

# **Cross-linguistic interaction in trilingual phonological development: The role of the input in the acquisition of the voicing contrast**

**Authors: Robert Mayr & Simona Montanari**

**This is a pre-publication version (post-peer review).**

**To cite this article:**

**ROBERT MAYR and SIMONA MONTANARI (2015). Cross-linguistic interaction in trilingual phonological development: the role of the input in the acquisition of the voicing contrast. *Journal of Child Language*, 42, pp 1006-1035 doi:10.1017/S0305000914000592, first published online 21 October 2014.**

**Link to this article:**

**[http://journals.cambridge.org/abstract\\_S0305000914000592](http://journals.cambridge.org/abstract_S0305000914000592)**

## **Abstract**

This paper examines the production of word-initial stops by two simultaneous trilingual sisters, aged 6;8 and 8;1, who receive regular input in Italian and English from multiple speakers, but in Spanish from only one person. The children's productions in each language were analysed acoustically and compared to those of their main input providers. The results revealed consistent cross-linguistic differences by both children, including between Italian and Spanish stops, although the latter have identical properties in the speech of Italian- and Spanish-speaking adults. While the children's English stops were largely target-like, their Italian stops exhibited non-target-like realisations in the direction of English, suggesting interactions. Interestingly, their Spanish productions were largely unaffected by cross-linguistic interactions, with target-like voiceless stops, and voiced stops predominantly realised as spirants. These findings raise interesting questions about phonological development in multilingual settings and demonstrate that the number and type of input providers may crucially affect cross-linguistic interactions. (150 words)

## **Introduction**

The acquisition of the voicing contrast in word-initial stops is a challenging task for bilingual children. Like their monolingual counterparts, they need to become responsive to the fine phonetic distinctions in voice onset time (VOT) that signal phonemic contrasts. This is a protracted process that takes several years to complete and involves a number of developmental stages (Macken & Barton, 1979, 1980). The task is especially complex for bilinguals since they need to learn to differentiate two sets of VOT distributions, one in each language. This is particularly difficult if the voicing distinction is implemented differently in the two languages. Not surprisingly, studies which have examined this scenario (e.g., Deuchar & Clark, 1996; Fabiano-Smith & Bunta, 2012; Kehoe, Lleó & Rakow, 2004; Khattab, 2000; ) have shown that the VOT patterns produced by bilingual children in each language influence each other, exhibiting cross-linguistic interactions. No previous study has, however, explored how children cope if the demands are even higher, i.e. where they have regular exposure to more than two languages. This paper is the first to address this issue by acoustically investigating word-initial stop productions by two simultaneous trilingual sisters growing up in California, aged 6;8 and 8;1. The children hear Italian from their mother and in their Italian-medium elementary school, English from their father, the school and the wider community, but Spanish only from their Mexican nanny. The aim of this study is to provide a first account of the acquisition of the voicing contrast by trilingual children, thereby addressing important questions, such as “How do different input settings affect phonetic/phonological acquisition?” and “Are cross-linguistic interactions less likely if the input in one language is provided by a single speaker?”

### ***The voicing contrast in English, Italian and Spanish***

English, Italian and Spanish distinguish the voiced stops /b d g/ and the voiceless stops /p t k/. However, the voicing distinction is implemented differently in these languages. The best measure to capture stop consonant voicing in word-initial position is VOT, i.e. the timing relation between release of the stop and the onset of voicing of the following segment. This timing relation can be placed on a continuum. Following Lisker and Abramson's (1964) seminal work, three types of stop voicing categories are distinguished in terms of VOT: (1) prevoiced stops, in which voicing occurs before the release of the stop, also referred to as lead voicing, (2) short-lag unaspirated stops, in which voicing is simultaneous with the release or occurs shortly thereafter, and (3) long-lag aspirated stops, in which voicing occurs with a significant time lag after the release.

In English, voiceless stops are aspirated word-initially, and thus characterised by long-lag VOT values, while voiced stops are unaspirated and typically realised with short-lag VOT values (Docherty, 1992; Lisker & Abramson, 1964). Note, however, that the latter may occur with lead voicing. For example, in Docherty's (1992) study, 7% of English voiced stops involved voicing prior to the release, and in Simon's (2009) study 27.5%. In Italian and Spanish, on the other hand, voiceless stops are unaspirated, with short-lag VOT values, while voiced stops are consistently realised with a voicing lead (cf. Bortolini, Zmarich, Fior & Bonifacio (1995) and Vaggel, Ferrero, Magno-Caldognetto & Lavagnoli (1978) for Italian; Lisker & Abramson (1964) and Rosner, López-Bascuas, Garic-Albea & Fahey (2000) for Spanish). Languages that implement the voicing contrast like Italian or Spanish are sometimes referred to as voicing languages, while languages like English are referred to as aspirating languages (Jansen, 2004). Table 1 depicts typical VOT values for word-initial voiced and voiceless stops in English, Italian and Spanish.

[Insert Table 1 about here]

VOT not only differs cross-linguistically, but also according to place of articulation: the further back a stop is produced in the oral cavity, the longer its VOT value. This is because the differing cavity sizes behind the articulators result in differences in air pressure (Cho & Ladefoged, 1999).

Inspection of Table 1 suggests that Italian and Spanish implement the voicing distinction in the same way, i.e. by contrasting lead voice and short-lag categories. However, this is only partly accurate as the voiced stops /b d g/ may only be realised as the spirants [β ð γ] in Spanish, but not in Italian. Thus, in Standard Spanish, stop and spirant realisations are in complementary distribution, with stops occurring utterance-initially, after homorganic nasals, and in the case of /d/ after laterals, while spirants occur in all other contexts (Branstine, 1991). Importantly, however, the stop/ spirant alternation rule varies considerably across different dialects. Thus, in Mexican Spanish, the variety spoken by the only person providing Spanish input in the present study, spirantisation is common, even in otherwise compulsory stop environments (Amstae, 1995; Macken & Barton, 1980). Macken and Barton (1980), for instance, found that 30%-40% of word-initial /b d g/ were spirantised by Mexican Spanish-speaking adults.

### ***Monolingual acquisition of the voicing contrast***

Monolingual children undergo several developmental stages in the acquisition of the English voicing contrast (Macken & Barton, 1979). Initially, they produce target voiced and voiceless stops in much the same way within the short-lag VOT range. Subsequently, they learn to produce a voicing difference, however with both voiced and voiceless stops still within the short-lag range. Finally, they succeed in realising voiced stops with short-lag VOT values and

voiceless stops with long-lag ones. The latter may, however, be realised with VOT values that are longer than target voiceless stops. Children typically reach this final stage around two years of age (Macken & Barton, 1979).

The voicing contrast is generally acquired later in languages that distinguish lead voice and short-lag categories (Allen, 1985; Bortoloni et al., 1995; Gandour, Petty, Dardarananda, Dechongkit & Munkgoen, 1986; Khattab, 2000, Macken & Barton, 1980). Bortoloni et al. (1995), for instance, found that at age 1;9 only three out of 14 Italian-learning children managed to produce a significant difference in voicing at each place of articulation. Similarly, Macken and Barton (1980) showed that not even by 3;10 were monolingual Spanish-speaking children able to produce a consistent difference in VOT between voiced and voiceless stops. Other studies suggest that the voicing contrast might not be adult-like in voicing languages until at least the elementary school years. Thus, in Gandour et al. (1986) only the seven-year-old children, but not the five-year-old, managed to produce Thai voiced stops with target-like prevoicing patterns, and in Khattab (2000) only the ten-year old child, but not the five-year old and the seven-year-old, exhibited consistent prevoicing of Arabic voiced stops.

Most researchers have explained the later acquisition of the lead voice/ short-lag contrast compared with the short-lag/ long-lag contrast on the basis of perceptual and articulatory difficulties. Prevoicing is acoustically less salient (Van Alphen & Smits, 2004) and articulatorily more complex than aspiration (Ohala, 1997). Specifically, the closing oral cavity involved in stop production together with the raised velum leads to a rapid increase in intra-oral air pressure. When the level of the subglottal pressure is reached, transglottal airflow stops, rendering vocal fold vibration impossible. The shorter vocal tracts of children compared with adults, in addition, make sustained voicing even more difficult. This is

particularly true for velar stops since the cavity behind the articulators is smaller than that for bilabial or coronal stops (Cho & Ladefoged, 1999).

These difficulties may prompt children to make use of compensatory strategies. Allen (1985), for instance, showed that French-learning children made use of prenasalisation in the production of voiced stops. The Spanish-learning children in Macken and Barton (1980), in contrast, realised word-initial voiced stops as spirants.

### ***Bilingual acquisition of the voicing contrast***

A number of studies have examined VOT production in bilingual children (e.g., Deuchar & Clark, 1996; Fabiano-Smith & Bunta, 2012; Heselwood & McChrystal, 2000; Kehoe et al., 2004; Khattab, 2000; Mack, 1990; Simon, 2010; Yavaş, 2002). These studies revealed that in addition to developmental factors, children's VOT patterns are affected by cross-linguistic interactions. For example, the ten-year-old French-English bilingual child in Mack (1990) produced French voiceless stops with inaccurately long VOT values, and thus in the direction of English categories. Similarly, the L1 Dutch child studied by Simon (2010) between the ages of 3;6 and 4;1 realised Dutch /p/ and /t/ with long-lag VOT values instead of target short-lag ones, following extensive exposure to English from age 3;2 when he moved from the Netherlands to the United States. Interestingly, despite such non-target-like realisations, bilingual children sometimes manage to preserve distinctions across languages. Thus, Mack's (1990) subject achieved cross-linguistic differentiation by producing both French and English voiceless stops with inaccurately long VOT values (mean VOT for French: 66 ms; mean VOT for English: 108 ms).

Cross-linguistic interactions have also been documented in voiced stops. The ten-year-old French-English bilingual in Mack's (1990) study, for instance, failed to prevoice French voiced stops consistently. Similarly, Heselwood and McChrystal (2000) found that

their ten-year-old Panjabi-English bilingual subjects used prevoicing more often in English voiced stops than age-matched English monolinguals. Interestingly, however, they did not consistently produce Panjabi voiced stops with a target-like voicing lead, suggesting a general lack of systematicity in the use of lead voicing. Note that while cross-language transfer is often invoked to explain these findings, it remains a matter of debate whether such non-target-like patterns can always be attributed to cross-linguistic effects. Khatlab (2000), for instance, found that not only Arabic-English bilinguals, aged 5;0 to 10;0, but also age-matched Arabic monolinguals used prevoicing inconsistently in voiced Arabic stops.

### *Cross-linguistic interactions and input settings*

What then are the factors that contribute to cross-linguistic effects? To begin with, certain aspects of speech development may be more prone to interactions than others. One of these may be VOT. According to Kehoe et al. (2004), what makes VOT particularly difficult to acquire is that it requires phonetic fine-tuning and automatic timing coordination. Moreover, aspiration and lead voicing are phonologically marked phenomena, and may require comparatively more input for successful acquisition.

Furthermore, interactions may be more common if the information in the input is ambiguous (Döpke, 1998; Paradis, 2000). For example, while English voiced stops are typically realised within the short-lag VOT range, they may be prevoiced, and as a result, show some resemblance to the patterns found in voicing languages, such as Italian or Spanish. This superficial similarity may lead bilingual children to equate the patterns for voiced stops cross-linguistically, realising English voiced stops predominantly with a voicing lead, or Italian and Spanish ones with short-lag values. In his Speech Learning Model, Flege (1995) refers to this phenomenon as EQUIVALENCE CLASSIFICATION.



The input may also be ambiguous and less supportive of language development if children are exposed to foreign-accented speech alongside target-like patterns. It has been proposed indeed that the phonological properties of non-native speech, either alone or in combination with native speech, provide children with a less consistent signal from which to extract language-specific phonological information and hence further develop language (e.g., Liu, Kuhl, & Tsao, 2003; Thiessen & Saffran, 2003). Place and Hoff (2011) found, for instance, that non-native input was a negative predictor of language skills among Spanish-English bilingual children growing up in the US, suggesting that specific properties of language exposure, such as the amount of input provided by native speakers, influence bilingual development. Similarly, Manuela, the bilingual child in Deuchar and Clark (1996), may have realised all her Spanish categories within the short-lag range because she not only heard target-like forms from her Spanish-speaking father, but also English-accented, short-lag Spanish stops from her English-speaking mother.

Finally, contexts that allow or require full activation and use of a bilingual's two languages may be more conducive to interactions. First, there is some evidence that language mixing in the input may interfere with the processing and learning of language-specific properties (Byers-Heinlein, 2013). In addition, multilingual contexts may increase the likelihood of influence between language systems (Grosjean, 2001). That is, in conversations with monolinguals, in which only one language can be used, bilingual speakers may be in a MONOLINGUAL MODE, where only that language is fully activated. On the other hand, dual language activation may occur in conversations with other bilinguals, in which both languages are relevant and useful to conversational needs, such as during code-switching. A number of studies have shown that under these circumstances, cross-linguistic interactions may occur more commonly. De Leeuw, Schmid and Mennen (2010), for instance, found that native German speakers with long-term residence in an L2-speaking environment were more

likely to be perceived as non-native in their L1 if they regularly engaged in code-switching. Similarly, Bullock and Toribio (2009) reported significant effects of code-switching on VOT values for Spanish-English bilinguals, suggesting that bilingual speech may be particularly vulnerable to cross-linguistic interactions when both languages are activated and alternated in discourse.

### *The present study*

This study aimed to examine the extent to which trilingual pronunciation patterns are influenced by the number and type of speakers providing the input. Specifically, it focused on the stop consonant productions of two school-aged simultaneous trilingual sisters growing up in Los Angeles. We chose comparatively old children for this study since VOT acquisition is known to be a protracted process, even in monolinguals (cf. Gandour et al., 1986; Khattab, 2000), and little is known about more advanced stages of acquisition.

The children hear (a) English, the majority language, from their father and the larger community; (b) Italian from their native-speaking mother and teachers, but also from their peers at a dual-language school who are largely from English-speaking homes; and (c) Spanish from just one person, their Mexican, Spanish-speaking monolingual nanny. Given this scenario, the study sought to address the following questions: (1) Do the children's voiceless and voiced stop productions conform to those produced by the adults providing the input? (2) Are stop categories differentiated across the various places of articulation in all three languages? (3) Are voiceless and voiced stops differentiated across English, Spanish and Italian? (4) Are there signs of interaction among the three phonological systems? (5) And if this is the case, can the interaction patterns be explained with reference to the different input settings for each language?

We predicted three possible outcomes for cross-linguistic interactions. To begin with, based on evidence from previous studies (e.g. Heselwood & McChrystal, 2000; Kehoe et al., 2004; Khattab, 2000; Mack, 1990; Simon, 2010), we hypothesised that interactions may take place between typologically different stop consonant systems. Since both Italian and Spanish are voicing languages with virtually identical systems, while English is an aspirating language, it is reasonable to assume that Italian and Spanish might exhibit the same type and extent of interaction with English. In fact, the two voicing languages may even be mutually reinforcing. Alternatively, interactions may occur, not due to typological considerations, but as a function of input characteristics, that is, when the input is ambiguous, non-native and possibly mixed (see previous section). According to this hypothesis, Spanish may be less affected by interactions than Italian because the only input provider in Spanish is monolingual, while virtually all Italian speakers in the children's environment are also competent in English, making dual language activation and interactions more likely. English, in turn, may be largely unaffected by interactions, since it constitutes the majority language and the children hear it from multiple native speakers on a regular basis. Finally, considering the children's relatively advanced age, if they have had sufficient experience with the three languages to perceive cross-linguistic differences and acquire the relevant motor commands to produce them, it is possible that they may exhibit little or no interaction in their stop consonant productions.

## **Method**

### ***Participants***

The principal participants were two simultaneous trilingual sisters growing up with English, Italian and Spanish in Los Angeles, California: Maya, aged 6;8, and Sofia, aged 8;1. The study also includes the three main sources of input in the children's home: their mother,

father and nanny. The children have been consistently exposed to the three languages from birth. They hear Italian from their Italian-speaking mother, the second author, who moved from her native San Marino to the United States at the age of 26, English from their father, a native speaker of American English with limited proficiency in Italian, and Spanish from their nanny, a native speaker of Mexican Spanish from Guadalajara who moved to the United States shortly before Sofia was born. Surprisingly, despite being a long-term resident in the United States, the nanny has no proficiency in either Italian or English.

During their first four years of life, the girls' estimated exposure to English, Spanish and Italian was approximately 24%, 33% and 43%, respectively. This estimate is based on a typical twelve-hour day. During this period, Sofia and Maya were primarily taken care of by their mother and their nanny, with the latter spending an average of 36 hours a week with the family. Note that the nanny was the sole Spanish provider as the family did not have Spanish-speaking friends and did not regularly watch Spanish-language media. English, on the other hand, was limited to evening and weekend conversations with the children's father. More consistent exposure to English began at age 4;0 for Maya, when she started to attend an English-only preschool for six hours a day, and at age 5;0 for Sofia, when she started kindergarten.

At the time of the study, both girls attended an Italian-English dual language program, with Maya in second grade and Sofia in third grade. The program follows the 90:10 model, with 90% of instruction in Italian and 10% in English in kindergarten and first grade. In second grade, the model becomes 80:20, in third grade 70:30, and each year thereafter, the amount of English instruction increases by 10% until it reaches 50% by fifth grade. Note that despite being labelled a 'dual-language program,' the children who attend it are primarily English speakers, some of Italian descent and some of other origin. Very few children start the program speaking Italian natively (cf. Montanari, 2013, for details of the program and its

history). This means that although a great deal of instruction is delivered in Italian, outside of lessons the students tend to speak English with each other, or code-switch between the two languages.

With the beginning of schooling and more after-school activities in English, the children's language exposure patterns changed. Thus, from age 6;0 to the time of the study, their exposure to English, Spanish and Italian shifted to an estimated 46%, 16% and 38%, respectively. The children's daily life revolved indeed around Italian and English, and Sofia and Maya also typically spent their summer vacation in Italy. Spanish input, in turn, was limited to the few hours spent with the nanny at home. These input patterns were nonetheless sufficient for Maya and Sofia to become fluent in all three languages, with Italian and English their strongest.

### ***Materials and procedure***

Maya and Sofia are not only able to speak English, Italian and Spanish. They are also literate in all three languages. A reading task was therefore considered appropriate. Table 2 depicts the materials used in the study. They consist of bisyllabic real words of English, Italian and Spanish with a single bilabial (/p/ or /b/), coronal (/t/ or /d/) or velar (/k/ or /g/) stop in the onset.

[Insert Table 2 about here]

The children were recorded in individual sessions in a quiet room in their home, using a Zoom H2 Handy Recorder with a sampling rate of 44.1 kHz and 16-bit resolution. They each participated in three recording sessions per language over a two-month period, thus in nine sessions in total. The Italian recording sessions were administered by the girls' mother,

the English sessions by their father, and the Spanish sessions by their nanny. They took place on different days to avoid dual or triple language activation (Grosjean, 2001). For the same reason, each recording session commenced with a brief conversation in the target language. Subsequently, the children were asked to read each stimulus word at a natural pace in two contexts, first in isolation and then in a carrier phrase, e.g. *puppy*; *puppy is what daddy said* (English); *pece*; *pece ha detto mamma* (Italian: ‘sap; sap said mummy’); *perro*; *perro me dijo Patty* (Spanish: ‘dog; dog Patty told me’). Note that the carrier phrases were matched cross-linguistically in terms of their syllabic complexity. This procedure was repeated twice in each of the three recording sessions. Across the recording sessions, this yielded 6 (consonants) x 2 (words) x 2 (repetitions) x 3 (recording sessions) x 2 (contexts) = 144 tokens per child in each language. With eight tokens excluded for poor recording quality, 429 tokens produced by Maya and 427 tokens produced by Sofia were subjected to acoustic analysis.

The children’s parents and nanny also recorded themselves, completing the same reading task as the children, however only in their respective native language. This yielded 6 (consonants) x 2 (words) x 2 (repetitions) x 2 (contexts) = 48 tokens from each speaker in their respective native languages.

### ***Data analysis***

The digitised materials were transferred to a standard PC and analysed acoustically using PRAAT software (Boersma & Weenink, 2010). VOT was measured from the release burst, signalled by a sharp peak in waveform energy, to the zero crossing of the first glottal pulse which marks the onset of voicing of the following vowel (cf. Figure 1 (a) and (b)). If voicing started during the closure period, VOT was measured from the point at which vocal fold vibration could be detected in the waveform, alongside the presence of aperiodic wide-band energy in the spectrograms, up to the release burst (cf. Figure 1 (c)). All prevoiced tokens

exhibited continuous voicing. Finally, some tokens were produced as the spirants [β], [ð] or [ɣ], rather than as stops, and as a result, VOT was not an appropriate measure. These tokens were characterised by continuous voicing and the absence of a release burst (cf. Figure 1 (d)). For further details of the acoustic properties of spirants, see Martínez-Celdrán and Regueira (2008).

[Insert Figure 1 about here]

## Results

In the following sections, the realisations of the phonologically voiceless and voiced stops will be discussed. As the data were not normally distributed, non-parametric statistical tests were used. The children and the adult participants made no difference in VOT between words produced in isolation and in a carrier sentence (Maya (English):  $U= 2463.5$ ;  $Z= -.374$ ;  $p=.709$ ; Sofia (English):  $U= 2442$ ;  $Z= -.173$ ;  $p=.862$ ; father:  $U= 272$ ,  $Z= -.33$ ,  $p= .741$ ; Maya (Italian):  $U=2360.5$ ;  $Z= -.79$ ;  $p=.43$ ; Sofia (Italian):  $U=2376$ ;  $Z= -.863$ ;  $p=.388$ ; mother:  $U= 279.5$ ,  $Z= -.175$ ,  $p= .861$ ; Maya (Spanish):  $U= 1068.5$ ;  $Z= -.609$ ;  $p=.543$ ; Sofia (Spanish):  $U= 1064.5$ ;  $Z= -.288$ ;  $p=.773$ ; nanny:  $U= 227$ ,  $Z= -.097$ ,  $p= .922$ ). As a result, the two sets of data were pooled for subsequent analysis.

### *Voiceless stops*

Table 3 presents the child and adult participants' VOT patterns for English, Italian and Spanish voiceless stops. This table allows us to directly examine whether the children's productions conformed to the adults' (Research Question 1).

[Insert Table 3 about here]

To begin with, inspection of the table shows that the adult participants' VOT values for /p t k/ are in line with those reported for their respective languages (cf. Table 1), with the father producing English voiceless stops with long-lag values, and the mother and nanny producing /p t k/ with short-lag values in Italian and Spanish, respectively. Note also that, with the exception of the father's voiceless velar stop, the adult participants produced increasingly longer VOTs as the place of articulation changed from bilabial to more posterior positions, consistent with previous studies (Cho & Ladefoged, 1999).

The table further indicates that the children managed to produce many categories with target-like VOT values. Thus, Maya's and Sofia's Spanish voiceless stops are virtually identical to the nanny's. The children also managed to produce the English voiceless stops accurately with long-lag VOT values, except for some /p/ tokens with somewhat short realisations. Their Italian categories, on the other hand, have considerably longer VOT values than their mother's, in particular /k/, which both children produced consistently as long-lag aspirated stops. Interestingly, Sofia, the elder sister, distinguished between English and Italian /k/, while the younger Maya did not. Inspection of Table 3 shows that while both realised Italian /k/ with inaccurately long VOT values, Sofia produced English /k/ with extra long values, thereby making a cross-linguistic distinction within the long-lag range.

The children's productions were not only compared with the adult target, but also with themselves in order to determine whether the children are capable of differentiating stop categories within each language (Research Question 2). To answer this question, Kruskal-Wallis tests with subsequent post hoc Mann Whitney U-tests were carried out. The results revealed that the children produced significant differences between /p/, /t/ and /k/ in each language, with VOT values systematically increasing from bilabial to coronal to velar (Maya (English):  $\chi^2(2, N=72) = 35.507, p < .001$ ; Sofia (English):  $\chi^2(2, N=72) = 41.923, p < .001$ ; Maya (Italian):  $\chi^2(2, N=72) = 41.309, p < .001$ ; Sofia (Italian):  $\chi^2(2, N=72) = 52.296, p < .001$ ;



Maya (Spanish):  $\chi^2(2, N=72) = 50.322, p < .001$ ; Sofia (Spanish):  $\chi^2(2, N=71) = 51.623, p < .001$ ). Only Maya's productions of English /t/ and /k/ did not differ significantly ( $U = 271.5; Z = -.34; p = .734$ ).

Finally, in order to determine whether the children are capable of differentiating stop categories cross-linguistically (Research Question 3), their productions of /p/, /t/ and /k/ were compared across English, Italian and Spanish (cf. Figure 2).

[Insert Figure 2 about here]

The results revealed that Maya produced a significant difference across the three languages at each place of articulation (bilabial:  $\chi^2(2, N=72) = 41.407, p < .001$ ; coronal:  $\chi^2(2, N=71) = 50.941, p < .001$ ; velar:  $\chi^2(2, N=72) = 40.972, p < .001$ ). However, there was no significant difference in VOT between her English and Italian /k/ ( $U = 284.5; Z = -.072; p = .942$ ). Similarly, Sofia produced a significant cross-linguistic difference at each place of articulation (bilabial:  $\chi^2(2, N=71) = 45.721, p < .001$ ; coronal:  $\chi^2(2, N=72) = 52.504, p < .001$ ; velar:  $\chi^2(2, N=72) = 50.005, p < .001$ ), but failed to make a significant difference between Italian and Spanish /p/ ( $U = 188.5, Z = -1.867; p = .062$ ).

Taken together, the results for the voiceless stops suggest sophisticated acquisition patterns for both children with high degrees of accuracy and differentiation. Only the children's Italian /k/ was found to be clearly non-target-like with long-lag VOT values. In the discussion, we will consider whether this pattern may be due to cross-linguistic interaction with English (Research Question 4), and whether it can be explained on the basis of the different settings in which the children receive input in their three languages (Research Question 5).

### *Voiced stops*

Table 4 depicts the adult and child participants' realisations of the phonologically voiced stops in the three languages. This table allows us to directly examine whether the children's voiced stop realisations conformed to the adults' (Research Question 1).

[Insert Table 4 about here]

First, note that the adult participants produced /b d g/ accurately, following the patterns produced by adult monolinguals of English, Italian and Spanish elsewhere (cf., Table 1). Thus, the father's English voiced stops were realised consistently within the short-lag range. The mother's productions of Italian voiced stops also conform to typical values, with all tokens realised with a voicing lead. Finally, the nanny realised Spanish /b d g/ with a voicing lead, as well, and thus in line with typical values, except for one token of /b/, which she realised as the spirant [β]. Recall that while the stop-spirant alternation rule does not apply in word-initial position in Standard Spanish (Branstine, 1991; Carrasco, Hualde & Simonet, 2012), it has been attested in this position in Mexican Spanish (Amstae, 1995; Macken & Barton, 1980).

Inspection of Table 4 shows that the children's voiced stops differ systematically from the adult targets. Specifically, Maya and Sofia realised English and Italian /b d g/ with either a voicing lead or short-lag VOT values. While English voiced stops may be produced with a voicing lead (Docherty, 1992; Lisker & Abramson, 1964), short-lag realisations of Italian voiced stops are not target-like (Bortolini et al., 1995; MacKay, Flege, Piske & Schirru, 2001; Vaggel et al., 1978). Maya, however, produced 74% (53 tokens) of Italian /b d g/ with short-lag VOT values, and Sofia 31% (22 tokens). As a consequence, the children's voiced and

voiceless categories were not always clearly contrasted in Italian. Finally, the children produced fewer than 10% of their Spanish /b d g/ tokens accurately with a voicing lead. Instead, they predominantly realised these categories as spirants, in particular /b/ and /d/, while velars were mostly realised as stops within the short-lag VOT range.

To illustrate the children's voiced stop productions further, Figure 3 presents histograms of Maya's and Sofia's stop realisations for /b d g/ in the three languages. Since Maya spirantised all her Spanish /d/ tokens, only her English and Italian /d/ tokens are included here. The figure shows a bimodal distribution for both children, with scattered tokens with negative VOT values and a large number of tokens in the short-lag VOT range. Note that Sofia prevoiced substantially more tokens at each place of articulation than did Maya. This is particularly noticeable with the coronals and velars.

[Insert Figure 3 about here]

In order to determine whether the children are capable of differentiating between different voiced stop categories in each language (Research Question 2), their realisations of /b/, /d/ and /g/ were compared with each other in English, Italian and Spanish, using non-parametric statistical tests. Since Maya spirantised all her Spanish /d/ tokens, and Sofia all except three of her Spanish /b/ tokens and all except one of her Spanish /d/ tokens, these categories were excluded from formal comparisons. The results revealed that both girls produced a significant difference in VOT between the English voiced stops (Maya:  $\chi^2(2, N=71) = 47.448, p < .001$ ; Sofia:  $\chi^2(2, N=69) = 32.452, p < .001$ ), with increasing VOT values as the place of articulation became more posterior. Maya also produced a significant difference between Italian /b d g/ ( $\chi^2(2, N=72) = 35.97, p < .001$ ) and Spanish /b/ and /g/ ( $U = 3.5; Z = -3.941; p < .001$ ). On the other hand, the difference in Sofia's realisations of Italian

voiced stops just failed to reach significance ( $\chi^2(2, N=72) = 5.815, p=.055$ ), and the difference for her Spanish /b d g/ could not be computed. Finally, to determine if the children managed to produce cross-linguistic differences in VOT (Research Question 3), their voiced stop realisations were compared across the three languages. The results revealed that Maya did not make a significant cross-linguistic difference between English, Italian and Spanish /b/ ( $\chi^2(2, N=59) = 3.461, p= .177$ ), nor between English, Italian and Spanish /g/ ( $\chi^2(2, N=60) = .096, p= .953$ ). While her English and Italian /d/ did differ significantly ( $U= 100; Z= -3.885; p<.001$ ), inspection of Figure 3 suggests a large degree of overlapping values.

Sofia, in contrast, made considerably more cross-linguistic distinctions than her younger sister. Thus, the difference between her English and Italian /b/ was significant ( $U= 166.5; Z= -2.335, p= .02$ ), as was the difference between her English and Italian /d/ ( $U= 57; Z= -4.662; p<.001$ ). Inspection of Table 4 and Figure 3 shows that she prevoiced a substantially larger number of Italian than English tokens at both these places of articulation. In contrast, the difference between her English, Italian and Spanish /g/ was not significant ( $\chi^2(2, N=66) = 3.171, p= .205$ ). Note, however, that she produced a much larger number of Italian than English tokens with a voicing lead at this place of articulation, as well, i.e. eleven versus two.

Overall, the results for the children's voiced stops indicate considerable deviations from target-like patterns and a lack of differentiation in places, in particular between Italian and English categories. In the discussion, we will consider whether these patterns may be indicative of cross-linguistic interactions (Research Question 4), and if so, whether they can be explained with reference to the children's different input settings (Research Question 5).

## Discussion

This study investigated word-initial stop productions by two simultaneous trilingual children, aged 6;8 and 8;1, growing up with English, Italian and Spanish in California, and compared them with those of the main input providers in their home, i.e. their father, mother and nanny. The results revealed a high degree of differentiation across categories by both children. This is consistent with previous work on VOT acquisition in monolingual and bilingual children of a similar age (Gandour et al., 1986; Khattab, 2000; Mack, 1990). Thus, Maya and Sofia systematically contrasted voiced and voiceless stops in each language. They also produced differences across the various places of articulation, with longer VOT values for more posterior positions (Cho & Ladefoged, 1999). The children not only differentiated stop categories within each language, but also cross-linguistically, suggesting a high degree of sophistication in their acquisition patterns. Nevertheless, not all of their realisations were target-like, and the patterns observed were highly complex. In what follows, we will discuss the findings separately for each of the children's languages, and then consider their implications for the acquisition of trilingual sound systems. In so doing, we will focus specifically on the effects of different input settings on cross-linguistic interactions.

### ***English***

English constitutes the majority language in California, and the children hear it on a regular basis when conversing with their father as well as many other native speakers in the school and the wider community. An analysis of their English stop productions revealed target-like patterns. Thus, Maya and Sofia realised English voiceless stops accurately with long-lag VOT values, except for some short /p/ tokens. Their English voiced stops, in turn, were also target-like with a preponderance of short-lag realisations alongside some prevoiced tokens. Although English /b d g/ may have a voicing lead (Docherty, 1992; Lisker & Abramson, 1964), it is interesting that the children produced this pattern considering their father's voiced

stops only had short-lag VOTs, and prevoicing is articulatorily more complex and acoustically less salient (Ohala, 1997; Van Alphen & Smits, 2004). Heselwood and McChrystal (2000) found that the Panjabi-English bilingual children in their study prevoiced more English voiced stops than age-matched monolinguals, and argued that this may be a result of interaction with Panjabi, a language in which voiced stops are consistently produced with a voicing lead. Unlike Heselwood and McChrystal's (2000) study, the present investigation did not include age-matched English monolinguals for comparison, and hence it is impossible to establish whether Maya's and Sofia's use of prevoicing in English was the result of cross-linguistic interaction with Italian and/or Spanish. It is worth noting, however, that the incidence of prevoicing in the children's productions, i.e. 13% of Maya's and 20% of Sofia's English voiced stops, is not excessive and consistent with those of English monolinguals reported elsewhere (Docherty, 1992; Lisker & Abramson, 1964; Simon, 2009). It therefore seems reasonable to conclude that the children's English productions were native-like and relatively immune to interaction from other languages.

### *Italian*

Italian is an important minority language in California with an estimated 568,000 speakers in the Los Angeles metropolitan area (OSIA, 2002). Maya and Sofia hear the language on a regular basis from multiple native and heritage speakers. These include their native Italian-speaking mother and teachers as well as their friends and relatives in Italy who they visit in the summer. At the same time, the children also hear Italian from their peers at school, who predominantly come from English-speaking homes. An analysis of the children's Italian stops revealed target-like patterns for /p/ and /t/, with short-lag VOT values. In contrast, their Italian /k/ as well as /b d g/ differed from typical adult realisations. In what follows, we will discuss these results in more detail.

To begin with, Maya and Sofia produced the voiceless velar stop with long-lag aspirated realisations instead of target short-lag ones. This is in line with Mack's (1990) and Simon's (2010) studies. It seems unlikely that this pattern is developmental as both types of categories emerge early in development and have been attested in much younger monolingual and bilingual children (Deuchar & Clark, 1996; Kehoe et al., 2004; Macken & Barton, 1979). Instead, consistent with previous work (Mack, 1990; Simon, 2010), it is more likely that the pattern has arisen as a result of cross-linguistic interactions, with Italian /k/ attracted to its aspirated English counterpart. This may have occurred because the children received English-accented input in Italian from their English-dominant peers at school. It is, however, equally possible that native-like Italian patterns caused the interaction. Thus, as far as short-lag categories are concerned, target Italian /k/ is relatively long and may contain items that overlap with aspirated stops in English, rendering them ambiguous. The monolingual adults in Bortolini et al. (1995), for instance, produced Italian /k/ with VOT values as high as 72 ms. In contrast, their maximum value for Italian /p/ was 23 ms and for Italian /t/ 35 ms. This may explain why the children only aspirated their Italian /k/, but not /p/ and /t/.

While both children were inaccurate on Italian /k/, the cross-linguistic interaction affected them differently. Thus, Maya did not produce a difference between Italian and English /k/. This suggests that she may have a merged representation that encompasses both categories, consistent with Flege's (1995) notion of equivalence classification. In contrast, Sofia produced a consistent cross-linguistic contrast within the long-lag range by realising English /k/ with extra-long VOTs (Italian median: 64.5 ms; English median: 94 ms). This pattern suggests separate representations for Italian and English /k/. A similar pattern with contrasting cross-linguistic VOT categories within the long-lag area is reported in Mack's (1990) study of a ten-year-old French-English bilingual. Why the children's realisations differed in this way is not entirely clear. However, it stands to reason that the additional

experience that Sofia has had with both languages may have helped her perceive differences between the two categories.

In addition to /k/, the children's Italian /b d g/ realisations differed from typical patterns. Thus, instead of target-like lead voicing, they exhibited a bimodal distribution, with short-lag realisations alongside lead voicing. Similar patterns are reported in Heselwood and McChrystal (2000) and Mack (1990). As in the present study, their bilingual subjects failed to realise voiced stops consistently with a voicing lead. The authors explained these patterns on the basis of cross-linguistic interactions. The same may be true in the present study. Thus, Maya and Sofia may have related Italian and English /b d g/ to each other because of the superficial structural similarity that holds across the two languages: voiced stops in both Italian and English can be produced with a voicing lead. However, they may not have realised that only English /b d g/ may also occur with short-lag VOTs, and that the lead voice/short-lag contrast signals a phonological distinction in Italian, but not in English. In addition, the children's English-dominant peers may have produced English-accented realisations of Italian voiced stops, and thus the reason for the interaction may be input-related. In concert with this interactional explanation, developmental factors may also have underpinned the children's patterns. After all, lead voicing is articulatorily complex (Ohala, 1997) and acquired late in monolinguals and bilinguals. Khattab (2000), for instance, showed that not even the seven-year-old Arabic monolingual child in her study was able to use prevoicing consistently since his voiced velar stops were largely realised with short-lag VOTs. Maya and Sofia exhibited similar patterns, with fewer prevoiced tokens of the articulatorily more complex velar stop than of bilabial and coronal categories, consistent with a developmental explanation.

Interestingly, although both children produced Italian /b d g/ with short-lag and prevoiced VOTs, they differed in the proportionate use of these categories: Maya realised the



majority of her Italian voiced stops with short-lag VOT values, and thus largely outside target-like patterns, while Sofia was much more accurate, realising them predominantly with a voicing lead. Importantly, although Maya's Italian voiced stops overlapped substantially with English ones, she prevoiced twice as many stops in Italian than English, i.e. 19 (26%) versus 9 tokens (13%). This suggests that she may have started to become attentive to the different realisations of /b/, /d/ and /g/ in the two languages. By comparison, Sofia prevoiced as many as 50 Italian tokens (69%), but only 14 English ones (20%), indicating more advanced levels of cross-linguistic differentiation. It is likely that her superior performance is again a result of her greater linguistic experience.

### *Spanish*

Spanish constitutes by far the largest minority language in California, with a population of approximately 4.4 million speakers in the Los Angeles-Long Beach conurbation (United States Census Bureau, 2013). However, Maya and Sofia only interacted in the language with one person: their monolingual Mexican nanny. An analysis of the children's Spanish stop productions revealed target-like patterns for /p t k/ with short-lag realisations that closely resemble their nanny's. In contrast, they failed to produce Spanish voiced stops with consistent prevoicing, instead realising them predominantly as spirants. When /b d g/ did occur as stops, few tokens had target-like lead voicing, with the majority realised with short-lag VOT values. How can these patterns be explained?

The Spanish stop/spirant alternation rule is complex. Recall that stops occur utterance-initially, after homorganic nasals, and in the case of /d/ after laterals, while spirants occur in all other contexts, including word-initially in connected speech (Branstine, 1991). Given the complexity of the rule, it is not surprising that it takes children a long time to acquire adult-like patterns. Thus, none of the monolingual Spanish-learning children in

Macken and Barton (1980), aged 3;10, exhibited target-like patterns, instead producing spirants in compulsory stop contexts. Although Maya and Sofia are older than these children, their overall exposure to Spanish may not differ much. Both sets of children may hence not have had enough input to determine which contexts favour stops and which spirants.

Alternatively, or in addition, Maya and Sofia may have been exposed to ambiguous input, with their Mexican nanny producing spirants in utterance-initial position. The data reported in this study only include one token of her /b/ realised as [β]. However, the recording session might have been perceived as a relatively formal occasion, and the nanny may have adapted her speech style accordingly using more Standard Spanish forms than usual (Labov, 1972). Informal observation of her speech in casual contexts certainly suggests frequent use of spirants in utterance-initial position, in line with previous accounts of Mexican Spanish adults (Amstae, 1995; Macken & Barton, 1980).

The use of spirants in this position enabled the children to distinguish voiced and voiceless categories in Spanish. It also allowed them to differentiate voiced categories cross-linguistically, with spirantisation only occurring in Spanish but not in English or Italian. At the same time, Maya and Sofia also produced a substantial number of Spanish tokens as stops, 24 (34%) and 23 (32.4%), respectively, in particular velars. Of these, some had a target-like voicing lead. However, the majority of their productions was inaccurate, with short-lag VOTs. It is not entirely clear whether these patterns have arisen from interaction or as a consequence of the children's limited input in Spanish. Together with the results for Italian and English, they suggest a lack of systematicity in the use of prevoicing, similar to the patterns reported for the Panjabi-English bilinguals in Heselwood and McChrystal (2000).

### *Cross-linguistic interactions*

The present study revealed that the children have separate stop consonant systems in each of their languages. Nevertheless, they do not constitute entirely autonomous entities since, in line with previous work on bilingual children (Heselwood & McChrystal, 2000; Kehoe et al., 2004; Khattab, 2000; Mack, 1990; Simon, 2010), they were found to interact with each other. Interestingly, however, the interactions observed were more complex than those in studies of bilingual development. Thus, while the children's English and Italian VOT patterns were related to each other, with target-like realisations in English but not in Italian, their Spanish productions were largely unaffected by the other two languages. As a consequence, Maya's and Sofia's Italian and Spanish realisations were fundamentally different. This is remarkable considering they are virtually identical in monolingual Italian- and Spanish-speaking adults (cf. Table 1), and one might hence expect them to be mutually reinforcing in the children's realisations. Only by examining trilingual development was it possible to reveal these patterns. An investigation limited to bilingual stop consonant systems would not have been in a position to identify them.

To understand the findings obtained here, it is necessary to consider the different settings in which the languages occur. To begin with, Maya and Sofia hear English on a regular basis from multiple native speakers. Although the children may also be sporadically exposed to foreign-accented speech, this exposure is not systematic, and hence the vast majority of their input is native-like. English may also show a 'majority language effect' in that it is ubiquitous in the wider community and thus constitutes a more stable input setting (Gathercole & Thomas, 2009). This may explain why the children's English was target-like and immune to the influence of Italian and Spanish.

In contrast, the children are regularly exposed to non-native models in Italian via their peers at school who are heritage speakers or L2 learners of Italian. Thus, only two out of 18 children in Sofia's class and four out of 21 in Maya's class spoke Italian when commencing

the Italian-English dual-language program (Montanari, 2013). It is therefore not surprising that the majority of children with whom Sofia and Maya interact on a regular basis speak Italian with a distinct English accent. Exposure to these models may have been significant enough to affect the children's Italian accent. Similar results have been obtained in other studies involving children on dual-language programs. Caldas (2006), for instance, attributed his daughters' English-accented French to the contact that the children had with non-native speech in their French-English dual-language school in Louisiana. The author's son, in contrast, whose education was entirely through the medium of English, spoke French natively. Caldas argued that this was due to a lack of contact with foreign-accented speech. As Maya's and Sofia's classmates are largely dominant in English, they not only speak Italian with an English accent, but also frequently engage in code-switching. This requires dual language activation and further increases the likelihood of cross-linguistic interactions (de Leeuw et al., 2010). In sum, the context in which Sofia and Maya hear Italian explains why the language was affected by English, but not by Spanish.

Finally, unlike English and Italian, the children only hear Spanish from a single source, their native Spanish-speaking nanny. As the latter speaks no other languages, Maya and Sofia are required to adopt a monolingual mode (Grosjean, 2001) when communicating with her, thereby inhibiting the use of elements from their other languages. At the same time, input from a single source will be less variable and ambiguous than input from multiple speakers, facilitating the adoption of speaker-specific patterns. A few studies in areas other than phonology have documented such patterns in trilingual children (Barnes, 2011; Cruz-Ferreira, 2006; Wang, 2008). Barnes (2011), for instance, reports observing trilingual adolescent males interacting with each other in English in a female register since they had acquired this language solely from their mother. A single, unambiguous input source may hence be beneficial for phonological acquisition, as it may lead to more firmly entrenched

storage of speaker-specific phonetic information (Allen & Miller, 2004; Smith & Hawkins, 2012). This, in turn, may limit the effects of cross-linguistic interactions. Consistent with this hypothesis, Maya's and Sofia's Spanish /p t k/ productions were unaffected by their other languages, yet remarkably similar to the adult model, while their English and Italian /p t k/ realisations were not. The effect of single-speaker input on the children's voiced categories in Spanish, on the other hand, remains unclear since we do not know the extent to which the nanny spirantises word-initial stops in casual contexts. Nevertheless, the results provide an initial indication that input from a single speaker may be conducive to phonological acquisition and inhibit cross-linguistic interactions.

## **Conclusion**

This study investigated for the first time the production of word-initial stops in school-aged trilingual children. It revealed sophisticated acquisition patterns for both children in each language, but also some non-target-like realisations arising from a complex array of factors, including cross-linguistic interactions. The study demonstrated that the nature of these interactions cannot be predicted solely on the basis of adult values. Instead, they are contingent on the specific contexts in which input is provided in each language. Settings in which more than one language needs to be fully activated may be vulnerable to cross-linguistic interactions, in particular if they involve minority languages. In addition, interactions may be more likely if foreign-accented input is involved. On the other hand, the likelihood of interactions between phonological systems may be reduced if the input provided is limited to a single speaker. This is because this setting facilitates responsiveness to the specific phonetic properties of the input provider without the need to generalise to other speakers. In the present study, this effect may have been enhanced by the fact that the input provider was monolingual. It is important to point out in this context that we do not mean to

portray cross-linguistic interactions as hindering acquisition in general. As instances of positive transfer, they may lead to enhanced cue strength and result in accelerated acquisition under certain circumstances (cf., Mayr, Howell & Lewis, 2014).

More research is required to further elucidate how the number and type of input providers affect phonological acquisition in multilingual contexts. Future studies should include larger samples together with age-matched monolingual controls. This latter issue is particularly important as it is not always possible to establish whether observable patterns are due to interaction or developmental in nature. Finally, future studies should examine the development of multiple phonological systems in children speaking the same languages but living in different social contexts, where majority and minority languages are reversed. Studies of this kind are needed to shed new light on the relationship between input settings and the extent and direction of cross-language interactions in multilingual phonological development.

## References

- Allen, G. (1985). How the young French child avoids the prevoicing problem for word-initial voiced stops. *Journal of Child Language*, 12, 37–46.
- Allen, J.S., & Miller, J.L. (2004). Listener sensitivity to individual talker differences in voice-onset time. *Journal of the Acoustical Society of America*, 116, 3171-3183.
- Amastae, J. (1995). Variable spirantization: Constraint weighting in three dialects. *Hispanic Linguistics*, 6/7, 267–285.
- Barnes, J. (2011). The influence of child-directed speech on early trilingualism. *International Journal of Multilingualism*, 8, 42-62.

Boersma, P., & Weenink, D. (2010). *PRAAT: Doing phonetics by computer* (version 5.1.31).

Institute of Phonetic Sciences, University of Amsterdam, <http://www.fon.hum.uva.nl/praat>

(retrieved 5 April, 2010).

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

Branstine, Z. (1991). Stop-spirant alternations in Spanish: On the representations of contrast. *Studies in the Linguistic Sciences*, 21, 1-22.

Bullock, B. E., & Toribio, A. J. (2009). How to hit a moving target: On the sociophonetics of code-switching. In L. Isurin, D. Winford, & K. de Bot (eds.), *Multidisciplinary approaches to code switching*, pp. 189-206. Amsterdam: John Benjamins.

Byers-Heinlein, K. (2013). Parental language mixing: Its measurement and the relation of mixed input to young bilingual children's vocabulary size. *Bilingualism: Language and Cognition*, 16, 32-48.

Caldas, S. (2006). *Raising bilingual biliterate children in monolingual cultures*. Clevedon: Multilingual Matters.

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

Cruz-Ferreira, M. (2006). *Three is a crowd? Acquiring Portuguese in a trilingual environment*. Clevedon: Multilingual Matters.

de Leeuw, E., Schmid, M., & Mennen, I. (2010). The effects of contact on native language pronunciation in an L2 migrant context. *Bilingualism: Language and Cognition*, 13, 33-40.

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

- Docherty, G.J. (1992). *The timing of voicing in British English obstruents*. Berlin/New York: Foris Publications.
- Döpke, S. (1998). Competing language structures: The acquisition of verb placement by bilingual German-English children. *Journal of Child Language*, 25, 555-585.
- Fabiano-Smith, L., & Bunta, F. (2012). Voice onset time of voiceless bilabial and velar stops in 3-year-old bilingual children and their age-matched monolingual peers. *Clinical Linguistics and Phonetics*, 26, 148-163.
- Flege, J.E. (1995). Second language speech learning: theory, findings, and problems. In W. Strange (ed.), *Speech perception and linguistic experience: Issues in cross-language research*, pp. 233–277. Timonium, MD: York Press.
- Flege, J.E., & Eefting, W. (1987). Production and perception of English stops by native Spanish speakers. *Journal of Phonetics*, 15, 67-83.
- Gathercole, V., & Thomas, E. (2009). Bilingual first-language development: Dominant language takeover, threatened minority language take-up. *Bilingualism: Language and Cognition*, 12, 213-237.
- Gandour, J., Petty, S., Dardarananda, R., Dechongkit, S., & Munkgoen, S. (1986). The acquisition of the voicing contrast in Thai: A study of voice onset time in word-initial stop consonants. *Journal of Child Language*, 13, 561–572.
- Grosjean, F. (2001). The bilingual's language modes. In J. Nicol (ed.), *One mind, two languages: Bilingual language processing*, pp. 1-22. Oxford: Blackwell.
- Heselwood, B., & McChrystal, L. (2000). Gender, accent features and voicing in Panjabi-English bilingual children. In D. Nelson & P. Foulkes (eds.), *Leeds Working Papers in Linguistics and Phonetics*, 8, 45–70.
- Jansen, W. (2004). *Laryngeal contrast and phonetic voicing: A laboratory phonology approach to English, Hungarian, and Dutch*. Ph.D. dissertation, Groningen University.



- Kehoe, M.M., Lléo, C., & Rakow, M. (2004). Voice onset time in bilingual German–Spanish children. *Bilingualism: Language and Cognition*, 7, 71–88.
- Khattab, G. (2000). VOT in English and Arabic bilingual and monolingual children. In D. Nelson & P. Foulkes (eds.), *Leeds Working Papers in Linguistics and Phonetics*, 8, 95–122.
- Labov, W. (1972). *Sociolinguistic patterns*. Oxford: Blackwell.
- Lisker, L., & Abramson, A.S. (1964). A cross-language study of voicing in initial stops: Acoustical measurements. *Word*, 20, 384–422.
- Liu, H. M., Kuhl, P. K., & Tsao, F. M. (2003). An association between mother’s speech clarity and infants’ speech discrimination skills. *Developmental Science*, 6, F1–F10.
- Mack, M. (1990). Phonetic transfer in a French-English bilingual child. In P.H. Nelde (ed.), *Language attitude and language conflict*, pp.107-124. Bonn: Dümmler.
- MacKay, I., Flege, J.E., Piske, T., & Schirru, C. (2001). Category restructuring during second-language speech acquisition. *Journal of the Acoustical Society of America*, 110, 516-528.
- Macken, M.A., & Barton, D. (1979). The acquisition of the voicing contrast in English: A study of voice onset time in word-initial stop consonants. *Journal of Child Language*, 7, 41–74.
- Macken, M.A., & Barton, D. (1980). The acquisition of the voicing contrast in Spanish: A phonetic and phonological study of word-initial stop consonants. *Journal of Child Language*, 7, 433–58.
- Martínez-Celdrán, E., & Regueira, X.L. (2008). Spirant approximants in Galician. *Journal of the International Phonetic Association*, 38, 51-68.
- Mayr, R., Howells, G., & Lewis, R. (2014). Asymmetries in phonological development: The case of word-final cluster acquisition in Welsh-English bilingual children. *Journal of*

- Child Language*, doi: 10.1017/S0305000913000603. Published online by Cambridge University Press, 14 February, 2014.
- McCarthy, K.M., Evans, B. G., & Mahon, M. (2013). Acquiring a second language in an immigrant community: The production of Sylheti and English stops and vowels by London-Bengali speakers. *Journal of Phonetics*, 41, 344-358.
- Mennen, I. (2004). Bi-directional interference in the intonation of Dutch speakers of Greek. *Journal of Phonetics*, 32, 543–563.
- Montanari, S. (2013). A case study of bi-literacy development among children enrolled in an Italian–English dual language program in Southern California. *International Journal of Bilingual Education and Bilingualism*, doi: 10.1080/13670050.2013.833892. Published online by Taylor & Francis, September 23, 2013.
- Ohala, J.J. (1997). The relation between phonetics and phonology. In W.J. Hardcastle & J. Laver (eds.) *The handbook of phonetic sciences*, pp. 674-694, Oxford: Blackwell.
- Order Sons of Italy in America [OSIA]. (2002). A profile of today's Italian Americans, [www.osia.org/documents/IA\\_Profile.pdf](http://www.osia.org/documents/IA_Profile.pdf), accessed 7 January 2014.
- Paradis, J. (2000). Beyond 'One System or Two?' Degrees of separation between the languages of French-English bilingual children. In S. Döpke (ed.), *Cross-linguistic structures in simultaneous bilingualism*, pp. 175–200. Amsterdam & Philadelphia: John Benjamins.
- Place, S., & Hoff, E. (2011). Properties of dual language exposure that influence 2-year-olds' bilingual proficiency. *Child Development*, 82(6), 1834–1849.
- Rosner, B.S., López-Bascuas, L.E., Garic-Albea, J.E., & Fahey, R.P. (2000). Voice-onset times for Castilian Spanish initial stops. *Journal of Phonetics*, 28, 217-224.
- Simon, E. (2009). Acquiring a new L2 contrast : An analysis of the English laryngeal system of L1 Dutch speakers. *Second Language Research*, 25, 377-408.

- Simon, E. (2010). Child L2 development: A longitudinal case study on Voice Onset Times in word-initial stops. *Journal of Child Language*, 37, 159-173.
- Smith, R., & Hawkins, S. (2012). Production and perception of speaker-specific phonetic detail at word boundaries. *Journal of Phonetics*, 40, 213-233.
- Thiessen, E. D., & Saffran, J. R. (2003). When cues collide: Use of stress and statistical cues to word boundaries by 7- to 9-month-old infants. *Developmental Psychology*, 39, 706–716.
- United States Census Bureau. (2013). Language use in the United States 2011: American Community Survey Reports, [www.census.gov/prod/2013pubs/acs-22.pdf](http://www.census.gov/prod/2013pubs/acs-22.pdf), accessed 9 December 2013.
- Vagges, K., Ferrero, F.E., Magno-Caldognetto, E., & Lavagnoli, C. (1978). Some characteristics of Italian consonants. *Journal of Italian Linguistics*, 3, 69-85.
- Van Alphen, P., & Smits, R. (2004). Acoustical and perceptual analysis of the voicing distinction in Dutch initial plosives: the role of prevoicing. *Journal of Phonetics*, 32, 455-491.
- Wang, X. (2008). *Growing up with three languages*. Clevedon: Multilingual Matters.
- Yavaş, M. (2002). VOT patterns in bilingual phonological development. In F. Windsor, L. Kelly & N. Hewitt (eds.), *Themes in clinical linguistics*, pp. 341–350. Mahwah, New Jersey: Lawrence Erlbaum Associates.

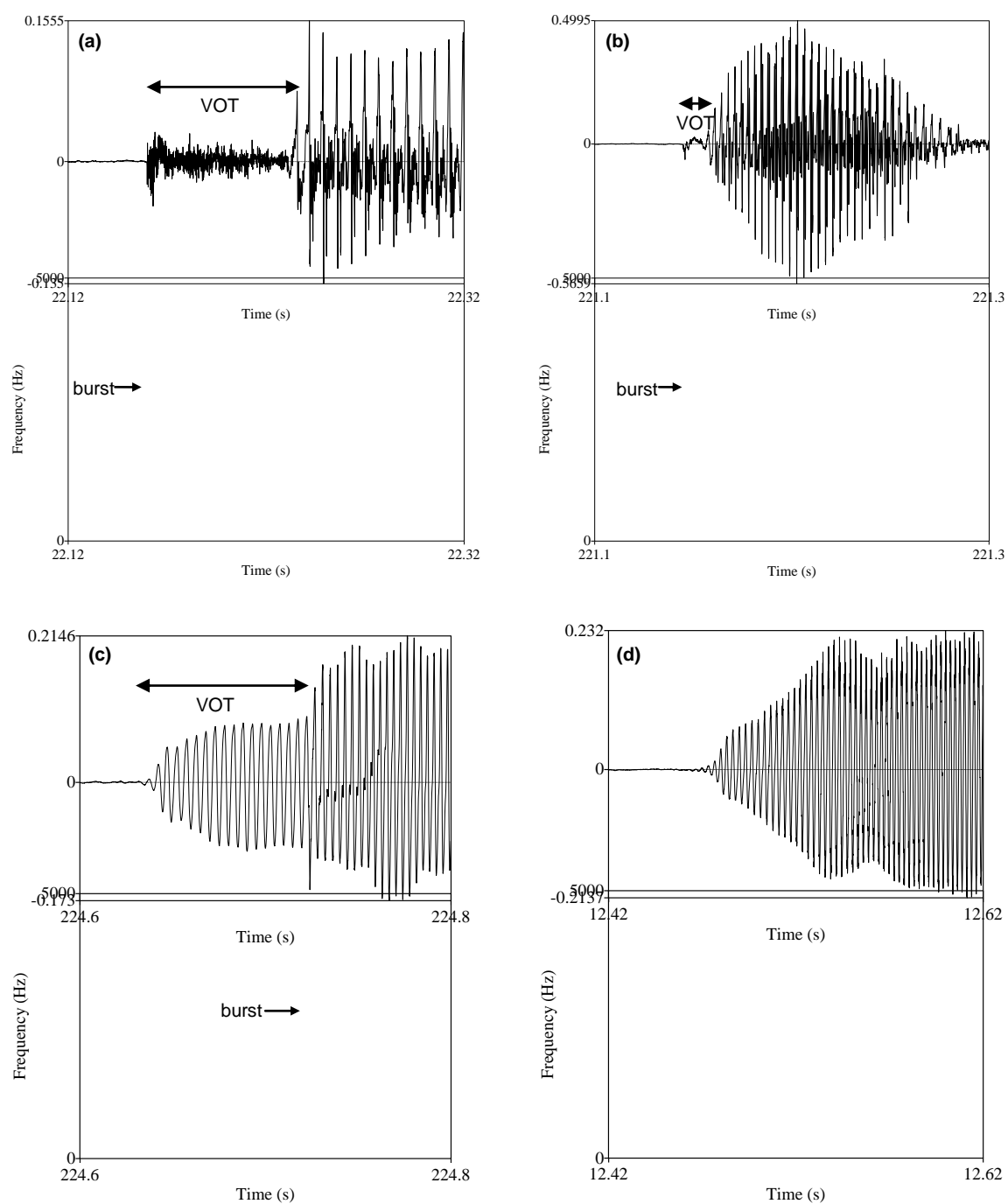


Figure 1. Waveform and spectrogram of plosive realised with long-lag VOT (a), short-lag VOT (b) and lead VOT (c), and waveform and spectrogram of spirant (d); all 200 ms in duration.

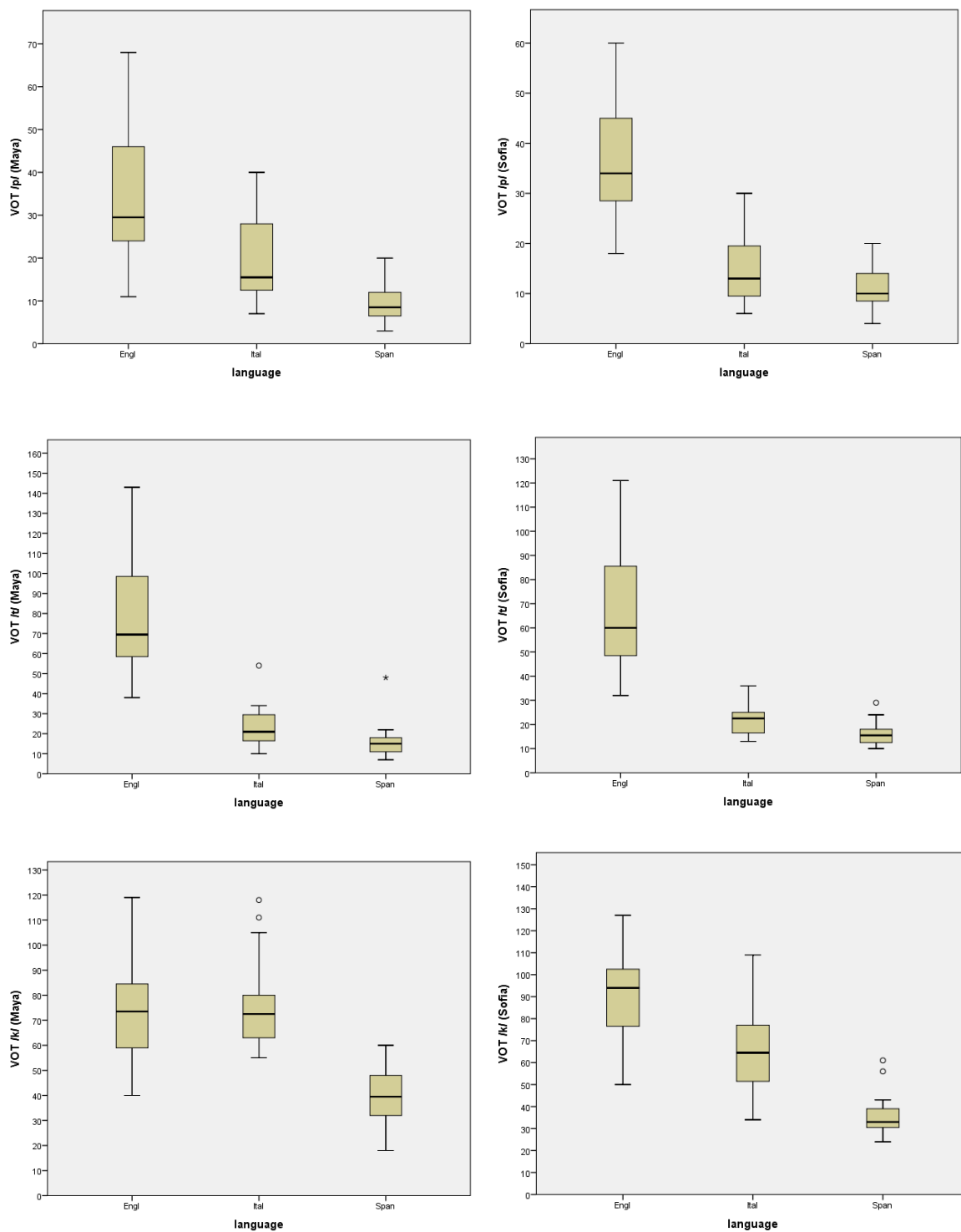
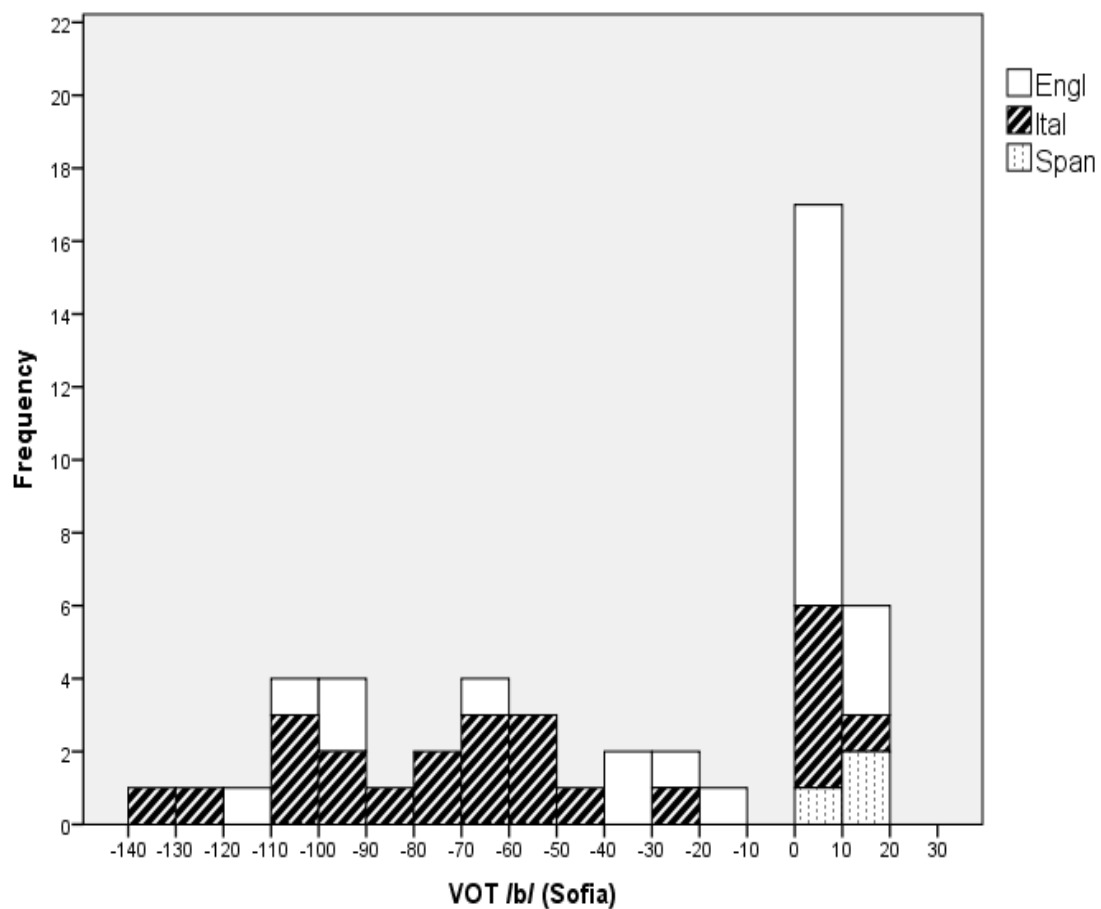
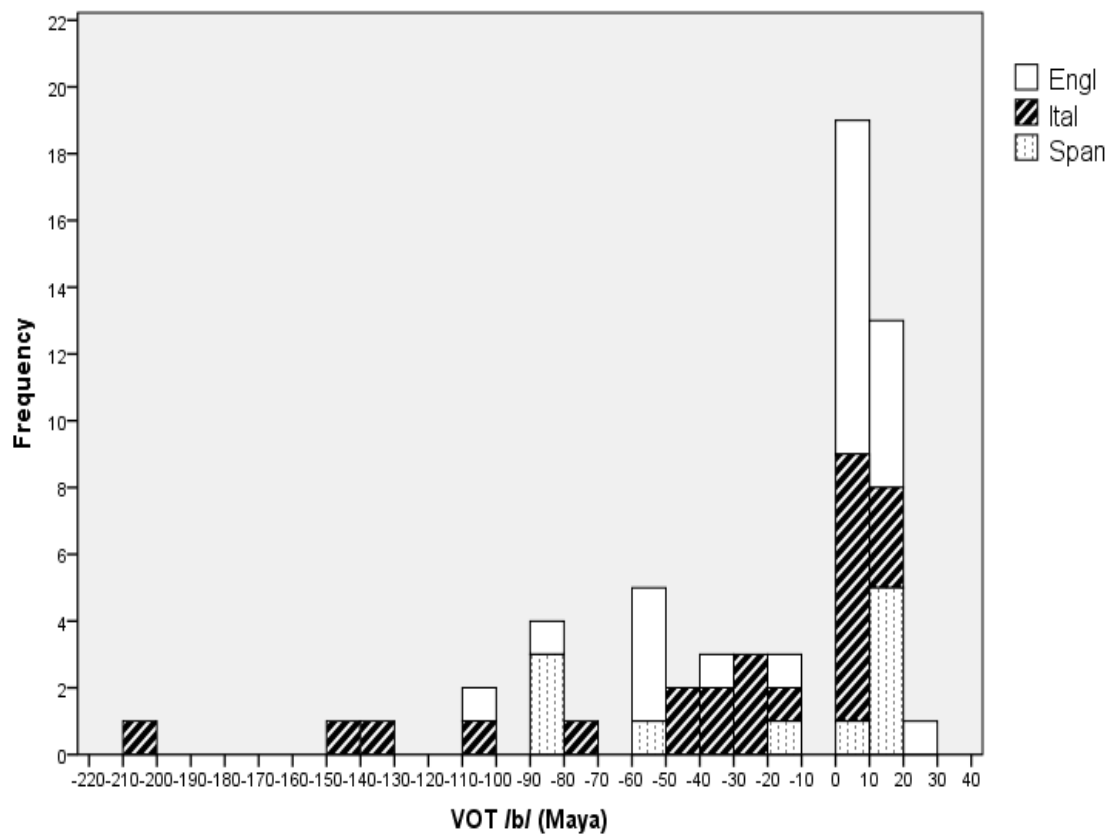
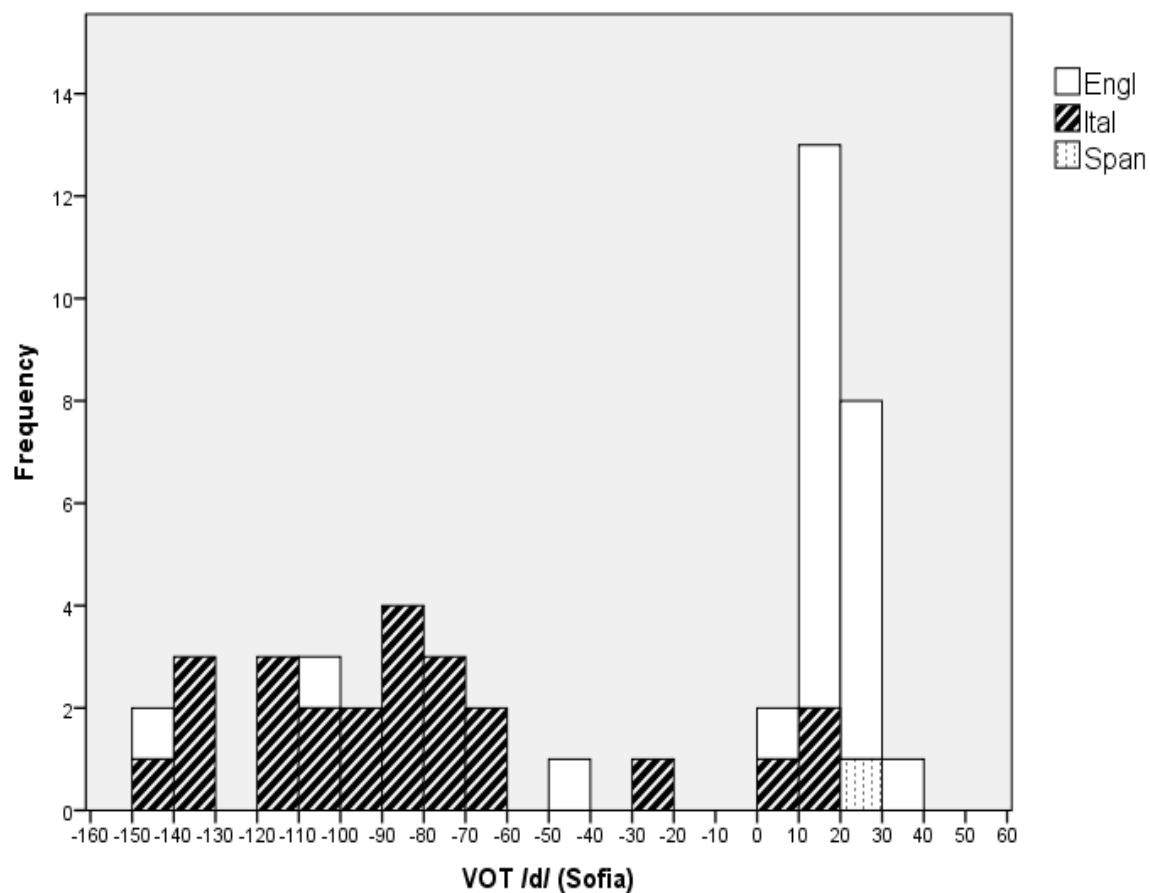
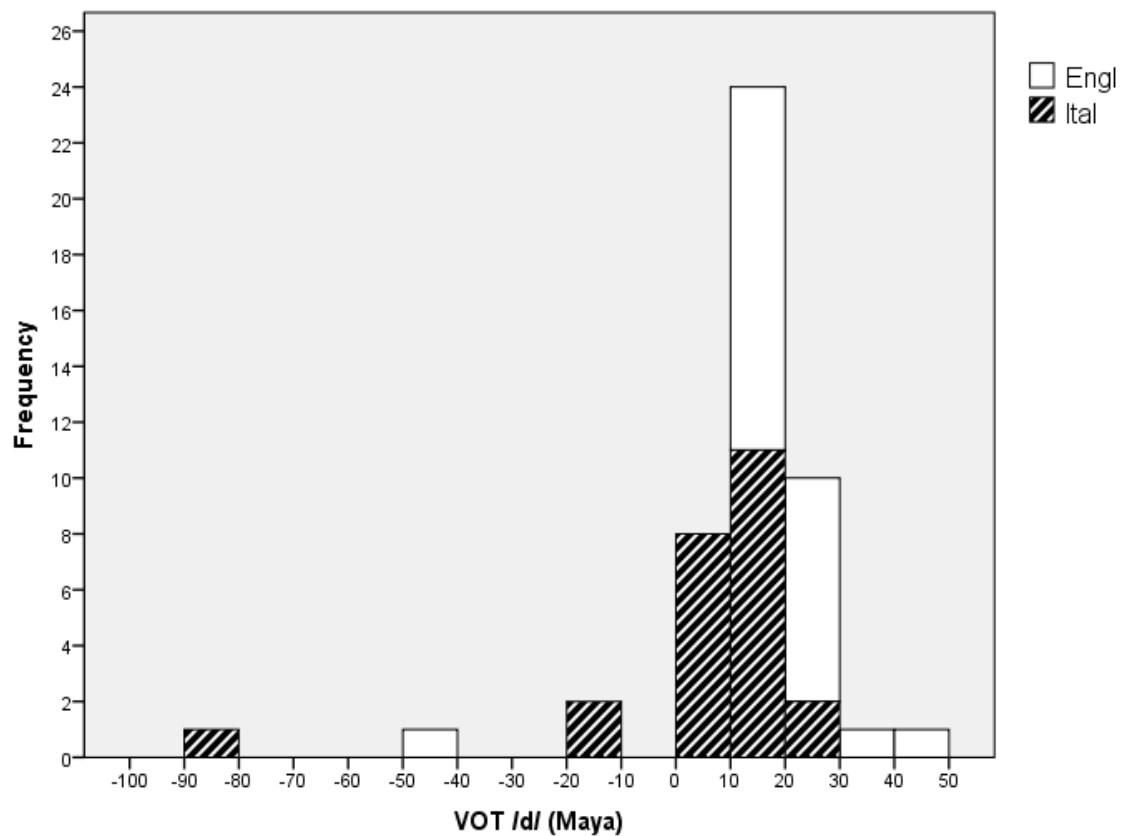


Figure 2. Maya's (left) and Sofia's (right) VOT distributions (in ms) for /p/ (top), /t/ (middle) and /k/ (bottom) in English, Italian and Spanish





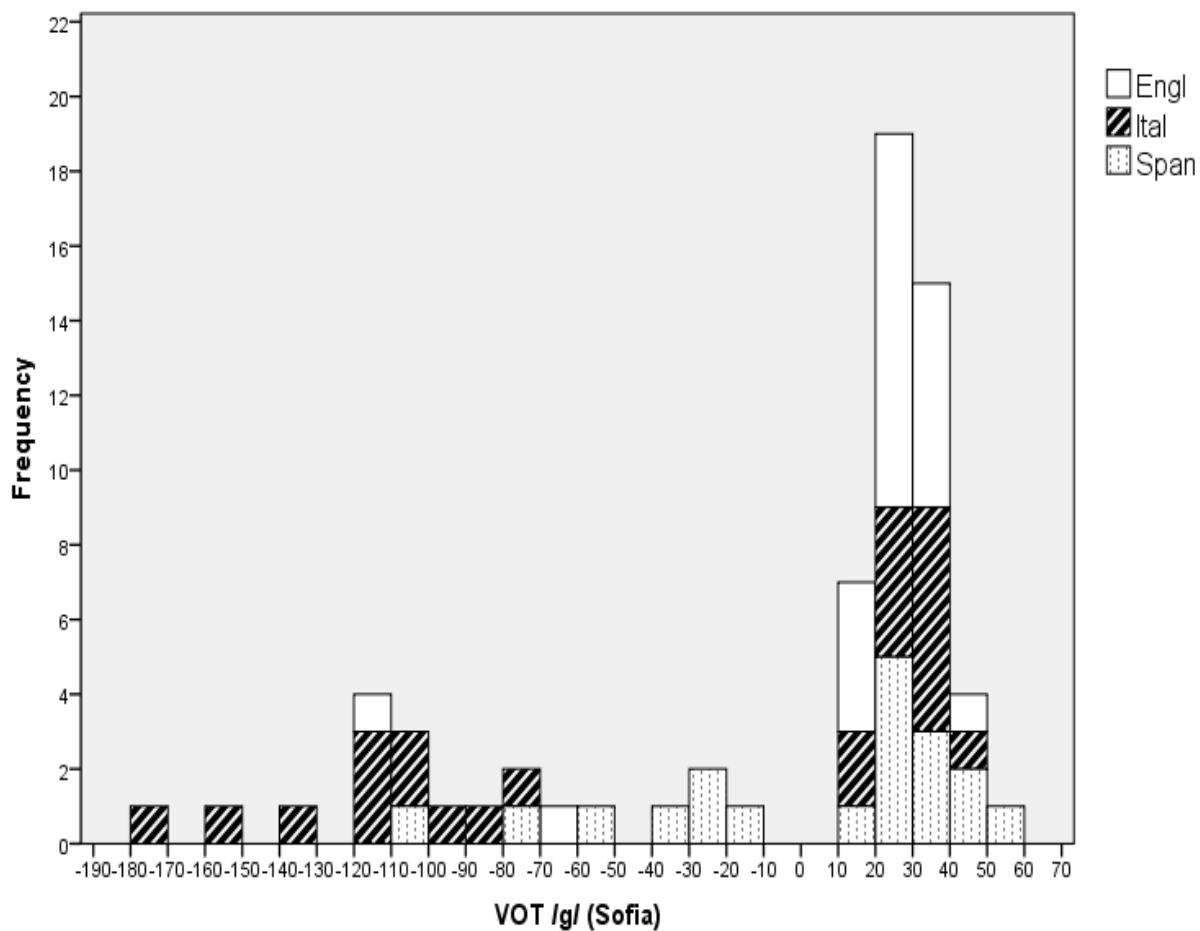
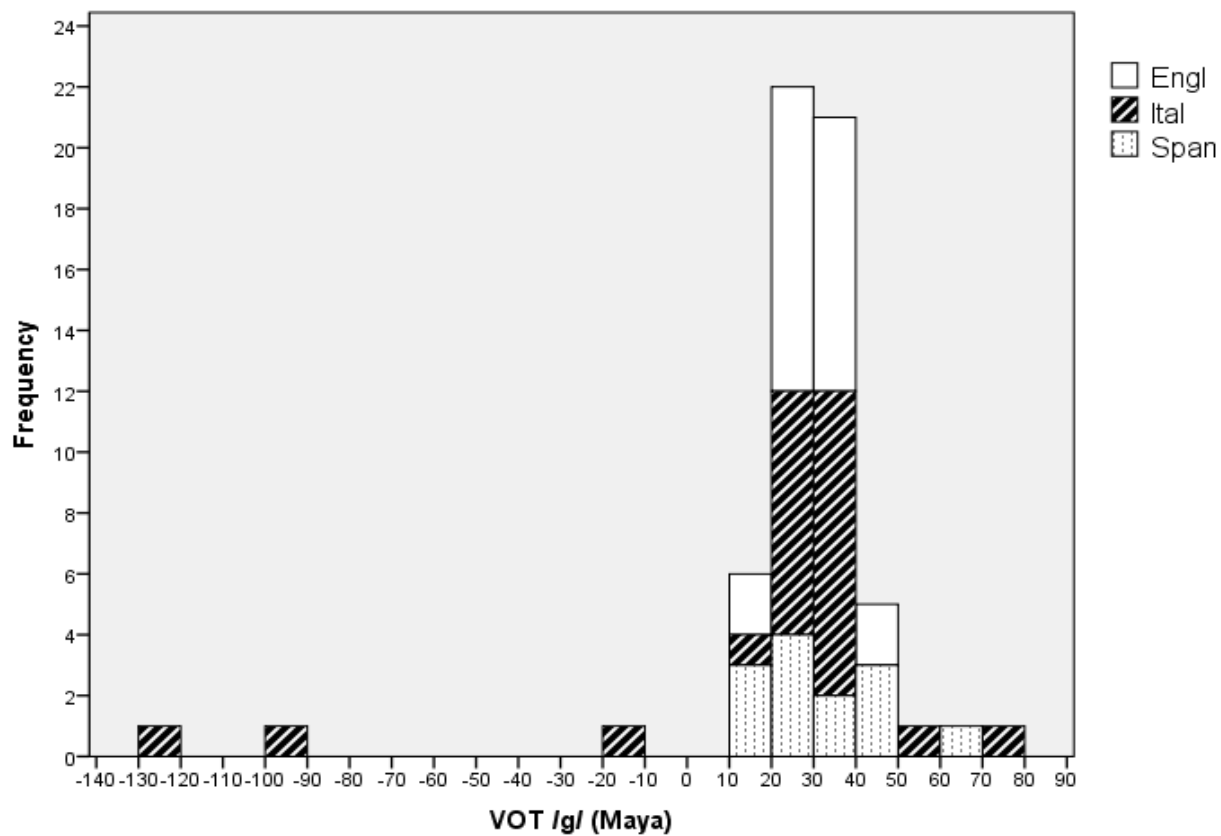




Figure 3. *Frequency of English, Italian and Spanish /b/, /d/ and /g/ in binary VOT ranges; Maya (top); Sofia (bottom)*

Table 1. Mean VOT (in ms) for word-initial stops in English, Italian and Spanish in adult speech;

	<i>bilabial</i>		<i>coronal</i>		<i>velar</i>	
	/p/	/b/	/t/	/d/	/k/	/g/
<b>English</b>						
Docherty* (1992)	42	18	63	26	73	31
Lisker & Abramson* (1964)	58	1	70	5	80	21
McCarthy et al. (2013)	39	13	58	11	65	18
<b>Italian</b>						
Bortolini et al. (1995)	11	-74	19	-80	34	-67
Vagges et al. (1978)	12	-100	17	-100	30	-105
<b>Spanish</b>						
Flege & Eefting (1987)	18	-82	22	-80	39	-68
Lisker & Abramson (1964);	4	-138	9	-110	29	-108
Rosner et al. (2000)	13	-92	14	-92	27	-74

\* The figures for lead voicing are not depicted here.

Table 2. *Stimulus material*

<b>English</b>	<b>Italian</b>	<b>Spanish</b>
/p/ puppy / <sup>1</sup> pʌpi/ puffy / <sup>1</sup> pʌfi/	pece / <sup>1</sup> petʃe/ “sap” pazzo / <sup>1</sup> pattso/ “crazy”	perro / <sup>1</sup> pero/ “dog” pata / <sup>1</sup> pata/ “paw”
/b/ belly / <sup>1</sup> bɛli/ baby / <sup>1</sup> bɛrbi/	buco / <sup>1</sup> buko/ “hole” babbo / <sup>1</sup> babbo/ “daddy”	bicho / <sup>1</sup> bitʃo/ “bug” barro / <sup>1</sup> baro/ “mud”
/t/ tummy / <sup>1</sup> tʌmi/ tiny / <sup>1</sup> tami/	taci / <sup>1</sup> tatʃi/ “be quiet” tuffo / <sup>1</sup> tuffo/ “dive”	taco / <sup>1</sup> tako/ “taco” toma / <sup>1</sup> toma/ “have a drink”
/d/ dizzy / <sup>1</sup> dɪzi/ doggy / <sup>1</sup> dɔgi/	dopo / <sup>1</sup> dɔpo/ “then” desto / <sup>1</sup> desto/ “awake”	deja / <sup>1</sup> dɛxa/ “let/leave” día / <sup>1</sup> di:a/ “day”
/k/ cookie / <sup>1</sup> kʊki/ kitty / <sup>1</sup> kɪti/	cane / <sup>1</sup> kane/ “dog” colla / <sup>1</sup> kɔlla/ “glue”	codo / <sup>1</sup> kɔðo/ “elbow” queso / <sup>1</sup> keso/ “cheese”
/g/ goofy / <sup>1</sup> ɡufi/ guppy / <sup>1</sup> ɡʌpi/	gola / <sup>1</sup> gola/ “throat” gufo / <sup>1</sup> ɡufo/ “owl”	gorro / <sup>1</sup> ɡoro/ “hat” ¡goza! / <sup>1</sup> ɡosa/ “enjoy!”

Table 3. Median VOT values (in ms) for voiceless stops; minimum and maximum values in parenthesis

	/p/	/t/	/k/	OVERALL
<b>English</b>				
FATHER	55 (37-63) n=8	70 (59-90) n=8	56 (49-95) n=8	59 (37-95) n=24
MAYA	29.5 (11-68) n=24	69.5 (38-143) n=24	73.5 (40-119) n=24	61.5 (11-143) n=72
SOFIA	34 (18-60) n= 24	60 (32-121) n= 24	94 (50-127) n= 24	58.5 (18-127) n=72
<b>Italian</b>				
MOTHER	10 (2-15) n=8	13 (8-18) n=8	40 (32-59) n=8	15 (2-59) n=24
MAYA	15.5 (7-40) n=24	21 (10-54) n=23	72.5 (55-118) n=24	31 (7-118) n=71
SOFIA	13 (6-30) n= 24	22.5 (13-36) n= 24	64.5 (34-109) n= 24	24 (6-109) n=72
<b>Spanish</b>				
NANNY	9 (4-9) n=8	15 (11-17) n=8	31 (19-51) n=8	15 (4-51) n=24
MAYA	8.5 (3-20) n=24	15 (7-48) n=24	39.5 (18-60) n=24	16 (3-60) n=72
SOFIA	10 (4-20) n= 23	15.5 (10-29) n= 24	33 (24-61) n= 24	16 (4-61) n=71

Table 4. Median VOT values (in ms) for voiced stops; minimum-maximum values in parenthesis

		English			Italian			Spanish		
		<i>Father</i>	<i>Maya</i>	<i>Sofia</i>	<i>Mother</i>	<i>Maya</i>	<i>Sofia</i>	<i>Nanny</i>	<i>Maya</i>	<i>Sofia</i>
<b>/b/</b>	short-lag VOT	6.5 (5-9) n=8	8 (2-20) n=16	4.5 (0-5) n=14	-	5 (0-17) n=11	5 (4-14) n=6	-	11.5 (2-14) n=6	12 (8-13) n=3
	lead VOT	-	-54.5 (-108 to -16) n=8	-64 (-117 to -17) n=9	-83.5 (-104 to -30) n=8	-42 (-201 to -12) n=13	-76.5 (-131 to -29) n=18	-39 (-71 to -32) n=7	-82 (-87 to -20) n=5	-
	spirant	-	-	-	-	-	-	[β] n=1	[β] n=13	[β] n=21
<b>/d/</b>	short-lag VOT	9.5 (5-14) n=8	17 (10-44) n=23	18 (9-31) n=20	-	10 (5-23) n=21	12 (6-19) n=3	-	-	27 n=1
	lead VOT	-	-47 n=1	-104 (-148 to -49) n=3	-75.5 (-110 to -53) n=8	-19 (-89 to -13) n=3	-91 (-145 to -28) n=21	-60 (-75 to -53) n=8	-	-
	spirant	-	-	-	-	-	-	-	[ð] n=24	[ð] n=23
<b>/g/</b>	short-lag VOT	16.5 (13-24) n=8	28 (13-45) n=23	26 (15-42) n=21	-	33 (13-72) n=21	32 (17-44) n=13	-	26 (11-65) n=13	29 (12-54) n=12
	lead VOT	-	-	-90.5 (-115 to -66) n=2	-85 (-121 to -60) n=8	-92 (-122 to -13) n=3	-111 (-177 to -78) n=11	-40.5 (-76 to -31) n=8	-	-38 (-106 to -20) n=7
	spirant	-	-	-	-	-	-	-	[ɣ] n=10	[ɣ] n=4
<b>OVERALL</b>	short-lag VOT	9.5 (5-24) n=24 (=100%)	20 (2-45) n=62 (=87%)	18 (0-42) n=55 (=80%)	-	13 (0-72) n=53 (=74%)	20.5 (4-44) n=22 (=31%)	-	24 (2-65) n=19 (27%)	27 (8-54) n=16 (=22.5%)
	lead VOT	-	-53 (-108 to -16) n=9 (=13%)	-81 (-148 to -17) n=14 (=20%)	-79.5 (-121 to -30) n=24 (100%)	-42 (-201 to -12) n=19 (=26%)	-93 (-177 to -28) n=50 (=69%)	-49 (-76 to -31) n=23 (=96%)	-82 (-87 to -20) n=5 (=7%)	-38 (-106 to -20) n=7 (=9.9%)
	spirant	-	-	-	-	-	-	n=1 (=4%)	n=47 (=66%)	n=48 (=67.6%)