VOT in French as a foreign language: A production and perception study with mono- and multilingual learners (German/Mandarin-Chinese)

Christoph Gabriel¹, Tanja Kupisch² & Jeanette Seoudy³

Johannes-Gutenberg-Universität Mainz¹, Universität Konstanz and The Arctic University of Norway², Universität Hamburg³

christoph.gabriel@uni-mainz.de, tanja.kupisch@uni-konstanz.de, j.seoudy@web.de

Abstract. This study addresses the production of Voice Onset Time (VOT) in French as a foreign language (FFL). The learners are young multilingual learners who speak Mandarin Chinese as a heritage language along with German. To determine potential effects of the learners' multilingual background, these multilingual learners were compared to three control groups: (1) monolingual German learners of FFL, (2) monolingual Mandarin Chinese learners of FFL, and (3) native speakers of French (L1). In addition, a perception experiment was carried out to find out to what extent the VOT values measured in the production data contribute to the perception of foreign accent. Our results show that multilingual learners perform similar to monolingual learners, suggesting that multilingual learners are neither advantaged nor disadvantaged when acquiring the phonology of a foreign language. Furthermore, the learners' perceived accent in the perception data study mirrors the results from the production study.
1 Introduction

Although the acquisition of non-native phonology is a well-established field of contemporary linguistic research, only little is known about phonological learning in learners who acquired more than one language during childhood and hence acquire the target language as an L3. This particularly holds for foreign language learning in migrant children who speak a so-called heritage language (HL, Valdés 2000, Montrul 2016) in addition to their dominant language, which is also the surrounding language and the language of school instruction. The goal of the present paper is to fill this research gap by investigating the production of Voice Onset Time (VOT) in voiced and voiceless stops in French as a foreign language (FFL). The participants of our study are young multilingual learners who speak Mandarin Chinese as their HL along with German. To determine potential effects of the learners’ multilingual background, the materials produced by the multilinguals are compared with data from three control groups: (1) monolingual German learners of FFL, (2) monolingual Mandarin Chinese learners of FFL, and (3) native speakers of French (L1). An additional experiment was carried out to find out to what extent the VOT values measured in the production data contribute to the perception of foreign accent. In this experiment, monolingual native speakers of French rated the learners’ perceived accent on a Likert scale, based on individual words containing stop consonants.

The paper is organized as follows. In Section 2, we summarize previous research on phonological learning in multilingual learners of French. Section 3 introduces the main characteristics of the languages of our sample, i.e. French, German, and Mandarin Chinese, with particular regard to their VOT characteristics. Section 4 introduces participants and methodology, presents the results of our production and perception study and discusses them in the broader context of the multilingual learning setting. Section 5 concludes with a short summary.

2 Exploring an understudied field: Phonological learning in multilingual learners of French

At least since the late 1970s and early 1980s, starting with seminal work on theoretical aspects of second language phonology by scholars like Eckman and Flege, research on the acquisition of foreign language phonology is a well-established field in linguistics. Based on a large sample of languages, among them French, Eckman (1977) proposed the influential Markedness Differential Hypothesis (MDH), which predicts that marked patterns of the target language are more difficult to learn than unmarked patterns. According to the Speech Learning Model, proposed by Flege (1981, 1995), phonological learning is essentially guided by the learners’ perception of the target language, which means that learners can easily acquire L2 sounds which are notably different from the L1 categories, while having severe difficulties in acquiring phonological categories that are similar to the phonological repertoire of the L1 and are thus not perceived as being different (cf. Archibald 1998 for an overview). During the past twenty years, there has been a considerable increase of studies devoted to the question of how a further foreign language is learned after a first one. By today, so-called L3 acquisition can be considered a fairly well established research field, although the domain of phonology has been comparatively less studied.
A central goal of the models proposed in the relevant literature consists in disentangling possible sources of cross-linguistic influence (CLI) or transfer (Odlin 1989, 2003)\(^1\) in learning settings that involve more than two linguistic systems and are thus more complex than the situation in which there is a single foreign language and the learners’ native language is the only possible source of CLI. Some recent L3 studies have highlighted the privileged status of transfer from the native (L1) to the target language (L3), assuming that the learners’ L1 in some way blocks their access to the system of the first foreign language (L2). For instance, Hermas (2014) investigated the acquisition of the pro-drop parameter in L3 English after L2 French by native speakers of Arabic and found that the source of transfer is the L1 (Arabic) rather than the L2 (French). Other authors have highlighted the cognitive similarity between two explicitly learned systems (L2 and L3) and consequently attributed more impact to the L2 than to the L1 regarding transfer effects in L3 learning (cf., e.g., Bardel & Falk 2012). As opposed to this view, Flynn et al. (2004), who proposed the so-called Cumulative Enhancement Model (CEM), postulated that all previously acquired languages, regardless of their status as an L1 or an L2, might have an impact on the learning of further foreign languages. Finally, the Typological Primacy Model (TPM), proposed by Rothman (2010, 2011), also assumes potential influence from any previously acquired language, but attributes more importance to the language which is typologically closer to the L3. Positive or negative transfer from the native to the target language is thus predicted to occur when the L1 shares structural properties with the L3; in turn, transfer from L2 to L3 is said to be likely when both foreign languages (L2 and L3) are typologically closely related. As mentioned above, most of the work done in field of L3 acquisition largely concentrates on the learning of the syntactic and morphosyntactic properties of the target language. As pointed out by Cabrelli Amaro (2012), L3 phonology can be considered an understudied issue, despite some outstanding work by scholars like Marx & Mehlhorn (2010), Gut (2010) or Llama et al. (2010), who looked at the production of stops in L3 Spanish.

It needs to be pointed out furthermore that research in L3 acquisition has largely focused on two learning settings. A first group of studies, e.g., Rothman (2011) or Cabrelli Amaro et al. (2015), addresses the ‘classical’ case of foreign language learning, where two (or more) foreign languages were acquired after puberty. Other scholars, e.g. Flynn et al. (2004), focused on the acquisition of a third language by (more or less balanced) bilingual learners who already grew up in a multilingual national environment (Russian/Kazakh). The particular case of foreign language learning in heritage speakers has largely been disregarded in the literature until now. Moreover, relevant studies have concentrated on the acquisition of selected morphosyntactic features of L3 English such as word order and use of determiners in German learners speaking Turkish, Russian or Vietnamese as a HL (cf. Sağın Şims ölç 2006; Kupisch et al. 2013; Siemund & Lechner 2015). An exception is a recent study by Lloyd-Smith et al. (2016), who investigated the perceived global accent of Turkish heritage speakers when speaking L3 English. These speakers were perceived to have a milder accent compared to monolingual Turkish speakers and the perceived transfer source was either German or Turkish, depending on the speakers’ relative proficiency in their early acquired languages.

Research on the acquisition of the phonological properties of L3 French in multilingual learners speaking a HL in addition to the dominant language is sparse.\(^2\) To our knowledge, the only study which directly addresses this particular case is by Gabriel et al. (2014, 2015), who investigated the acquisition of the syllable-timed rhythm of L3 French in multilingual learners who speak (syllable-timed) Mandarin Chinese as a HL in addition to their dominant language German, which is characterized by a typical stress-timed rhythm. Based on a corpus of read data, the authors found that multilingual learners who show a high degree of cross-linguistic and phonological awareness and a positive attitude towards both the foreign language (French) and the HL (Mandarin Chinese) produced the speech rhythm of FFL more target-like than their monolingual German counterparts, who had no experience with a syllable-timed language. This speaks in favor of positive transfer of the syllable-timed rhythm from the HL Mandarin Chinese to the target language French. A central goal of the present study is to test whether the positive effect of multilingualism also holds for phonological
aspects other than speech rhythm. More specifically, we ask whether multilingual learners of L3 French who speak Mandarin Chinese as a HL are advantaged over monolingual German or monolingual Chinese learners with respect to the production of voiced and voiceless stops in the target language.

3 VOT in French, German, and Mandarin Chinese

The sample of our study comprises a wide spectrum of typologically distinct languages. In this constellation, Mandarin Chinese is maximally distant from both French (target language) and German (surrounding language and language of instruction). An obvious feature of ‘otherness’ consists in the use of completely different writing systems: While the logographic script used in Mandarin Chinese represents meaningful units and thus refers to the semantic level, the Latin script used in both French and German refers to the phonological level of the languages concerned. In terms of morphosyntactic typology, Mandarin Chinese differs from the other languages in its isolating grammar, which is in contrast with the inflecting structures of French and German. The same holds, at least partly, for the phonological level: Regarding intonation, Mandarin Chinese is a tone language, expressing lexical contrasts in segmentally identical morphemes through contrasting F0 movements. French and German, by contrast, qualify as so-called “intonation-only languages” (Gussenhoven 2004: 12), which lack tonal marking of lexical contrasts, but make systematic use of F0 for prosodic phrasing and clause typing, among other things. As for the durational properties of the languages under discussion, the picture changes considerably, in so far as Mandarin Chinese patterns with French regarding its syllable-timed speech rhythm (cf. Section 2, above, and Lin & Wang 2007 for Mandarin Chinese), but contrasts with the stress-timed language German. As already stated, the rhythmic similarity of French and Mandarin Chinese has an impact on the acquisition of the durational properties of L3 French, in that Mandarin Chinese/German bilingual learners (who dispose of a syllable-timed language in their linguistic repertoire in addition to stress-timed German) have an advantage over monolingual German learners with respect to the production of French speech rhythm, at least under certain conditions (cf. Section 2, above).

Certain properties of the production of individual segments have a direct impact on speech rhythm, among them Voice Onset Time (VOT). VOT is considered to be the most salient cue differentiating the language-specific realizations of plosives. VOT refers to the interval between the release of the stop and the onset of voicing (Lisker & Abramson 1964: 389). There exist three different types of VOT:

(i) ‘voicing lead’ (voicing starts before the release),
(ii) ‘short voicing lag’ (voicing begins with the release or shortly after it), and
(iii) ‘long voicing lag’ (voicing starts late after the release).

Many of the world’s languages distinguish two categories of stops, voiced and voiceless, which, depending on the language, are associated with different types of VOT.

There are several factors that influence VOT. Place of articulation is the most important one with a hierarchy of shorter to longer VOTs ranging from /p/ over /t/ to /k/ (Lisker & Abramson 1964). VOT can further be influenced by syllable stress, speech rate (Kessinger & Blumstein 1997), word length (Lisker & Abramson 1964) and the quality of the following vowel (cf. Fischer-Jørgensen 1979: 98). Within a given language, there is also considerable regional variation. Finally, stops in isolated words are said to present longer VOTs than those in spoken sentences and in spontaneous speech (Baran et al. 1977), although this is not necessarily confirmed by all existing studies.

In German, voiced stops are produced with (ii) a short voicing lag, while voiceless stops are produced with (iii) a long lag. The voiced plosives /b d g/ are commonly produced without aspiration, while their voiceless counterparts are normally aspirated. Furthermore, aspiration does not occur in consonant clusters and is strongest before a stressed vowel (Wiese 1996: 270). For /k/, where VOT is highest, previous research has reported VOTs between 37 and 67 ms (Stock 1971; Fischer-Jørgensen 1978), but it is likely that these values were influences by dialectal variation and methodology, e.g., not always has a
distinction between stressed and unstressed syllables been made (cf. Lein et al. 2016 for discussion). More recent studies suggest much higher VOTs. Van de Weijer & Kupisch (2015) reported 76 ms on average for the spontaneous production data of seven speakers of standard German producing /k/ in stressed word-initial syllables. The potential effects of word type (function vs. content word), number of syllables (mono- vs. disyllabic) and vowel context (low vs. high) were also tested. Only vocalic context turned out to be a significant factor with the VOT values being higher after high vowels.

Mandarin Chinese has often been characterized as marking a distinction between (voiceless) unaspirated /p t k/ vs. (voiceless) aspirated /b d ɡ/ stops (Lin 2007: 38, 42f, 124; Duanmu 2000: 24). This suggests that the difference between the two Mandarin categories is less pronounced as compared to the difference in German and French. However, previous studies measuring VOT in these languages have shown that the VOTs for the unaspirated stops correspond to /b d g/ realizations in German and French. The values reported by Li (2013) for Mandarin Chinese are similar, although they are characterized as voiced (short lag) and voiceless (long lag) rather than [±aspirated]. The author reported 12.5-29.5 ms for voiced plosives and 77.9-90.8 ms for voiceless ones based on a word repetition test, including alveolar and velar stops.4 In addition, female speakers had a tendency to produce slightly longer VOTs than male speakers. According to Duanmu (2000: 24f), voiceless [p t k] can become voice in unstressed syllables: “The unaspirated stops … [p t k] can become voiced [b d ɡ] when they occur in an unstressed syllable, such as [ˈtsʰai.pa] > [ˈtsʰai.ba] ‘mouth’ … It is possible to represent [p t k] as [b d ɡ] instead”. Crucially for us, whether or not the unaspirated category in Mandarin Chinese is represented as voiced or voiceless, it is within the short lag range, just like in German.

In French, voicing lead with negative VOTs characterizes voiced stops /b d ɡ/, and short voicing lag (with VOT values defined as < 30ms) characterizes voiceless stops /p t k/. Lein et al. (2016) summarize ten previous studies on French with voiceless plosives. For /t/, they report a range of VOTs between 18 and 35 ms, and for /k/ between 32 and 49, suggesting substantially lower values than those reported for German.5 They also discuss potential confounds, such as including stressed and unstressed syllables and different methodologies. This points to the need to collect native speaker control data in new experiments to ensure that the monolingual baseline values are comparable to the experimental data.

In summary, all languages have a phonological two-way contrast for stops, but differ in its phonetic realization. German and Mandarin Chinese are similar, while differing from French. As VOT is a durational property linked to the production of plosives, it should have a direct impact on speech rhythm, in that the longer vs. shorter VOT values of aspirated vs. unaspirated stops should be mirrored in higher values for consonantal variability as compared to languages without aspirated stops. We thus ask whether the results of previous rhythmic analyses with the same speakers (Gabriel et al. 2014, 2015) can be replicated for VOT.

4 Empirical study

In this section, we present an empirical study of heritage speakers of Mandarin Chinese in Germany who acquired Mandarin Chinese and German early in life. The goal is to determine whether these learners have advantages in acquiring FFL over speakers who have acquired either only Mandarin Chinese or only German during their early years. We will present a production and a perception experiment with the same speakers. The questions we sought to answer were the following:

(i) Do multilingual learners have advantages over monolingual learners with respect to certain (supra)segmental aspects of pronunciation?
(ii) Can the results of previous rhythmic analyses performed on data produced by the same speakers (Gabriel et al. 2014, 2015) be replicated for VOT?
(iii) Can measurable differences in VOT be perceived by native speakers, thus confirming or disconfirming the results of the production data?
To answer our research questions (i) and (ii), we analyzed the production data described in Section 4.2; research question (iii) will be answered based on the results from a perception study outlined in Section 4.3.

4.1 Participants

The participants of this study were 13 Mandarin Chinese/German learners of French, heritage speakers of Chinese who attended a senior high school in Germany (Gymnasium). They were aged between 14 and 18 years and had had 2-6 years of schooling in French. In addition, we used control data from ten monolingual native speakers of French (baseline data), as well as control data from ten monolingual (L1) German learners (without any knowledge of Mandarin Chinese), and from ten monolingual (L1) Mandarin Chinese learners who did not know any German. Comparison of the multilingual group on the one hand and the L1 German and L1 Mandarin Chinese control groups on the other hand will show whether there is a difference in foreign language performance between speakers who had acquired two languages early in life vs. speakers who had acquired only one language early in life. The participant information is summarized in Table 1.

<table>
<thead>
<tr>
<th>Group</th>
<th>Purpose</th>
<th>n</th>
<th>Ages</th>
<th>Years of learning French at school</th>
<th>Educational setting</th>
</tr>
</thead>
<tbody>
<tr>
<td>German / Mandarin Chinese (HS) learners of FLL</td>
<td>Experimental</td>
<td>13</td>
<td>14-18</td>
<td>2–6</td>
<td>Senior High School (Gymnasium) in Hamburg</td>
</tr>
<tr>
<td>German learners of FFL</td>
<td>Control</td>
<td>10</td>
<td>15</td>
<td>4</td>
<td>Senior High School (Gymnasium) in Hamburg</td>
</tr>
<tr>
<td>Mandarin Chinese learners of FLL</td>
<td>Control</td>
<td>10</td>
<td>17-21</td>
<td>1–6</td>
<td>Senior High School in Beijing</td>
</tr>
<tr>
<td>Native speakers of French (L1)</td>
<td>Baseline Control</td>
<td>10</td>
<td>22-28</td>
<td>n.a.</td>
<td>University of Bordeaux</td>
</tr>
</tbody>
</table>

Table 1. Participant information.

Both our production and our perception experiment were based on reading samples from the above participants. Three types of read data were gathered from all speakers.

(i) a short text (*Nordwind und Sonne* for German, 北风和太阳 *Běifēng hé tàiyáng* for Mandarin Chinese, and *Amandine fait du sport*, taken from a textbook, for French),

(ii) 13 sentences consisting of CV syllables, and

(iii) 10 nonce words embedded in carrier dialogues.

The bilinguals were tested in all three languages. The German and Chinese monolinguals were tested in their native languages and in French, and the L1 French controls were recorded only in French.

4.2 Production data

For the analysis of VOT, we determined all stops in onset positions of stressed syllables. Since the data were not collected for the purpose of investigating VOT, the different types of stops were not equally distributed over the three types of materials and languages. We chose to analyze only alveolar stops /d/ and /t/ because they were the only ones appearing at least three times across materials and languages. The words that we eventually selected are summarized in Table 2, below. For these items, we measured VOT
using "Praat" (Boersma & Weenink 2011). We begin by comparing the VOT values for German and Mandarin Chinese, produced by the multilingual speakers, with the ones produced by the monolingual controls (Section 4.2.1), before turning to French in Section 4.2.2.

<table>
<thead>
<tr>
<th></th>
<th>German</th>
<th>Mandarin Chinese</th>
<th>French</th>
</tr>
</thead>
<tbody>
<tr>
<td>/t/</td>
<td>Tasche ‘bag’</td>
<td>là tī mò (nonce word)</td>
<td>Visiter ‘visit’</td>
</tr>
<tr>
<td></td>
<td>Lokomotive ‘locomotive’</td>
<td>nù tè kè (nonce word)</td>
<td>fauteuil ‘armchair’</td>
</tr>
<tr>
<td></td>
<td>Latino (nonce word)</td>
<td>太阳 ‘sun’</td>
<td>matin ‘morning’</td>
</tr>
<tr>
<td></td>
<td>dieser ‘this one’</td>
<td>le gú de mì (nonce word)</td>
<td>cadeau ‘gift’</td>
</tr>
<tr>
<td></td>
<td>legudemí (nonce word)</td>
<td>德国 ‘Germany’</td>
<td>les BDs ‘comics’</td>
</tr>
<tr>
<td></td>
<td>derjenige ‘the one (who)’</td>
<td>斗篷 ‘cloak’</td>
<td>dans ‘in’</td>
</tr>
</tbody>
</table>

Table 2. Stimuli measures for VOT production in the three languages.

4.2.1 VOT in monolingual German and Mandarin Chinese

The results of the analyses performed on the data are illustrated in Figure 1. The VOT values in both German and Mandarin Chinese largely confirm those previously found in the literature (cf. Section 3, above) for both monolingual and bilingual speakers.

Figure 1. VOT values for /t/ and /d/ in monolingual and bilingual German (left panel) and for /t/ and /tʰ/ in monolingual and bilingual Mandarin Chinese (right panel).

Both groups show a clear contrast between voiced /d/ and voiceless /t/ in German and between unaspirated /t/, roughly corresponding to German /d/ and aspirated /tʰ/ in Mandarin Chinese. There is no noticeably difference between monolinguals and bilinguals. Finally, the values for voiced /d/ in German are very similar to those for unaspirated stop /t/ in Mandarin Chinese, and those for voiceless /t/ in German roughly correspond to those for aspirated /tʰ/ in Mandarin Chinese. In there is cross-linguistic
influence from the early acquired languages onto French as a foreign language (FLL), it is to be expected that French /d/ is produced with short lag instead of lead voicing and French /t/ with long lag instead of short lag. The prediction is the same for monolingual and bilingual learners.

4.2.2 VOT in French as a foreign language (FFL): Comparing monolingual and bilingual speakers of Mandarin Chinese and German

Figure 2 shows the results of the VOT production of our bilingual learners in French as a foreign language, as compared to L1 speakers of Mandarin Chinese and German when speaking L2 French. The leftmost panel compares their productions of /t/, which should be produced with short lag in French, and the rightmost panel those of /d/, which should be produced with lead lag in French. The native controls produce these two stops as expected with VOT values of around 40ms for /t/ and of between -60 and -100 in French. The learners of FFL show a uniform pattern for French /d/ (panel on the right), where their values are in the short lag range, as expected under the influence of their native language(s). In this respect, they differ substantially from the native French controls. The picture for French /t/ is less uniform. The heritage speakers pattern with the monolingual German learners of French in producing VOTs in the long lag range, as expected under the influence of German. The monolingual Mandarin learners, by contrast, produce VOTs in the short lag range, with values even shorter than those of the native French controls, thus “undershooting” the target values.

Figure 2. VOT values for /t/ (left panel) and /d/ (right panel) in native and non-native French. From left to right: L1 French (monoling. French), French as a foreign language produced by the monolingual German learners (monoling. German), by the multilingual learners (heritage speakers of Chinese, biling. Ger/Man), and by the monolingual Chinese learners (monoling. Mandarin).
4.3 Perception data

In this part of our study, data from the production experiment were rated for perceived foreign accent. The raters who judged the participant data for the degree of foreign accentedness were adult native speakers of French from three French-speaking countries including France, Canada and Switzerland (mean age: 38.3 years; sex: 22 male, 42 female). We found no noticeable inter-rater discrepancies depending on their country of origin, which is why we included them all. Almost half of them (n=29) had expert knowledge in phonetics and/or phonology. Only three raters claimed to speak some Mandarin Chinese, while almost half of them (n=30) spoke German at different proficiency levels. The test was carried out using the online form building software ™Formstack.

4.3.1 Preparation, materials and methods

We used the recordings of two trisyllabic sequences, one containing the voiced stop [d], *les BDs* [le.be.ˈde] ‘the comics’, and one containing the voiceless stop [t], *visiter* [vi.zi.ˈte] ‘to visit’. The decision in favor of these two stimuli was based on the criteria that (a) we intended to compare sequences in which the stop occurred in the onset position of a stressed syllable and (b) that the vocalic environment was the same for both. In this case, both stops were followed by the mid front vowel [e]. We were primarily interested in judgments that were caused by a different production of the stops. Since it was obvious that many stimuli also sounded foreign due to the intonational contour and rhythmic patterns that were not typical of French, we manipulated the F0 contour and the duration of the stressed vowel, making the stimuli sound more “French” where necessary.

For the experiment, we then created a task with stimuli from a subgroup of the participants in the production task. We included five speakers from each of the four groups, presenting them to the raters in a randomized order. The raters’ task was to rate each of the 20 stimuli for degree of foreign accentedness on a seven-point Likert scale (0=no foreign accent, 6=strong foreign accent). The experiment was carried out in two parts, i.e., separately for [d] and [t].

4.3.2 Results

Figure 3 shows the results of the perception study, on the left for the stimulus containing voiceless [t], and on the right for the stimulus containing voiced [d], for all four speaker groups. The y-axis shows the degree of perceived accent, ranging from 0 (no accent) to 6 (strong foreign accent).

![Figure 3](image-url)
With respect to *visiter* (voiceless stop), the L1 French speakers were perceived as different from the other three groups (i.e. as more native-like), indicating that the raters were able to distinguish the native and the non-native speech samples. Like in the production study, the monolingual Mandarin Chinese learners stand out, as they were perceived as relatively more native-like (median around 2). Monolingual German speakers and heritage speakers of Chinese, by contrast, were perceived as having relatively similar accents, stronger compared to those of the native French group (median around 4), which again mirrors the production data. With respect to *les BDs* (voiced stop), L1 French speakers were again perceived as different from the other groups, although this time with a slightly stronger accent as in the other experiment. In other words, the native/non-native contrast is less clear for the sequence containing [d] than for the one containing [t]. The accent strength of the non-native groups was perceived as being similar (medians of 3–4), with a larger degree of overall variation. In summary, the perception data mirror the production data, suggesting that the stops contributed to the perceived of foreign accent.

### 4.4 Discussion

The results of our empirical study indicated that the VOT values are transferred from the learners’ early acquired language (Mandarin Chinese and/or German) to the target language (French). This holds for the multilinguals, who speak Mandarin Chinese as a HL along with German, as well as for the monolingual German control group. The results for /t/ in FFL from the multilingual group are only slightly more target-like than those from monolingual German learners (cf. Table 2). The monolingual Mandarin Chinese controls, by contrast, do not transfer from their L1 categorically. They pattern with the other learners in terms of the voiced stop /d/, but in the case of /t/, they appear to undershoot the target values, thus neutralizing the phonological contrast of the French plosives /t/ vs. /d/ (i.e. [+-voiced]) in their productions. This result is surprising, since these learners’ voiceless stops in Mandarin Chinese, i.e. in their L1, are clearly produced within the long lag range (cf. Table 1). As a consequence, the undershooting of the target values for /t/ cannot be interpreted in terms of transfer from the learners’ L1. To explain this result, we need to refer to extra-linguistic data instead: It became clear from the semi-focused interviews that we conducted with the learners prior to the production experiments that the monolingual Mandarin learners were explicitly instructed by their teachers to avoid aspiration in the target language. However, we suspect that they were not told to produce pre-voicing with French /b d g/.

With regard to the question of whether multilingual learners are in general advantaged over monolingual learners when acquiring foreign languages (cf. Section 4, research question (i)), our results suggest neither a disadvantage nor an advantage. Instead, the more or less target-like production of a specific feature in a foreign language (in our case VOT) depends on either linguistic factors (i.e. the VOT characteristics of the L1) or on extra-linguistic aspects such as particular instructions given by the teachers in the foreign language classes (cf. above).

As for research question (ii), our results show that the potential impact of VOT—a durational property linked to consonant production—on speech rhythm is not reflected in the data: The VOT values for German and Mandarin Chinese pattern alike for both the voiceless (aspirated) and the voiced (unaspirated) stops (cf. Fig. 1, above), while French is set apart from the other language in presenting considerably lower VOTs for the segments in question (cf. Table 2, leftmost column in both panels). With regard to speech rhythm, by contrast, there is a clear difference between German and Mandarin Chinese, which, in turn, patterns alike with French. As can be seen in Fig. 4, the values expressing consonantal variability are higher in German as compared to both Mandarin Chinese and French; cf. the values for the Pairwise Variability Index for consonantal intervals (CrPVI), depicted on the y-axis). This may explain why a positive effect of Mandarin Chinese as a HL was detected with respect to speech rhythm in previous studies on FFL (cf. Section 2 and Gabriel et al. 2014, 2015), but not for VOT in the current study. The results of previous rhythmic analyses with the same speakers thus cannot be directly replicated for VOT. The higher PVI values for German must consequently be explained by resorting to other factors, such as consonant clusters (which may be complex in German, but are virtually absent in Mandarin Chinese)

Our findings thus support the critical view of the PVI put forward by Arvaniti’s (2009), who showed that the PVI values obtained for a given language crucially depend on the choice of
the materials analyzed, which, in turn, makes the interpretation of fine-grained phonetic details such as VOT in terms of rhythmic values very difficult.

Regarding research question (iii), finally, the results of the perception study show that the native speaker raters perceived the VOT differences that we measured. The ratings are thus consistent with the results of the production study: More target-like values in stop production are reflected in better ratings; cf. the values given in Fig. 3 (Section 4.3.2) correspond with the results from the production study illustrated in Fig. 2 (Section 4.2.2). This suggests that VOT is a characteristic of consonant production, which should be taught in foreign language instruction to ensure a perceivable improvement of pronunciation.

As already stated by Fellbaum (1986), who addresses the transfer of aspiration from L1 (US) American English to L2 Portuguese, the fact that native speakers of a language with voiceless aspirated plosives transfer their long lag VOTs to the target language cannot be explained by Eckman’s (1977) markedness theory. Unaspirated plosives (as in Portuguese or French) are less marked than their aspirated counterparts in languages such as English, German or Mandarin Chinese, and should therefore be easily acquired. Many studies on consonant production in non-native speech, including our own study, have shown that this is not the case. Transfer of aspiration can better be explained by referring to Flege’s (1981, 1995) Speech Learning Model, according to which similar sounds that are perceived as being the same as in the L1 are difficult to acquire in foreign language learning. Seen within the context of current theories of L3 acquisition, finally, our results suggest that typological similarity plays a role, as predicted by Rothman’s (2010, 2011) Typological Primacy Model (cf. Section 2), if typological proximity is seen from the perspective of structural similarity rather than being based on genealogical relatedness or perceived proximity. However, since the VOT values obtained for the languages that make up the learners’ linguistic background, i.e. German and Mandarin Chinese, are quite similar, it is hard to tell whether the multilinguals rather rely on their HL (Mandarin Chinese) or on their dominant language (German) as a transfer base.

5 Concluding remarks

The results of our study on stop production in FFL in three groups of learners – (1) monolingual German learners of FFL, (2) monolingual Mandarin Chinese learners of FFL, and (3) multilingual learners of FFL speaking Mandarin Chinese as a HL along with German – suggest that multilingual learners are neither advantaged nor disadvantages when acquiring the phonology of a foreign language. We have shown furthermore that the results of the production study are reflected in the perception data.
References


1 We follow Odlin in defining transfer as “the influence resulting from the similarities and differences between the target language and any other language that has been previously (and perhaps imperfectly) acquired” (1989: 27).

2 The learning of FFL by Turkish/German bilinguals has been addressed in a few recent pedagogical studies (cf. Rück 2011; Thiele 2015). These studies, however, neither address phonological issues nor do they include empirical linguistic data.

3 Typical stress-timed languages, among them German, allow for complex syllable structures and present reduction of unstressed vowels, while syllable-timed languages such as French or Mandarin Chinese are characterized by a strong preference for unmarked CV syllables and by the lack of vowel reduction. These phonological properties have an impact on the distribution of vocal and consonant material on the phonetic surface: As compared to stress-timed languages, syllable-timed languages present a higher amount of vocalic material in the speech signal; they furthermore exhibit a higher degree a durational variability of vocalic intervals, stressed vowels being considerably longer in duration than unstressed (reduced) ones; cf. Ramus et al. (1999), White & Mattys (2007).

4 This largely confirms the results of Chao & Chen (1998: 223), who measured 14-27 ms for voiced and 82-92 ms for voiceless plosives. Lisker & Abramson (1964) reported comparable VOT values for Cantonese, i.e. 9-34 ms for unaspirated stops and 75-87 ms for aspirated ones.

5 Cf. also Splendido (2014: 71f) who gives an overview of previous VOT studies and reports average values of between 24 and 33 ms for French voiceless stops.

6 All participants have also learnt English at school. This could potentially affect their French VOTs, but the effect would be similar to that caused by the learners’ native languages, which is why we will not discuss it further here. English would, of course, be relevant if we were to test whether influence into L3 comes primarily from an L1 or an L2, but that is not the major focus of this contribution.

7 The Pairwise Variability Index (PVI), proposed by Grabe & Low (2002) computes the durational variability in successive vocalic or consonantal intervals. In the literature, it has been applied in both its non-normalized or ‘raw’ form (nPVI) and its speech rate normalized version (nPVI). See also Arvaniti (2009) for a critical view of PVI measurements.

8 For instance, German allows for up to five consonantal segments in the coda (e.g., du schimpfst [du.ʃɪmpf] ‘you grumble’), which may result in large consonantal intervals when the following syllable starts with a complex onset (e.g., du schimpfst jüngend [du.ʃɪmpfʃt.ʃtɪn.du] ‘you always grumble’). In Mandarin Chinese, by contrast, the maximum of consonants in a sequence is restricted to two segments and only occurs across syllable boundaries (e.g. 蛋糕 dàng‘ā [tan.kaʊ] ‘cake’).