage of prevoiced plosives. Figure 3 visualizes the percentages of prevoiced plosives by language and consonantal place of articulation across the groups. The devoiced voiced plosives had VOT values close to 10 ms in both languages and all groups (Table 4).

Table 4. Voice onset time (VOT) in ms of short lag voiced plosives by place of articulation over participants.

		Dutch			German		
		L1G–L2D	L1D–L2G	MonoD	L1D–L2G	L1G–L2D	MonoG
/b/	M	9	11	5	8	7	6
	SD	3	2	2	2	3	3
	Tokens	48	10	23	14	87	116
/d/	M	12	14	13	13	12	12
	SD	7	3	9	5	4	4
	Tokens	53	17	32	37	108	139
	Overall M	11	13	9	11	10	9

Notes. L1G-L2D = bilinguals with German as first language and Dutch as second language; L1D-L2G = bilinguals with Dutch as first language and German as second language; MonoD = Dutch monolinguals; MonoG = German monolinguals.

1 Description of the statistical models

Statistical analyses using mixed effects regression were performed in *R* (R Core Team, 2013). An alpha level of .05 was adopted throughout. VOT of the voiceless plosives /p/, /t/ and /k/ was analysed as a continuous variable using mixed effects linear regression. VOT of the voiced plosives /b/ and /d/ was analysed as a categorical variable using mixed effects logistic regression to address the aforementioned bimodal distribution of VOT. Negative VOT values were coded as 'prevoiced' and values equal to or greater than zero were coded as 'short lag'. Due to the use of different regression types, each research question was addressed with separate models for voiceless and voiced plosives. Each research question was furthermore addressed with specific between-group or within-group comparisons, which are outlined below.

The bilinguals' differentiation of L1 and L2 VOT was assessed with within-group comparisons of the bilinguals' Dutch and German. This L1–L2 comparison was conducted separately for the L1G–L2D speakers and the L1D–L2G speakers, and the independent variable (IV) of main interest was *Language* (Dutch vs. German).

Two between-group analyses addressed nativelikeness of the bilinguals' VOT in the two languages. L2 attainment was assessed by comparing the bilinguals' L2 VOT to the other bilinguals' L1 VOT. L1 attrition was assessed by comparing the bilinguals' L1 VOT to the VOT of an independent sample of monolinguals. The IV of main interest in all between-group analyses was *Language Background* (the bilinguals' L2 vs. the other bilinguals' L1; the bilinguals' L1 vs. the monolinguals' L1).

Additional IVs were used in all models to account for item-related and participant-related variance due to factors that are known to impact on VOT. Item-related IVs for analyses on voiceless plosives were *Place of Articulation* of the plosive (/p/ vs. /t/ and /t/ vs. /k/) and *Word Length* (monosyllabic vs. disyllabic). The item-related IV for analyses on voiced plosives was *Place of Articulation* (/b/ vs. /d/). The participant-related IV in all analyses was *Gender*.

Table 5 provides an overview of the model specifications for each group comparison. All models comprised interactions between the IV of main interest and the other IVs, except for the models on L2 attainment, where simplification due to model convergence problems was required. Significant interactions were explored in separate follow-up analyses for each level of the IVs.

2   Results of the statistical models

This section presents the main findings of the three research questions. The first two analyses addressed the bilinguals' differentiation of VOT in the L1 and L2. Subsequent analyses addressed the bilinguals' L2 attainment and potential L1 attrition. Lastly, we present findings on variability specific to the target words and participants that did not contribute to the main results.

a   Differentiation between L1 and L2 VOT within the bilinguals. The analyses on language differentiation in the L1G–L2D speakers showed that they produced VOT differently when speaking German compared to when speaking Dutch. The L1G–L2D speakers specifically produced longer VOT in voiceless plosives when speaking German ($\beta = 16.22$, $SE = 2.41$, $t = 6.72$, $p < .001$), and a higher percentage of voiced plosives with prevoicing when speaking Dutch ($\beta = 0.95$, $SE = 0.34$, $z = 2.84$, $p < .005$). In addition, an interaction between *Language* and *Place of Articulation* ($\beta = -6.37$, $SE = 2.87$, $t = -2.22$, $p = .026$) revealed that the L1G–L2D speakers produced longer VOT in /k/ than in /t/ in Dutch ($\beta = 12.31$, $SE = 3.32$, $t = 3.70$, $p < .001$), but not in German ($\beta = -0.45$, $SE = 4.91$, $t = -0.09$, $p > .250$).

The L1D–L2G speakers produced distinct VOT for Dutch and German voiceless plosives, but not for voiced plosives. They produced voiceless plosives with longer VOT in German than in Dutch ($\beta = 13.83$, $SE = 2.44$, $t = 5.68$, $p < .001$), but no difference in the percentage of voiced plosives produced with prevoicing in Dutch and in German was detected ($\beta = 0.43$, $SE = 0.28$, $z = 1.54$, $p = .124$). An interaction between *Language* and *Word Length* ($\beta = 2.60$, $SE = 1.25$, $t = 2.07$, $p = .038$) revealed that the L1D–L2G speakers produced voiceless plosives with longer VOT in monosyllabic than in disyllabic words in German ($\beta = 4.62$, $SE = 2.05$, $t = 2.25$, $p = .024$), but not in Dutch ($\beta = -0.39$, $SE = 0.98$, $t = -0.40$, $p > .250$). Overall, the results on phonetic differentiation between L1 and L2 suggest that Dutch–German late bilinguals produced VOT differently in L1 and L2 with the exception of the L1D–L2G speakers' production of voiced plosives.

b   L2 attainment and L1 attrition. The following four analyses concerned the bilinguals' VOT production in both their L2 and their L1. The reference point for L2 attainment was the other bilinguals' L1. The reference point for L1 attrition was the speech of monolingual native-speakers.

L1G–L2D speakers. The analyses on L2 attainment in the L1G–L2D speakers showed that they attained native-like VOT in L2-Dutch for /p/ and /t/, but not for /k/ or voiced plosives. In L2-Dutch voiceless plosives, no overall VOT differences were detected between the L1G–L2D speakers and the L1D–L2G speakers ($\beta = -2.10$, $SE = 1.45$, $t = -1.45$, $p = .147$), but an interaction between *Language Background* and *Place of*

Table 5. Model specifications.

Groups	Analysis	Fixed effects	Interactions	Random effects & intercept	Nesting	Random slopes
Bilingual L1 vs. bilingual L2 (L1G-L2D speakers & L1D-L2G speakers)	voiceless	Language, Gender, PoA-LC, PoA-CD, Word length	Language*Gender, Language*PoA-LC, Language*PoA-CD, Language*Word length	Participant — Item		Language, PoA-LC, PoA-CD, Word length — none
	voiced	Language, Gender, PoA	Language*Gender, Language*PoA	Participant — Item		Language, PoA — none
Bilingual L2 vs. bilingual native speakers (Dutch & German)	voiceless	LangBackgr., Gender, PoA-LC, PoA-CD, Word length	LangBackgr.*Gender, LangBackgr.*PoA-LC, LangBackgr.*PoA-CD, LangBackgr.*Word length[2]	Participant — Item	Couple	PoA-LC, PoA-CD, Word length, LangBackgr.
	voiced	LangBackgr., Gender, PoA	LangBackgr.*Gender, LangBackgr.*PoA	Participant — Item	Couple	PoA, LangBackgr.
Bilingual L1 vs. monolingual native speakers (Dutch & German)	voiceless	LangBackgr., Gender, PoA-LC, PoA-CD, Word length	LangBackgr.*Gender, LangBackgr.*PoA-LC, LangBackgr.*PoA-CD, LangBackgr.*Word length	Participant — Item		PoA-LC, PoA-CD, Word length, LangBackgr.
	voiced	LangBackgr., Gender, PoA	LangBackgr.*Gender, LangBackgr.*PoA	Participant — Item		PoA, LangBackgr.

LangBackgr. = Language Background; PoA-LC = Place of Articulation: Labial vs. Coronal; PoA-CD = Place of Articulation: Coronal vs. Dorsal.
[1]only in Dutch model due to convergence problems; [2]only in German model due to convergence problems.

Articulation ($\beta = -2.30$, *SE* = 1.13, $t = -2.04$, $p = .041$) revealed that the L1G–L2D speakers produced in fact longer VOT in /k/ than the L1D–L2G speakers ($\beta = -4.91$, *SE* = 1.68, $t = -2.92$, $p = .004$). In L2-Dutch voiced plosives, the L1G–L2D produced a lower percentage of prevoiced plosives than native speakers ($\beta = -0.95$, *SE* = 0.46, $z = -2.06$, $p = .039$).[3]

The analyses on L1 attrition in the L1G–L2D speakers showed that their L1-German VOT of voiceless but not voiced plosives is affected by L1 attrition. The L1G–L2D speakers produced L1-German voiceless plosives with shorter VOT than monolinguals ($\beta = -6.94$, *SE* = 3.10, $t = -2.24$, $p = .025$). By contrast, no differences in the percentage of prevoicing between the L1G–L2D speakers and monolinguals were observed ($\beta = -0.13$, *SE* = 0.50, $z = -0.25$, $p > .250$).

L1D–L2G speakers. The analyses on L2 attainment in the L1D–L2G speakers showed that they produced non-native VOT in L2-German. The L1D–L2G speakers produced L2-German voiceless plosives with shorter VOT than the L1G–L2D speakers ($\beta = -6.57$, *SE* = 1.65, $t = -3.97$, $p < .001$). Similarly, they produced a higher percentage of German voiced plosives with prevoicing than the L1G–L2D speakers ($\beta = -1.06$, *SE* = 0.28, $z = -3.79$, $p < .001$). An interaction between *Language Background* and *Gender* ($\beta = -0.92$, *SE* = 0.37, $z = -2.49$, $p = .013$) did not reveal any gender differences in the L1D–L2G group ($\beta = -0.50$, *SE* = 0.41, $z = -1.20$, $p = .230$), but rather revealed that males in the L1G–L2D group produced a higher percentage of prevoiced voiced plosives than females ($\beta = 1.67$, *SE* = 0.51, $z = 3.30$, $p < .001$).

The analyses on L1 attrition in the L1D–L2G speakers did not find evidence for attrition of L1-Dutch VOT. The L1D–L2G speakers neither produced L1-Dutch voiceless plosives ($\beta = 1.86$, *SE* = 1.16, $t = 1.60$, $p = .110$) nor voiced plosives ($\beta = -0.06$, *SE* = 0.44, $z = -0.13$, $p > .250$) detectably different from Dutch monolinguals.

In sum, the results on L2 attainment and L1 attrition show that only the L1G–L2D bilinguals who were immersed in the L2 country partially attained native-like L2 VOT. Similarly, only the L1D–L2G bilinguals who were immersed in the L1 country maintained native-like L1 VOT.

c Variability related to the words and participants. In the following, we present the significant findings on the IVs relating to the target words and participants. As the bilinguals were part of three analyses, the results of an IV for a group was considered significant when at least one analysis including the group yielded significance for an IV. The complete model output of all models is presented in Appendices 2–4.

In analyses on voiceless plosives, all groups produced shorter VOT for /p/ than for /t/ in Dutch and in German, and all groups produced longer VOT for /k/ than for /t/ only in Dutch, but not in German. In addition, all groups produced longer VOT in monosyllabic than in disyllabic words in German, but not in Dutch. In analyses on voiced plosives, all groups prevoiced /b/ more frequently than /d/ in both languages. In all groups except the Dutch monolinguals, males prevoiced more frequently than females. Late bilinguals thus produce language-specific within-category VOT variability related to consonantal place of articulation and word length.

Table 6. Results overview.

			Language	Language background
Research question 1	L1G–L2D speakers	voiceless	Longer VOT in German ***	—
		voiced	Higher percentage of prevoicing in Dutch **	—
	L1D–L2G speakers	voiceless	Longer VOT in German ***	—
		voiced	non-significant	—
Research question 2	L1G–L2D speakers	voiceless	—	non-significant
		voiced	—	L2 speakers: lower percentage of prevoicing than L1 speakers *
	L1D–L2G speakers	voiceless	—	L2 speakers: shorter VOT than L1 speakers ***
		voiced	—	L2 speakers: higher percentage of prevoicing than L1 speakers ***
Research question 3	L1G–L2D speakers	voiceless	—	Bilingual L1 speakers: shorter VOT than monolinguals *
		voiced	—	non-significant
	L1D–L2G speakers	voiceless	—	non-significant
		voiced	—	non-significant

Notes. L1G–L2D = bilinguals with German as first language and Dutch as second language; L1D–L2G = bilinguals with Dutch as first language and German as second language; *** $p < .001$; ** $p < .01$; * $p < .05$; non-significant $p > .05$.

IV Summary

The present study investigated how two groups of Dutch–German late bilinguals in the Netherlands realize the voicing contrast in both Dutch and German by means of voice onset time (VOT). The bilinguals who speak Dutch as native language and German as the L2 are referred to as L1D–L2G speakers, and the bilinguals who speak German as native language and Dutch as the L2 are referred to as L1G–L2D speakers. To achieve native-like L2 VOT, the L1D–L2G speakers need to acquire aspiration for L2-German voiceless plosives and suppress prevoicing for L2-German voiced plosives. The L1G–L2D speakers need to suppress aspiration in L2-Dutch voiceless plosives and consistently prevoice L2-Dutch voiced plosives. We investigated whether (1) both groups of late bilinguals produced VOT differently in L1 and L2; (2) both groups of bilinguals achieved native-like L2 VOT; and (3) both groups of bilinguals maintained native-like L1 VOT.

The L1G–L2D speakers produced voiceless plosives with short lag VOT in L2-Dutch /p/ ($M = 21$ ms) and /t/ ($M = 31$ ms), and slight aspiration in Dutch /k/ ($M = 43$ ms), while they aspirated L1-German voiceless plosives ($M = 52$ ms). Similarly, the L1G–L2D speakers prevoiced a higher percentage of voiced plosives in L2-Dutch (65%) than in L1-German (32%). The L1G–L2D speakers produced the remaining voiced plosives with short lag VOT that was virtually alike in L2-Dutch ($M = 11$ ms) and L1-German ($M = 10$ ms), and considerably shorter than their VOT of L2-Dutch voiceless plosives ($M = 32$ ms). However, the L1G–L2D speakers did not acquire new VOT ranges, as aspiration, short lag and prevoicing are all observed in monolinguals' speech as well. Instead, the acquisition task they accomplished was redefining their phonetic space. In addition to the pre-existing aspirated category (German /p/, /t/, /k/), the L1G–L2D speakers restructured their 'prevoicing to short lag' phonetic space into three individual categories: short lag > 20 ms (Dutch /p/, /t/, /k/), short lag ~10 ms (German /b/, /d/ and sometimes Dutch /b/, /d/), and prevoicing (Dutch /b/, /d/ and sometimes German /b/, /d/). This L1-German–L2-Dutch phonetic system displays absolute phonological differentiation between voiceless and voiced plosives, as well as absolute by-language differentiation between Dutch and German voiceless plosives, but gradient by-language differentiation between Dutch and German voiced plosives.

The L1G–L2D speakers seem to have attained native-like Dutch short lag VOT, at least for /p/ and /t/, but they did not yet reach native-like consistent prevoicing. In German, their VOT partly seems to be affected by language attrition, as revealed by shorter than monolingual-like VOT in voiceless plosives. Voiced plosives, by contrast, seem to remain unaffected by language attrition.

The L1D–L2G speakers produced voiceless plosives with longer VOT in L2-German ($M = 38$ ms) than in L1-Dutch ($M = 21$ ms), but they prevoiced the majority of voiced plosives in both L2-German (76%) and L1-Dutch (87%). The L1D–L2G speakers seem to have three phonetic categories: a new L2 long lag category ~40 ms (German /p/, /t/, /k/), their pre-existing L1 short lag category ~20 ms (Dutch /p/, /t/, /k/), and a prevoiced category that merges L2 with L1 voiced plosives (Dutch and German /b/, /d/). Their L1-Dutch–L2-German phonetic space displays absolute phonological differentiation between voiceless and voiced plosives, whereas by-language differentiation between Dutch and German is present for voiceless plosives, but absent for voiced plosives.

The L1D–L2G speakers' differentiation between voiceless plosives between Dutch and German does not go hand in hand with attainment of native-like VOT in German. They hardly aspirate /p/ ($M = 23$ ms) and produce less aspiration in /t/ ($M = 48$ ms) and /k/ ($M = 44$ ms) than the L1G–L2D speakers. Similarly, they prevoiced a higher percentage of voiced plosives in L2-German (76%) compared to the L1G–L2D speakers (32%). Despite the L1D–L2G speakers' exposure to German at home, their Dutch VOT was not affected by attrition and remained similar to that of monolingual native speakers of Dutch.

V   Discussion

In the following, we first interpret the results in light of the Speech Learning Model's (SLM) equivalence classification and contrast maintenance hypotheses (Flege, 1995). We then discuss immersion and language use, articulatory constraints, and foreign accentedness as additional explanations of the results.

1   *Equivalence classification and contrast maintenance*

The SLM (Flege, 1995) attempts to explain L2 phonetic attainment in relation to the L1 phonetic system. The two main concepts applicable to this study are equivalence classification and contrast maintenance. Differential acquisition, that is deviation from native norms, was observed in the L1D–L2G speakers for both L2-German voiceless and voiced plosives, and in the L1G–L2D speakers for L2-Dutch voiceless /k/ and voiced plosives.

One account within the SLM to explain such differential acquisition is equivalence classification (Flege, 1987, 1995): L2 speakers perceive L2 sounds into their preexisting L1 categories, and thus produce them in line with their L1 categories. However, equivalence classification cannot explain the specific patterns of differential acquisition in the present results. The L1G–L2D speakers prevoiced less frequently in Dutch than native speakers, but they prevoiced more frequently in L2-Dutch than in L1-German. Similarly, the L1D–L2G speakers did not produce native-like aspiration in L2-German, but they produced voiceless plosives with longer VOT in L2-German than in L1-Dutch. The observed differences between Dutch and German in the L1G–L2D speakers and the L1D–L2G speakers indicate that they perceive differences between the respective Dutch and German plosives. An alternative account for the differential acquisition of Dutch prevoicing and German aspiration lies in articulatory constraints, as discussed in detail below.

Equivalence classification has further limitations explaining the L1D–L2G speakers' transfer of prevoicing from L1-Dutch to L2-German. Prevoicing is the main cue for Dutch native listeners' voicing perception (Van Alphen and Smits, 2004). Equivalence classification would thus predict that the L1D–L2G speakers perceive German short lag plosives into their equivalent Dutch short lag voiceless category and thus produce German voiced plosives without any prevoicing. The need to maintain contrast between L2-German voiceless and voiced plosives offers an alternative explanation for the L1D–L2G speakers transfer of prevoicing to German.

Contrast maintenance is a second hypothesis within the SLM to explain differential L2 phonetic acquisition, and suggests acquisition of deviating phonetic categories in L2 to maintain contrast with already existing phonetic categories. The L1D–L2G speakers may need to produce prevoicing in L2-German to maintain a distinction between their voiced and voiceless categories. The VOT of their German voiceless plosives, especially in /p/, is perhaps too short to be contrasted with target-like short lag voiced plosives (Flege and Eefting, 1987a; Keating, 1984).

In contrast to the SLM's predictions of differential acquisition, the L1G–L2D speakers reached native-like VOT in L2-Dutch /p/ and /t/. Their short lag space was initially occupied by L1-German voiced plosives, and therefore acquiring L2-Dutch short lag voiceless plosives constitutes an intricate task: keeping L2-Dutch voiceless short lag plosives separate from L1-German voiced short lag plosives requires restructuring of L1 phonetic categories. Native-like L2 phonetic categories can thus be acquired under favorable conditions, including long-term L2 immersion with diverse L2 use, simple articulatory gestures, and the social need to reduce a potential foreign accent. The effect of these conditions on L2 attainment and L1 attrition is discussed in detail below.

2   Immersion and language use

The two investigated immersion contexts, full immersion in an L2 environment and immersion in the L2 at home, are comparable in that both contexts involve natural and frequent use of the L2. Full L2 immersion is inherently tied with L2 use in a variety of contexts and also with numerous speakers, whereas it largely limits L1 use to conversations within the family. By contrast, L2 immersion at home limits L2 use to interactions within the family, while the L1 is continuously used outside the home in a variety of contexts and with numerous speakers. Successful L2 acquisition as well as L1 attrition seem to be limited to an immersion context that involves drastic reduction of native L1 contact due to extensive L2 use, as is the case for the L1G–L2D speakers.

One aspect of full immersion that may influence the outcomes of L2 acquisition is exposure to multiple speakers, which is beneficial in monolingual and heritage L1 acquisition (Gollan et al., 2015; Seidl et al., 2014). Such diverse L2 exposure was experienced by the L1G–L2D speakers (exposed to Dutch in and outside the home), who acquired target L2-Dutch voiceless plosives, but not by the L1D–L2G speakers (exposed to German in the home) who did not acquire target L2-German plosives.

Conversely, frequent L1 contact and use in diverse contexts and with multiple speakers may be necessary to prevent phonetic L1 attrition, as has previously been suggested by Mayr et al. (2012). This hypothesis is in line with previous research that found quality and quantity of native language input to play a crucial role in L1 maintenance (De Leeuw et al., 2010). Only the L1D–L2G speakers, who were exposed to L1-Dutch outside the home, maintain native-like L1 VOT. Without frequent and diverse exposure to the L1, the more prominent L2 is likely to impact on the L1 phonetic categories. The L1G–L2D speakers, whose L1-German use was limited to the family context, were affected by L1 phonetic attrition surfacing as shorter than native-like aspiration in L1-German voiceless plosives. Diversity of language use and exposure are important topics for future research into the circumstances that lead to successful L2 acquisition and L1 maintenance.

3 Articulatory constraints

Articulatory constraints seem to be at play when it comes to successful L2 acquisition and L1 maintenance of VOT. In comparison to short lag VOT, aspiration requires an additional timing component, as the glottis must remain open during burst release and be closed shortly after. Prevoicing requires complete glottal closure, and initiation and sustainment of vocal fold vibration before burst release (Kewley-Port and Preston, 1974).

Articulatory least complex short lag VOT was successfully acquired for L2-Dutch /p/ and /t/ by the L1G–L2D bilinguals. L1 short lag VOT was furthermore successfully maintained by the L1D–L2G speakers for L1-Dutch voiceless plosives and also by the L1G–L2D speakers for L1-German voiced plosives. Despite the articulatory simplicity of short lag VOT, it is still remarkable that the L1G–L2D speakers were able to suppress their L1-German aspiration and produce short lag VOT in /p/ and /t/ in L2-Dutch. To our knowledge, such suppression of aspiration in an L2 with target short lag voiceless plosives has never been reported in late L2 learners, and instead aspiration was carried over from L1 to L2 (Flege, 1987).

Although short lag VOT is allegedly easy to produce (Kewley-Port and Preston, 1974), the L1D–L2G speakers produced L2-German voiced plosives with prevoicing instead of short lag VOT. As discussed above, the production of prevoiced voiced plosives in L2-German may be caused by the need to maintain phonetic contrast with the L2-German voiceless plosives, which were produced with shorter than target-like VOT.

Articulatory more complex aspiration was not completely acquired by the L1D–L2G speakers in L2-German. Similarly, the target aspirated L1-German voiceless plosives of the L1G–L2D speakers appear to be affected by phonetic attrition.

The articulatorily most complex Dutch prevoicing was not completely acquired by the L1G–L2D speakers, but was successfully maintained by the L1D–L2G speakers. Despite the complex velopharyngeal activity involved in the production of prevoicing, the L1G–L2D speakers, and also the German monolinguals, are well capable of initiating velopharyngeal adjustments to close the glottis prior to oral release of the consonant, as evidenced by occasional occurrences of prevoicing in their speech. They may, however, not necessarily be able to control the required muscular activities to a similar extent as native speakers of a prevoicing language, which results in overall fewer productions of prevoicing in their speech.

4 Foreign accent

Another factor contributing to successful L2 acquisition and L1 maintenance may be accentedness and the associated social stigmatization (Fuertes et al., 2012; Kinzler et al., 2007). Production of aspiration in a language without aspiration, such as Dutch, is associated with a foreign accent (Flege, 1984; Major, 1987; Riney and Takagi, 1999; Sancier and Fowler, 1997; Schoonmaker-Gates, 2015). Dutch short lag voiceless plosives were successfully acquired by the L1G–L2D speakers and maintained by the L1D–L2G speakers. The social need to avoid stigmatization may be advantageous for the suppression of aspiration in L2-Dutch and the maintenance of short lag VOT in L1-Dutch.

Not all non-native VOT productions are associated with a perceived foreign accent: when target short lag voiced plosives are prevoiced, listeners do not perceive this as

foreign accented (Hazan and Boulakia, 1993). This may explain why the L1D–L2G speakers did not suppress prevoicing in L2-German. The finding that the L1G–L2D speakers did not acquire consistent prevoicing in Dutch asks for additional explanations that can be related to articulatory complexity, as discussed in detail above.

5 Limitations

The present study comes with two limitations. First, the amount and contexts of L2 exposure are confounded with the speakers' L1: as a result of the couples living in the Netherlands, all L1-German bilinguals were exposed more to Dutch than all L1-Dutch bilinguals were exposed to German. Second, the genders were not well balanced across groups: more L1-German bilinguals were female, and more L1-Dutch bilinguals were male. Although all analyses included the variable *Gender*, the uneven distribution of males and females across groups limits statistical power for this variable, as well as for the interactions between *Gender* and *Language* or *Gender* and *Language Background*. These limitations do not affect the main conclusions we can draw from the present study because the relation between the degree of immersion and the degree of nativelikeness is not dependent on whether a bilingual speaks Dutch or German as L1. In addition, we focused on the two bilingual groups individually with respect to both their specific acquisition tasks (acquiring a prevoicing or aspirating L2) and the circumstances of their language learning and use (immersed in the society and the home or exclusively in the home). This allowed us to better understand the way in which each group extended or restructured their phonetic space to accommodate L1 and L2 plosives. As we followed this approach for each group individually, the interpretation is not dependent on the above-mentioned confounding variables. Fully disentangling the effects of the language-learning task and the language-learning circumstances will be a task for future research and would require testing an additional group of Dutch–German couples living in Germany.

VI Conclusions

The present study provided new insight into phonetic differentiation between L1 and L2, as well as L2 attainment and L1 attrition by comparing VOT productions of two groups of L2 speakers who differed in their degree of L2 immersion. Both groups used their L1 and L2 at home, but differed in their L1 vs. L2 use outside the home. Referencing the L2 speakers' speech to L1 speech of their immediate environment, rather than to a monolingual reference group, addressed the question to what extent the L2 speakers had been able to acquire the L2 from the input that is available to them. The results show that both immersion contexts allowed L2 speakers to restructure their phonetic space to accommodate old L1 and new L2 phonetic categories for voiceless plosives. Only the L1G–L2D speakers who were frequently exposed to Dutch in a variety of contexts and by multiple speakers in their country of residence restructured their phonetic space to accommodate new L2-Dutch VOT for both voiceless and voiced plosives. The acquisition of language-specific VOT did not automatically go hand-in-hand with native-like L2 acquisition. Even when the L2 plays a crucial role in everyday life, L1 phonetic attrition seems to be prevented by frequent use of and exposure to the L1 in a variety of contexts and multiple speakers, for example at the workplace. Combining speech data of

bilinguals with L1-Dutch and bilinguals with L1-German for both voiceless and voiced plosives revealed that success in acquiring native-like VOT in L2 and maintaining native-like VOT in L1 may be limited to VOT in the short lag range.

Declaration of conflicting interests

Funding

Notes

1. The participant who was first exposed to German at age 1;0 grew up close to the German border and started overhearing German as an infant, for example when his parents ran errands in Germany.
2. The Dutch monosyllabic words were sometimes complemented by the diminutive suffix *-tje*. In these cases, they were also coded as 'disyllabic' in the analyses.
3. A potential caveat in the comparison of prevoicing in Dutch between the L1G–L2D speakers and the L1D–L2G speakers is that the latter group has a higher number of males. The higher percentage of prevoicing in the L1D–L2G speakers could thus be ascribed to the gender difference rather than to the language background (Ryalls et al., 1997), even though gender is taken into account in the model. To investigate further whether the L1G–L2D speakers prevoiced less in Dutch than native speakers, a second model was run in which the Dutch monolinguals served as a reference. This model showed that the differences in the percentage of prevoicing between the L1G–L2D speakers and Dutch monolinguals was significant in a one-tailed, but not a two-tailed comparison ($\beta = 0.77$, $SE = 0.43$, $z = 1.80$, $p = .073$). The effect of language background on voiced plosives is interpreted in the following section, but the reader is asked to recall that the effect is small to marginal, depending on the reference group.

References

Abrahamsson N and Hyltenstam K (2009) Age of onset and nativelikeness in a second language: Listener perception versus linguistic scrutiny. *Language Learning* 59: 249–306.

Abramson AS and Lisker L (1973) Voice-timing perception in Spanish word-initial stops. *Journal of Phonetics* 1: 1–8.

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

Bergmann C, Nota A, Sprenger SA, and Schmid MS (2016) L2 immersion causes non-native-like L1 pronunciation in German attriters. *Journal of Phonetics* 58: 71–86.

Boersma P and Weenink D (2014) *Praat: Doing phonetics by computer [Computer program]. Version 5.4.04*. Available at: http://www.praat.org (accessed March 2014).

Bongaerts T, Van Summeren C, Planken P, and Schils E (1997) Age and ultimate attainment in the pronunciation of a foreign language. *Studies in Second Language Acquisition* 19: 447–65.

Bortolini U, Zmarich C, Fior R, and 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.

Brennan E, Ryan E, and Dawson W (1975) Scaling of apparent accentedness by magnitude estimation and sensory modality matching. *Journal of Psycholinguistic Research* 4: 27–36.

Chang CB (2012) Rapid and multifaceted effects of second-language learning on first-language speech production. *Journal of Phonetics* 40: 249–268.

Cook V (2007) The goals of ELT: Reproducing native-speakers or promoting multicompetence among second language users? In: Cummins J and Davison C (eds) *International handbook of English language teaching, Part 1, Volume 15*. New York: Springer, pp. 237–48.

De Leeuw E, Mennen I, and Scobbie JM (2012) Singing a different tune in your native language: First language attrition of prosody. *International Journal of Bilingualism* 16: 101–16.

De Leeuw E, Mennen I, and Scobbie JM (2013) Dynamic systems, maturational constraints, and phonetic attrition. *International Journal of Bilingualism* 17: 683–700.

De Leeuw E, Schmid MS, and Mennen I (2010) The effects of contact on native language pronunciation in an L2 migrant setting. *Bilingualism: Language and Cognition* 13: 33–40.

Fabiano-Smith L and 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 and Phonetics* 26: 148–63.

Ferguson C and Garnica O (1975) Theories of phonological development. In: Lenneberg EH and Lenneberg E (eds) *Foundations of language development, Vol. 1*. New York: Academic Press, pp. 153–80.

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

Fitch WT and Giedd J (1999) Morphology and development of the human vocal tract: A study using magnetic resonance imaging. *Journal of the Acoustical Society of America* 106: 1511–22.

Flege JE (1980) Phonetic approximation in second language acquisition. *Language Learning* 30: 117–34.

Flege JE (1981) The phonological basis of foreign accent: A hypothesis. *TESOL Quarterly* 15: 443–45.

Flege JE (1984) The detection of French accent by American listeners. *Journal of the Acoustical Society of America* 76: 692–707.

Flege JE (1987) The production of 'new' and 'similar' phones in a foreign language: Evidence for the effect of equivalence classification. *Journal of Phonetics* 15: 47–64.

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

Flege JE (1995) Second language speech learning: Theory, findings, and problems. In: Strange W (ed) *Speech perception and linguistic experience: Issues in cross-language research*. Timonium, MD: York Press, pp. 233–77.

Flege JE and Eefting W (1987a) Production and perception of English stops by native Spanish speakers. *Journal of Phonetics* 15: 67–83.

Flege JE and Eefting W (1987b) Cross-language switching in stop consonant perception and production by Dutch speakers of English. *Speech Communication* 6: 185–202.

Flege JE and Hillenbrand J (1984) Limits on phonetic accuracy in foreign language speech production. *Journal of the Acoustical Society of America* 76: 707–21.

Flege JE and Port R (1981) Cross-language phonetic interference: Arabic to English. *Language and Speech* 24: 125–46.

Flege JE, Frieda EM, Walley AC, and Randazza LA (1998) Lexical factors and segmental accuracy in second language speech production. *Studies in Second Language Acquisition* 20: 155–87.

Francis AL, Ciocca V, and Man Ching Yu J (2003) Accuracy and variability of acoustic measures of voicing onset. *Journal of the Acoustical Society of America* 113: 1025–32.

Freed BF (1982) Language loss: Current thoughts and future directions. In: Lambert RD and Freed BF (eds) *The loss of language skills*. Rowley, MA: Newbury House, pp. 1–5.

Fuertes JN, Gottdiener WH, Martin H, Gilbert TC, and Giles H (2012) A meta-analysis of the effects of speakers' accents on interpersonal evaluations. *European Journal of Social Psychology* 42: 120–33.

Gollan TH, Starr J, and Ferreira VS (2015) More than use it or lose it: The number of speakers effect on heritage language proficiency. *Psychonomic Bulletin and Review* 22: 147–55.

Hamann S and Seinhorst K (2016) Prevoicing in Standard German plosives: Implications for phonological representations? *Unpublished paper presented at the Thirteenth Old World Conference in Phonology (OCP), Budapest, Hungary*, January.

Hazan VL and Boulakia G (1993) Perception and production of a voicing contrast by French–English bilinguals. *Language and Speech* 36: 17–38.

Hopp H and Schmid MS (2013) Perceived foreign accent in first language attrition and second language acquisition: The impact of age of acquisition and bilingualism. *Applied Psycholinguistics* 3: 361–94.

Jessen M (1998) *Phonetics and phonology of tense and lax obstruents in German*. Amsterdam: Benjamins.

Kager R, Van Der Feest S, Fikkert P, Kerkhoff A, and Zamuner T (2007) Representations of [voice]: Evidence from acquisition. In: Van De Weijer J and Van Der Torre EJ (eds) *Voicing in Dutch: (De)voicing: phonology, phonetics, and psycholinguistics*. Amsterdam: Benjamins, pp. 41–80.

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

Kewley-Port D and Preston MS (1974) Early apical stop production: A voice onset time analysis. *Journal of Phonetics* 2: 194–210.

Khattab G (2000) VOT production in English and Arabic bilingual and monolingual children. In: Nelson D and Foulkes P (eds) *Leeds Working Papers in Linguistics and Phonetics* 8: 95–122.

Kinzler KD, Dupoux E, and Spelke ES (2007) The native language of social cognition. *Proceedings of the National Academy of Sciences of the United States of America* 104: 12577–80.

Kohler K (1977) *Einführung in die Phonetik des Deutschen [Introduction to German phonetics]*. Berlin: Erich Schmidt Verlag.

Kroll JF, Bobb SC, and Wodniecka Z (2006) Language selectivity is the exception, not the rule: Arguments against a fixed locus of language selection in bilingual speech. *Bilingualism: Language and Cognition* 9: 119–35.

Kupisch T, Akpınar D, and Stöhr A (2013) Gender assignment and gender agreement in adult bilinguals and second language learners of French. *Linguistic Approaches to Bilingualism* 3: 150–79.

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

Macken MA and Barton D (1980a) 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 MA and Barton D (1980b) The acquisition of the voicing contrast in Spanish: A phonetic and phonological study of word-initial stop consonants. *Journal of Child Language* 7: 433–58.

MacLeod AA (2016) Phonetic and phonological perspectives on the acquisition of voice onset time by French-speaking children. *Clinical Linguistics and Phonetics* 30: 584–98.

MacLeod AA and Stoel-Gammon C (2009) The use of voice onset time by early bilinguals to distinguish homorganic stops in Canadian English and Canadian French. *Applied Psycholinguistics* 30: 53–77.

Major RC (1987) English voiceless stop production by speakers of Brazilian Portuguese. *Journal of Phonetics* 15: 197–202.

Major RC (1992) Losing English as a first language. *The Modern Language Journal* 76: 190–208.

Mayr R, Price S, and Mennen I (2012) First language attrition in the speech of Dutch–English bilinguals: The case of monozygotic twin sisters. *Bilingualism: Language and Cognition* 15: 687–700.

Mennen I (2004) Bi-directional interference in the intonation of Dutch speakers of Greek. *Journal of Phonetics* 32: 543–63.

R Core Team (2013) *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing. Available at: http://www.R-project.org (accessed March 2014).

Riney T and Takagi Y (1999) Global foreign accent and voice onset time among Japanese EFL speakers. *Language Learning* 49: 275–302.

Rothman J (2007) Heritage speaker competence differences, language change, and input type: Inflected infinitives in heritage Brazilian Portuguese. *International Journal of Bilingualism* 11: 359–89.

Rothman J and Treffers-Daller J (2014) A prolegomenon to the construct of the native speaker: Heritage speaker bilinguals are natives too! *Applied Linguistics* 35: 93–98.

Ryalls J, Zipprer A, and Baldauff P (1997) A preliminary investigation of the effects of gender and race on voice onset time. *Journal of Speech, Language, and Hearing Research* 40: 642–45.

Sancier ML and Fowler CA (1997) Gestural drift in a bilingual speaker of Brazilian Portuguese and English. *Journal of Phonetics* 25: 421–36.

Schmid MS (2004) First language attrition: The methodology revised. *International Journal of Bilingualism* 8: 239–55.

Schmid MS, Gilbers S, and Nota A (2014) Ultimate attainment in late second language acquisition: Phonetic and grammatical challenges in advanced Dutch–English bilingualism. *Second Language Research* 30: 129–57.

Schmid MS and Köpke B (2007) Bilingualism and attrition. In: Köpke B, Schmid MS, Keijzer M, and Dostert S (eds) *Language attrition: Theoretical perspectives*. Amsterdam: Benjamins, pp. 1–7.

Schoonmaker-Gates E (2015) On voice-onset time as a cue to foreign accent in Spanish: Native and nonnative perceptions. *Hispania* 98: 779–91.

Scovel T (1969) Foreign accent, language acquisition and cerebral dominance. *Language Learning* 19: 245–54.

Seidl A, Onishi KH, and Cristia A (2014) Talker variation aids young infants' phonotactic learning. *Language Learning and Development* 10: 297–307.

Simon E (2009) Acquiring a new second language contrast: An analysis of the English laryngeal system of native speakers of Dutch. *Second Language Research* 25: 377–408.

Simon E and Leuschner T (2010) Laryngeal systems in Dutch, English, and German: A contrastive phonological study on second and third language acquisition. *Journal of Germanic Linguistics* 22: 403–24.

Stock D (1971) *Untersuchungen zur Stimmhaftigkeit hochdeutscher Phonemrealisationen [Investigations on the voicing of high German phoneme realizations]*. Hamburg: Buske.

Stoehr A, Benders T, van Hell JG, and Fikkert P (2017) Heritage language exposure impacts voice onset time of Dutch–German simultaneous bilingual preschoolers. *Bilingualism: Language and Cognition*. doi: 10.1017/S1366728917000116.

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

Ulbrich C and Ordin M (2014) Can L2-English influence L1-German? The case of post-vocalic /r/. *Journal of Phonetics* 45: 26–42.

Van Alphen PM and Smits R (2004) Acoustical and perceptual analysis of the voicing distinction in Dutch initial plosives: The role of prevoicing. *Journal of Phonetics* 32: 455–91.

Ventureyra VAG, Pallier C, and Yoo H (2004) The loss of first language phonetic perception in adopted Koreans. *Journal of Neurolinguistics* 17: 79–91.

Yu VY, De Nil LF, and Pang EW (2015) Effects of age, sex and syllable number on voice onset time: Evidence from children's aspirated stops. *Language and Speech* 58: 152–67.

Appendix 1

Target words

Table 7. Dutch target words.

Word	Pronunciation	Translation
bal	[ˈbɑl]	ball
bed	[ˈbɛt]	bed
beer	[ˈbeːr]	bear
boom	[ˈboːm]	tree
boot	[ˈboːt]	boat
buik	[ˈbœyk]	tummy
deur	[ˈdøːr]	door
dieren	[ˈdiːrə]	animals
dokter	[ˈdɔktər]	doctor/physician
doos	[ˈdoːs]	cardboard box
douche	[ˈduʃ]	shower
duim	[ˈdœym]	thumb
kaas	[ˈkaːs]	cheese
kast	[ˈkɑst]	cupboard
kikker	[ˈkɪkər]	frog
kip	[ˈkɪp]	chicken
koe	[ˈku]	cow
koning	[ˈkoːnɪŋ]	king
paard	[ˈpaːrt]	horse
pan	[ˈpɑn]	pot
peer	[ˈpeːr]	pear
pink	[ˈpɪŋk]	little finger
pizza	[ˈpidza]	pizza
pop	[ˈpɔp]	doll
taart	[ˈtaːrt]	pie
tafel	[ˈtaːfəl]	table
tak	[ˈtɑk]	branch
tas	[ˈtɑs]	bag
tent	[ˈtɛnt]	tent
tijger	[ˈtɛiɣər]	tiger

Table 8. German target words.

Word	Pronunciation	Translation
Ball	[ˈbal]	ball
Bär	[ˈbɛːɐ̯]	bear
Baum	[ˈbaʊm]	tree
Bett	[ˈbɛt]	bed
Biene	[ˈbiːnə]	bee
Birne	[ˈbɪɐ̯nə]	pear
Dach	[ˈdax]	roof
Daumen	[ˈdaʊmən]	thumb
Decke	[ˈdɛkə]	blanket
Doktor	[ˈdɔktoːɐ̯]	doctor/physician
Dose	[ˈdoːzə]	box
Dusche	[ˈduːʃə]	shower
Käse	[ˈkɛːzə]	cheese
Katze	[ˈkat͡sə]	cat
Kette	[ˈkɛtə]	necklace
Korb	[ˈkɔɐ̯p]	basket
Kuh	[ˈkuː]	cow
Küken	[ˈkyːkən]	chick
Pilz	[ˈpɪlt͡s]	mushroom
Pinsel	[ˈpɪnzəl]	paintbrush
Pizza	[ˈpɪt͡sa]	pizza
Pommes	[ˈpɔməs]	French fries
Puppe	[ˈpʊpə]	doll
Puzzle	[ˈpʊzəl]	jigsaw
Tasse	[ˈtasə]	cup
Teller	[ˈtɛlɐ]	plate
Tiere	[ˈtiːʀə]	animals
Tiger	[ˈtiːgɐ]	tiger
Tisch	[ˈtɪʃ]	table
Tür	[ˈtyːɐ̯]	door

Appendix 2

Research question 1: Differentiation of VOT in L1 and L2

Table 9. L1G–L2D speakers.

	β	SE	t	p
Voiceless plosives:				
Intercept	44.23	3.31	13.36	<.001
Language	16.22	2.41	6.72	<.001
Gender	2.69	2.48	1.08	>.250
WordLength	2.49	1.32	1.88	.060
PoA_LC	−15.05	3.11	−4.84	<.001
PoA_CD	5.93	3.02	1.97	.049
Language*Gender	−1.65	1.28	−1.29	.197
Language*WordLength	2.25	1.30	1.73	.084
Language*PoA_LC	−4.98	2.90	−1.72	.085
Language*PoA_CD	−6.37	2.87	−2.22	.026

	β	SE	z	p
Voiced plosives:				
Intercept	−0.69	0.49	−1.39	.165
Language	0.95	0.34	2.84	.005
Gender	1.41	0.48	2.94	.003
PoA	−0.34	0.16	−2.11	.035
Language*Gender	0.37	0.32	1.15	.250
Language*PoA	−0.21	0.16	−1.32	.187

Notes. PoA-LC = Place of Articulation: Labial vs. Coronal. PoA-CD = Place of Articulation: Coronal vs. Dorsal.

Table 10. L1D–L2G speakers.

	β	SE	t	p
Voiceless plosives:				
Intercept	37.46	2.54	14.74	<.001
Language	13.83	2.44	5.68	<.001
Gender	2.58	1.42	1.82	.069
WordLength	2.28	1.26	1.81	.070
PoA_LC	−17.63	3.23	−5.46	<.001
PoA_CD	2.06	3.12	0.66	>.250
Language*Gender	2.09	1.13	1.85	.064
Language*WordLength	2.60	1.25	2.07	.039
Language*PoA_LC	−4.75	2.97	−1.60	.110
Language*PoA_CD	−5.68	2.93	−1.94	.052

	β	SE	z	p
Voiced plosives:				
Intercept	−2.76	0.49	−5.59	<.001
Language	0.43	0.28	1.54	.124
PoA	−1.12	0.29	−3.86	<.001
Language*Gender	−0.16	0.26	−0.63	>.250
Language*PoA	−0.25	0.18	−1.36	.174

Notes. PoA-LC = Place of Articulation: Labial vs. Coronal. PoA-CD = Place of Articulation: Coronal vs. Dorsal.

Appendix 3

Research question 2: L2 VOT attainment

Table 11. L1G–L2D speakers.

	β	SE	t	p
Voiceless plosives:				
Intercept	25.91	2.40	10.80	<.001
LanguageBackground	−2.10	1.45	−1.45	.147
Gender	2.27	1.30	1.74	.082
WordLength	−0.40	1.00	−0.40	>.250
PoA_LC	−11.40	2.64	−4.32	<.001
PoA_CD	9.92	2.66	3.73	<.001
LanguageBackground*Gender	−2.23	1.46	−1.53	.126
LanguageBackground*WordLength	Missing due to convergence problems			
LanguageBackground*PoA_LC	−1.44	1.06	−1.36	.174
LanguageBackground*PoA_CD	−2.30	1.13	−2.04	.041

	β	SE	z	p
Voiced plosives:				
Intercept	−2.42	0.49	−4.93	<.001
LanguageBackground	−0.95	0.46	−2.06	.039
Gender	0.38	0.38	1.00	>.250
PoA	−0.68	0.28	−2.43	.015
LanguageBackground*Gender	−0.34	0.38	−0.89	>.250
LanguageBackground*PoA	−0.42	0.22	−1.87	.062

Notes. PoA-LC = Place of Articulation: Labial vs. Coronal. PoA-CD = Place of Articulation: Coronal vs. Dorsal.

Table 12. L1D–L2G speakers.

	β	SE	t	p
Voiceless plosives:				
Intercept	54.44	4.19	12.98	<.001
LanguageBackground	−6.57	1.65	−3.97	<.001
Gender	2.14	1.62	1.32	.187
WordLength	5.27	1.76	3.00	.003
PoA_LC	−20.11	5.13	−3.92	<.001
PoA_CD	−0.49	4.88	−0.10	>.250
LanguageBackground*Gender	0.95	2.27	0.40	>.250
LanguageBackground*WordLength	0.13	0.52	0.25	>.250
LanguageBackground*PoA_LC	0.19	1.17	0.16	>.250
LanguageBackground*PoA_CD	Missing due to convergence problems			

	β	SE	z	p
Voiced plosives:				
Intercept	−0.87	0.37	−2.34	.019
LanguageBackground	−1.06	0.28	−3.79	<.001
Gender	0.68	0.28	2.44	.015
PoA	−0.66	0.13	−4.93	<.001
LanguageBackground*Gender	−0.92	0.37	−2.49	.013
LanguageBackground*PoA	−0.20	0.13	−1.46	.144

Notes. PoA-LC = Place of Articulation: Labial vs. Coronal. PoA-CD = Place of Articulation: Coronal vs. Dorsal.

Appendix 4

Research question 3: L1 VOT maintenance

Table 13. L1G–L2D speakers.

	β	SE	t	p
Voiceless plosives:				
Intercept	67.23	4.69	14.34	<.001
LanguageBackground	−6.94	3.10	−2.24	.025
Gender	−1.25	2.89	−0.43	>.250
WordLength	4.18	1.74	2.41	.016
PoA_LC	−21.52	5.20	−4.14	<.001
PoA_CD	0.75	5.13	0.15	>.250
LanguageBackground*Gender	2.06	2.89	0.71	>.250
LanguageBackground*WordLength	−0.28	0.58	−0.49	>.250
LanguageBackground*PoA_LC	1.27	1.68	0.75	>.250
LanguageBackground*PoA_CD	−1.20	1.52	−0.79	>.250

	β	SE	z	p
Voiced plosives:				
Intercept	0.46	0.51	0.90	>.250
LanguageBackground	−0.13	0.50	−0.25	>.250
Gender	1.37	0.50	2.72	.007
PoA	−0.32	0.13	2.53	.011
LanguageBackground*Gender	0.14	0.50	0.28	>.250
LanguageBackground*PoA	−0.12	0.12	−0.94	>.250

Notes. PoA-LC = Place of Articulation: Labial vs. Coronal. PoA-CD = Place of Articulation: Coronal vs. Dorsal.

Table 14. L1D-L2G speakers.

	β	SE	t	p
Voiceless plosives:				
Intercept	21.37	1.65	12.93	<.001
LanguageBackground	1.86	1.16	1.60	.110
Gender	0.59	0.62	0.95	>.250
WordLength	−0.27	0.80	−0.33	>.250
PoA_LC	−12.95	1.93	−6.73	<.001
PoA_CD	7.59	1.95	3.90	<.001
Language*Gender	−1.16	0.62	−1.87	.061
Language*WordLength	−0.02	0.32	−0.07	>.250
Language*PoA_LC	−0.12	1.00	−0.12	>.250
Language*PoA_CD	0.18	1.07	0.17	>.250

	β	SE	z	p
Voiced plosives:				
Intercept	−3.16	0.45	−7.02	<.001
LanguageBackground	−0.06	0.44	−0.13	>.250
Gender	0.37	0.37	1.00	>.250
PoA	−0.94	0.25	−3.81	<.001
LanguageBackground*Gender	−0.37	0.37	−1.01	>.250
LanguageBackground*PoA	−0.17	0.23	−0.74	>.250

Notes. PoA-LC = Place of Articulation: Labial vs. Coronal. PoA-CD = Place of Articulation: Coronal vs. Dorsal.