

Vowel perception and production in Turkish children acquiring L2 German

Isabelle Darcy¹ & Franziska Krüger²

¹ Department of Second Language Studies, Indiana University, USA

² Department of Linguistics, University of Potsdam, Germany

Abstract

First language (L1) phonological categories strongly influence adult learners' perception and production of second language (L2) categories. For learners who start learning L2 early in life ("early learners"), this L1 influence is assumed to be substantially reduced (or at least more variable). In this paper, we examine the question of the age at which L1 vowel categories are already powerful enough to influence the acquisition of L2 vowels. We test a child population with a very narrow range of age of first exposure, controlled use of L1 vs. L2, and various naturally produced (not synthetic) contrasts that are not allophonic in the L1 of the children. An oddity perception task provides evidence that Turkish children learning German as an L2 since kindergarten categorized difficult German contrasts differently from age-matched native speakers. At the same time, their vowel productions of these same contrasts (un-cued object naming) were mostly target-like. We observe a pattern similar to Pallier et al.'s data (1997), but with a child population, offering strong support for the claim that children's L1 categories are already powerful enough to influence L2 acquisition even if exposure starts at a very early age (2-4). The advantage we observed in the production task extends data from Tsukada et al. (2005) who observed native-like performance in early learners with a cued naming task (with audio model), but is methodologically closer to Baker et al. (2008), who observed differences in early learners compared to native speakers with an un-cued production method.

Keywords: Turkish, German, sequential bilingualism, child second language acquisition, perception-production link, vowel categorization, vowel production

Dr. Isabelle Darcy (Corresponding Author)

Department of Second Language Studies, Indiana University

1021 E. Third Street

Bloomington, IN 47405

USA

e-mail: idarcy@indiana.edu

tel: 001 812 855 0033

fax: 001 812 855

Franziska Krüger (will be moving to Bloomington, IN, in the Fall 2011)

Institut für Linguistik

Postfach 601553

14415 Potsdam

Germany

e-mail: franzi@riessen.de

Tel: 0331 977 2401

Fax: 0331 977 2087

1. Introduction

Learning a second language early in life will usually yield excellent results, sometimes even leading to native-like competence in specific domains of the second language grammar (Birdsong & Molis, 2001; Guillelmon & Grosjean, 2001; Patkowski 1980; Johnson & Newport 1989; Flege, Yeni-Komshian & Liu 1999). Among the first to link young age and successful language learning were Penfield and Roberts (1959), who attributed the apparent ease with which children learn languages to the plasticity of the developing child's brain. With the publication of Lenneberg's 1967 book in which he argued that language acquisition must occur within a critical period ending around puberty in order to be fully successful, many researchers sought to establish whether second language acquisition would obey the same constraints. It has been subsequently commonly expressed – maybe too optimistically – that if L2 learning occurs early, before puberty, chances are good that learners will achieve native or near-native competence in their L2 (e.g. Johnson & Newport, 1989).

Early learners (or early bilinguals) are mainly defined by undergoing sequential bilingual acquisition, different from simultaneous acquisition. Sequential acquisition occurs when the second language is introduced after the first language is established. Some researchers use age 3 as the age when a child has basic communicative competence in L1 (Kessler, 1984). By contrast, learners are usually considered “late” when the onset of L2 acquisition starts after puberty or in adulthood. Based on this definition, an “early learner” is anyone whose exposure to L2 starts between 2-3 and 14, roughly. In the remainder of this article, the onset of L2 acquisition or exposure will be referred to as “age of learning” or “age of arrival” (AoL or AoA).

Research on early learner acquisition unanimously suggests that early learners are globally better than late learners in their L2, in all domains of language acquisition (e.g. Montrul, 2005). Regarding phonological acquisition, and in particular the acquisition of phonetic categories, studies have shown that early learners have less foreign accent than late learners (Flege, Yeni-Komshian, & Liu, 1999), better production (Baker & Trofimovitch, 2005), better perception of vowels and consonant categories, and recognize words in noise better (Meador, Flege MacKay, 2000; Mayo, L. H., Florentine, M., & Buus, S., 1997).

And yet, explaining age effects has proven difficult, and their source is still controversial (Flege & MacKay, 2010). The most well-known hypothesis to explain age effects is the “critical period hypothesis” according to which any learner exposed to L2 after puberty would – in the domain of phonology – retain an indelible foreign accent, whereas an exposure before the end of the critical period (before puberty) would lead to native-like pronunciation. The underlying mechanism put forward is a loss of neural flexibility – or plasticity. According to Scovel (1969, 2000), plasticity explanations would apply particularly well to phonological acquisition because of the link between brain plasticity and the neuromotor basis of pronunciation. Foreign accent is seen as the result of a reduction in plasticity, emerging around puberty (Scovel, 2000).

However, this view from the brain plasticity hypothesis that the critical period allows unrestricted learning up to puberty has been called into question on the basis of a growing body of evidence. Indeed, its implication would be a generalized nativelikeness (including for phonological acquisition) in early learners. The question whether early learners are similar to native speakers for phonological acquisition has sparked an intensive research effort, but has led to inconclusive results. Many studies suggest that pronunciation ability is subject to native language influence (hence yielding a foreign accent) well before puberty, leading to attempts to situate the “threshold” of a critical or sensitive period at ages much earlier than puberty (Asher & Garcia, 1969; Bever 1981; Johnson & Newport, 1989; Oyama, 1976). The data from both Johnson and Newport (1989) and Oyama (1976) suggest a cut-off point at age 6-7.

An alternative hypothesis is related to the quality and the quantity of the native speaker input received by early learners compared to late learners. While this hypothesis correctly predicts that late learners have more pronounced difficulties in phonological acquisition than early learners, the input hypothesis has difficulties explaining why certain studies find that some early learners exposed to the L2 from age 4 (like in the Barcelona early bilingualism setting, e.g. Pallier et al., 1997) still have difficulties acquiring vowel contrasts specific to the L2. It seems difficult to find a situation where more and better L2 input would be provided to learners than in such an early bilingualism setting without including simultaneous bilinguals (but see below 1.3).

Another hypothesis for age effects is the “Interaction Hypothesis” (Flege, 1992; Baker et al., 2008). According to this hypothesis, a suggested explanation for age effects

is that the interference or the interaction between L1 and L2 is different in adults compared to children. In particular, the L1 in children would not act as such a strong “attractor” to the L2 speech sounds. The L1 creates a stronger interference in adults than in children. For phonetic categories, more fully developed categories in L1 will exert a stronger attraction on L2 sounds. Children would be less likely to perceptually assimilate L2 sounds to L1 categories, which would allow for a faster acquisition of categories. In other words, the nature of the interaction between L1 and L2 categories might be different in adults vs. children, in perception as well as in production. This hypothesis also incorporates the notion (similarly to the Critical period hypothesis) that if L2 exposure starts early enough, the L1-L2 interference will be null, or at least not strong enough to produce a non-native pattern in production or perception.

While the “Interaction hypothesis” receives support from various studies comparing children and adults’ acquisition of sound contrasts (Tsukada et al., 2005; Baker et al., 2002; Baker et al., 2008), the question of how early the L1-L2 interaction becomes large enough to produce such interference effects remains insufficiently understood. In other words: can early learners – if they start early enough – be native-like? Baker et al. (2008) find some limited support for the Interaction hypothesis. However, this reduced interference was limited to production data. Korean children’s English vowels were correctly identified more often than adults’, but they were still identified less well than age matched native speakers’. Perception data were similar for both children and adults, and not native-like. Of note, the children had a mean AoL of 10 years (ranging between 6-13 years), which might not be “early enough” to observe the absence of L1 influence.

Many studies have examined early learners in order to understand whether or not they can be native-like in perception or production of non-native phonological dimensions. As seen above, while most agree that early learners are better than late learners, the state of knowledge regarding the differences between *early and native speakers* is far less clear-cut. Compared to late learners, the influence of L1 on acquisition of L2 in early learners has been described as ranging from absent (= native-like acquisition) to strong – comparable to the degree present in late-learners. It is still an open question when exactly the influence of the L1 phonological system starts to be strong enough as to impede native-like acquisition of both L1 and L2 phonologies, when L2 is acquired very early but after L1, keeping in mind that acquisition of L1 is

not completed by the time L2 acquisition starts.¹ In the studies we review below, it seems that if there is interference between L1 and L2 in early learners, it is much more variable and unpredictable than for late learners.

The review begins with studies that have reported no difference between early learners and native speakers, and then considers studies that have identified differences. Specific tasks will be highlighted to both show the range of tasks used and to point out possible task effects when they occur. It ends with a summary.

1.1 Early learners equal native-speakers

Some studies have found no difference between native speakers and early learners for vowel or consonant perception and production, or word stress patterns (Baker et al., 2002; Flege, MacKay, Meador, 1999; Mack, 1989; Oturan, 2002; Guion, Harada & Clark, 2004). In Oturan's study, Turkish early learners (AoL around 4 years) performed close to native speaker levels as measured through a vowel confusion score. Late Turkish-German bilinguals performed very differently from the early group or from the native German speaker group; unfortunately, the lack of detailed analysis of significance levels makes the interpretation of this study difficult.

Mack (1989) examined the perception and production of two English contrasts (/d-t/ and /i-ɪ/) by monolingual English and early bilingual English-French speakers (AoL before 8, ranging from 0-8). Results show that in discrimination and identification tasks for a synthetic VOT continuum between /d-t/, bilinguals did not differ from the monolinguals. For a synthetic /i-ɪ/ continuum, the category boundary was slightly shifted in the bilingual group, but all bilingual listeners clearly had acquired both categories. There was no difference in production for either contrast. However, the interpretation of these results for or against the question of nativelikeness in early bilinguals should be made with care, since the bilingual group was not homogeneous with respect to the L1 acquired (this was not crucial for the research question). Of 10 bilinguals, half of the bilinguals learned English as their L1, four learned French as their

¹ This question is different from asking when speech perception becomes language-specific. This has been shown to happen around 10-12 months of age (Werker & Tees, 1984, and many others). It is now rather undebated that the native categories are established enough to guide perception to be language-specific by the end of the first year. A different case occurs when exposure to a second language takes place early but *after* this initial set-up where a child establishes language-specific categories; it is unclear when those L1 categories are robust enough to interfere and prevent the establishment of L2 categories.

L1, and one was a simultaneous bilingual. The purpose of the study was to test whether the dominant language of early bilinguals is “monolingual-like” (= ‘intact’). Therefore, all bilinguals had to be dominant in English (and were all rated “native speakers” by judges). It is unclear when exactly L2 exposure started for each bilingual. The lack of statistically significant difference compared to the monolinguals could be explained by the fact that half of the bilingual group would qualify as “L1 listeners” (even if not monolingual) whereas the other half as “L2 listeners”. So the shift found in the boundary of the /i-ɪ/ continuum is difficult to interpret, leaving open the question of whether it comes mainly from the “L2 listeners”, or whether it is homogeneously shared among all bilinguals. Thus, we cannot definitely say to what extent the early L2 learners (French L1) group behaves like English monolinguals.

Flege, MacKay and Meador (1999) examined vowel perception and production in 72 Italian-English bilinguals. The early bilingual groups had a mean AoA of 7 years but differed in L1 use. The early-low bilinguals used Italian roughly 8% of the time, whereas the early-high bilinguals used it around 32% of the time. They did not differ on other measures like mean age (47-48 years) or length of residence (LoR, 40 years), and were highly experienced in English. Both early bilingual groups were found to be undistinguishable from a matched English native speaker group in a vowel production task and in a vowel perception task containing eight English and five Italian vowels; the effect of L1 use was not significant. This study suggests that when L2 acquisition starts early (7 years) and exposure goes on for long enough (mean of 40 years of residence), native-like performance is attainable.

Baker et al. (2002) studied a group of child (mean age 18 years) and adult (mean age 28 years) Korean immigrants. The children’s AoA was on average 8, adults’ was 19, and therefore, the children can be considered “early” learners, while the adults are “late” learners. The children did not differ significantly from a group of native English speaker adults in a vowel discrimination task. The adults were native-like only for the contrast that did have a Korean counterpart. Production data (vowel intelligibility) parallel the perception data: the children were native-like, whereas the adults were not.

Another study (Tsukada et al. 2005) which did find differences in perception (reviewed below) failed to find differences in production between early learners (Korean children) and age-matched native speakers. In a picture naming task, Korean

children and adults produced English words three times after first hearing an auditory model. Only the first (“cued”) production was analyzed; comparisons didn’t reveal any differences between the Korean children and the native speakers. The child group had variable chronological ages (9-17) and ages of first exposure (AoA range: 6-14). The lack of differences may be due, as noted by the authors, to the early learners imitating the model more accurately than the Korean adults.

These studies suggest that in some cases, early bilinguals can perform at native-like levels on perception and/or production tasks. It would be tempting to conclude at this point that if acquisition starts roughly around age 7, native-likeness is attainable. However, it is premature to conclude that early learners are generally native-like or that L1 interference is insignificant in early childhood. In particular, the amount of L1 use can influence outcomes and might have been at play in some of the studies just reviewed. As the following studies show, the high use of L2 and low use of L1 produces native-like or close to native-like results, especially if the early years are advantaging the L2 over the L1 (see also Højen & Flege, 2006).

Two studies (Flege & MacKay, 2004; Piske, Flege, MacKay & Meador, 2002) provide evidence that native-likeness is observed mostly in early bilinguals who use their L1 seldom and their L2 most of the time. For instance, Flege and MacKay (2004) found that two groups of early Italian-English bilinguals (mean AoA 7-8, range 2-13, mean LoR 40 years) that differed only in L1 use (Italian use of 7% “low” vs. 43% “high”) also differed in their ability to detect mispronunciations in L2. While the low-early group was exactly like a group of native English speakers, the high-early group behaved differently, and similarly to a group of late Italian-English bilinguals. However, task effects may be an additional confound. A second experiment employing an oddity discrimination task showed that the same two groups of early bilinguals were not significantly different from native speakers on 8 out of 9 vowel contrasts (the high-early group was significantly different from the natives, but not the low-early group, on one single vowel contrast: /ɒ/ - /ʌ/) (p. 17).

Piske, Flege, MacKay and Meador (2002) similarly found a slight difference between two groups of early bilinguals (the same tested by Flege MacKay and Meador, 1999) in their vowel productions, obtained through nonword reading, and rated with a measure of goodness by native speaker judges (rather than intelligibility as used

in Flege et al. 1999). Some of the high-early participants' vowels received lower ratings than vowels spoken by native speakers. None of the vowels spoken by the low-early group differed from the ratings obtained by native speakers' vowels. As the authors point out, the higher activation of Italian in the high-early group might have led to a stronger influence of Italian sound-spelling correspondence in reading nonwords.

1.2 Early learners differ from native-speakers

A number of studies report differences between early bilinguals and native speakers. Global accentedness studies show that even in early learners with an AoA earlier than 10 years, a foreign accent can be detected (Flege, Birdsong, Bialystok, Mack et al., 2006; Flege, Munro & MacKay, 1995; Flege, Yeni-Komshian & Liu, 1999). The tasks used in these studies (global accentedness ratings) are likely more sensitive to differences, but they don't point specifically to the elements of the speech that differ from native speakers.

Studies that measure phonological acquisition in terms of realization of phonetic dimensions or processing of phonetic categories, however, also often report the influence of L1 categories to be strong enough to impede native-like production and processing of L2 categories. Early and intensive exposure to a second language is not enough to build native-like phonemic categories (Bosch, Costa & Sebastian-Gallés, 2000), to perform like native speakers in discrimination tasks (Pallier, Bosch, & Sebastián-Gallés, 1997; Højen & Flege, 2006; Sebastian-Gallés & Soto-Faraco, 1999; Navarra et al., 2005; Tsukada et al. 2005) or in vowel production tasks (Baker & Trofimovitch, 2005; Baker et al. 2008). The learners in those studies can be roughly split into two groups: the "broad age range" groups and the "narrow age range" groups, depending on how narrow their age of learning ranges are.

Tsukada and colleagues (2005) report that Korean children exposed early to English perform better than adults but still differently from monolingual children in an oddity vowel discrimination task. Groups varied widely in chronological age and age of first exposure ("broad age range" AoA range: 6-17), but were controlled for length of residence (LoR: 2-4, vs. 4-6 years). Results show that the group of children with the shorter LoR (2-4) behaved differently from the native English children group on all test contrasts. The group of children with longer LoR (4-6) patterned more closely with the native English children, showing a significantly lower discrimination on only two

contrasts out of four. Of note, all Korean children were attending an English-speaking school and reported using English more often than Korean adults. Recall that this study did not find differences between the early learners and native speakers in production (measured via vowel identification rates by native speaker judges), unlike Baker et al. (2008) who report differences using a similar method, but with one difference: They analyzed the *last* token produced (spontaneously), instead of the first (cued) production used by Tsukada et al. (2005). The early learners (Korean children, AoA 10;0, range 6-13;5, LoR 9 months) were more intelligible than the late learners (Korean adults, AoA 25;1, range 19-31, LoR 6 months), but were still clearly different from native speakers. The children and the adults did not differ significantly in terms of self-rated L2 proficiency or use.

Højen and Flege (2006) show that adult Spanish early learners of English, with a mean AoL of 6 years (range 2-10) behave similarly to native speakers on a vowel discrimination task using a phonetically sensitive categorial ABX. Spanish monolinguals are very different from English monolinguals. However, on a shorter inter-stimulus interval (ISI=0 ms) for two of three difficult vowel contrasts, the early bilinguals' discrimination was lower (unfortunately, no value of statistical significance is given) than the native speakers' performance, leading the authors to conclude that the early learners' perception of English vowels was not "functionally equivalent" to that of the native speakers (p. 3080).

There are two possible reasons why these studies found differences between early bilinguals and native-speakers. First, the amount of L2 use varies a lot between learners, and Højen and Flege (2006) explicitly mention the possibility that this factor together with an earlier AoL (2-5) is responsible for individual differences found in the data. Second, the broad range in ages of first exposure considered in these studies makes group behavior potentially less homogeneous, and perhaps also too variable to allow for definite conclusions. The differences between early learners and native speakers observed in those studies could arguably be due to those large ranges in AoA, leading the statistical analyses to show differences mainly due to the later arrivals. The groups are not homogenous enough to be able to tell with certainty how the learners with earlier AoAs are doing. In addition, the age of first exposure (earliest AoL was 6) chosen in Tsukada et al's study might already have been too late to allow native-like

perception abilities, given that earlier studies suggested a cut-off age for the maturational window around 6 (Oyama, 1976, Johnson & Newport, 1989).

Yet some studies do report data from learners groups with a “narrow age range”. To our knowledge, however, the only studies offering a clearly narrow range of AoA / AoL have been conducted in Barcelona (Bosch, Costa, & Sebastian-Gallés, 2000; Navarra, Sebastian-Gallés & Soto-Faraco, 2005; Pallier, Bosch, & Sebastián-Gallés, 1997; Ramon-Casas, Swingley, Sebastian-Gallés & Bosch, 2009; Sebastian-Gallés, Echeverria, & Bosch, 2005; Sebastian-Gallés & Soto-Faraco, 1999).

Pallier et al. (1997) find that adult early Spanish-dominant bilinguals (who learned Spanish first), exposed at age 4 to L2 (Catalan) do not perform as a group like Catalan dominant bilinguals (for whom Catalan is considered their native language) on a synthetic continuum discrimination task. The examined vowel contrast is “difficult,” and probably falls within the “single category assimilation” pattern (according to PAM; Best, 1995): [e-ɛ] (like [o-ɔ] and [s-z] examined in Pallier et al, 2001), is not present in Spanish, and discrimination between both vowels is very difficult. The authors did not collect production data. Interestingly, however, individual results show large differences between individuals, with some of the Spanish early bilinguals performing just like the Catalan native speakers. Other studies have also examined the acquisition and encoding of difficult contrasts such as /e/-/ɛ/ by early bilinguals in different tasks (Gating: Sebastian-Gallés & Soto-Faraco, 1999; Looking times: Ramon-Casas et al., 2009; implicit ABX: Navarra et al., 2005; lexical decision: Sebastian-Gallés et al., 2005). All results consistently revealed that early Spanish dominant learners of Catalan do not acquire the contrast to the same degree as Catalan-dominant (native speakers). These studies converge towards the conclusion that even early and intensive exposure to a second language is not enough – at least for some learners – to prevent L1 representations from influencing processing, acquisition and encoding at the lexical level (see also Pallier et al., 2001).

1.3 Summary

The contradiction is clear: Although some studies with wide-ranging ages of first exposure are sometimes unable to, and sometimes do find differences between early learners and native speakers, the “Barcelona studies” with a very early and narrow AoL

range consistently find large differences between Catalan-dominant and Spanish-dominant early bilinguals. The variability can be explained by attention to four variables: age, allophony, L1 use and task differences.

Age range. The first reason for which studies find differences between early learners and native speakers is possibly a matter of how tightly the age ranges are controlled. In some studies where differences are found, age ranges vary greatly, so there is no way to know without analyzing the individual data whether the early learners exposed later may be influencing the data.

L1 use. The degree to which the L1 is activated, or how much pressure there is to retain the L1 may also influence the results and may be one additional difference between the Barcelona studies and those who do not find any difference between early learners and native speakers. In Barcelona (where differences are consistently observed), the pressure to keep both languages activated is very high, unlike in other settings where studies have mainly examined immigrant children and adults (and where no differences have sometimes been observed, e.g. Flege, MacKay & Meador, 1999). In addition, all learners in Barcelona live in a fully bilingual society. Every resident can correctly assume that regardless of the language spoken, they will be understood. Similarly, the exposure of Barcelona residents to accented Catalan or accented Spanish is likely high. Thus, the pressure to resemble native speakers in perception or production might be less strong than in immigrant situations, where there is high pressure from the environment to acquire the L2 spoken in the place of residence, and conversely less pressure to keep the L1 intact, even though this does differ from individual to individual. Consequently, if the Barcelona studies were conducted in a different setting with different contrasts (see below), they may not find such consistent differences. When L1 use is taken into account as a variable, it turns out to be a major factor in explaining variability between groups (Flege & MacKay, 2004). In particular for those groups where the L1 is spoken less than the L2 (the “early-low” groups, who are also more likely to be native-like), some of the native-like early bilinguals also report that they are more proficient in their L2 than in their L1 (Højen & Flege, 2006). The status of what language is the dominant one is unclear in those early bilinguals. When the differential L1 use is examined in participants with a very long length of residence (~40 years), however, the effect seems difficult to observe with reliability, as

shown in the studies by Flege and colleagues, testing the same population but arriving at different results with slightly different tasks (Flege, MacKay & Meador, 1999 and Piske, Flege, MacKay & Meador, 2002).

Allophony. The consistency with which the Barcelona studies find differences between early bilinguals and native speakers even with a narrow age range calls for a careful consideration of phonetic details: In all the observed cases of differences, the vowels examined are very close to the Spanish prototypes in the acoustic space. For example, the Catalan contrast [ɛ] – [e] falls in the range of allophonic variation attached to the Spanish vowel [e] (Bosch et al., 2000). This fact alone might contribute to the consistent finding of differences in the Barcelona setting, and would jeopardize the generalizability of the findings (see also Flege & MacKay, 2004).

Task. Finally, the task used (perception with low vs. high task demand and low vs. high acoustic variability, or production with vs. without aural model) can also contribute to a lack of sensitivity in certain studies. As shown by Højen and Flege (2006), a sensitive task is necessary in order to show possibly well-hidden differences (if they in fact exist) between early learners and native speakers. Similarly, in production, collecting spontaneous (or not cued) data seems important in order to uncover possible differences, as shown by the different findings from Tsukada et al. (2005; cued, no differences) compared to Baker et al. (2008, un-cued, significant differences). To summarize, there are four major factors that may explain the large variability observed across the different studies: variable age-ranges, differences in L1 use of the participants, differences due to the contrasts examined, and methodological differences.

In order to be sure that observed differences (if they exist) between early learners and native speakers are due to L1 interference and not to the combined action of other confounded factors, it seems important to conduct studies which a) control for ,early age of first exposure, b) with participants learning L2 in a setting where pressure to maintain L1 is high, c) where the contrasts examined are not allophonic in the L1 of the speakers, and d) using tasks sensitive enough to detect possible differences.

2. The present study

In order to add to our understanding of early age effects and L1-L2 interactions, and specifically focusing on the question of “how early is already too late”, this study controls for age of first exposure among learners in environments in which L1 use is high, examines four different contrasts and collects both perception and production data. Our goal is to gain a better understanding of the development of phonetic categories in children who learn a second language (sequentially) very early in life. We tested 10 years-old early sequential bilingual Turkish-German children, who were first exposed to German between the ages of 2-4. This section first reviews the German and Turkish vowel systems (2.1), describes the expected perceptual proximity (2.2), and then presents our specific hypotheses and predictions (2.3).

2.1 German and Turkish vowels

The German vowel inventory is larger than the Turkish vowel system. As the Figures 1a and 1b show, Turkish does not distinguish between tense and lax vowels at the phonemic level as clearly as German does. To distinguish vowels, German uses the distinctive features of tenseness and duration (Féry, 2004, Wiese, 1996), both of which are not used to this end in Turkish. Turkish vowels are generally more lax than the German vowels, and the cardinal vowels appear centralized. Duration as a contrastive feature is not recognizable anymore in contemporary Standard Istanbul-Turkish, but is present phonetically through compensatory lengthening and in borrowed words (Menges, 1994, Kornfilt, 1997, Oturan, 2002, Topbaş & Yavaş, 2006).

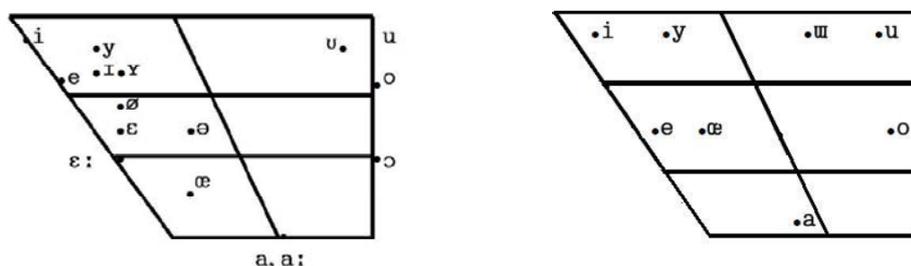


Figure 1a,b: German (a) and Turkish (b) vowel system, from Kohler (1999); Zimmer & Orgun (1999)

We selected six German vowels [i:], [ɪ], [e:], [ɛ], [a:] and [ɑ]. The contrasts (described in section 2.3) were chosen according to the predictions made by the perceptual assimilation results discussed below (2.2), which are based on the study by

Oturan (2002). In an attempt to increase the generalizability of findings, the contrasts chosen vary with respect to being in the range of allophonic variation in the L1 of the participants (Goksel & Kerslake, 2005). In particular, our most difficult contrast is *not* within the range of allophonic variation in the L1 of the participants (see below).

Figure 2 compares the localization of the German and Turkish vowels (after Wängler, 1981, and Selen, 1979) in the F1/F2 vowel space dimensions. The Turkish phones are marked with a subscript “t”, the German with a subscript “d”.

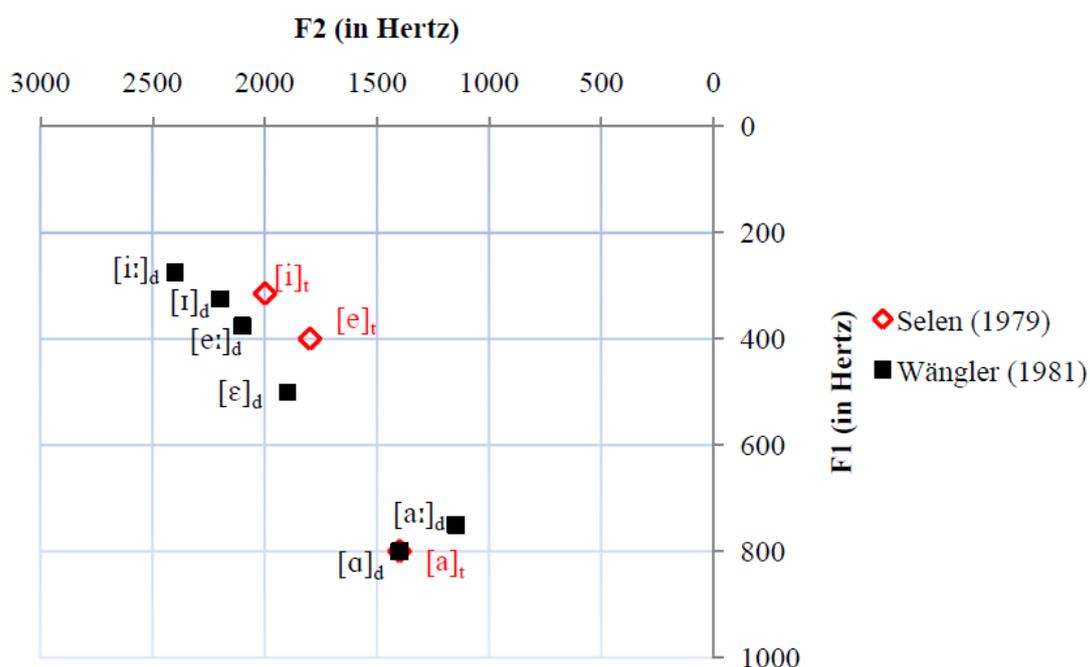


Figure 2: Comparison of the formant values for F1 and F2 of German tense and lax vowels (Wängler, 1981) with Turkish vowels (Selen, 1979)

Both German and Turkish have a phonetic category defined as high front unrounded vowel. Turkish [i] however is articulated more centrally than German [i:] (Kornfilt, 1997, Oturan, 2002; see Figure 1). Comparing the measures of Sendlmeier (1981) and Wängler (1981), it seems that the Turkish [i] (value from Selen 1979) is closer to the German lax [ɪ] than to the German tense [i:]. What is clear is that in this high/front F1/F2 area, German distinguishes two unrounded phonetic categories, whereas Turkish has only one. An allophone of Turkish [i] is [ɪ], usually occurring in word-final position (Goksel & Kerslake, 2005).

For mid vowels, we observe a similar pattern. In both languages, there is a phonetic category described as a mid front unrounded vowel, but German again distinguishes two where Turkish has only one. The formant values (in particular those given by Sendlmeier, 1981) illustrate that Turkish [e] is located in close proximity to the German lax [ɛ]. Allophones of Turkish [e] are [ɛ] and [æ] (Goksel & Kerslake, 2005), but not [i]. Yet, Turkish [e] is more central than German [e:] and always lax, articulated with a more open jaw, and close to German [ɛ]. German tense [e:] is – as a result – often confused with the German high vowel [i:] by Turkish listeners (Cimilli & Lieber-Harkort, 1979).

Low vowels in both languages are usually centrally articulated. Turkish [a] is described a little bit higher and more front than in German (Figure 1, Zimmer & Orgun, 1999). The comparison of formant values given by Selen (1979, Figure 2), however, indicates a slightly lower articulation than Zimmer and Orgun's description. Both German low vowels [ɑ] and [a:] are in close proximity to or overlap with Turkish [a].

In sum, the comparison the six German vowels (long tense vowels [i:], [e:], [a:] (as in *schief* 'inclined', *Schnee* 'snow', *Hahn* 'rooster') and short lax vowels [ɪ], [ɛ], and [ɑ] (*Schiff* 'ship', *schnell* 'fast', *Hand* 'hand')) with the three Turkish categories [i], [e] and [a] reveals that the Turkish phones are acoustically closer to the respective German lax vowel, and do not have a tense counterpart.

The acoustic and articulatory comparison of German and Turkish vowels is informative, but it is insufficient with regard to establishing cross-linguistic perceptual distance and perceptual assimilation patterns, especially because the F1-F2 values obtained from the literature can't be directly compared across studies (Bohn, 2002). Perceptual distance is frequently evaluated through collecting perceptual assimilation data (see Tsukada et al., 2005 for an example), which in turns serves to predict discrimination (and possibly also acquisition) performance.

2.2 Perceptual proximity of German and Turkish vowels

To obtain a benchmark evaluation of perceptual proximity between German and Turkish vowels, perceptual assimilation patterns are usually collected from naive participants. For the present study, we report the results of perceptual assimilation patterns presented in Oturan (2002), collected for Turkish and German vowels. The participants were 31 Turkish students (mean age = 19;7 years). They were asked to listen to German vocalic monophthongs and to provide a Turkish equivalent for those vowels (i.e. classify them according to their native Turkish categories). As well, they were required to judge the perceptual similarity between the German sound and their Turkish representations on a scale of 1 through 6 (1= identical, 6= different vowel – see Figure 3). Participants had never had any contact with German. All were born and grew up in a monolingual environment.

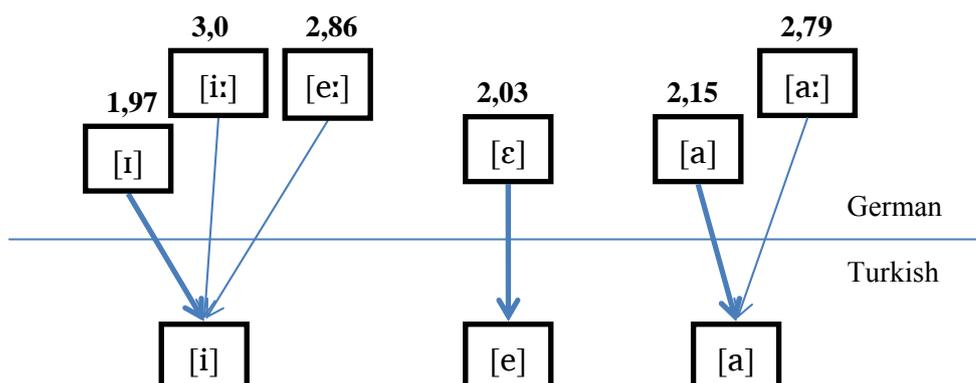


Figure 3. Results of the perceptual assimilation and goodness ratings summarized from Oturan (2002). (1 = identical; 2 = very similar; 3 = quite similar; 4 = different; 5 = very different; 6 = other vowel). Length of arrows roughly approximates goodness ratings.

Figure 3 visualizes the mapping and shows that both German tense and lax [i:] and [ɪ] are mapped onto the /i/ category in Turkish, as well as the German tense [e:]. The German lax short [ɛ] alone as in *Bett* ‘bed’ is mapped onto the Turkish front mid vowel [e]. Both German [a] vowels are mapped onto Turkish [a]. The average goodness ratings reported range for most from “very” to “quite” similar. These results corroborate the phonetic descriptions seen above, as well as the perceptual proximity of the German lax vowels [ɛ]_d, [ɪ]_d and [a]_d to the Turkish vowels [e]_t, [i]_t and [a]_t. Their goodness ratings is generally “very similar”, while the tense counterparts [i:]_d and [a:]_d received a rating of “somewhat similar”, but were still categorized as [i]_t and [a]_t. German tense [e:] is categorized differently, to 95%, as Turkish [i].

2.3 Hypotheses and predictions

In terms of predicting discrimination ability, such mappings have been shown in several studies to predict quite reliably how adult non-native listeners will be able to discriminate between those sounds (Best et al. 1988, Best, 1995, Levy & Strange 2008; Levy 2009a, b; Best et al., 2001, Tsukada et al. 2005). Stimuli for this study were chosen according to the way vowels were categorized in terms of the Turkish vowel categories (Oturán 2002) described above. Based on the pattern observed in Figure 3, the following predictions can be made *for naïve listeners of German*.

- (1) [e:]~[i:] is our test contrast, differing only in spectral cues. [i:] and [e:] are both equally “good” exemplars of the same Turkish category [i]: this is a case of single-category assimilation, and discrimination is predicted to be very difficult.
- (2) and (3) [i:]~[ɪ] and [e:]~[ɛ] are acoustically quite close but still distinguished by both length and spectral cues. Turkish adults categorized both contrasts differently: German tense [i:] and lax [ɪ] are both also mapped onto the same category, but differ in category goodness, with [i:] being a slightly less good exemplar for Turkish [i]. This is a case of category goodness difference, and the discrimination is predicted to be better than in single-category cases. German tense [e:] and lax [ɛ] are clearly mapped onto two different categories, a pattern which usually predicts a better discrimination than in both previous cases.
- (4) [a:]~[i:] is our control contrast, because of their mapping onto two different categories which are even further apart in the vocalic space, for which very good discrimination is expected (see Best, 1995).

We hypothesize that the native categories of the Turkish early bilingual children already have been established well enough to impede the exact acquisition of the German vowels and their specific features – despite an early AoA. If the L1 phonetic categories already influence L2 perception despite early exposure, we expect participants to

behave in a similar way to naïve listeners and to conform to the predictions above. However, given that our participants can be considered already advanced early learners of German (with an average of 7 years of exposure), an alternative hypothesis is that their perceptual similarity patterns and hence their discrimination ability has evolved, particularly in the case of [e:]~[i:] and [i:]~[ɪ], and would be better than what is predicted for naïve listeners (see Levy 2009a and b; Tsukada et al., 2005). They may, for instance, be able to pick up on the duration feature, useful in addition to spectral differences to distinguish both high vowels [i:]~[ɪ], even though it would not help for the most difficult contrast [e:]~[i:]. However, since it exists but is not contrastive in the L1, the children might be able to use it only to a limited extent for the discrimination task (McAllister, Flege & Piske, 2002).

Because of restrictions on the number of tests we could reasonably expect the children to perform, it was virtually impossible to collect perceptual assimilation data directly. In addition, the difficulty of ensuring that children understood the task in the same way as adults did was a concern.² Discrimination performance even without perceptual assimilation data can be revealing of the kind of interaction between categories at work in our participants: Depending on their performance with the two more difficult contrasts ([e:]~[i:] and [i:]~[ɪ]), it will be possible to see whether L1 influence is likely to have had an effect despite the early age of L2 exposure. It will not, however, be possible to say whether or not this possible L1 influence is stronger or weaker than in adult learners.

The predictions in the domain of production are less clear-cut. In some studies, differences in productive patterns have been reported for early bilingual children (Flege et al., 2006, Baker & Trofimovich, 2005). The Speech Learning Model (Flege, 1995) states that L2 categories which are more distant from the corresponding L1 category will be acquired more accurately than those L2 categories which are close to the L1 category, and for which the L1 production specifications are used. Even though this model applies to advanced adult late learners, we would expect accordingly that the spectral specifications of the Turkish categories to which the contrasts are mapped

² We don't compare adults and children in this study, so that a comparison of their performance/perceptual similarity patterns is not crucial.

perceptually to be used in production. Regarding duration, if we consider it as “new” feature, we may expect that both long front vowels [i:] and [e:] could benefit from less clear overlap with the Turkish category [i] (as opposed to German [ɪ]), and could therefore be produced with better accuracy than German [ɪ]. The same could apply to [a:], which may be produced accurately, as opposed to German [a]. According to the Feature hypothesis (McAllister, Flege & Piske, 2002), the fact that long vowels exist in Turkish in certain environments could be an alternative reason for this feature to be acquired easily (see Nimz, 2011).

In order to examine these hypotheses, we tested two groups of children (early Turkish-German bilinguals and German monolinguals) in an oddity vowel categorization task and an uncued word naming production task.

3. Experiment 1: Vowel categorization

3.1 Method

3.1.1 Participants

Twenty-eight children were tested. They were either native speakers of German, or early sequential Turkish - German bilinguals. Table 1 summarizes participant data. All demographic data were determined through parental questionnaire.

Table 1: Participant information (age, gender, L2 exposure)

Participant	Age	gender	age exposed to L2	Language spoken	
				with friends/ siblings	at home
b7	11;5	f	2;6	German	Turkish
b8	11;1	m	3	German	Turkish
b9	11;0	m	3	German	Turkish
b11	11;5	f	3-4 (kindergarden)	German	Turkish
b12	12;0	m	4	German	Turkish
b13	10;11	m	3	German	Turkish
b14	12;3	f	2;6	German	Turkish
b15	11;6	f	daycare (2;6 - 4)	German	Turkish
b16	11;4	f	3;6	German	Turkish
b17	11;0	f	3-4	German	Turkish
b18	11;1	m	3	German	Turkish
b19	11;1	f	3	German	Turkish
b20	9;8	f	2;6	German	Turkish
b22	11;6	f	2;6	German	Turkish
m1	11;7	m	-	German	German

m2	11;7	f	-	German	German
m3	10;7	f	-	German	German
m4	11;3	f	-	German	German
m5	11;5	f	-	German	German
m6	11;11	m	-	German	German
m7	11;0	f	-	German	German
m8	11;10	f	-	German	German
m9	11;4	m	-	German	German
m11	11;6	f	-	German	German
m12	10;10	f	-	German	German
m13	9;7	f	-	German	German
m14	11;1	f	-	German	German
m15	11;3	f	-	German	German

Note: *b*= bilingual; *m* = monolingual

None of the children presented evidence of hearing problems, irregular early speech development or developmental dyslexia. All bilingual participants (N = 14, 9 girls, 5 boys) were growing up in a Turkish-German environment in the Berlin area in Germany. The dialect of Turkish they were exposed to was restricted to standard Istanbul-Turkish (Menges, 1994, see 2.1 above). They were schooled in a bilingual Turkish-German elementary school, where alphabetization takes place in both languages. No participants in the bilingual group were exposed to any language other than Turkish prior to entering daycare or kindergarten.³ The mean age of the participants is 11.2 years, ranging between 9;8-12;3 (134.7 months, SD 7.1). Their average age of first exposure to L2 is 2.9 years, ranging between 2;6-4;0 (35.9 months, SD 5.6); they were therefore exposed to German on average for 7 years. A linguistic separation according to social context was also visible in parents' and children's responses: they indicated that the bilingual children speak German with siblings and friends, but speak only Turkish at home. This situation often resulted in or stemmed from one or both parent not knowing any German. In interactions with the experimenter (who only spoke German), no problems emerged in understanding the task, and no clearly perceptible accent was detected in most. In some children however, the short [a] vowel was perceptibly different.

The monolingual participants (N=14, 11 girls, 3 boys) were recruited from a German-only school in the same Berlin area. The mean age of the monolingual group is 11.1 years (134.3 months, SD 7.1, range 9;7-11;11). All children spoke "Hochdeutsch",

³ Children who according to the answers were exposed to German since birth are not included in this report, since we are here interested in early sequential bilingualism.

standard German. None of them had any contact with another language before entering school.

3.1.2 Stimuli

Stimuli are nonword syllables containing the six vowels chosen for this experiment. The nonwords were read in a short sentence context several times by three adult German native speakers (2 females, one male) in a sound-isolated recording room, digitized, and manually cut from each sentence to be presented in isolation. The qualitatively best recording for each item was chosen as stimulus for the experiment.

Given evidence that the consonantal context of vowels can interact with perceptual performance in vowel categorization (see Strange, Weber, Levy, Shafiro, et al. 2007), we used two contexts for the nonwords: velar (k_k) and bilabial (p_p), which provides a more detailed understanding of perceptual patterns.

3.1.3 Procedure

Children were tested in a quiet room in their school. They started with the perception task, and then moved on to the production task (described in Experiment 2 below). Approximate total testing time was 20 minutes (for both tasks).

For the perception task, we used an oddity vowel categorization task (“pick the odd one out”) similar to the one used by Tsukada and colleagues (2005). Nonword syllables were presented auditorily as triads in “same” (N=48) or “change” (N=48) trials on a computer.

For each contrast ([a:]~[i:], [i:]~[ɪ], [e:]~[ɛ], [i:]~[e:]), there were six possible orderings for a “change” trial (AAB, ABA, BAA, BBA, BAB, ABB). With four contrasts, this yields 24 change trials. In addition, we created another 24 “same” trials, four with each vowel. In order to keep the experiment short enough for the children, we created two lists of stimuli: A contrast presented in pVp context in List 1 appeared in kVk context in List 2 and vice-versa. A given context remains the same for a given contrast (for instance, [a:]~[i:] is presented in kVk in all 6 trials associated with this contrast in a given list). Both pVp and kVk contexts were varied across “change” and “same” trials in roughly equal proportion. Children were assigned randomly to each list.

The experiment was organized in two blocks. In the first block ISI was 500 ms, in the second block it was 0 ms (see Højen & Flege, 2006). All children had the same block order. The order of trials within each block was automatically randomized. The total number of trials was 96 for each child (48 trials in each ISI). Stimuli presentation was controlled by custom presentation software in form of a game using a display picturing three robots on the computer screen (see Figure 4).

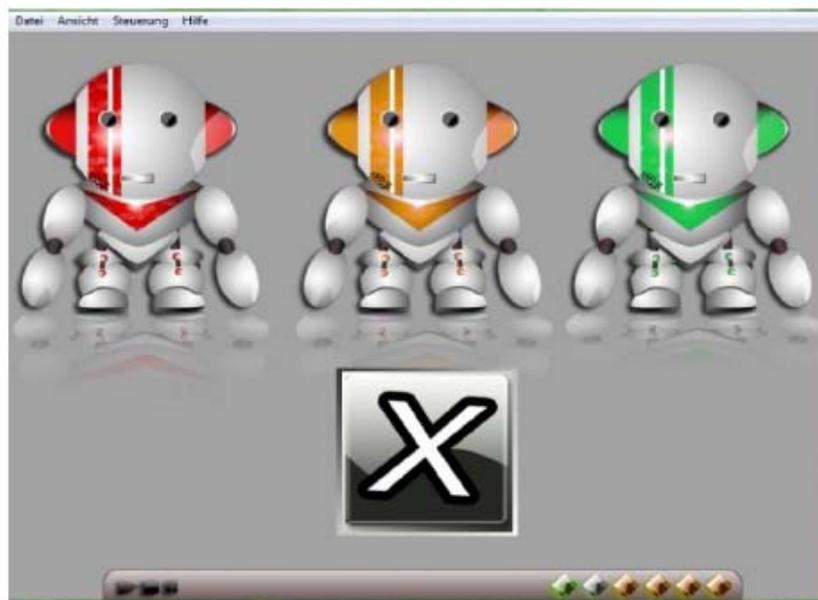


Figure 4. Layout of the categorization task; Robot 1 (left): female voice; 2 (center): female voice; 3 (right): male voice

Children were seated in front of a laptop computer, equipped with high-quality Sennheiser headphones and a mouse. They were instructed to listen to what each robot said at each trial, and to click on the robot saying something different. If the child thought that all robots said the same thing, she was instructed to click on the X box in the lower half of the screen. Each token within a trial was spoken in a different voice. To reduce confusion, each robot always spoke in the same voice. Prior to the test, children had to pass a familiarization phase containing five trials with other stimuli than those used in the test (context bVf). During the familiarization phase only, children were allowed to listen several times to the stimuli, and received feedback about their answers from the experimenter. Both feedback and the repetition option were absent in the test phase. Children's answers were not speeded, but the next trial initiated 1000 ms after they gave their answer.

As explained above, the activation of L1 Turkish is high in our participants because of the bilingual setting they are evolving in. In order to avoid inflating artificially the presence of L1 interference effects, and to maximize their significance if we find them, every effort was made to favor a monolingual mode during the experiment for the participants (Grosjean, 1989), in this case German. The experiment was conducted entirely in German, with an experimenter who didn't know any Turkish. The children were aware of this fact, which likely contributes to reduce the activation of the language that is not shared (see Khattab 2007), in this case, Turkish. As a result, the setting we chose is expected to reduce the activation of the L1 (even though it will not be completely deactivated, Grosjean, 2001).

3.2 Results

Participants' raw detection values of the control contrasts were first screened in order to establish that all participants correctly understood the task. Both vowels ([a:]~[i:]) are articulatory and perceptually very different and should be categorized as "different" easily by all participants. Examination of the performance (correct answers) in the monolingual group reveals that all were within 1 SD of the group median. In the bilingual group, the performance of one participant was below 3 SD from the group median ($z = -3.1254$). For this participant, we cannot assume that the task was correctly understood, and hence her results were excluded from further analyses. Table 2 summarizes the correct detection rates for each group for each contrast.

Table 2: Average correct answers (%) per group (bil=13, mon=14)

Group	[a:]~[i:]	[e:]~[ɛ]	[i:]~[ɪ]	[i:]~[e:]
Monolingual	97.92	87.8	98.29	85.42
Bilingual	98.08	82.05	76.92	56.73

From the raw correct answers, we calculated a d' measure of sensitivity based on hits (H) and false alarms (FA), following Macmillan and Creelman (2005). A hit results from correctly detecting a difference in a "change" trial (i.e. any robot was clicked), regardless of whether the odd one was correctly located.⁵ A false alarm occurred when a

⁴ z is the standard deviation of a value from the group median (Spiegel & Stephens, 1998).

⁵ The experiment design necessitates first to detect a difference (signal detection) and then to identify the correct robot from which the difference comes. It was therefore necessary to compute two performances (detection performance = any robot is clicked in a change trial, and identification performance = the right robot is clicked in a change trial). The differences between groups in detection and identification

participant clicked a robot in a “no-change” trial (Schwarz, 2008, Macmillan & Creelman, 2005). The computation of d' additionally incorporates the adjustment proposed by Macmillan & Creelman (2005) for perfect values ($H=1, FA=0$)⁶. Given our small sample, non-parametric Mann-Whitney U-tests were conducted on the average d' for each contrast comparing both groups (in case of paired samples such as for the ISI variable, we used the non-parametric Wilcoxon test).

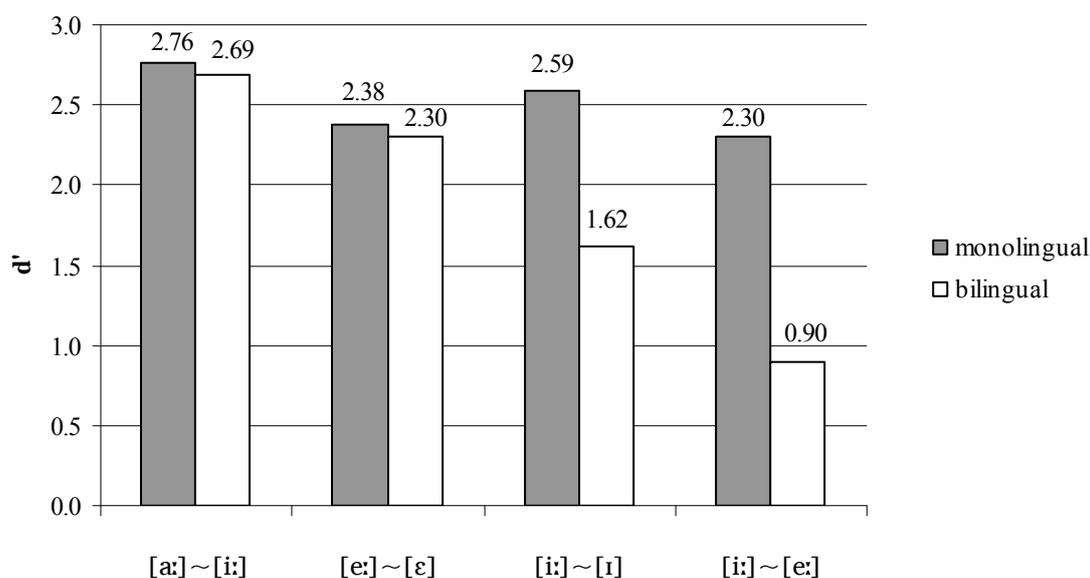


Figure 5. Average d' for the bilingual and monolingual groups (N bil=13, N mon=14)

As shown in Figure 5, L2 learners are significantly worse at discriminating the [i:]~[ɪ] and [i:]~[e:] contrasts – predicted to be difficult – than monolinguals. The group difference is significant for both contrasts (Mann-Whitney for [i:]~[ɪ]: $U=52.5, p<.05$; Mann-Whitney for [e:]~[ɛ]: $U=16.5, p<.0001$). By comparison, L2 learners were like the monolinguals on both other contrasts ([a:]~[i:], Mann-Whitney $U=97.5, p>.05$; [e:]~[ɛ] Mann-Whitney $U=87.5, p>.05$). The effect of group for the [i:]~[e:] pair was significantly larger than all others, indicating that the lack of the salient duration cue

performance were not significantly different. The results here therefore only present the “detection” performance.

⁶ With the adjustment, d' for a perfect detection performance is 2,7659. At a value of 0, a participant is not able to perceive a difference between change and no-change trials.

may represent an additional hindrance for L2 learners to discriminate this pair (McAllister, Flege & Piske, 2002).

Bias calculations for the control contrast (X_c , based on d' values) revealed that bias was very limited in all participants. The averages for each group for x_c were $\bar{X}_{bil}=0$ and $\bar{X}_{mon}=0$. The one participant who had been excluded based on the detection performance also exhibited a larger bias, which was about $\frac{1}{2}$ SD away from the group median. No other participant showed a bias that differed from the group median. The influence of bias is therefore considered minimal.

An analysis of the effect of context reveals the pattern presented in Figure 6. For the control contrast, no effect of context is visible in any group. The bilingual children identified the difference between [i:] and [ɪ] in the bilabial (pVp) context significantly less accurately than when it was presented in the velar (kVk) context (Mann-Whitney-U-Test: $U=1$, $p<0.01$). The same comparison between contexts for the contrast [e:]~[ɛ] showed that the bilinguals were better at discriminating it in a velar context than in a bilabial context (Mann-Whitney-U-Test: $U=6$, $p<0.05$). The comparison with the monolingual group is however necessary. For the [e:]~[ɛ] contrast, despite that d' is more variable in the bilingual group when the contrast occurs in the bilabial context, the comparison between bilinguals and monolinguals reveals that the groups do not differ significantly in each context (**pVp** Mann-Whitney-U-Test: $U=14,5$, $p>0.05$; **kVk** Mann-Whitney-U-Test: $U=10$, $p>0.05$). In the case of [i:]~[ɪ], the group comparisons shows that in the velar context, both groups ($N_{bil}=6$, $N_{mon}=7$) are not different (Mann-Whitney-U-Test: $U=18,5$, $p>0.05$), but in the bilabial context, their accuracy does differ ($N_{bil}=7$, $N_{mon}=7$, Mann-Whitney-U-Test: $U=2$, $p<0.01$). For the [i:]~[e:] contrast, the bilingual group's performance was comparably low regardless of the context (Mann-Whitney-U-Test: $U=14$, $p>0.05$).

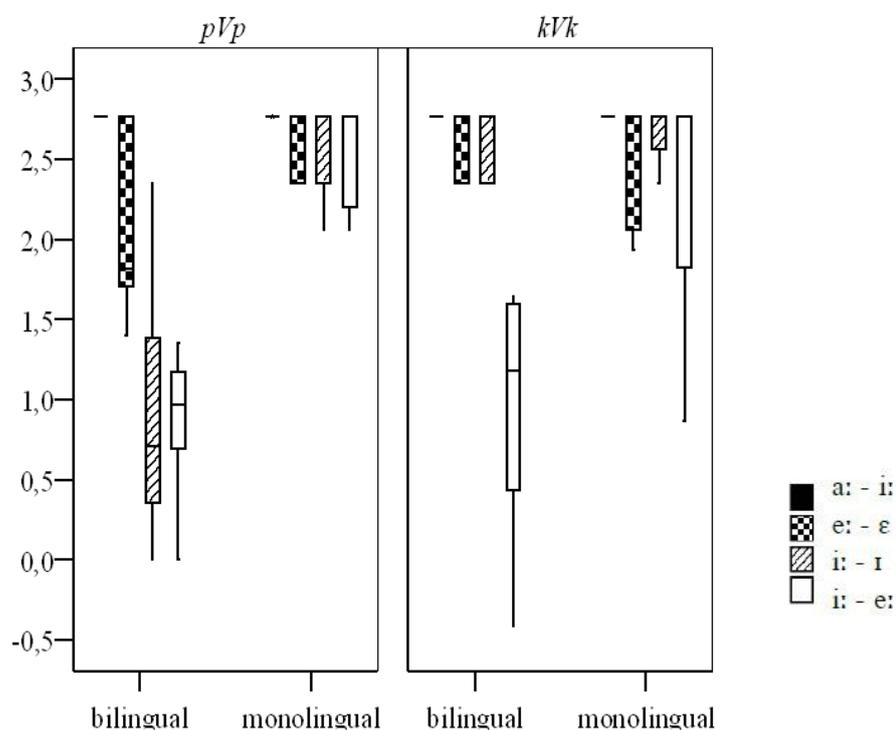


Figure 6: Average d' for the bilingual and monolingual groups as a function of context condition (N bil=13, N mon=14)

The ISI conditions were also examined. The only contrast for which an ISI effect was significant was [i:]~[e:]. In both groups, performance was better at the shorter ISI interval of 0 ms. (bilingual: Wilcoxon $Z=-2.3$, $p=.019$; monolingual: Wilcoxon $Z = -2.19$, $p=.028$).

3.3 Discussion

The results of Experiment 1 show a clear pattern of discrimination that parallels the perceptual proximity as defined through the adult perceptual assimilation data. With increasing perceptual proximity, discrimination ability declines. The bilingual children's experienced difficulties on those contrasts that were also predicted to yield most confusions (according to the perceptual similarity obtained with naïve adults). These difficulties can therefore be attributed to an influence of the L1 phonological structure. The children in our study behaved as naïve adult listeners without experience of German would have been expected to. We can therefore safely conclude that if L1 did not have a strong influence in these children's discrimination abilities, the pattern observed would have been different and likely closer to the German monolingual pattern. In this sense, our data corroborate the patterns observed by Pallier et al. (1997)

with adults, and also other studies conducted in Barcelona with early Spanish-Catalan bilinguals. The fact that these children were exposed to German very early seems to have had little influence on their discrimination abilities, even if this is difficult to quantify with precision, since late and early bilingual adult data using these contrasts and methods are still lacking (Darcy & Krüger, *in preparation*). In any case, early exposure has not been enough to cause their discrimination performance to equal that of the monolinguals.

4. Experiment 2: Vowel production

We now examine the production patterns of this group of early bilinguals. We collected production data along with perception data, in order to investigate the possible advantage observed in production over perception (Tsukada et al., 2005, Baker et al., 2008). Unlike Tsukada et al. (2005), and more like Baker et al. (2008), we measured production using spontaneous uncued production (no auditory model was given).

4.1 Method

4.1.1 Participants

Participants were the same children tested in Experiment 1. They first took part in experiment 1, and then in experiment 2 (see also section 3.1.3 for general procedure).

4.1.2 Stimuli

For each of the six vowels, three common German words were selected (see Table 3). To increase picturability and frequency, both minimal pairs and near-minimal pairs were selected. The consonantal context was kept as similar as possible for each vowel pair. In order to allow for the [i:]~[e:] comparison, the context surrounding the vowels [i:] and [e:] was also held as close as possible: [bi:st], [be:t], [ʃti:l], [ʃne:] (*Biest – Beet*; *Stiel – Schnee*).

Table 3. Overview of the German words chosen for elicitation

i:	<i>Stiel</i> 'handle'	[ʃti:l]	<i>schief</i> 'inclined'	[ʃi:f]	<i>Biest</i> 'beast'	[bi:st]
ɪ	<i>still</i> 'quiet'	[ʃtɪl]	<i>Schiff</i> 'ship'	[ʃɪf]	<i>Biss</i> 'bite'	[bɪs]
e:	<i>Beet</i> 'flowerbed'	[be:t]	<i>Fee</i> 'fairy'	[fe:]	<i>Schnee</i> 'snow'	[ʃne:]
ɛ	<i>Bett</i> 'bed'	[bet]	<i>Fell</i> 'fur'	[fel]	<i>schnell</i> 'fast'	[ʃnɛl]
a:	<i>Saat</i> 'seed'	[za:t]	<i>Hahn</i> 'chicken'	[ha:n]	<i>Schwan</i> 'swan'	[ʃfa:n]
ɑ	<i>satt</i> 'satisfied'	[zat]	<i>Hand</i> 'hand'	[hant]	<i>Schwanz</i> 'tail'	[ʃfants]

For the elicitation of the words, children engaged in a *Memory* game. Using a custom display that showed the picture upon a mouse click, the children's task was to find the pairs of the same picture. They were asked to name the pictures aloud upon turning each card. The children's productions were recorded from a Sennheiser microphone on a Marantz *PMD 670* solid state recorder at 44.1 kHz, and digitized for further acoustic analysis.

Prior to the game in a training phase, children were shown the pictures accompanied by the written form of the words, and produced the word associated with each picture. To verify that they remembered the words associated with each picture, they were then asked to name each picture without the written form. This step was repeated in case of difficulties. No auditory cues (such as the initial) or any auditory modeling of the stimuli was provided to facilitate the remembering of the words, in order to avoid any influence of the children's pronunciation through a phonological model (Kuhl & Meltzoff, 1996, Tsukada et al., 2005). No child had notable difficulties with this task.

4.2 Analysis

Children's vowel productions (N=1008) (28 participants x 6 vowels x 3 contexts x 2 Items, or 168 per vowel) were acoustically measured. Items that had a low recording quality or which contained noise or clicks were excluded from the analysis (2.9 % of the data, 979 items were analyzed). Vowel duration and formants F1 and F2, as well as f0 were extracted. Given the large inter- and intravariability of the children's production, we used a normalization procedure using the Bark conversion (Syrdal & Gopal, 1986; Bohn & Flege 1992). The f0-F1 difference corresponds to the height dimension, while the F1-F2 difference corresponds to the horizontal (back-front) dimension.

Despite the normalization procedure, some items could be considered outliers (their Bark-difference values being beyond 2SD from the mean) and were not included in the statistical analysis. This was the case for 77 datapoints, out of a total of 2016 datapoints (3.8%).

4.3 Results

4.2.1 Spectral data

The Bark values for each vowel in each group are summarized in Table 4, and graphically represented in Figure 7. A higher F1-f0 value represents a lower vowel, and a higher F2-F1 value indicates that the vowel is more front. This analysis is independent of the consonantal context of the specific words.

Table 4. Average Bark differences for 6 German vowels

F1-f0 (vertical)	bilingual	SD	monolingual	SD	U-value	p-value
e:	2.089	0.203	2.296	0.187	39	0.006**
ε	3.807	0.698	4.076	0.336	73	0.25
i:	1.337	0.442	1.36	0.368	94	0.854
ɪ	2.463	0.41	2.594	0.393	77	0.334
a:	5.821	1.384	6.134	0.741	75	0.29
ɑ	5.35	0.845	6.306	0.694	37	0.005**

F2-F1 (horizontal)	bilingual	SD	monolingual	SD	U-value	p-value
e:	10.671	0.52	10.5176	0.372	84	0.747
ε	7.707	0.529	7.441	0.462	68,5	0.093
i:	11.547	0.346	11.569	0.414	97,5	0.505
ɪ	8.201	0.309	7.977	0.624	73	0.098
a:	3.902	0.603	3.096	0.848	40	0.001**
ɑ	4.345	0.64	3.518	0.652	36	0.003**

Note. ** significant at the $p < 0.01$ level

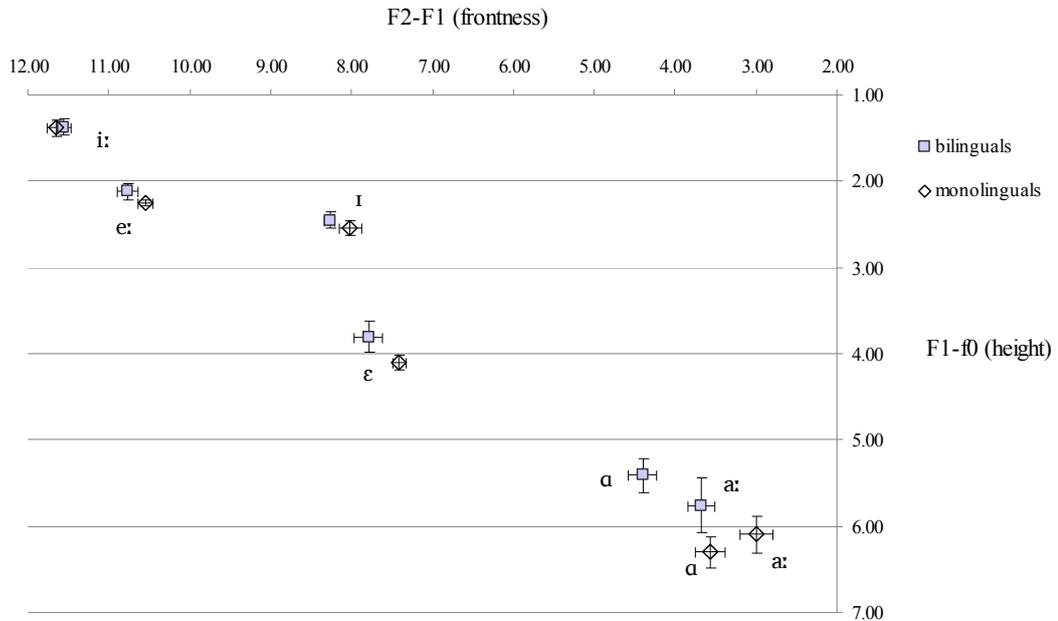


Figure 7. Mean values (Bark) per vowel for each group.

The group comparison revealed that both groups differ in their production of the lower vowels [a:] and [a]. Both are produced more centrally by the bilingual group. The lax [a] is also significantly higher in the bilingual productions. Another group difference, even if only marginal, is indicated for [e:] and [ɪ], which are articulated higher and more front respectively by the bilinguals than by the monolinguals.

Considering the vowel pairs used in the perception task, the bilingual group produced adequately differentiated vowels, comparable to the monolingual group. The spectral difference between [i:]~[ɪ] was significant for both the bilingual and monolingual group. In both groups is [i:] higher and more front than its counterpart [ɪ] (F1-F0 Wilcoxon: $Z = -3,296$, $p < 0.01$; F2-F1 Wilcoxon: $Z = -3,296$, $p < 0.01$). All children produced the expected contrast difference between the tense and lax vowels. For [e:]~[ɛ], the contrast analysis also showed that the difference between both vowels is produced adequately by all children. The [e:] is significantly higher and less centralized than the [ɛ] (both groups: F1-F0 Wilcoxon: $Z = -3,296$, $p < 0.01$; F2-F1 Wilcoxon: $Z = -3,296$, $p < 0.01$). For the [i:]~[e:] contrast, the analysis revealed that in both groups, [i:] is higher and more front than its counterpart [e:] (both groups: F1-F0 Wilcoxon: $Z = -3,296$, $p < 0.01$; F2-F1 Wilcoxon: $Z = 3,296$, $p < 0.01$). Both groups of children produced the expected contrast difference between both tense vowels. For the low vowels [a:] and [a], the bark values from Table 4 indicate that spectrally, both vowels are much less clearly differentiated than the other pairs. As Sendlmeier (1981) suggested, these vowels are mainly differentiated by the duration feature. As expected, neither group showed a significant difference in the vertical (F1-f0) dimension. Both vowels are low. Horizontally (F2-F1), however, both groups produced the lax [a] more centrally than its counterpart [a:] (monolingual Wilcoxon-Test: $Z = -1,381$, $p > 0.05$; bilingual Wilcoxon-Test: $Z = -1,852$, $p > 0.05$).

4.2.2 Duration analysis

Table 5 summarizes the vowel durations. The comparison of each minimal pair confirms that all children in both groups produced long tense vowels with significantly longer duration than short lax vowels.

Table 5. Duration differences (in ms) between tense and lax vowels

	Item	Duration (in ms)	Item	Duration (in ms)	Z-value	p-value
bilingual						
[e:] [ɛ]	Beet	251.45	Bett	154.34	-2.934	0.002**
	Fee	267.35	Fell	160.25	-3.296	0.001**
	Schnee	268.03	schnell	176.16	-3.179	0.001**
[i:] [ɪ]	Biest	237.65	Biss	166.25	-3.179	0.001**
	schief	217.16	Schiff	149.93	-3.296	0.001**
	Stiel	180.32	still	141.93	-2.605	0.009**
[a:] [ɑ]	Hahn	255.93	Hand	117.20	-3.296	0.001**
	Saat	277.15	satt	146.99	-2.934	0.003**
	Schwan	244.00	Schwanz	150.88	-3.296	0.001**
monolingual						
[e:] [ɛ]	Beet	255.33	Bett	141.06	-3.296	0.001**
	Fee	320.52	Fell	190.20	-3.296	0.001**
	Schnee	300.14	schnell	201.07	-3.180	0.001**
[i:] [ɪ]	Biest	220.57	Biss	149.14	-3.296	0.001**
	schief	224.62	Schiff	148.19	-3.296	0.001**
	Stiel	202.66	still	147.42	-3.045	0.002**
[a:] [ɑ]	Hahn	297.18	Hand	119.66	-3.296	0.001**
	Saat	281.69	satt	141.30	-3.296	0.001**
	Schwan	286.55	Schwanz	149.70	-3.180	0.001**

Note. ** significant at the $p < 0.01$ level

Even though Table 5 shows that duration differences are appropriately implemented, a group comparison in terms of the duration ratio between long and short vowels reveals a significant difference (Table 6). The bilingual vowel duration ratio is on average shorter than the monolingual's suggesting that the duration differences are more clearly realized in the monolingual group. The monolingual speakers produced short and long vowels with a duration ratio of 1:1,74, shorter than for adults (Withworth 2000, Antoniadis & Strube 1984). Similarly to Lee et al. (1999) for English, we observe here that children produce a shorter duration ratio compared to adults. The bilingual group had a ratio of 1:1,63, which is significantly smaller than the ratio of the German monolingual children (Mann-Whitney-U-Test: $U=54$, $p<0.05$).

Table 6. Vowel duration ratio for the three tense-lax vowel pairs

	ratio [e:]~[ɛ]	ratio [i:]~[ɪ]	ratio [a:]~[ɑ]
--	----------------	----------------	----------------

bilingual	1: 1.68	1: 1.48	1: 1.97
monolingual	1: 1.74	1: 1.50	1: 2.16
comparison	U=82, p>0.05	U=85, p>0.05	U=42, p<0.01

The comparison of the ratios for each vowel contrast (Table 6) shows that the significant effect is mainly due to the low vowel contrast.

4.3 Discussion

Appropriate tense-lax distinction and duration differences observed in production data indicate that for most vowels under scrutiny, German-specific articulatory patterns were adequately acquired. Between-group-differences were significant for [a:]~[ɑ] and [e:] only. Subtle differences observed in the localization of those vowels, along with context effects observed in perception, suggest however that L2 learners lack the flexibility found in native speakers to adjust for coarticulatory influences, and generally display greater variability across tokens than the monolinguals (see also Levy & Strange 2008). While it is hardly surprising to find more variability in bilinguals (bilinguals after all have two phonetic systems to acquire and accommodate), we note that the bilingual children were able to realize a difference for words which was not clearly distinguished perceptually, at least in the oddball discrimination task we used. Not all comparisons were made in the perceptual task, so we do not have any perceptual data for the [a:]~[ɑ] contrast. But we have production data for all three other perceptual contrasts that were also examined. In particular, the [i:]~[ɪ] difference and the [i:]~[e:] difference (for which the bilinguals had a significantly lower sensitivity than the monolinguals) were adequately contrasted, both for spectral and duration characteristics. There was, however, a difference concerning the spectral realization of [e:], which was higher in the bilinguals (hence closer to its counterpart in the difficult perceptual contrast [i:]) compared to the monolinguals. It therefore seems that [e:] does pose a problem, but not to the extent that bilinguals would not be able to distinguish it appropriately in production.

5. General Discussion

Given the range of contradictory findings in the literature about early successive bilinguals, our foremost goal was to test the perception and production of a set of carefully chosen stimuli by participants with a narrow age range of L2 exposure. We reasoned that such a design would offer a way to reconcile the disparate findings in light of a better understanding of the factors that contribute to conflicting data in the first place. This in turn allows for better generalization of findings, and adds to our understanding of how the interaction between L1 and L2 categories develops.

Our results corroborate previous results showing that in children and adults alike, L2 vowel categorization is subject to L1 influence very early (possibly earlier than 3-4) and is difficult to modify despite early intensive exposure to L2 (e.g. Pallier et al., 1997, Højen & Flege, 2006).

This L1 influence seems to be reduced in the production data. In sharp contrast to the perception results, the two groups (monolinguals and bilinguals) were globally undistinguishable in production. Differences emerged only in analyses comparing the exact location of vowels, showing for instance that the bilinguals' [e:] was slightly closer to [i:] than in the monolingual group. When analyzed in terms of *contrast between* vowel pairs, in both groups of children, spectral and duration differences were adequately implemented, as evaluated by the presence of a significant difference between long and short (duration) as well as tense vs. lax (spectral) vowels. Even if the exact phonetic detail of their productions exhibits a larger variability than the monolinguals', we can say that the early bilingual children are able to produce a distinction in production which they are not necessarily able to show in perception, especially for [i:]~[ɪ] and [i:]~[e:]. Generally speaking, our results support previous findings of an advantage in production over perception (Tsukada et al., 2005; Baker et al., 2008).

The seemingly clear difference between perception and production we observed can be due to either of two possibilities. Perception difficulties as observed in tasks such as the oddball task may artificially underestimate the full extent of perceptual abilities, since the poor performance on distinguishing [i:] from [e:] or from [ɪ] for example does not seem to impede the children's ability to correctly realize the vocalic contrasts in

each of the words *Biest* [bi:st], *Biss* [bɪs] or *Beet* [be:t]. In other words, they have correctly encoded this difference in the lexical representations for those words, and were able to spontaneously retrieve them (recall that they were not given any auditory model for the words). When faced with nonwords in a somewhat artificial task (clicking on a robot who speaks “differently”), they may not be able to apply their (possibly less stable) knowledge of the [i:] and [e:] categories to the task (unlike the monolinguals). It remains however possible that the bilinguals have encoded this contrast lexically without being able to generalize this difference to an abstract category, which would be necessary for performing well in the perception task. Monolingual children, on the contrary, have succeeded in having both lexically encoded a contrast and displaying stable categories in the perception task. This phenomenon parallels dissociations observed in other studies between lexical encoding vs. categorization of categories, which are encoded contrastively at the lexical level even though their categorization is not always completely stable, or vice-versa (see Darcy *et al.*, 2009, *under review*; Hayes-Harb & Masuda, 2008; Ota *et al.*, 2009). As said, this effect may be due in part to the fact that in production, children were asked to name words with which they were familiar, whereas in perception, they were listening to nonwords. Therefore, the contrast between perception difficulties and production accuracy may be due to our design which confounds lexical status and processing modus (perception - nonlexical vs. production - lexical), and it is difficult to conclude with enough assurance that perception difficulties are not directly reflected in production, or to tease apart the contribution of lexical knowledge vs. production / perception differences to the performance in the task.

The other possibility is that the perception task is not suited to show how they differentiate [i:] from [e:], for instance. Testing them on a larger variety of items or with a different perceptual task may reveal that bilinguals employ a different strategy than monolinguals to perceptually distinguish [i:] from [e:]. Hence, this doesn't necessarily mean that they would suffer communication breakdowns in daily conversations. But the researchers' task is to know how categories are formed and represented, both in terms of their existence and their stability against coarticulatory or other variability. So, even if we conclude that our perception task may be not suited to reflect the full extent of

perceptual abilities, it is still able to show that native speakers perform well on this task, and that therefore, L2 learners who do not perform well may still have something to learn in order to approach native speakers' categorization behavior. It may be that they still need to form abstract and stable categories from a set of words where it is encoded. This could stem from the fact that the input of the early bilingual children is overall less consistent than that of the monolinguals, given the fact that the bilinguals are, to some extent, probably also exposed to other late Turkish-German bilinguals whose German is more accented.

Whatever it is that native speakers have learned that allows them to perform so well in this task, this has not been yet reached by the bilinguals, even though the bilingual group had been exposed to German since an early age and was being exposed extensively through schooling on a daily basis: We observed a clear difference between this group and the monolingual participants in perceptual behavior. In perception, their performance was very much like that predicted by assimilation patterns and perceptual proximity for adults (Oturán 2002). The presence of context effects in the perception experiment, especially for the [i:]~[ɪ] contrast, which is more difficult to distinguish in the bilabial context, and indicates a higher susceptibility to context effects on the part of the bilinguals. These effects are similar to those reported for adult American English listeners' perception of French front rounded vowels (Levy & Strange, 2008; Darcy et al., *under review*) – however, in both studies, the group that was more sensitive to context was the inexperienced bilinguals whereas the more advanced listeners had recovered from this effect. It may be the case that in our experiment, things are different because we are testing children, who may need more time than adults to stabilize categories. Still, the influence of the L1 in the children's perceptual patterns cannot be doubted.

The results of our experiments allow for a reconciliation of previously conflicting results. Our study has controlled for a narrow range of the age of exposure: All our participants were first exposed to L2 between 2 and 4 years of age. In the Barcelona studies (Pallier et al., 1997), similarly controlled age ranges also led to differences in perceptual behavior between early Spanish-dominant (aka L2 learners) vs. Catalan-dominant (aka native speakers) participants. It has been argued that the contrast examined in those studies ([e]-[ɛ]) could be a reason for the finding of differences,

because it was in the range of allophonic variation in the L1 of the L2 learners (Spanish). Our findings solved the allophonic debate by contrasting two categories [i:] and [e:] that are not in allophonic variation with each other in Turkish. We still observed a clear difference in behavior between the two groups, even with a very narrow range of first exposure. As a result, L1 influence is visible at such an early age for any contrast that is expected to be difficult.

Of course, our findings leave open the possibility that the social setting in which our participants were living is responsible for the observed differences between groups in our experiment, compared to the absence of differences in other research. The social setting of our experiments is in a way similar to the bilingual society of Barcelona, but there are central differences. It is similar because children are schooled in a bilingual school, and interact on a daily basis in both languages. However, they cannot assume to be understood in either language at any moment in time, as shown by the clear situational separation in their use of each language. Because of this separation, we expected a higher functional pressure to acquire L2 than in the Barcelona studies, and this may be at work in the accurate production performance we find. But at the same time, our participants have a high pressure to maintain the L1 intact (again, this seems similar to the case in Barcelona), which is likely an important factor in our observing differences between groups in perception. In order to test this possibility, it would be necessary to test a comparable pool of participants in a situation where the L1 doesn't need to be maintained intact as much, or where the L1 is not spoken widely, or, also, where the parents are fluent in both L1 and L2. A manipulation of the experimental setting for language mode could also be undertaken.

This also allows us to suggest that those studies that *failed* to find differences between early learners and native speakers (usually with wide age ranges) either used tasks that were not sensitive enough to uncover the influence of L1, or indeed, tested participants whose L1 was not highly activated (Baker et al., 2002; Flege, MacKay & Meador, 1999; Mack, 1989, where all participants had to be English dominant; Tsukada et al., 2005, for production, using cued repetition. See also Yeni-Komshian, Flege & Liu, 2000). In the case of studies that do find differences despite a wide age range, however, we rather surmise that it is probably not just an artifact of the later arrivals

(unlike what we suggested earlier, see 1.2). It seems more likely that all participants experience L1 influence (It is not possible to tell for the earlier arrivals because individual results were not presented). In sum, it seems that it is not really age that matters most to resemble native speakers. It is likely more the effect of the continued and sustained use of L1, as soon as that L1 is already established as L1 in its own right (unlike for simultaneous bilinguals where both languages are established in parallel). The reduced influence of L1 claimed to characterize children compared to adults (Baker et al., 2008, or Tsukada et al., 2005) is also likely to be dependent on a reduced use of L1.

We tested children in order to have a better understanding of the early stages of category development and of the development of the relationship between L1 and L2 categories. Since children exposed early to L2 are sometimes said to have less strong L1-L2 interference (compared to an adult-onset of exposure, see Flege, 1992), it is a logical possibility that this interaction between L1 and L2 categories changes over time. It could therefore be different in early *adults* bilinguals compared to early *children* bilinguals. It is not possible to infer children's competence from adult data. Even in case early bilingual children are different from native speakers, it is still a possibility that they over time gradually resemble native speakers as they become early bilingual *adults* and reinforce the categories they form in both languages, and/or also maybe in combination with a changing use of the L1 compared to the L2. Conversely, interference could be stronger in early bilingual adults tested as adults than if they were tested as children. The lesser interference is usually attributed to the unfinished L1 categories, which does not prevent the establishment of L2 categories. But as children continue to become adults, their L1 categories also continue to stabilize, and therefore, interference may appear in adults even though it was not apparent as children. This may be again determined by the use of L1 compared to L2. In this paper, we were not testing this hypothesis specifically, since adult early bilinguals are not compared to children early bilinguals (Darcy & Krüger, *in preparation*). Nevertheless, it is important to know how children who are still in the process of acquiring both languages “deal” with the interference from both languages.

In perception, our data show that early bilingual *children* behave differently from age-matched monolinguals on both contrasts that were expected to be difficult as a

result of a possible L1 influence. However, Turkish early bilingual *adults* who were asked to distinguish the same contrasted sounds we used did not differ from a group of adult German monolinguals (Oturán, 2002) – even if this claim is based on a different methodology (vowel confusion scores, see 1.1). If early bilingual adult learners indeed behave differently, and closer to German monolinguals (as is suggested by data in Oturan, 2002), there is a need to understand better how this influence of L1 early on – as evidenced in our group of 10 year-old children – can be reduced or reversed as development progresses. If this phenomenon were to be confirmed, it maybe linked to a slowly decreasing use of the L1 (Turkish) over time – something that could amount to L1 attrition. In Oturan’s work, unfortunately, the extent of the early learners’ Turkish and German use is not specified. They all learned German in kindergarten or preschool between 3 and 6, and they had lived in Germany for about 20 years, but half were tested in a Turkish cultural center in Germany by a Turkish experimenter. But it is nevertheless a possibility (they all were working or studying, and using German for professional reasons) that, in their daily lives, Turkish plays a less important role than in the lives of the participants we tested. It is important to understand how an early bilingual child who seems to behave like a late learner (in perception, our data) can “turn into” a “typical early bilingual” whose performance is close to that of natives, as described by many studies (e.g. Højen & Flege 2006, with the exception of Pallier et al. 1997 and the Barcelona studies), and what factors contribute to this transformation.

The amount of L2/L1 use could be one of these factors. Use seems to have a strong effect – sometimes even larger than “age” of exposure – and it seems to be able to “overturn” the effects of L1, and yield (in those studies who controlled for it) more native like results in perception and production (Flege et al., 1999; Flege & MacKay, 2004; Piske, Flege, MacKay & Meador, 2002; for those studies, it should be noted that the LoR was very long, about 40 years, just like the early learners of Oturan, 2002, which was about 20 years). In our study, use of L1 and L2 was very consistent in all children; Turkish was mostly reserved for use at home with the parents, German for school and friends. The bilingual children were recruited in schools that were participating in a project of Turkish-German alphabetization. As a result, the teachers were fluent in both German and Turkish, and both German native speakers and Turkish native speakers form the input of the children (even though, as mentioned above, they

are likely exposed to accented German too). Because of the age of the participants, we refrained from asking them for their self-evaluation of their L2 competence (as did e.g. Sebastián-Gallés and Soto-Faraco, 1999). An exact determination of their perceived dominance was therefore not possible; it is however very likely that both languages were heard daily and had a comparable importance. For such cases, because of the high activation/use of the L1, some studies predict a strong influence of the first language that may possibly undermine the success of L2 acquisition, with possible effects both in perception and production (Flege, 1998, Piske et al., 2002). In our case, the pressure to speak German would be high, but the pressure to maintain the L1 is very high too. This seems to be the most significant impediment to acquisition of L2 categories: If the phonetic space is overlapping, the pressure to maintain L1 adequately is determinant for the modifying of perceptual boundaries.

The observations made in our experiments therefore must be interpreted accordingly in terms of the use of L1. If – as in the present study – it is high, and despite the experimental setting favoring activation of German, it is not the case that the phonological acquisition of early learners will be free of L1 influences, even when this acquisition starts very early. It seems that controlling for the use of L1 and L2 will be determinant in future studies if we want to understand whether early learners who have a low-activated L1 then really “benefit” from a reduced L1 influence and can acquire the L2 structures more like native speakers of that target language; alternatively, they might still show the same pattern of L1 influence in perception, as observed in this study and in others such as Sebastián-Gallés and Soto-Faraco (1999), Pallier et al., (1997). If L1 use turns out to be a central factor in predicting L2 phonological acquisition levels, it will be of equal importance to evaluate the L1 phonological competence as well. It may well be that what is at work here is a compromise between maintaining two functioning systems, and incurring the influence of L1 onto L2 processing, the alternative being losing one system (L1) to fully acquire the L2 system. This view, which is represented in studies like Ventureyra et al. (2004), and in other attrition work (see the review in Bardovi-Harlig & Stringer, 2010), may also be limited to segmentals and be linked to the hypothesis that L1 and L2 segmental categories coexist in the same phonetic space (Flege, 1995). Maintaining both at an equal level of activation and accessibility would then necessarily reduce the attainment possible for

the acquisition of L2 categories. If on the other hand the first language is used less, this would then allow for the space to be “taken over” by the category specifications pertaining to the L2. In other words, the best way to limit the influence of L1 may be to stop using it (see Ventureyra et al., 2004). For many bilinguals, of course, this will not be an option. If at all possible, an approaching situation will likely be more easily observed in immigration situations where the pressure to acquire the L2 is considerable, and where maintaining the L1 may seem less appealing. In such cases, L2 phonological acquisition will be boosted by an early start. Of note, the possibility exists that such L1 interference problems will not have any measurable impact on the encoding of lexical items in early learners (as suggested by the dissociation between the perception and production data in this study). The problems we observed could be restricted to highly specific test situations of phoneme categorization.

We conclude with a mention of the limitations of this study. The perception and production data were obtained with different kinds of stimuli. Even though monolingual participants excelled at both, and even though we used spontaneous production, this design doesn't allow for very strong conclusions regarding a possible advantage of production over perception.

The conclusions regarding production performance are also only based on acoustic analysis, which is necessarily incomplete. We have not measured every possible aspect of the production, and we only had three words for each vowel. We concluded that the contrasts examined were accurately acquired in production because all children produced significant duration and spectral differences. This test might be complemented by asking German age-matched children to judge the productions of the bilinguals: If they can correctly identify the referent intended by each word or the vowel intended in each word, this would provide additional evidence for our interpretation of the acoustic analysis.

An important aspect of the discussion has been the use of L1 vs. L2 in the participants tested. Even if the frequency of use of each language was deemed comparable, in order to obtain a more detailed picture of this situation in our participants, we could also inspect their Turkish proficiency, or collect perception and production data in Turkish, which then would be compared to a group of age-matched monolingual Turkish children (Yeni-Komshian, Flege & Liu, 2000). Further

understanding of the variability found across studies could indeed potentially be obtained by examining systematically the L1 of the early learners. If it is in the process of being attrited, it is to be expected that the L2 phonological acquisition performance will closely resemble the target language of monolinguals also in phonetic characteristics. If the L1 is highly used, as in the present experiments, we expect that L2 phonological acquisition will be negatively affected, regardless of the age at which L2 acquisition starts.

In conclusion, this study reinforces previous claims that L1 influence will strongly affect L2 vowel perception even when exposure starts very early. In contrast, L1 influence does not lead to significant production difficulties. These data support the possibility that early sequential L2 phonological acquisition parallels later L2 acquisition for perception, but follows a different path for production.

Statement about authors' contribution and acknowledgements

These results are based on the second author's master thesis advised by the first author. I.D. and F.K. worked on the design and the stimuli, F.K. conducted the recruitment and testing, I.D. and F.K. worked on the statistical analysis, the interpretation and wrote the paper. We are indebted to R. Krüger who has designed and programmed the computer scripts used for experiments 1 and 2. We also thank Kathleen Bardovi-Harlig as well as audiences at Indiana University and at the 12th LabPhon Conference in Albuquerque for insightful feedback.

References

- Antoniadis, Z., & Strube, H. W. (1984). Untersuchungen zur spezifischen Dauer deutscher Vokale. *Phonetica*, 41(2), 72-87.
- Asher, J., & Garcia, R. (1969). The Optimal Age to Learn a Language. *The Modern Language Journal*, 53, S. 334-341.
- Baker, W., & Trofimovich, P. (2005). Interaction of Native- and Second-Language Vowel System(s) in Early and Late Bilinguals. *Language and speech*, 48(1), 1-27.
- Baker, W., Trofimovich, P., Flege, J. E., Mack, M., & Halter, R. (2008). Child--Adult Differences in Second-Language Phonological Learning: The Role of Cross-Language Similarity. *Language and Speech*, 51(4), 317-342.
- Baker, W., Trofimovich, P., Mack, M., & Flege, J. E. (2002). The Effect of Perceived Phonetic Similarity on Non-Native Sound Learning by Children and Adults. In B. Skarabela, S. Fish & A. H.-J. Do (Eds.), *BUCLD 26: Proceedings of the 26th Annual Boston University Conference on Language Development* (pp. 36-47). Somerville, MA: Cascadilla Press.
- Bardovi-Harlig, K. & Stringer, D. (2010). Variables in second language attrition: Advancing the state of the art. *Studies in Second Language Acquisition*. 32(1), 1-45.
- Best, C. T. (1995). A direct realist view of cross-language speech perception. In W. Strange (Ed.), *Speech perception and linguistic experience. Issues in cross-language research* (pp. 171-204). Timonium (MD): York Press.
- Best, C. T., McRoberts, G. W., & Goodell, E. (2001). Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system. *The Journal of the Acoustical Society of America*, 109(2), 775-794.

- Best, C. T., McRoberts, G. W., & Sithole, N. M. (1988). Examination of perceptual reorganization for nonnative speech contrasts: Zulu click discrimination by English-speaking adults and infants. *Journal of Experimental Psychology: Human Perception and Performance*, 14(3), 345-360.
- Bever, T. (1981). Normal acquisition processes explain the critical period for language learning. In K. Diller (Ed.), *Individual differences and universals in language learning aptitude* (pp. 176-198). Rowley, MA: Newbury House.
- Birdsong, D., & Molis, M. (2001). On the Evidence for Maturational Constraints in Second-Language Acquisition. *Journal of Memory and Language*, 44(2), 235-249.
- Bohn, O.-S. (2002). On Phonetic Similarity. In P. Burmeister, T. Piske & A. Rohde (Eds.), *An Integrated View of Language Development: Papers in Honor of Henning Wode* (pp. 191-216). Trier: Wissenschaftlicher Verlag.
- Bohn, O.-S., & Flege, J. E. (1992). The production of new and similar vowels by adult German learners of English. *Studies in Second Language Acquisition*, 14, 131-158.
- Bosch, L., Costa, A., & Sebastian-Galles, N. (2000). First and second language vowel perception in early bilinguals. *European Journal of Cognitive Psychology*, 12(2), 189-221.
- Cimilli, N., & Liebe-Harkort, K. (1979). *Sprachvergleich Türkisch-Deutsch* (2. edition). Düsseldorf: Pädagogischer Verlag Schwann.
- Darcy, I., Dekydtspotter, L., Sprouse, R., Scott, J. H. G., Glover, J., McGuire, M., et al. (2009). Lexical encoding of new phonological categories. *Second Language Research Forum*. University of Michigan, East Lansing.
- Féry, C. (2004). *Phonologie des Deutschen: Eine Optimalitätstheoretische Einführung - Teil 1* (3. edition). Potsdam: Universitäts-Verlag.
- Flege, J. & MacKay, I. (2010). "Age" effects on second language acquisition. *Proceeding of the 6th International Symposia on the Acquisition of Second-language Speech*, May 1-3, 2010, Poznan, Poland [<http://www.ifa.amu.edu.pl/newsounds/introduction>], retrieved October 4, 2010
- Flege, J. E. (1992). Speech learning in a second language. In C. A. Ferguson, L. Menn, & C. Stoel-Gammon (Eds.), *Phonological development: models, research, implications* (pp. 565-604). Timonium, MD: York Press.
- 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. (1998). The Role of Subject and Phonetic Variables in Second-Language Learning. In M. Gruber, D. Higgins, K. Olsen, & T. Wysocki (Hrsg.), *Papers From the 34th Annual Meeting of the Chicago Linguistic Society, Volume II, The Panels* (S. 213-232). Chicago: Chicago Linguistic Society.
- Flege, J. E., & MacKay, I. R. A. (2004). Perceiving vowels in a second language. *Studies in Second Language Acquisition*, 26(01), 1-34.
- Flege, J. E., Birdsong, D., Bialystok, E., Mack, M., Sung, H., & Tsukada, K. (2006). Degree of Foreign Accent in English Sentences Produced by Korean Children and Adults. *Journal of Phonetics*, 34, S. 153-175.
- Flege, J. E., MacKay, I. R. A., & Meador, D. (1999). Native Italian speakers' perception and production of English vowels. *The Journal of the Acoustical Society of America*, 106(5), 2973-2987.
- Flege, J. E., Munro, M. J., & MacKay, I. R. A. (1995). Factors affecting strength of perceived foreign accent in a second language. *The Journal of the Acoustical Society of America*, 97(5), 3125-3134.
- Flege, J. E., Yeni-Komshian, G. H., & Liu, S. (1999). Age constraints on second-language acquisition. *Journal of Memory and Language*, 41, 78-104.
- Göksel, A., & Kerslake, C. (2005). *Turkish: A Comprehensive Grammar*. London: Routledge.
- Grosjean, F. (1989). Neurolinguists, Beware! The Bilingual is Not Two Monolinguals in One Person. *Brain and Language*, 36, S. 3-15.
- Grosjean, F. (2001). The Bilingual's Language Modes. In J. Nicol (Hrsg.), *One Mind Two Languages - Bilingual Language Processing* (S. 1-22). Walden, Mass.: Blackwell Publishers.
- Guillelmon, D., & Grosjean, F. (2001). The gender marking effect in spoken word recognition: The case of bilinguals. *Memory & Cognition*, 29(3), 503-511.
- Guion, S. G., Harada, T., & Clark, J. J. (2004). Early and late Spanish-English bilinguals' acquisition of English word stress patterns. *Bilingualism: Language and Cognition*, 7(03), 207-226.
- Hayes-Harb, R., & Masuda, K. (2008). Development of the ability to lexically encode novel second language phonemic contrasts. *Second Language Research*, 24(1), 5-33.

- Højen, A., & Flege, J. E. (2006). Early learners' discrimination of second-language vowels. *Journal of the Acoustical Society of America*, 119(5), 3072-3084.
- Johnson, J. S., & Newport, E. L. (1989). Critical Period Effects in Second Language Learning: The Influence of Maturational State on the Acquisition of English as a Second Language. *Cognitive Psychology*, 21(1), 60-99.
- Kessler, C. (1984). Language acquisition in bilingual children. In N. Miller (Ed.), *Bilingualism and language disability: Assessment and remediation* (pp. 26-54). London: Croom Helm.
- Khattab, G. (2007). Variation in Vowel production by English-Arabic bilinguals. In J. Cole & J. I. Hualde (Eds.), *Laboratory Phonology 9* (pp. 383-410). Berlin: Mouton de Gruyter.
- Kohler, K. J. (1999). German. In *Handbook of the International Phonetic Association: A Guide to the Use of the International Phonetic Alphabet* (pp. 86-89). Cambridge: Cambridge University Press.
- Kornfilt, J. (1997) *Turkish. A descriptive grammar*. London: Routledge.
- Kuhl, P. K., & Meltzoff, A. N. (1996). Infant vocalizations in response to speech: Vocal imitation and developmental change. *The Journal of the Acoustical Society of America*, 100(4), 2425-2438.
- Lee, S., Potamianos, A., & Narayanan, S. (1999). Acoustics of children's speech: Developmental changes of temporal and spectral parameters. *The Journal of the Acoustical Society of America*, 105(3), 1455-1468.
- Lenneberg, E. H. (1967). *Biological Foundations of Language*. New York: Wiley.
- Levy, E. S. 2009a: Language experience and consonantal context effects on perceptual assimilation of French vowels by American-English learners of French. *Journal of the Acoustical Society of America*, 125, 1138-1152.
- Levy, E. S. 2009b: On the assimilation-discrimination relationship in American English adults' French vowel learning. *Journal of the Acoustical Society of America*, 126, 2670-2682.
- Levy, E. S., & Strange, W. (2008). Perception of French vowels by American English adults with and without French language experience. *Journal of Phonetics*, 36, 141-157.
- Mack, M. (1989). Consonant and vowel perception and production: Early English-French bilinguals and English monolinguals. *Perception & Psychophysics*, 46(2), 187-200.
- Macmillan, N. A., & Creelman, C. D. (2005). *Detection Theory: A User's Guide* (2. edition). New Jersey: Lawrence Erlbaum Associates.
- Mayo, L. H., Florentine, M., & Buus, S. (1997). Age of Second-Language Acquisition and Perception of Speech in Noise. *Journal of Speech, Language and Hearing Research*, 40(3), 686-693.
- McAllister, R., Flege, J. E., & Piske, T. (2002). The influence of L1 on the acquisition of Swedish quantity by native speakers of Spanish, English and Estonian. *Journal of Phonetics*, 30(2), 229-258.
- Meador, D., Flege, J. E., & Mackay, I. R. A. (2000). Factors affecting the recognition of words in a second language. *Bilingualism: Language and Cognition*, 3(01), 55-67.
- Menges, K. H. (1994). *Turkic Languages and Peoples: An Introduction to Turkic Studies* (2. edition). Wiesbaden: Harrassowitz.
- Montrul, S. (2005). Second language acquisition and first language loss in adult early bilinguals: exploring some differences and similarities. *Second Language Research*, 199-249.
- Navarra, J., Sebastián-Gallés, N., & Soto-Faraco, S. (2005). The Perception of Second Language Sounds in Early Bilinguals: New Evidence From an Implicit Measure. *Journal of Experimental Psychology: Human Perception and Performance*, 31(5), 912-918.
- Nimz, K. (2011) Vokalperzeption und -produktion türkischer DaF-Lerner. Poster presented at the symposium "Empirische Methoden in Deutsch als Fremd- und Zweitsprache" Universität Bielefeld, February.
- Ota, M., Hartsuiker, R. J., & Haywood, S. L. (2009). The KEY to the ROCK: Near-homophony in nonnative visual word recognition. *Cognition*, 111(2), 263-269.
- Oturan, K. (2002). *Vokalperzeption und -produktion in der Zweitsprache : eine Untersuchung mit türkischen Muttersprachlern*. Aachen: Shaker.
- Oyama, S. (1976). A sensitive period for the acquisition of a nonnative phonological system. *Journal of Psycholinguistic Research*, 5(3), 261-283.
- Pallier, C., Bosch, L., & Sebastian-Gallés, N. (1997). A limit on behavioral plasticity in speech perception. *Cognition*, 64(3), B9-B17.
- Pallier, C., Colomé, A., & Sebastian-Gallés, N. (2001). The influence of native-language phonology on lexical access: exemplar-based versus abstract lexical entries. *Psychological Science*, 12(6), 445-449.

- Patkowski, M. S. (1980). The sensitive period for the acquisition of syntax in a second language. *Language Learning*, 30(2), 449-468.
- Penfield, W., and Roberts, L. (1959). *Speech and Brain Mechanisms*, Princeton University Press, Princeton.
- Piske, T., Flege, J. E., MacKay, I. R. A., & Meador, D. (2002). The production of English vowels by fluent early and late Italian-English bilinguals. *Phonetica*, 59, 49-71.
- Ramon-Casas, M., Swingle, D., Sebastián-Gallés, N., & Bosch, L. (2009). Vowel categorization during word recognition in bilingual toddlers. *Cognitive Psychology*, 59(1), 96-121.
- Schwarz, W. (2008). *Skript zu LV: Die Signalentdeckungstheorie in der Psychologie*. Grundbegriffe der Signalentdeckungstheorie, Universität Potsdam.
- Scovel, T. (1969). Foreign accents, language acquisition, and cerebral dominance. *Language Learning*, 19, 245-253.
- Scovel, T. (2000). A critical review of the critical period research. *Annual Review of Applied Linguistics*, 20, 213-223.
- Sebastián-Gallés, N., & Soto-Faraco, S. (1999). Online processing of native and non-native phonemic contrasts in early bilinguals. *Cognition*, 72(2), 111-123.
- Sebastián-Gallés, N., Echeverría, S., & Bosch, L. (2005). The influence of initial exposure on lexical representation: Comparing early and simultaneous bilinguals. *Journal of Memory and Language*, 52(2), 240-255.
- Selen, N. (1979). Söyleyiş sesbilimi, akustik sesbilim ve Türkiye Türkçesi. Türk Dil Kurumu, Ankara.
- Sendlmeier, W. F. (1981). Der Einfluss von Qualität und Quantität auf die Perzeption Betonter Vokale des Deutschen. *Phonetica*, 38, S. 291-308.
- Spiegel, M. R., & Stephens, L. J. (1998). *Schaum's Outline of Theory and Problems of Statistics* (3. edition). New York: MacGraw-Hill.
- Strange, W., Weber, A., Levy, E., S., Shafiro, V., Hisagi, M., & Nishi, K. (2007). Acoustic variability within and across German, French, and American English vowels: Phonetic context effects. *The Journal of the Acoustical Society of America*, 122(2), 1111-1129.
- Syrdal, A. K., & Gopal, H. S. (1986). A perceptual model of vowel recognition based on the auditory representation of American English vowels. *The Journal of the Acoustical Society of America*, 79(4), 1086-1100.
- Topbaş, S., & Yavaş, M. (2006). Phonological Acquisition and Disorders in Turkish. In H. Zhu, & B. Dodd (Hrsg.), *Phonological Development and Disorders in Children: A Multilingual Perspective* (S. 233-261). Great Britain: Cromwell Press Ltd.
- Tsukada, K., Birdsong, D., Bialystok, E., Mack, M., Sung, H., & Flege, J. (2005). A developmental study of English vowel production and perception by native Korean adults and children. *Journal of Phonetics*, 33(3), 263-290.
- Ventureyra, V. A. G., Pallier, C., & Yoo, H.-Y. (2004). The loss of first language phonetic perception in adopted Koreans. *Journal of Neurolinguistics*, 17(1), 79-91.
- Wängler, H. H. (1981). *Atlas Deutscher Sprachlaute* (7. Ausg.). Berlin: Akademischer Verlag.
- Whitworth, N. (2000). Acquisition of VOT and Vowel Length by English-German Bilinguals: A Pilot Study. In D. Nelson, & P. Foulkes (Hrsg.), *Leeds Working Papers in Linguistics and Phonetics* (Bd. 8, S. 229-243). Leeds: University of Leeds.
- Wiese, R. (1996). *The Phonology of German*. Oxford: Clarendon.
- Yeni-Komshian, G. H., Flege, J. E. and Liu, S. (2000). Pronunciation proficiency in the first and second languages of Korean-English bilinguals. *Bilingualism: Language and Cognition*, 3, 131-149.
- Zimmer, K., & Orgun, O. (1999). Turkish. In *Handbook of the International Phonetic Association: A Guide to the Use of the International Phonetic Alphabet* (pp. 154-156). Cambridge: Cambridge University Press.