Vowel perception and production in Turkish children acquiring L2 German

Isabelle Darcy [a,*], Franziska Krüger [b,c]

[a] Department of Second Language Studies, Indiana University, 1021 E. Third Street, Bloomington, IN 47405, USA
[b] Department of Linguistics, University of Potsdam, Germany
[c] Department of Germanic Studies, Indiana University, Bloomington, IN 47405, USA

A R T I C L E   I N F O

A B S T R A C T

First language (L1) phonological categories strongly influence late learners' perception and production of second language (L2) categories. For learners who start learning an L2 early in life ("early learners"), this L1 influence appears to be substantially reduced or at least more variable. In this paper, we examine the age at which L1 vowel categories influence the acquisition of L2 vowels. We tested a child population with a very narrow range of age of first exposure, controlling for the use of L1 vs. L2, and various naturally produced contrasts that are not allophonic in the L1 of the children. An oddity discrimination task provided evidence that children who are native speakers of Turkish and began learning German as an L2 in kindergarten categorized difficult German contrasts differently from age-matched native speakers. Their vowel productions of these same contrasts (un-cued object naming) were mostly target-like.

© 2012 Elsevier Ltd. All rights reserved.

1. Introduction

Research on second language (L2) acquisition suggests that the performance of early learners is globally more native-like than late (adult) learners' in all domains of language acquisition (e.g. Montrul, 2005). Regarding phonological acquisition, specifically the acquisition of phonetic categories, studies have shown that early learners have a weaker foreign accent than late learners (Flege, Yeni-Komshian, & Liu, 1999), that they produce and perceive L2 vowels and consonant categories more accurately (Baker & Trofimovich, 2005), and that they recognize more L2 words in noise (Mayo, Florentine, & Buus, 1997; Meador, Flege, & Mackay, 2000).

However, the source of age effects is still controversial (Flege & MacKay, 2010). A well-known hypothesis to explain age effects is the Critical Period Hypothesis, according to which any learner exposed to an L2 after puberty would – in the domain of phonology – retain a foreign accent, whereas an exposure before the end of the critical period (before puberty) would lead to native-like pronunciation. The underlying mechanism is a loss of neural flexibility or plasticity (Scovel, 2000). However, studies examining pronunciation abilities have found evidence of the influence of the native language (through the presence of a foreign accent) in learners who began L2 acquisition well before puberty (Abrahamsson & Hyltenstam, 2009; Flege et al., 1999; Oyama, 1976).

An alternative hypothesis relates to the higher quality and quantity of native speaker input received by early learners in comparison to late learners. However, even early learners who are exposed to an L2 extensively from early childhood through adulthood (Pallier, Bosch, & Sebastián-Gallés, 1997) still experience difficulties when acquiring L2-specific vowel contrasts, suggesting that more input in itself may not entirely preclude the first language from interfering with L2 phonological acquisition.

Yet another hypothesis about the source of age effects is the Interaction Hypothesis (Baker, Trofimovich, Flege, Mack, & Halter, 2008; Flege, 1992). It interprets age effects in terms of the intensity of the interaction between L1 and L2. Since L1 categories are more developed in adults than in children, it is assumed that the interaction between L1 and L2 categories will be stronger in adults. The L1 phonological categories of children are still developing, and only mildly interfere (if at all) with the acquisition of L2 speech sounds, thus allowing for children's (or early learners') more accurate perception and production of the sounds of the L2 (Baker et al., 2008; Baker, Trofimovich, Mack, & Flege, 2002; Tsukada et al., 2005) as compared to adults (or late learners).

While most studies agree that early learners outperform late learners in various phonological tasks of perception and production, the state of knowledge regarding the differences between *early learners and native speakers* is far less clear-cut. The influence of L1 on the acquisition of an L2 in early learners has been described as ranging from absent (i.e. early learners exhibit native-like acquisition), to strong (i.e. the L1 influence is comparable to the degree

* Corresponding author. Tel.: +1 812 855 0033; fax: +1 812 855 5605.
  E-mail addresses: idarcy@indiana.edu (I. Darcy),
fkruger@indiana.edu (F. Krüger).

present in late learners). It is still debated when exactly the influence of the L1 phonological system starts impeding native-like acquisition of both L1 and L2 phonologies in sequential bilingual speakers, because L1 acquisition is not completed by the time L2 acquisition starts. Thus, the main goal of this study is to identify whether L1 categories can interact with L2 category acquisition when L2 acquisition begins early in life, and at what age of acquisition interaction effects might emerge. In addition, this study examines which factors contribute to the emergence of these effects. The studies reviewed below indicate that if there is interaction between L1 and L2 in early learners, it is much more variable and unpredictable than for late learners.

1.1. Studies suggesting early learners' performance equals native speakers'

Many studies have found no difference between native speakers' and early learners' performance in vowel and consonant perception and production, or in word stress patterns (Baker et al., 2002; Flege, MacKay, & Meador, 1999; Guion, Harada, & Clark, 2004; Mack, 1989; Oturan, 2002; Tsukada et al., 2005). Flege et al. (1999) examined Italian–English early bilinguals whose performance was found to be indistinguishable from an English native speaker group in an L2 vowel production and perception task, suggesting that when L2 acquisition starts early (Age of Acquisition, AoA $M=7$ years), native-like performance is attainable. Baker et al. (2002) also found the performance of a group of Korean early learners of English who had immigrated to the United States (AoA $M=8.2$ years) to be indistinguishable from a group of native English speakers in a vowel discrimination task, whereas the results of a group of late learners were different from the native speakers' on English-only contrasts. Production accuracy measured through vowel intelligibility paralleled the perception data: the early Korean–English bilinguals' accuracy was native-like, whereas the late learners' was not. In a picture naming task, Tsukada et al. (2005) also found no differences between the performance of early learners (Korean children, AoA $M=9$ years, range: 6–14) and age-matched native speakers.

While there is some evidence that nativelikeness is attainable if L2 learning starts early, it is premature to conclude that early learners are generally native-like, or that L1–L2 interaction is insignificant in early childhood. The amount of L1/L2 use can also influence outcomes. Studies by Flege and MacKay (2004) and Piske, Flege, MacKay, and Meador (2002) provide evidence that nativelikeness is observed mostly in early bilinguals who rely primarily on their L2 and seldom use their L1, but especially in individuals who used their L2 more than their L1 during early childhood (see also Højen & Flege, 2006). For instance, Flege and MacKay (2004) found that two groups of early Italian–English bilinguals (AoA $M=7.5$ years, range: 2–13) who differed only in L1 use (low vs. high) also differed in their ability to detect mispronunciations in their L2. While the results of the low-early group were exactly like a group of native English speakers, those of the high-early group indicated differences. In fact, they were similar to a group of late Italian–English bilinguals. By having native speakers rate the goodness of nonwords read by bilinguals, Piske et al. (2002) likewise found a slight difference between two groups of early bilinguals (AoA $M=7$ years) in their vowel productions. Several of the high-early participants' vowels received lower ratings than vowels spoken by native speakers. None of the ratings of the vowels spoken by the low-early group differed from the ratings obtained by native speakers' vowels.

1.2. Studies suggesting early learners' performance differs from native speakers'

A number of studies report differences in performance between early bilinguals and native speakers. Global accentedness studies show that even in early learners with an AoA of less than 10 years, a foreign accent can be detected (Flege et al., 1999, 2006; Flege, Munro, & MacKay, 1995). Studies that specifically measure phonological acquisition in terms of the production and perception of phonetic categories also report the influence of L1 categories to be strong enough to impede a native-like processing of L2 categories, indicating that early and intensive exposure to a second language may not necessarily be enough to build native-like phonemic categories (Bosch, Costa, and Sebastián-Gallés, 2000), or to perform like native speakers in discrimination (Højen & Flege, 2006; Navarra, Sebastián-Gallés, & Soto-Faraco, 2005; Pallier et al., 1997; Sebastián-Gallés & Soto-Faraco, 1999; Tsukada et al., 2005) and vowel production tasks (Baker & Trofimovich, 2005; Baker et al., 2008).

Tsukada et al. (2005) report that Korean children exposed at an early age to English outperformed Korean late learners, but differed from English monolingual children in an oddity vowel discrimination task. The early learners varied widely in chronological age and age of first exposure (AoA range: 6–14 years), and reported using English more often than the Korean adults. This study did not find differences between the early learners and native speakers in production, unlike Baker et al. (2008) who observed differences using a similar production method, with the exception that they analyzed a spontaneously produced token, instead of an auditorily cued one as in Tsukada et al. (2005). The early learners in Baker et al. (2008) (AoA range, years;months: 6;0-13;5) were more intelligible than the late learners, but their production was still clearly different from native speakers. The learner groups did not differ in terms of self-rated L2 proficiency or use. Højen and Flege (2006) showed that adult Spanish early learners of English, with a mean AoA of 6 years (range: 2–10 years) behaved similarly to native speakers on a vowel discrimination task using a phonetically sensitive categorical ABX design. However, on a shorter inter-stimulus interval (ISI=0 ms) for two of three difficult vowel contrasts, the early bilinguals' discrimination was lower 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.

One possible reason why these studies found differences between early bilinguals' and native speakers' performance might have been the heterogeneity in their learner groups. The amount of L2 use varied considerably between learners, and Højen and Flege (2006) proposed that this factor together with an earlier AoA (between 2 and 5 years) might be responsible for individual differences found in the data. Also, the broad range in the ages of first exposure in these studies makes group behavior potentially less homogeneous, and perhaps also too variable to allow for definite conclusions. The differences between early learner and native speaker performance observed in those studies could arguably be due to the later arrivals, given the large ranges in AoA (up to 12 years). In addition, the earliest age of first exposure chosen in some studies (such as age 6 in Tsukada et al.'s study) might already have been too late to yield native-like perception.

Yet some studies do report data from learner groups with a "narrow age range". To our knowledge, the only studies offering a clearly narrow range of AoA have been conducted in Barcelona (e.g. Bosch et al., 2000; Navarra et al., 2005; Pallier et al., 1997; Sebastián-Gallés, Echeverría, & Bosch, 2005; Sebastián-Gallés & Soto-Faraco, 1999).

Pallier et al. (1997) found that adult early Spanish-dominant bilinguals (who learned Spanish first), exposed at age 4 to L2 Catalan, did not discriminate the contrast [e]~[ɛ] like Catalan-dominant bilinguals (for whom Catalan is considered their native language). However, results showed large differences between individuals, with some of the Spanish early bilinguals performing similarly to the Catalan native speakers. Other studies have also examined the acquisition and encoding of difficult contrasts such

as [e]∼[ɛ] by early bilinguals on different tasks (gating: Sebastián-Gallés & Soto-Faraco, 1999, AoA range: 3–4 years; implicit ABX: Navarra et al., 2005, AoA=3 years; lexical decision: Sebastián-Gallés et al., 2005, AoA=4 years). Results consistently revealed that Spanish-dominant early learners of Catalan do not acquire the contrast to the same degree as Catalan-dominant (native) speakers. These studies conclude 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, Colomé, & Sebastián-Gallés, 2001).

On the whole, whether or not studies find differences between early learners and native speakers might be related to four variables: age, allophony, L1 use and task differences.

1.2.1. Age range

Observed performance differences between early learners and native speakers are possibly a matter of how tightly the age ranges are controlled. In some studies where differences are found, age ranges vary greatly, yielding less conclusive data about the onset of L1 interaction with L2 category development.

1.2.2. 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. When L1 use is taken into account as a variable, it is consistently identified as a major factor in explaining variability between groups (Flege & MacKay, 2004; Mora & Nadeu, 2012). Whether or not there is pressure to maintain and actively use the L1 is important to ascertain, because a more active L1 is more likely to show evidence of interaction with the L2. In a bilingual community like Barcelona, where differences in performance between early learners and native speakers are consistently observed, the pressure to keep both languages activated is likely higher than in settings with mostly immigrant children and adults, where sometimes no differences are found (e.g. Flege et al., 1999). Typically, immigration to a country where the L1 is not widely spoken does not impart the same pressure to keep the L1 highly activated (e.g. Flege et al., 1995), especially if the L2 is also spoken at home. In addition, as a result of prolonged language contact, the exposure of Barcelona residents to accented Catalan or accented Spanish is high, hence the pressure to resemble native speakers in perception or production might be less strong than in immigrant situations (see also Mora, Keidel, & Flege, 2011; Mora & Nadeu, 2012). The case we examine here is similar to an immigration situation, since Turkish is not widely spoken in German society. Exposure to German, and the pressure to socialize in German in school, is therefore high. However, the bilingual children in our study speak only Turkish with their parents, and Turkish is the dominant language in the household; the children also have large Turkish peer groups, and they are schooled in a dual-language school. Therefore, there is pressure to keep the L1 active as well (see also Section 3.1.1).

1.2.3. Allophony

The consistency with which studies conducted on Catalan–Spanish bilinguals find differences in performance 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 type of contrast probably falls within the "single category assimilation" pattern, according to PAM (Best, 1995) or PAM-L2 (Best & Tyler,

2007), and discrimination between both vowels, as well as establishing separate L2 categories, is predicted to be very difficult. Contrasts that are not allophonic in the L1 may fall into a "category goodness difference" assimilation pattern, and may be easier to acquire. It is therefore possible that performance differences between early learners and native speakers are attenuated when examining contrasts that are not allophonic in the L1 (see also Flege & MacKay, 2004).

1.2.4. Task

Finally, the tasks 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 in performance between early learners and native speakers (see also Abrahamsson & Hyltenstam, 2009). The different findings in Tsukada et al. (2005; cued, no differences) compared to Baker et al. (2008, un-cued, significant differences) also indicate that the elicitation method (spontaneous vs. cued) might be crucial for uncovering possible differences in performance.

There are four major factors that may explain the large variability observed across the different studies: variable age-ranges, variations in the L1 use of the participants, differences due to the contrasts examined, and methodological differences. In order to be sure that differences in performance between early learners and native speakers are due to the early onset of interaction between languages and not to the combined action of other confounded factors, it is important to conduct studies in which (a) early age of first exposure is controlled for, (b) participants are learning L2 in a setting in which the pressure to maintain L1 is high, (c) contrasts examined are not allophonic in the L1 of the speakers, and (d) tasks sensitive enough to detect possible differences are used.

2. The present study

In this study we investigated the effects of early onset of L2 acquisition on the interaction of L1 and L2. We controlled for age of first exposure among learners in an environment where L1 use is high, and examined four different contrasts while collecting both perception and production data. Our goal was 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 ten-year-old early sequential bilingual Turkish–German children, who were first exposed to German between the ages of two and four. In Experiment 1, we tested their ability to discriminate four German vowel contrasts using an oddity task. In Experiment 2, we examined their production of the same vowels with an uncued word naming task.

This section begins with a review of the German and Turkish vowels systems (Section 2.1), followed by a description of the expected cross-language perceptual similarity (Section 2.2), and then presents our specific hypotheses and predictions (Section 2.3) for the perception task. Stimuli and predictions for the production task are presented in Section 4.

2.1. German and Turkish vowels

The German vowel inventory is larger than the Turkish vowel system. This comparison focuses on the six German vowels [iː], [ɪ], [eː], [ɛ], [aː] and [a] that were selected for the stimuli and the three Turkish vowels [i], [e], and [a]. To distinguish vowels, German uses the distinctive features of tenseness and duration (Féry, 2004; Wiese, 1996), neither of which is used to this end in

Turkish. Turkish vowels generally are more centralized than the German vowels, and unlike German, Turkish does not distinguish between tense and lax vowels at the phonemic level (Kohler, 1999; Zimmer & Orgun, 1999). 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 (Kornfilt, 1997; Menges, 1994; Topbaş & Yavaş, 2006).

Both German and Turkish have a phonetic category defined as a high front unrounded vowel. Turkish [i], however, is articulated more centrally than German [iː] (Kornfilt, 1997). Based on formant values reported in Selen (1979) for Turkish, and in Sendlmeier (1981), Wängler (1976) and Valaczkai (1998) for German, Turkish [i] is closer to the German lax [ɪ] than to the German tense [iː].[1] 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 (Göksel & Kerslake, 2005).

In both languages, there is a phonetic category described as a mid-front unrounded vowel, but German again distinguishes two where Turkish only has one. Turkish [e] is only slightly more open, but clearly more central than German [eː], and at comparable acoustic distance to both German [ɛ] and [eː]. As a result, German tense [eː] is often confused with the German high vowel [iː] by Turkish listeners (Cimilli & Liebe-Harkort, 1979). Allophones of Turkish [e] are [ɛ] and [æ], but not [i] (Göksel & Kerslake, 2005).

Low vowels in both languages are usually centrally articulated. Turkish [a] is described as somewhat higher and more fronted than the German low vowels (Zimmer & Orgun, 1999). The German long, low, slightly retracted vowel [aː] (as well as the short, low, slightly fronted [a]) is in close similarity to or overlaps with Turkish [a]. In sum, the comparison of the German long vowels [iː], [eː], [aː] (as in *schief* 'inclined', *Schnee* 'snow', *Hahn* 'rooster') and short vowels [ɪ], [ɛ], and [a] (*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 short (lax) vowels, and do not have a tense counterpart.

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

2.2. Perceptual similarity of German and Turkish vowels

To obtain information about the cross-linguistic perceptual similarity of vowels, perceptual assimilation patterns are usually collected from naïve participants, i.e. listeners who have had no prior exposure to the relevant speech sounds. For the present study, we report the results of perceptual assimilation patterns for German vowels as presented in Oturan (2002). The participants

---

Fig. 1. Results of the perceptual assimilation and average 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.

were 31 Turkish students (mean age=19;7 years) who had grown up in a monolingual environment; none of them had ever had any contact with German. Stimuli were /b+V/ syllables cut from the German words [biːtn̩] 'to offer', [bɪtn̩] 'to beg', [bɛtn̩] 'beds', [beːtn̩] 'to pray', [baːtn̩] 'to beg, pret. pl.' and [debaːtə] 'debate', which were produced by a phonetically trained male native German speaker from Cologne in the sentential context "Ich habe … gesagt" ('I said …'). The participants were asked to listen to the syllables and determine a Turkish equivalent for the vowels (i.e. classify them according to their native Turkish categories). They were then required to judge the perceptual similarity between the German speech sounds and their Turkish representations on a scale of 1 through 6 (1=identical, 6=different vowel—see Fig. 1).

In Fig. 1, which illustrates the mapping and perceived similarity between German and Turkish vowels, we can see that both German tense and lax [iː] and [ɪ], as well as the German tense [eː], were mapped onto the /i/ category in Turkish. The German short lax [ɛ] as in *Bett* /bɛt/ 'bed' was alone in mapping onto the Turkish front mid-vowel [e]. Both German [a] vowels were mapped onto Turkish [a]. The average goodness ratings reported ranged for most vowels from "very" to "quite" similar. These results corroborate the phonetic descriptions seen above, as well as the perceptual similarity of the German lax vowels [ɛ], [ɪ] and [a] to the Turkish vowels [e], [i] and [a]. Their goodness ratings were generally "very similar", while the tense counterparts [iː] and [aː] received a rating of "somewhat similar", but were still categorized as Turkish [i] and [a]. German tense [eː] was categorized differently: in 95% of the cases [eː] was assigned to Turkish [i].

2.3. Hypotheses and predictions

Perceptual similarity mappings have been shown to predict quite reliably how adult non-native listeners and L2 learners will discriminate between non-native phones (Best, 1995; Best, McRoberts, & Goodell, 2001; Best, McRoberts, & Sithole, 1988; Best & Tyler, 2007; Levy, 2009a, 2009b; Levy & Strange, 2008; Tsukada et al., 2005). The four German vowel pairs serving as stimuli for this study were chosen according to the way the vowels were categorized in terms of the Turkish vowel categories just reviewed. Based on the pattern observed in Fig. 1, the following predictions for discrimination, based on Best and Tyler (2007) for L2 learners, were made:

(1) [eː]~[iː], our test contrast, differs only in spectral cues. [iː] and [eː] are both equally good exemplars of the same Turkish category [i]. In this case of single category assimilation, the discrimination and establishment of two separate L2 categories was predicted to be very difficult.

(2 and 3) [iː]~[ɪ] and [eː]~[ɛ] are acoustically quite close but still distinguished by both length and spectral cues. Turkish

---

[1] The F1 and F2 values in Wängler (1976), Sendlmeier (1981) and Valaczkai (1998) were obtained from one male German speaker speaking Standard German. No indication of larger prosodic contexts is given. Vowels were embedded in monosyllabic words (minimal pairs read in randomized order for Sendlmeier; different words for Wängler: "wie", "Wind"; "See", "nett", "haben", "hatte"; in Valaczkai, several words for each vowel were recorded). Selen (1979) reports F1 and F2 values for isolated vowels, but does not report on the gender of the speaker.
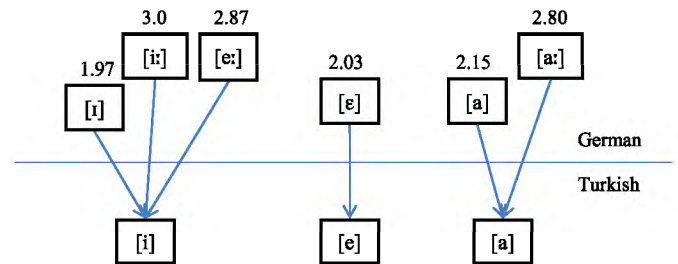
adults categorized the contrasts differently: German tense [iː] and lax [ɪ] were both mapped onto the same category, but differed in category goodness, with [iː] being a slightly less good exemplar than [ɪ] for Turkish [i]. In this case of category goodness difference, the discrimination was predicted to be more accurate than in single-category cases, but not as accurate as in the case of two-category assimilations. German tense [eː] and lax [ɛ] were clearly mapped onto two different Turkish categories, a pattern that predicted a more accurate discrimination than in both previous cases.

(4) [aː]~[iː] is our control contrast, because these vowels map onto two different categories which are further apart in the vocalic space than the mappings of the vowels of the three other contrasts. For the control contrast, very good discrimination was expected.

We hypothesized that the native categories of the Turkish early bilingual children have already been established well enough to interfere with the establishment of native-like categories for 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 conform to the predictions above. However, given that our participants can be considered experienced 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 [eː]~[iː] and [iː]~[ɪ], and would be more accurate than what is predicted for naïve listeners (see Levy, 2009a, 2009b; Tsukada et al., 2005). For instance, they may be able to attend to the duration feature (Bohn, 1995) in addition to spectral differences, which is useful in distinguishing both high vowels [iː]~[ɪ], even though duration would not help to discriminate the most difficult contrast [eː]~[iː]. On the other hand, since duration exists phonetically but is not contrastive in the L1, the children might be able to use it only to a limited extent for the discrimination task.

Because of restrictions on the number of items and the length of the task we could reasonably expect the children to perform, the contrast [aː]~[a] was eliminated in the perception task. We did not collect perceptual assimilation data from an age-matched Turkish monolingual group directly because a sufficient number of children could not be recruited in Berlin.[2] Categorization performance even without perceptual assimilation data can reveal the types of interaction between categories at work in our participants: Depending on their performance with the two more difficult contrasts ([eː]~[iː] and [iː]~[ɪ]), it would be possible to see whether L1 influence has an effect despite the early age of L2 exposure. In order to examine this hypothesis, we tested two groups of children (early Turkish–German bilinguals and German monolinguals) in an oddity vowel categorization task.

3. Experiment 1: Vowel categorization

3.1. Method

3.1.1. Participants

Twenty-eight children were tested. Fourteen were native speakers of German and fourteen were 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		Age exposed to L2	Language spoken	
				with peers	in the family
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 (daycare)	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	2;6–4 (daycare)	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, disordered 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 of Germany. The dialect of Turkish they were exposed to was restricted to standard Istanbul–Turkish (Menges, 1994, see Section 2.1). No participants in the bilingual group were significantly exposed to any language other than Turkish prior to entering daycare or kindergarten.[3] The mean age of the participants at the time of testing was 11;2 years ($M=134.7$ months, $SD=7.1$, range: 9;8–12;3 years). Their mean age of first exposure to the L2 was 2.9 years ($M=35.9$ months, $SD=5.6$, range: 2;6–4;0 years); they were therefore exposed to German for 7 years on average. Children were schooled in a dual-language Turkish–German elementary school, where learning takes place in both languages in comparable amounts. Teachers are native speakers of Turkish or German and teach in their respective L1s. Turkish is therefore used for about half of the instructional day. The school also keeps a roughly equal ratio of native Turkish and native German students. All teachers can also communicate in the other language. A linguistic separation according to social context was also visible in parents' and children's responses in the questionnaires: they indicated that the bilingual children spoke German with peers, including friends and siblings at school, but spoke only Turkish with their parents in family settings, in which case Turkish was also spoken with siblings. This situation often resulted in or stemmed from one or both parents not knowing any German. In interactions with the experimenter (who only spoke German), no problems emerged in the understanding of the task, and no clearly perceptible foreign accent was detected in most children.

---

[2] We do not compare adults and children in this study, so a comparison of their performance/perceptual similarity patterns is not crucial.

[3] Children who were exposed to German since birth were not included.

It was not possible to exactly measure the percentage of the time each child spent using their L1 and L2. In school, we assume that German was used slightly more than Turkish because of the predominant use of German during recess and with peers; however, given that Turkish was predominantly the language used within the family outside of school and on the weekends, we consider the use of German and Turkish to be relatively balanced for the bilingual children.

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 was 11.1 years ($M=134.3$ months, $SD=7.1$, range: 9;7–11;11 years). All children spoke Standard German. None of them had any significant contact with another language before entering school.

3.1.2. Stimuli

Given evidence that the consonantal context of vowels affects perceptual performance in vowel categorization (see Strange et al., 2007), we used two consonantal contexts for the five German vowels chosen for this experiment: bilabial (p_p) and velar (k_k). Stimuli consisted of the ten syllables [paːp], [kaːk], [piːp], [pɪp], [kiːk], [kɪk], [peːp], [pɛp], [keːk], [kɛk]. Two ([kɛk] 'perky' and [piːp] 'cheep, peep') are listed as words in the Celex database for German (Max Planck Institute for Psycholinguistics, 2001), each with a spoken frequency of 0 in the Mannheimer Corpus (MannS, MannSLog, MannSMln). Two other syllables can be considered loanwords from English ([kɪk] 'kick' and [pɛp] 'pep') and one a regional word form ([kiːk] 'look'), which is used in the variety spoken in Berlin, but neither is listed in the Celex database.

Three adult German native speakers from the Berlin area (two females, one male) were recorded in a quiet environment with a *Marantz PMD 670* solid state recorder at 44.1 kHz. The target items were read several times in the sentential context "Hier ist ein …" ('*Here is a …*'), and then manually cut from each sentence using the software Praat (Boersma & Weenik, 2011), to be later presented in isolation to the participants. For each item, the qualitatively best recording, as determined by voice volume, speed and surrounding noise, was chosen as a stimulus for the experiment.

3.1.3. Procedure

Children were tested in a quiet room in their school. They started with the perception task, and then performed the production task (described in Experiment 2 below). Approximate total testing time was 20 min (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 et al. (2005). Nonword syllables were presented auditorily as triads in "same" ($N=48$) or "change" ($N=48$) trials on a computer.

For each contrast ([aː]∼[iː], [iː]∼[ɪ], [eː]∼[ɛ], [iː]∼[eː]), 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 remained the same for a given contrast (for instance, [aː]∼[iː] was 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 randomly assigned to each list.

Højen and Flege (2006) argue that the AXB design might yield ceiling effects in early learners; they added a testing condition
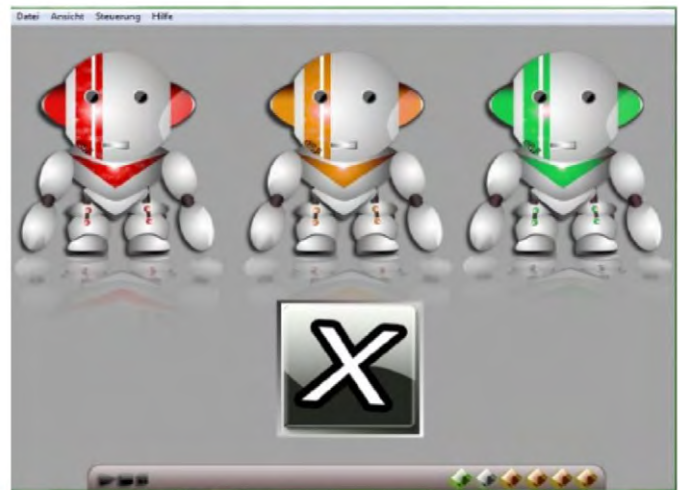

Fig. 2. Layout of the categorization task; Robot 1 (left): female voice; 2 (center): female voice; 3 (right): male voice.

with a shorter ISI, which they found to be more adequate in revealing differences between L2 learners and native speakers. Following this suggestion, the present experiment also was organized into two blocks. In the first block the ISI was 500ms; in the second block it was 0ms. All children had the same block order. The order of trials within each block was randomized for each participant. The total number of trials was 96 for each child (48 trials in each ISI). Custom software presented stimuli in the form of a game, in which three robots were displayed on a computer screen (see Fig. 2).

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 stimuli other than those used in the test (context bVf). During the familiarization phase only, children were allowed to listen to the stimuli several times, and received feedback about their answers from the experimenter. Both feedback and the repetition option were absent in the test phase. No time limit was set for the children to answer; the next trial began 1000 ms after their response.

As explained above, the activation of L1 Turkish was high in our participants because of the bilingual setting they were living in. In order to avoid artificially inflating the presence of L1–L2 interaction 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). The experiment was thus conducted entirely in German, with an experimenter who did not know any Turkish. The children were aware of this fact, which likely contributed to reducing the activation of the language that was not shared (see Khattab, 2007), in this case, Turkish. As a result, the setting we chose was expected to reduce the activation of the L1 (Grosjean, 2001).

3.2. Results

Participants' percent correct discrimination scores for the control contrast were first screened in order to establish that all participants correctly understood the task. Both vowels ([aː]∼[iː])

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

Group	[aː]~[iː]	[eː]~[ɛ]	[iː]~[ɪ]	[iː]~[eː]
Monolingual	97.92	87.8	98.29	85.42
Bilingual	98.08	82.05	76.92	56.73

are articulatorily and perceptually very different and were expected to be easily categorized as "different" by all participants. Examination of the performance (correct answers) in the monolingual group revealed that all scores 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 ($Z=-3.125$[4]). 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.

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

As shown in Fig. 3, L2 learners were significantly less accurate at discriminating the [iː]~[ɪ] and [iː]~[eː] contrasts than monolinguals. Both contrasts were predicted to be difficult. The group difference was statistically significant for both contrasts (Mann–Whitney for [iː]~[ɪ]: $U=52.5$, $p<0.05$; Mann–Whitney for [eː]~[iː]: $U=16.5$, $p<0.0001$). By comparison, L2 learners were like the monolinguals on both other contrasts ([aː]~[iː], Mann–Whitney $U=97.5$, $p>0.05$; [eː]~[ɛ] Mann–Whitney $U=87.5$, $p>0.05$).

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 $M(\text{bil})=0$ and $M(\text{mon})=0$. The one participant who had been excluded based on the detection performance also exhibited a larger bias, which was -0.48, more than 2 SD away from the group median. No other participant showed a bias that differed from the group median. The influence of bias was therefore considered minimal.

An analysis of the effect of labial vs. velar context revealed the pattern presented in Fig. 4. For the control contrast, no effect of phonetic context was visible in any group. The bilingual children identified the difference between [iː] and [ɪ] in the bilabial (pVp) context significantly less accurately than when it was presented
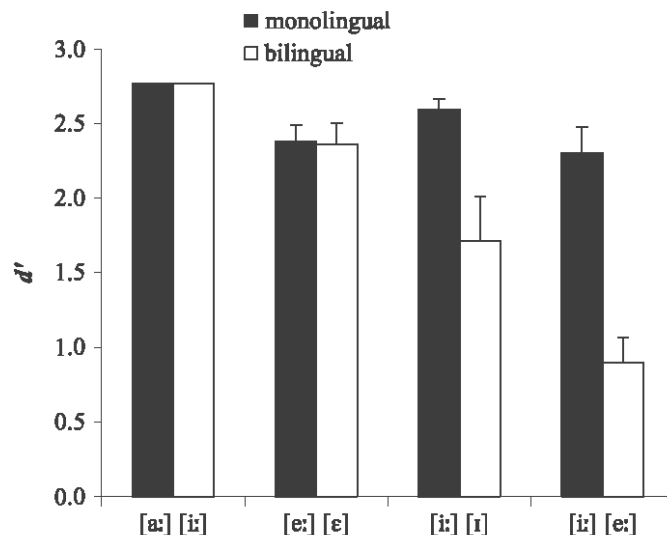

Fig. 3. Average d' values for the bilingual and monolingual groups (N bil$=13$, N mon$=14$). Error bars enclose $+$ one standard error.
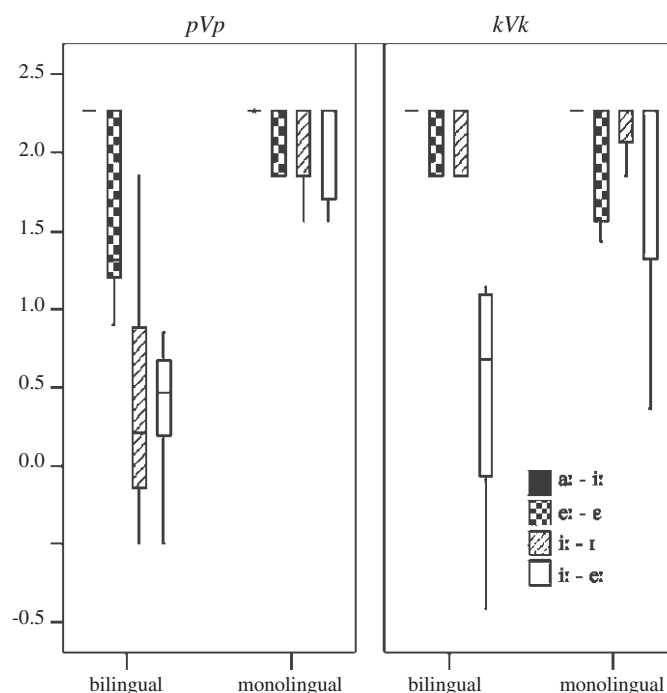

Fig. 4. Average d' values for the bilingual and monolingual groups as a function of context condition (N bil$=13$, N mon$=14$).

in the velar (kVk) context (Mann–Whitney $U=1$, $p<0.01$). The same comparison between contexts for the contrast [eː]~[ɛ] yielded a similar result: the bilinguals were more accurate at discriminating these vowels in a velar context than in a bilabial context (Mann–Whitney $U=6$, $p<0.05$). For the [eː]~[ɛ] contrast, despite d' values being more variable in the bilingual group when the contrast occurs in the bilabial context, the comparison between bilinguals and monolinguals revealed that the groups' performance did not differ significantly in each context (***pVp*** Mann–Whitney $U=14.5$, $p>0.05$; ***kVk*** Mann–Whitney $U=10$, $p>0.05$). In the case of [iː]~[ɪ], the group comparisons showed that in the velar context, both groups' performance (N bil$=6$, N mon$=7$) was not different (Mann–Whitney $U=18.5$, $p>0.05$), but in the bilabial context, their accuracy did differ (N bil$=7$, N

---

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

[5] The experiment design necessitates first detecting a difference (signal detection) and then identifying 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 correct robot is clicked in a change trial). The differences within the groups in detection and identification performance were not significant. The results here therefore only present the detection performance.

[6] With the adjustment, the d' value 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.

mon=7, Mann–Whitney $U=2$, $p < 0.01$). For the [iː]~[eː] contrast, the bilingual group's performance was comparably low regardless of the context (Mann–Whitney $U=14$, $p > 0.05$).

The ISI conditions were also examined. The only contrast for which an ISI effect was significant was [iː]~[eː]. In both groups, performance was more accurate at the shorter ISI interval of 0 ms. (bilingual: Wilcoxon $Z=-2.3$, $p=0.019$; monolingual: Wilcoxon $Z=-2.19$, $p=0.028$).

3.3. Discussion

The results of Experiment 1 showed a clear pattern of discrimination that paralleled the categorization patterns as defined through the perceptual assimilation data. With increasing perceptual similarity, discrimination ability declined. The bilingual children experienced difficulties on those contrasts that were predicted to yield the most confusion 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 in the way naïve adult listeners without experience of German would have been expected to. In this sense, our data corroborate the patterns observed by Pallier et al. (1997) with adults, and also other studies conducted with early Spanish–Catalan bilinguals who are Catalan-dominant (e.g. Bosch et al., 2000, AoA range: 4–6 years; Navarra et al., 2005; Sebastián-Gallés & Soto-Faraco, 1999). The presence of context effects, especially for the [iː]~[ɪ] contrast, which was more difficult to distinguish in the bilabial consonantal context, indicated a higher susceptibility to contextual variation on the part of the bilinguals. These effects are similar to those reported for American late learners' perception of French front rounded vowels (Levy & Law, 2010; Darcy et al., 2012)—however, in those studies, inexperienced bilinguals were more sensitive to context whereas more advanced listeners no longer showed this effect.

We infer from our results that, even though these children were exposed to German very early in life, their perception for the vowels tested here was somewhat different from that of the monolingual children of the same age. Currently, a precise quantification of these findings is difficult, since we did not test naïve adults, and late and early bilingual adult data using these contrasts and methods are still lacking (Darcy & Krüger, *in preparation*). At the current stage we can draw the tentative conclusion that early exposure was not enough for our group of bilinguals to show performance equal to the monolinguals on the vowel contrasts examined. If bilingual children perceive vowel contrasts less accurately than monolingual children, it is possible that bilinguals also produce words involving these contrasts less accurately than monolinguals (Baker et al., 2008; Baker & Trofimovich, 2005). Alternatively, because of the early age of exposure to German, it is possible that the groups do not differ in their realization of vowel categories (Khattab, 2007). The next experiment was conducted to evaluate these possibilities: Experiment 2 examined the production of the five German vowels examined in perception as well as the short low vowel [a], in order to investigate the effects of bilingualism on production.

4. Experiment 2: Vowel production

While phonetic category formation for adult bilinguals has been addressed extensively in the literature (Bohn & Flege, 1992; Flege, 1995), comparatively less attention has been devoted to bilingual children's production. Several studies suggest that despite a very early age of acquisition, transfer from the L1 into the developing L2 is present, especially for vowels (Fabiano-Smith & Goldstein, 2010; Gildersleeve-Neumann & Wright, 2010; Hecht & Mulford, 1982;

Wenzel, 2000). However, these studies are based on phonetic transcription, and children are generally younger than in our sample. Older children were examined in Tsukada et al. (2005), who used a picture naming task in which Korean children and adults produced English words three times after first hearing an auditory model. Only the first ("cued") production was analyzed; comparisons did not reveal any differences between the utterances of 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, as noted by the authors, may have been due to the early learners imitating the model more accurately than the Korean adults. Two other comparable studies (Baker & Trofimovich, 2005; Baker et al., 2008) reported acoustic differences in vowel production for early bilingual children compared to age-matched monolinguals using a a more spontaneous elicitation technique very similar to the one used in the present study. However, these studies examined children who arrived in the U.S. after the age of 7. It is still not clear to what extent bilingual children with an earlier age of acquisition can accurately produce L2 phonetic categories.

The Speech Learning Model (Flege, 1995) states that L2 categories that 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 late learners, we would expect the spectral specifications of the Turkish categories to which the contrasts are perceptually mapped to be used in production. Regarding duration, if we consider this a "new" feature, we may expect that both long front vowels [iː] and [eː] could benefit from less clear overlap with the Turkish category [i] (as opposed to German [ɪ]), and could therefore be produced more accurately than German [ɪ]. The same could apply to [aː], 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 also Bohn, 1995; Nimz, 2011).

4.1. Method

4.1.1. Participants
Participants were the same children tested in Experiment 1.

4.1.2. Stimuli
A repetition task involving the same tokens as in Experiment 1 was deemed non-suitable because it would have necessitated an auditory model. A reading task would also potentially be confounded by reading ability. To elicit spontaneous production data, we therefore chose to use a naming task (Baker et al., 2008) in form of a game without auditory modeling. To ensure that all children were familiar with the words used, frequent and imageable words were chosen. The consonantal context was kept as similar as possible for each vowel pair. It was not possible to use the same contexts as in the perception task stimuli which involved nonwords or very low frequency words, but every effort was made to hold consonantal context constant using both minimal pairs and near-minimal pairs.

For each of the five vowels tested in perception, three common German words were selected (see Table 3). In addition, three words with the low short vowel [a], which was not tested in perception, were included here. In order to allow for the [iː]~[eː] comparison, the context surrounding the vowels [iː] and [eː] was also held as close as possible: [biːst], [beːt], [ʃtiːl], [ʃneː] (*Biest—Beet*; *Stiel—Schnee*).

Table 3
Overview of the German words chosen for elicitation.

i:	*Stiel* 'handle'	[ʃtiːl]	*schief* 'inclined'	[ʃiːf]	*Biest* 'beast'	[biːst]
ɪ	*still* 'quiet'	[ʃtɪl]	*Schiff* 'ship'	[ʃɪf]	*Biss* 'bite'	[bɪs]
e:	*Beet* 'flowerbed'	[beːt]	*Fee* 'fairy'	[feː]	*Schnee* 'snow'	[ʃneː]
ɛ	*Bett* 'bed'	[bɛt]	*Fell* 'fur'	[fɛl]	*schnell* 'fast'	[ʃnɛl]
a:	*Saat* 'seed'	[zaːt]	*Hahn* 'rooster'	[haːn]	*Schwan* 'swan'	[ʃvaːn]
a	*satt* 'full'	[zat]	*Hand* 'hand'	[hant]	*Schwanz* 'tail'	[ʃvants]

For the elicitation of the words, children played 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 with a *Sennheiser* microphone on a *Marantz PMD 670* solid state recorder at 44.1 kHz.

In a training phase prior to the game, children were shown the pictures accompanied by the written form of the words, and asked to produce 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; this time the written form was not provided. This step was repeated in case of difficulties. No auditory cues (such as the word onset) or 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 × 6 vowels × 3 contexts × 2 items, or 168 per vowel) were acoustically measured. Items that had low recording quality or that contained too much background noise to ensure reliable analysis were excluded (2.9% of the data). In total, 979 items were analyzed. Acoustic measurements were carried out by inspection of wide-band spectrograms and time domain waveforms by the second author. The beginning and end of each vowel was identified by manually placing one boundary at the onset of periodicity accompanied by a steep intensity increase (corresponding to the vowel transition) and the other at the point where noticeable periodicity diminished along with a clear intensity decrease. Vowel durations were measured from the onset of periodicity up to the end of periodic energy in the waveform. For spectral measurements, the steady-state vowel mid-point was chosen to minimize possible coarticulatory effects of adjacent consonants. The mid-point was computed by a Praat script (Boersma & Weenik, 2011) based on the vowel duration, and the formants F1 and F2, as well as f0, were automatically extracted at this point. The reliability of automated measurements was verified manually by the second author on 10% of the data chosen randomly. No deviations from the script measurements were detected. Given the large inter- and intravariability of children's production, we used a normalization procedure utilizing the Bark conversion (Syrdal & Gopal, 1986; Bohn & Flege, 1992). The $F1-f0$ difference corresponds to the vowel height dimension (lower vowels have a higher value), while the $F2-F1$ difference corresponds to the horizontal dimension (front vowels have a higher value).

Out of 979 items, 35 were considered outliers, their Bark-difference values being beyond 2 *SD* from the mean of their respective group. These items were excluded from further statistical analysis.

Table 4
Average Bark difference scores for German vowels.

	Bilingual	SD	Monolingual	SD	U-value	p-value
F1 − f0						
e:	2.089	0.203	2.296	0.187	39	0.006**
ɛ	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
a	5.35	0.845	6.306	0.694	37	0.005**
F2 − F1						
e:	10.671	0.52	10.518	0.372	84	0.747
ɛ	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**
a	4.345	0.64	3.518	0.652	36	0.003**

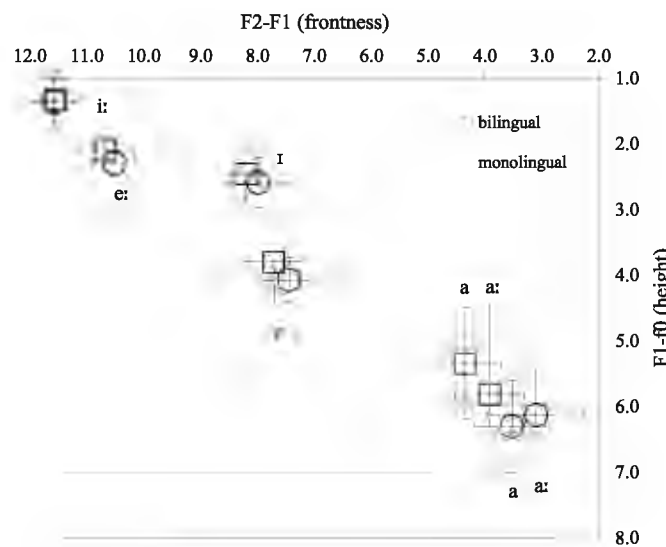** *Significant at the $p < 0.01$ level.*

Fig. 5. Mean values for each group, in Bark, for each vowel. Error bars enclose one standard deviation.

4.3. Results

4.3.1. Spectral data

The Bark values for each vowel in each group are summarized in Table 4, and graphically represented in Fig. 5. The analysis does not take the consonantal context into consideration.

The group comparison revealed that the bilinguals differed from the monolinguals in their production of the lower vowels [aː] and [a], which were both realized more front by the bilingual group. The lax [a] was also significantly higher in the bilingual productions. Another group difference was seen for [eː], which was produced higher by bilinguals than by monolinguals. Minor differences in the horizontal dimension were visible for [ɛ] and

[ɪ], articulated farther forward by the bilinguals than by the monolinguals; for both vowels, the two groups did not differ significantly with regard to the vertical dimension.

Considering the test vowel pairs used in the perception task, the bilingual group produced adequately differentiated vowels, comparable to the monolingual group. The spectral difference between [iː]∼[ɪ] was significant for both the bilingual and monolingual group. In both groups was [iː] higher and more fronted 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 [eː]∼[ɛ], the contrast analysis also showed that the difference between both vowels was produced adequately by all children. The [eː] was 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 [iː]∼[eː] contrast, both groups of children produced the expected spectral contrast difference between these tense vowels. The analysis revealed that in both groups, [iː] was higher and more fronted than its counterpart [eː] (both groups: F1−f0 Wilcoxon: $Z=-3.296$, $p < 0.01$; F2−F1 Wilcoxon: $Z=3.296$, $p < 0.01$).

For the low vowels [aː] and [a], the Bark values from Table 4 indicate that spectrally, these vowels were 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 front than its counterpart [aː] (monolingual Wilcoxon: $Z=-1.381$, $p < 0.05$; bilingual Wilcoxon: $Z=-1.852$, $p < 0.05$).

4.3.2. Duration analysis

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

Even though Table 5 shows that duration differences were appropriately implemented, a group comparison in terms of the duration ratio between long and short vowels revealed a significant difference (Table 6). The bilinguals' vowel duration ratio was on average smaller than the monolinguals', suggesting that the duration differences were more clearly realized in the monolingual group. The monolingual speakers produced short and long vowels with a duration ratio of 1:1.74.[7] The bilingual group had a ratio of 1:1.63, which was significantly smaller than the ratio of the monolingual children (Mann–Whitney $U=54$, $p < 0.05$).

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

4.4. Discussion

Appropriate tense–lax distinction and duration differences observed in the production data indicated that for most vowels under scrutiny, German-specific articulatory patterns were adequately acquired by the bilingual group. Significant production differences between the bilingual and the monolingual group were found for only [aː], [a] and [eː]. The position of these vowels in the vowel space (see Fig. 5) differed slightly for the two groups. Together with the fact that bilingual perception data showed context effects,

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

Item	Duration (in ms)	Item	Duration (in ms)	Z-value	p-value
Bilingual					
[eː] [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**
[iː] [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**
[aː] [a]					
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					
[eː] [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**
[iː] [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**
[aː] [a]					
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**

** Significant at the $p < 0.01$ level.

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

	ratio [eː]∼[ɛ]	ratio [iː]∼[ɪ]	ratio [aː]∼[a]
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 production data suggest that L2 learners might lack the flexibility found in native speakers to adjust for coarticulatory influences, and overall, their vowels tended to display greater variability across tokens than the monolinguals (see also Levy & Law, 2010). Since bilinguals have two phonetic systems to acquire and accommodate, it is hardly surprising to find more variability in this group. We note that the bilingual children were able to realize an articulatory difference which they could not clearly distinguish perceptually, at least in the oddball discrimination task we used. In particular, the [iː]∼[ɪ] and the [iː]∼[eː] differences, for which the bilinguals had significantly lower sensitivity, were adequately contrasted using spectral and duration characteristics. Nevertheless, there was a difference concerning the spectral realization of [eː], which was higher in the bilinguals (hence closer to the vowel with which it was paired in the difficult perceptual contrast: [iː]) compared to the monolinguals. It therefore seems that [eː] does pose a problem for the bilingual group, although presumably not to the extent that they would not be able to distinguish it appropriately from [ɛ] and [iː] in production.

5. General discussion

Given the range of contradictory findings in the literature about early sequential bilinguals, our foremost goal was to test

the perception and production of a set of carefully chosen vowels by participants with a narrow age range of L2 exposure. We reasoned that such a design would clarify which factors contribute to the conflicting data in the first place and thus offer a way to reconcile the disparate findings. This might allow for better generalization of findings, and add 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 and is difficult to modify despite early intensive exposure to L2 (e.g. Pallier et al., 1997; Højen & Flege, 2006).

In contrast to the perception results, the monolinguals and bilinguals were globally indistinguishable in production. Both groups implemented spectral and durational differences adequately to realize a contrast between vowels. Differences emerged only in analyses comparing the exact location of vowels, for example showing that [eː] was slightly closer to [iː] in the bilinguals than in the monolinguals. Even if the exact phonetic detail of bilinguals' productions exhibits a larger variability than the monolinguals', we see that the early bilingual children could produce a distinction in production that they could not necessarily discriminate successfully, especially for [iː]∼[ɪ] and [iː]∼[eː].

Observations of perceptual difficulties in oddball tasks, such as ours, may artificially underestimate the full extent of a bilingual's competence, since the poor performance on distinguishing [iː] from [eː] or from [ɪ] did not seem to impede the children's ability to correctly realize the vocalic contrasts in each of the words *Biest* [biːst], *Biss* [bɪs] or *Beet* [beːt]. The bilinguals have correctly encoded the difference between the vowels in the lexical representations they have for these words, and were able to retrieve them during the naming task. When faced with nonlexical or very rare items in the somewhat artificial task of clicking on the robot that spoke differently, it is possible that bilinguals were not able to apply to the task their knowledge of the [iː], [ɪ] and [eː] categories, which may be less stable than the categories of the monolinguals. Alternatively, it is possible that the bilinguals have encoded this contrast lexically first, without having generalized the difference to an abstract category definition, which would be necessary for performing well in the perception task. Bilinguals may still need to form abstract and stable categories from the set of words where the difference between [iː], [ɪ] and [eː] is encoded. Monolingual children, on the other hand, have succeeded in encoding a contrast lexically and moreover, displayed stable categories in the perception task. Evidence that phonetic differences can be encoded first in lexical representations without category definitions being completely stable in categorization tasks comes from late learner studies (e.g. Darcy et al., 2012; Hayes-Harb & Masuda, 2008; Weber & Cutler, 2004). The present results suggest that such dissociations may also be found in child L2 acquisition. The relationship between perception and production in L2 acquisition is complex and still insufficiently understood, particularly in the case of children. Our findings echo some previous studies that have examined the relationship between L2 speech production and perception in adults and in which perception accuracy lagged behind production accuracy (Bohn & Flege, 1997, for more experienced learners; Goto, 1971; Gottfried & Beddor, 1988; Sheldon & Strange, 1982; Tsukada et al., 2005, for children). However, other findings suggest that perception accuracy is (moderately) correlated with production accuracy, despite high inter-individual variability (Bradlow, Pisoni, Akahane-Yamada, & Tohkura, 1997; Flege, Bohn, & Jang, 1997; Levy & Law, 2010). As discussed elsewhere (e.g. Levy & Law, 2010; Mack, 1989; Tsukada et al., 2005), there are nevertheless several reasons for caution in concluding that production precedes perception.

The apparent dissociation between production and perception in our data 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 nonlexical syllables. The fact that lexical status and familiarity interacts with phonetic categorization has been repeatedly shown since Ganong (1980) who reported that an ambiguous sound is categorized differently if it forms a real word. Similar effects have also been observed in children (Walley & Flege, 1999). Therefore, the contrast between perception difficulties and production accuracy may be due to our design which confounds lexical status and processing modus (perception, mostly nonlexical vs. production, lexical). It is therefore difficult to conclusively claim that perception difficulties are not directly reflected in production, or to clearly tease apart the contribution of lexical knowledge vs. production/perception differences to the performance in our task.

In addition, the words selected for naming were mostly (near-) minimal pairs, which may have enhanced the possibility of correctly producing vowels in this task. As shown by Port and Crawford (1989), awareness about the presence of minimal pairs in a task can increase the clarity with which disambiguating cues are realized. In their study, German minimal pairs that become homophonous under the effect of final devoicing were not completely neutralized, and even less so in task manipulations where participants' awareness about the presence of these contrasting pairs was highest. In our experiment, even though we also used (near-)minimal pairs, no auditory model was given, thereby reducing the likelihood that the presence of minimal pairs was quickly noticed. Furthermore, during the task itself, the members of a (near-)minimal pair never followed each other directly in the list, due to the nature of the task. The ordering was randomized for each child. This is likely to have effectively prevented participants from producing an emphatic difference between words as they would if reading or dictating them in pairs (Warner, Jongman, Sereno, & Kemps, 2004). The written form of each word was presented once, together with the picture, in a training phase prior to the game. Children's attention was not drawn to it, and the written form was never part of the task itself. However, since the phonetic difference between our test vowels is reflected to some extent in the orthography, with consistently different graphemes for four out of six different sounds (⟨ie⟩ for [iː], ⟨i⟩ for [ɪ], ⟨ee⟩ for [eː], ⟨e⟩ for [ɛ]), we cannot exclude the possibility that this influenced their productions (see also Warner et al., 2004).

Additionally, our perception task may not be suited to reflect the full extent of perceptual abilities. The fact that bilingual children were less accurate than monolinguals in discriminating [iː] from [eː] or from [ɪ] in the perception task does not necessarily mean that they fully lack the ability to differentiate these speech sounds, or that they would suffer communication breakdowns in daily conversations. Testing bilingual children on a larger variety of items or with a different perceptual task may reveal that they are able to employ a different strategy from monolinguals to perceptually distinguish [iː] from [eː] and [ɪ]. In particular, the observations of context effects in perception leave open this possibility: for bilinguals, [iː]∼[ɪ] was more difficult to distinguish in the bilabial context, but performance in the velar context equaled that of monolinguals'.

Our results nevertheless reflect the possibility that bilingual and monolingual children's category definitions are not yet functionally equivalent. This state of affairs could stem from the fact that the input of the bilingual children was presumably overall less consistent than that of the monolinguals, given the fact that the bilinguals also interacted to some extent with late Turkish-German bilinguals whose German is more accented. The bilingual children also have to accommodate two phonological inventories. In production, their L2 vowels (at least the front mid- and high vowels) did not seem to be affected by a less consistent input, yet we do not know from these results whether other

aspects of their production might be affected. Our conclusions regarding production performance are only based on acoustic analysis, which is necessarily incomplete. We concluded that the contrasts examined were adequately acquired in production because all children produced significant duration and spectral differences. This might be complemented by asking German age-matched native speakers 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 (see Bosch & Ramon-Casas, 2011).

Even though the Turkish-German children had been first exposed to German at an early age and received ongoing and frequent input through their enrollment in a dual-language school, we observed a clear difference between them and the monolingual participants in perceptual behavior. Overall, the performance of the bilingual children in the perception task conformed to the predictions derived by cross-language assimilation patterns and perceptual similarity for Turkish monolingual adults (Oturan, 2002), and thus suggests the presence of L1–L2 interaction effects during early sequential L2 acquisition.

Our findings leave open the possibility that the social setting in which our participants were living is responsible for the observed differences between monolinguals and bilinguals in perception. The bilingual children were schooled in a dual-language school and interacted on a daily basis in both languages. They also had high pressure to maintain the L1 – similar to the participants in Pallier et al. (1997) or Bosch et al. (2000) – which might be related to the differences in performance between groups in the perception task. Assuming L1 use and its activation as an influential factor might explain why several studies did not find significant differences between early learners and native speakers. For instance, in Mack's (1989) study, participants were included only if they were English (L2) dominant, while Tsukada et al. (2005) used cued repetition to elicit utterances, reducing L1-activation in the experimental setting. Therefore, variables other than age alone might be found to be more influential in achieving native speaker resemblance. The reduced influence of the L1 on L2 categories which is often invoked to account for the advantage of bilingual children compared to bilingual adults (Baker et al., 2008, or Tsukada et al., 2005) thus might be a reflection of a reduced *use* of L1 in children, compared to adults. In order to test this possibility, it would be necessary to compare a pool of participants in different L1-use situations.

If L1 use turns out to be an important factor in predicting L2 phonological acquisition levels, it will be of equal importance to evaluate the L1 phonological competence. It may well be that sequential bilinguals have to choose between maintaining two functioning phonological systems and fully acquiring or retaining only one (see also Yeni-Komshian, Flege, & Liu, 2000). In the first scenario, incurring the mutual interaction of the two languages would be a necessary compromise. This possibility, evoked by studies like Ventureyra, Pallier, and Yoo (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). If this were the case, the best way to limit the influence of the L1 would be to stop using it (see Ventureyra et al., 2004). Consequently, immigrant situations where the pressure to acquire the L2 is considerable and maintaining the L1 seems less appealing might offer a better starting position for L2 segmental acquisition, which would be boosted by an early age of learning. The observations made in our experiments must be interpreted accordingly, in terms of the use of L1. If – as in the present study – it is high, it is unlikely that early learners' acquisition of L2 segments will be free of L1 influences, even when acquisition starts very early.

The results of our experiments further allow for clarification of previously conflicting results. Our study has controlled for a narrow range of the age of exposure: All participants were first exposed to the L2 between 2 and 4 years of age. In studies such as Pallier et al. (1997) or Bosch et al. (2000), similarly controlled age ranges also revealed differences in Catalan perceptual behavior between early Spanish-dominant (i.e. L2 learners) vs. Catalan-dominant (i.e. native speakers) participants. It has been argued that the contrast examined in Catalan ([e]∼[ɛ]) is undergoing a merger process, and is also likely in the range of allophonic variation in the Spanish L1 of the L2 learners (Bosch et al., 2000; Mora & Nadeu, 2012). This particularity of the contrast examined could be a reason for the finding of differences in performance between groups. Our findings elucidate the allophonic debate by contrasting two categories [iː] and [eː] that are not in allophonic variation with each other in Turkish. We observed a clear difference in behavior between the two groups, even with a very narrow range of first exposure and therefore propose that L1 influence may be detectable at an early age for a contrast that is expected to be difficult.

In light of these results, the question arises as to how children who are still in the process of acquiring both their native and their second language manage the interaction of both languages during later development. It is crucial to test the phonological representations of children who are acquiring an L2, to better understand the early stages of category development and the evolving relationship between L1 and L2 categories. As shown by Højen and Flege (2006), early learners' performance can closely resemble native speakers'. Hence, there is a need to establish whether and how this early L1 influence can be reduced or reversed as development progresses, taking into account the developing patterns of L1 and L2 use over time as well.

Based on our data, claims that L1–L2 interaction might be less strong if early learners are tested during childhood (*child* early learners) as opposed to adulthood (*adult* early learners) must be carefully evaluated considering L1–L2 use patterns. Indeed, this claim is usually based on the fact that L1 categories in children are not yet fully established. L1 categories presumably stabilize with increasing age and might yield interaction patterns in adults that were not apparent in children. However, it is also a logical possibility that strong L1–L2 interaction in adults can be attenuated by a specific pattern of L1–L2 use. Therefore, if interaction is detected as in our study, it might still decline with age, as L1 and L2 categories are both being further stabilized, and L1–L2 use might change over time.

References

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

Antoniadis, Z., & Strube, H. W. (1984). Untersuchungen zur spezifischen Dauer deutscher Vokale. *Phonetica, 41*(2), 72–87 [Duration analysis of German vowels].

Baker, W., & Trofimovich, P. (2005). Interaction of native- and second-language vowel system(s) in early and late bilinguals. *Language and Speech*, 48, 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, 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, 1–45.

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., & Tyler, M. D. (2007). Nonnative and second language speech perception. Commonalities and complementarities. In: O.-S. Bohn, & M. J. Munro (Eds.), *Language experience in second language speech learning: In honor of James Emil Flege* (pp. 13–34). Philadelphia: John Benjamins.

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. *Journal of the Acoustical Society of America*, 109, 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, 345–360.

Boersma, P., & Weenik, D. (2011). *Praat: Doing phonetics by computer, 5.2.35*. Amsterdam: University of Amsterdam.

Bohn, O.-S. (1995). Cross-language speech perception in adults: First language transfer doesn't tell it all. In: W. Strange (Ed.), *Speech perception and linguistic experience. Issues in cross-language research* (pp. 273–304). Baltimore: York Press.

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.

Bohn, O.-S., & Flege, J. E. (1997). Perception and production of a new vowel category by adult second language learners. In: A. James, & J. Leather (Eds.), *Second-language speech: Structure and process* (pp. 53–73). Berlin: Mouton de Gruyter.

Bosch, L., & Ramon-Casas, M. (2011). Variability in vowel production by bilingual speakers: Can input properties hinder the early stabilization of contrastive categories?. *Journal of Phonetics*, 39, 514–526.

Bosch, L., Costa, A., & Sebastián-Gallés, N. (2000). First and second language vowel perception in early bilinguals. *European Journal of Cognitive Psychology*, 12, 189–221.

Bradlow, A. R., Pisoni, D. B., Akahane-Yamada, R., & Tohkura, Y. (1997). Training Japanese listeners to identify English /r/ and /l/: IV. Some effects of perceptual learning on speech production. *Journal of the Acoustical Society of America*, 101, 2299–2310.

Cimilli, N., & Liebe-Harkort, K. (1979). *Sprachvergleich Türkisch–Deutsch* (2nd ed.). Düsseldorf: Pädagogischer Verlag Schwann [Turkish–German contrastive analysis].

Darcy, I., Dekydtspotter, L., Sprouse, R. A., Glover, J., Kaden, C., McGuire, M., et al. (2012). Direct mapping of acoustics to phonology: On the lexical encoding of front rounded vowels in L1 English–L2 French acquisition. *Second Language Research*, 28, 5–40.

Fabiano-Smith, L., & Goldstein, B. A. (2010). Phonological acquisition in bilingual Spanish–English speaking children. *Journal of Speech, Language and Hearing Research*, 53, 160–178.

Féry, C. (2004). *Phonologie des Deutschen: Eine Optimalitätstheoretische Einführung—Teil 1* (3rd ed.). Potsdam: Universitäts-Verlag [Phonology of German: An introduction with Optimality Theory, Part 1].

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., 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, 153–175.

Flege, J. E., Bohn, O.-S., & Jang, S. (1997). Effects of experience on non-native speakers' production and perception of English vowels. *Journal of Phonetics*, 25, 437–470.

Flege, J. E., & MacKay, I. R. A. (2004). Perceiving vowels in a second language. *Studies in Second Language Acquisition*, 26, 1–34.

Flege, J. E., & MacKay, I. R. A. (2010). "Age" effects on second language acquisition. In *Proceedings of the sixth 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., MacKay, I. R. A., & Meador, D. (1999). Native Italian speakers' perception and production of English vowels. *Journal of the Acoustical Society of America*, 106, 2973–2987.

Flege, J. E., Munro, M. J., & MacKay, I. R. A. (1995). Factors affecting strength of perceived foreign accent in a second language. *Journal of the Acoustical Society of America*, 97, 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.

Ganong, W. F., III (1980). Phonetic categorization in auditory word perception. *Journal of Experimental Psychology: Human Perception and Performance*, 6, 110–125.

Gildersleeve-Neumann, C. E., & Wright, K. L. (2010). English speech acquisition in 3- to 5-year-old children learning Russian and English. *Language, Speech & Hearing Services in Schools*, 41, 429–444.

Göksel, A., & Kerslake, C. (2005). *Turkish: A comprehensive grammar*. London: Routledge.

Goto, H. (1971). Auditory perception by normal Japanese adults of the sounds "l" and "r". *Neuropsychologia*, 9, 317–323.

Gottfried, T. L., & Beddor, P. S. (1988). Perception of temporal and spectral information in French vowels. *Language and Speech*, 31, 57–75.

Grosjean, F. (1989). Neurolinguists, Beware! The bilingual is not two monolinguals in one person. *Brain and Language*, 36, 3–15.

Grosjean, F. (2001). The bilingual's language modes. In: J. Nicol (Ed.), *One mind two languages—Bilingual language processing* (pp. 1–22). Walden, MA: Blackwell Publishers.

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, 207–226.

Hayes-Harb, R., & Masuda, K. (2008). Development of the ability to lexically encode novel second language phonemic contrasts. *Second Language Research*, 24, 5–33.

Hecht, B. F., & Mulford, R. (1982). The acquisition of a second language phonology: Interaction of transfer and developmental factors. *Applied Psycholinguistics*, 3, 313–328.

Højen, A., & Flege, J. E. (2006). Early learners' discrimination of second-language vowels. *Journal of the Acoustical Society of America*, 119, 3072–3084.

Khattab, G. (2007). Variation in vowel production by English–Arabic bilinguals. In: J. Cole, & J. I. Hualde (Eds.), *Laboratory phonology*, Vol. 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. *Journal of the Acoustical Society of America*, 100, 2425–2438.

Lee, S., Potamianos, A., & Narayanan, S. (1999). Acoustics of children's speech: Developmental changes of temporal and spectral parameters. *Journal of the Acoustical Society of America*, 105, 1455–1468.

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., & Law, F. F., II (2010). Production of French vowels by American-English learners of French: Language experience, consonantal context, and the perception–production relationship. *Journal of the Acoustical Society of America*, 128, 1290–1305.

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, 187–200.

Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory: A user's guide* (2 ed.). New Jersey: Lawrence Erlbaum Associates.

Max Planck Institute for Psycholinguistics (2001). *WebCelex*. Online at: ⟨http://celex.mpi.nl⟩ (consulted on 11.18.2011).

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, 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, 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, 55–67.

Menges, K. H. (1994). *Turkic languages and peoples: An introduction to Turkic studies* (2nd ed.). Wiesbaden: Harrassowitz.

Montrul, S. (2005). Second language acquisition and first language loss in adult early bilinguals: Exploring some differences and similarities. *Second Language Research*, 21, 199–249.

Mora, J. C., & Nadeu, M. (2012). L2 effects on the perception and production of a native vowel contrast in early bilinguals. *International Journal of Bilingualism*. http://dx.doi.org/10.1177/1367006911429518.

Mora, J. C., Keidel, J. L., & Flege, J. E. (2011). Why are the Catalan contrasts between /e/–/ɛ/ and /o/–/ɔ/ so difficult for even early Spanish–Catalan bilinguals to perceive?. In: M. Wrembel, M. Kul, & K. Dziubalska-Kołaczyk (Eds.), *Achievements and perspectives in SLA of speech: New Sounds 2010*, Vol. 2 (pp. 183–193). Frankfurt am Main: Peter Lang.

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*, 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. [Vowel perception and production in Turkish learners of German as a second language; Poster presented at the conference "Empirical methods in German as a Foreign and Second Language"].

Oturan, K. (2002). *Vokalperzeption und -produktion in der Zweitsprache: Eine Untersuchung mit türkischen Muttersprachlern.* Aachen: Shaker [Vowel perception and production in a second language: A study of Turkish native speakers].

Oyama, S. (1976). A sensitive period for the acquisition of a nonnative phonological system. *Journal of Psycholinguistic Research, 5*, 261–283.

Pallier, C., Bosch, L., & Sebastián-Gallés, N. (1997). A limit on behavioral plasticity in speech perception. *Cognition, 64*, B9–B17.

Pallier, C., Colomé, A., & Sebastián-Gallés, N. (2001). The influence of native-language phonology on lexical access: Exemplar-based versus abstract lexical entries. *Psychological Science, 12*, 445–449.

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.

Port, R., & Crawford, P. (1989). Incomplete neutralization and pragmatics in German. *Journal of Phonetics, 17*, 257–282.

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*, 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*, 240–255.

Selen, N. (1979). *Söyleyiş sesbilimi, akustik sesbilim ve Türkiye Türkçesi.* Ankara: Türk Dil Kurumu [Articulatory phonetics, acoustic phonetics and Turkey Turkish].

Sendlmeier, W. F. (1981). Der Einfluss von Qualität und Quantität auf die Perzeption Betonter Vokale des Deutschen. *Phonetica, 38*, 291–308 [The influence of quality and quantity on the perception of stressed vowels in German].

Sheldon, A., & Strange, W. (1982). The acquisition of /r/ and /l/ by Japanese learners of English: Evidence that speech production can precede speech perception. *Applied Psycholinguistics, 3*, 243–261.

Spiegel, M. R., & Stephens, L. J. (1998). *Schaum's outline of theory and problems of statistics* (3rd ed.). 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. *Journal of the Acoustical Society of America, 122*, 1111–1129.

Syrdal, A. K., & Gopal, H. S. (1986). A perceptual model of vowel recognition based on the auditory representation of American English vowels. *Journal of the Acoustical Society of America, 79*, 1086–1100.

Topbaş, S., & Yavaş, M. (2006). Phonological Acquisition and Disorders in Turkish. In: H. Zhu, & B. Dodd (Eds.), *Phonological development and disorders in children: A multilingual perspective* (pp. 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*, 263–290.

Valaczkai, L. (1998). *Atlas deutscher Sprachlaute: Instrumentalphonetische Untersuchung der Realisierung deutscher Phoneme als Sprachlaute.* Wien: Edition Praesens [Atlas of German speech sounds: Instrumental phonetic analysis of the realization of German phonemes as speech sounds].

Ventureyra, V. A. G., Pallier, C., & Yoo, H.-Y. (2004). The loss of first language phonetic perception in adopted Koreans. *Journal of Neurolinguistics, 17*, 79–91.

Walley, A. C., & Flege, J. E. (1999). Effect of lexical status on children's and adults' perception of native and non-native vowels. *Journal of Phonetics, 27*, 307–332.

Wängler, H. H. (1976). *Atlas Deutscher Sprachlaute (6 Ausg.).* Berlin: Akademischer Verlag [Atlas of German speech sounds].

Warner, N., Jongman, A., Sereno, J., & Kemps, R. R. (2004). Incomplete neutralization and other sub-phonemic durational differences in production and perception: Evidence from Dutch. *Journal of Phonetics, 32*, 251–276.

Weber, A., & Cutler, A. (2004). Lexical competition in non-native spoken-word recognition. *Journal of Memory and Language, 50*, 1–25.

Wenzel, V. (2000). "Ich sag allebei": Strategien beim frühen Erwerb einer verwandten Zweitsprache. *IRAL—International Review of Applied Linguistics in Language Teaching, 38*(3–4), 247–260 ["I say both": strategies during the early second language acquisition of a typologically related language].

Whitworth, N. (2000). Acquisition of VOT and vowel length by English–German bilinguals: A pilot study. In: D. Nelson, & P. Foulkes (Eds.), *Leeds working papers in linguistics and phonetics*, Vol. 8 (pp. 229–243). Leeds: University of Leeds.

Wiese, R. (1996). *The phonology of German.* Oxford: Clarendon.

Yeni-Komshian, G. H., Flege, J. E., & 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.