

## Cross-linguistic perceptual learning in advanced second language listeners

Katharina S. Schuhmann\*

**Abstract.** An issue of long-standing interest is whether phones, which correspond to each other phonologically but differ phonetically across two languages, are distinct in the minds of bilinguals and second language learners. Here, we ask whether listeners maintain separate representations for phonetically highly similar phonemes common to two languages even though this might seem inefficient. One way to address this question is to induce a change in the representation of phonemes in one language, and to test whether the other language is also affected. We carried out two perceptual learning studies, one with intermediate to advanced L1 English learners of German and one with intermediate to advanced L1 German learners of English, to probe the relationship between listeners' representations for the /f-s/ contrast in English and in German. We found cross-linguistic perceptual learning effects extending from English to German in both studies, though the effect sizes for perceptual learning in English and German differed. Based on these findings, we propose that phonetically highly similar phonemes common to listeners' L1 and L2 have separate yet interconnected mental representations for speech perception.

**Keywords.** speech perception; second language listeners; mental representations; highly similar phones across languages; cross-linguistic effects; perceptual learning

**1. Introduction.** A long-standing question in the field of bilingualism has been whether languages are stored separately in bilinguals and second-language (L2) learners. In the studies presented here, we ask how knowledge of sound is structured in intermediate to advanced L2 learners. Past and ongoing research on the question of phonological representations in bilinguals and L2 learners tends to focus on L2 phones that *differ* from L1 phones. Besides differences in the phoneme inventories between two languages, languages can also diverge in terms of how phonemes common to both languages are realized at the phonetic-acoustic level, such as the voiceless stops in English versus Spanish. Research on these dissimilar phonemes in two languages has focused on determining under what circumstances phonologically similar but phonetically different non-native and native phones are “equated” (Flege 1995, 2003’s *Speech Learning Model*) or “perceptually assimilated at the phonological level” (Best & Tyler 2007’s *Perceptual Assimilation Model-L2*).

Hitherto, research on bilingualism and L2 acquisition has not typically tackled the question of how bilinguals represent phonemes that are common to both languages and do *not* differ markedly in terms of their phonetic-acoustic realization in the two languages, such as /f/ or /s/ in German and in English. One might assume a priori that such phonetically *highly similar* phonemes, e.g. /f/ in English and /f/ in German, correspond to each other across languages and

---

\* I am very grateful to Marie K. Huffman for feedback on an earlier version of this paper and on my overall work on perceptual learning. I am also thankful to Arty Samuel, Ellen Broselow, and Rachel Hayes-Harb for helpful discussions and suggestions relating to my dissertation research on perceptual learning; some of the data reported here (L1 German–L2 English listeners) are also reported in Schuhmann (2014). Finally, a heartfelt thank you to Rachel Hayes-Harb and Laura C. Smith for going above and beyond in helping me with the logistics related to collecting data in Utah. This work has been supported in part by NSF grant #1127718, by an ETS TOEFL Small Grant, and the DAAD (a DAAD Short-Term Research Grant and a DAAD Travel Grant to the LSA 2016). Author: Katharina S. Schuhmann, The University of Bonn, Germany ([katharina.schuhmann@uni-bonn.de](mailto:katharina.schuhmann@uni-bonn.de)).

are equated (cf. Flege 2007:367). This assumption seems especially appropriate if even phonetically *dissimilar* phonemes, such as the voiceless stops in Spanish and in English referred to above, may be equated by bilinguals or L2 learners despite the “noticeable” (Hazan et al. 2009) differences in their phonetic-acoustic realizations.<sup>1</sup> Initial research into bilinguals’ cross-linguistic perception suggested that French–English (Caramazza et al. 1973) and Spanish–English (Williams 1977, 1979) bilinguals did not show significantly different responses to VOT continua in two different language contexts. However, using different methodologies to establish the respective language contexts, other researchers have been able to provide evidence for a “double-phonemic boundary” for the voiceless:voiced stop contrast in English–Spanish bilinguals (e.g., Garcia-Sierra et al. 2009). In effect, bilingual listeners – in particular highly proficient bilinguals – perceive VOT continua differently in different language settings, provided language contexts are well established in the experimental design (Garcia-Sierra et al. 2009). Thus, highly proficient bilinguals can be shown to have language-specific perceptual categories for phonetically dissimilar phoneme contrasts.

We now ask whether the representations for phonetically *highly similar* phonemes in two languages, such as /f/ and /s/ in English and in German, are separate or shared between languages in proficient non-native speakers. Although it is generally understood that the “same” phones are not necessarily identical at the phonetic-acoustic level across languages (Ladefoged 1990:70f, International Phonetic Association 1999:18), the /f/ and /s/ phonemes in English and German are phonetically *highly similar*, particularly in comparison to the VOT differences between stops in English and Spanish referred to above. Given these slight phonetic differences between phonologically corresponding phonemes in two languages, as well as language-specific differences in the precise phonological patterns and phonotactic rules the phonemes are involved in, the differentiation of *highly similar* phonemes in the minds of bilinguals and L2 learners does not seem implausible, but rather befitting. Indeed, it has been proposed that L2 phones start out as copies of corresponding L1 phones (e.g., Escudero 2009).

For the studies reported here, we set out to determine how the /f-s/ contrast in English and the /f-s/ contrast in German are represented in the minds of relatively proficient L2 listeners. Our approach is to first induce a change in the representation of the /f-s/ phoneme contrast in one language (English), and to then test whether the other language (German) is also affected. We hypothesized that if the representations of phonetically *highly similar* sounds are fully independent, change in one language would not extend to the other language. On the other hand, if the representations of the sounds are fully shared between the two languages, the change in one language should be identical to the change in the other language. Lastly, the change induced in one language might affect the other language, albeit not to the same degree. The latter outcome would provide evidence that the sounds are interconnected but separate, though not fully independent and autonomous.

Previous research has established the ability of perceptual learning effects to extend to another phoneme contrast within the same language. Perceptual learning effects on VOICE have been shown to generalize from the /t-d/ to the /p-b/ contrast in native English speakers (Kraljic & Samuel 2006). Moreover, perceptual learning effects on PLACE have been shown to generalize

---

<sup>1</sup> For illustration, the phoneme /k/ in English has a long-lag VOT (voice onset time), in which aspiration follows the release of the stop, while the phoneme /k/ in Spanish has a short-lag VOT (e.g., Lisker and Abramson 1964). The phoneme /g/ in English has a short-lag or lead VOT (pre-voicing), while /g/ in Spanish has only a lead VOT (e.g., Lisker and Abramson 1964; Flege and Eefting 1987).

from the voiceless fricative contrast /f-s/ to the voiced fricative contrast /v-z/, but not across manner to the voiceless stop contrast /p-t/ in native English speakers (Schuhmann 2014).

Non-native listeners have been shown to manifest perceptual learning effects in a non-native language (L1 German–L2 Dutch in Reinisch et al. 2013; L1 Dutch–L2 English in Cutler et al. 2015), but only for non-native listeners who are in an L2 immersion context at the time of the study. However, the studies presented in this paper involve non-native listeners in their native-language environment. Non-native English listeners have further been shown to generalize perceptual learning from their L2 English to their L1 Dutch (Reinisch et al. 2013). In that case, however, the listeners heard L1 (Dutch)–accented L2 (English) speech, which might have led to the observed cross-linguistic generalization effects (cf. Reinisch & Holt 2014). Moreover, Reinisch et al. 2013 did not directly test for perceptual learning in L2 English in addition to perceptual learning in L1 Dutch.

In the two studies reported here, we ask whether relatively proficient second-language learners in their L1 environment have a) shared, b) separate, or c) separate yet interconnected representations for phonetically *highly similar* phonemes in their L1 and their L2. We set out to test both L1 English–L2 German speakers and L1 German–L2 English speakers to be able to test for effects of directionality (L1 influencing L2 more than L2 influencing L1, for example). Most importantly, we attempt to induce a change in the representation of the /f-s/ contrast in English in both listener groups, and then test for changes in the representation of the /f-s/ contrast in English, as well as in the representation of the /f-s/ contrast in German, in order to be able to compare the nature and size of possible changes in both languages.

**2. Methodology.** In this work, we employed the perceptual learning paradigm (e.g., Norris et al. 2003, Eisner & McQueen 2005, Kraljic & Samuel 2006, Reinisch et al. 2013). In this approach, listeners first hear real words and non-words produced by a native English speaker in the context of an auditory lexical decision task (LDT). A subset of the existing English words heard in this lexical decision task – the critical stimuli – contain fricative sounds that are phonetically somewhat intermediate between the two contrasting phonemes /f/ and /s/. This exposure to atypical phonetic realizations of the /f/ or /s/ phoneme can lead to a change in the /f-s/ phoneme category boundary, as evidenced by a subsequent phoneme categorization task. The entire perceptual learning paradigm was computerized by means of the software SuperLab (Cedrus, Phoenix, Arizona). Listeners took part in the study individually; each participant completed the lexical decision task and the subsequent phoneme categorization task using the same equipment (a MacBookPro laptop and Sennheiser headphones).

**2.1. EXPOSURE PHASE.** All participants from both studies, L1 English–L2 German and L1 German–L2 English listeners, first took part in an exposure phase conducted entirely in English. Here, listeners participated in an auditory lexical decision task during which they were exposed to 100 English words and 100 English non-words (i.e., non-words that conform to the phonology and the phonotactics of the English language). The 100 English words contained 20 critical /s/-stimuli, e.g. <dinosour>, <recononcile>, or <embasso>, which each contained exactly one /s/ phoneme. Moreover, the 100 English words also contained 20 critical /f/-stimuli, e.g. <daffodil>, <microophone>, <qualolify>, which each contained exactly one /f/ phoneme. (A list of the frequency-matched critical stimuli is available in the Appendix.) The remainder of the word and non-word stimuli did not contain any other instances of /f/ or /s/ phonemes, nor any /v, z, p, t/ phonemes.

The primary goal of this phase of the study was to expose listeners to unusual pronunciations of one of the members of the critical /f-s/ phoneme contrast. To this end, listeners

were randomly assigned to one of two groups (conditions), in which either all of the critical /s/-stimuli or all of the critical /f/-stimuli contained fricatives with unusual pronunciations. In particular, participants in condition A were exposed to the 20 critical /s/-stimuli in which each [s] was replaced with a [ʔfs] mixture. In this condition, the 20 critical /f/-stimuli each contained an unmodified [f] sound. In condition B, on the other hand, participants were exposed to the 20 critical /f/-stimuli in which each [f] was replaced with an [ʔfs] mixture, while the 20 critical /s/-stimuli each contained an unmodified [s] sound. (See the Appendix for details about the creation of [ʔfs] mixtures.)

2.2. CATEGORIZATION PHASE. Subsequent to the exposure phase, all listeners participated in a categorization phase in which they identified phones on a fricative continuum as either /f/ or /s/, both in English and in German. All listeners heard randomized items from a 7-step /f-s/ continuum in German and, separately, randomized items from a 7-step continuum in English. (The [ʔfs] mixtures were created using the same basic processes as for the critical stimuli. See Appendix for details about the creation of the [ʔfs] mixtures.) These intermediate fricatives on each /f-s/ continuum occurred within the following non-word (frame): /ri\_i/. These non-word frames were the same for both languages at the phonological level, but differed crucially in the realization of the /r/ phoneme. While the English /ri\_i/ non-word contained an initial retroflex approximant [ɻ] phone, the German /ri\_i/ non-word contained an initial apical trill [r]. This non-word frame for the categorization task was chosen so that the /r/ realizations would provide a (minimal) language context to the listeners. Additionally, listeners were told which language they would be hearing; instructions were provided in English for all parts, and additionally in German for the German sections of the studies. The experimental block with the German continuum occurred first, followed by the block with the English continuum.

For each language, participants heard each of the seven fricative steps of the continuum (embedded within the non-word frame /ri\_i/) ten times, for a total of 70 items per continuum per language. The seven steps were randomized within each of the ten blocks for each participant by the stimulus presentation software (SuperLab). The phoneme categorization phase was a forced-choice task. The participants were asked to judge for each stimulus item, whether the sound between the vowels in the /ri\_i/ non-word frame was an /f/ or an /s/ sound. For each stimulus item, participants had up to 2.5 seconds to respond.<sup>2</sup>

2.3. PARTICIPANTS. Fifty-eight English learners of German (L1 English–L2 German listeners) participated in study 1 in the US (Utah), and 52 German learners of English (L1 German–L2 English listeners) participated in study 2 in Germany (Berlin). All participants were paid for taking part in the study. As a group, the participants in each study had a self-reported intermediate to advanced proficiency level in their non-native language. Table 1 below provides additional information about their average age of L2 acquisition, years of studying the L2, time spent in an L2 speaking country, current usage of the L2, as well as overall self-identified proficiency level in L2 (on a scale from 1 (*beginner*), 2 (*intermediate*), 3 (*advanced*) to 4 (*native speaker*)).<sup>3</sup>

---

<sup>2</sup> This paper reports parts of a larger study, in which listeners completed the categorization task also for the /v-z/ phoneme contrast in both German and English.

<sup>3</sup> Note that in part due to differences in when and how foreign languages are introduced in the educational context in Germany vs. in the US, the participants are not perfectly matched for their L2 language background.

	L1 English-L2 German listeners	L1 German-L2 English listeners
Age of L2 acquisition	14.7 (4.0)	8.6 (2.4)
Years of studying L2	4.7 (2.2)	11.4 (2.2)
Time spent in L2 country (months)	16.2 (16.7)	8.1 (13.2)
Current usage of L2 (hours/week)	8.8 (10.7)	18.7 (12.8)
Proficiency ( <i>self-identified</i> )	2.55/4 (0.6)	2.95/4 (0.4)

Table 1. Participants' language background (*mean* and (*SD*))

**3. Results.** We first present the results of the exposure phase, followed by the results of the categorization phase. The analysis of the exposure phase is mainly used to check whether native and non-native listeners performed well during the auditory lexical decision task (LDT), and whether the crucial modified stimuli were accepted to a high degree by the native and non-native listeners.

3.1. RESULTS: EXPOSURE PHASE. The participants in the L1 English–L2 German study performed very well overall in the lexical decision task, with a mean overall accuracy rate of 95.04% (SD 3.78%), ranging from 83.0% to 100.0%. Table 2 below illustrates the mean accuracy for all words (critical stimuli and filler words): 95.49% (SD 3.09%); individual scores ranged from 85.0% to 100.0%. This table further shows that participants were also faster to correctly accept words (277.4, SD 123.9) than to correctly reject non-words (411.6, SD 136.5). The analysis excludes three L1 English–L2 German participants who accepted fewer than 14 (70%) of the twenty ambiguous critical stimuli during the lexical decision task.

	Words	Non-words
% correct* (SD)	95.49 (3.09)	94.58 (4.48)
RT (in ms)** (SD)	277.4 (123.9)	411.6 (136.5)

\* Mean accuracy (*i.e.*, "yes" responses to critical items)

\*\* Mean RT for correct items (*i.e.*, "yes" responses to critical items) measured from word offset

Table 2. L1 English listeners: Results for all stimuli in the LDT

Table 3 illustrates the L1 English–L2 German listeners' response time (RT) and acceptance rates of the critical stimuli during the lexical decision task. Overall, the RT and acceptance data suggest that the critical stimuli sounded acceptable to the native listeners, although they accepted slightly fewer of the ambiguous critical stimuli and responded slightly more slowly to the ambiguous critical stimuli compared to the natural critical stimuli. Participants accepted fewer (12.29%) of the ambiguous [ʔf] compared to natural [f] items (82.88% compared to 95.17%, respectively), and responded about 40.58 ms more slowly to ambiguous [ʔf] items. Participants accepted about the same number of ambiguous [ʔs] items compared to natural [s] items (98.62% compared to 98.65%, respectively), but responded slightly more slowly to ambiguous [ʔs] items (45.40 ms more slowly). (Note again, however, that each participant heard either only ambiguous [ʔf] and natural [s] words, *or* ambiguous [ʔs] words and natural [f] words.)

	Natural Stimuli		Ambiguous Stimuli	
	[f]	[s]	[?f]	[?s]
% correct* (SD)	95.17% (4.12)	98.65% (2.26)	82.88% (8.96)	98.62% (2.27)
RT (in ms)** (SD)	267.0 (144.9)	242.1 (119.1)	307.6 (142.6)	287.5 (147.9)

\* Mean accuracy (i.e., "yes" responses to critical items)

\*\* Mean RT for correct items (i.e., "yes" responses to critical items) measured from word offset

Table 3. L1 English listeners: Results for critical stimuli in the LDT

In the second study, the L1 German–L2 English participants also performed relatively well overall in the lexical decision task, which they completed in their L2 English. They achieved a mean overall accuracy rate of 85.35% (SD 5.39%), ranging from 71.00% to 94.50%. Table 4 below illustrates the mean accuracy for all words (critical stimuli and filler words): 84.6% (SD 7.1%); individual scores ranged from 63.00% to 98.00%. This table further shows that participants were also faster to correctly accept words (355.7 ms, SD 124.0) than to correctly reject non-words (541.3 ms, SD 210.2). The analysis excludes one L1 German-L2 English participant who accepted fewer than 10 (50%) of the ambiguous critical stimuli during the lexical decision task.

	Words	Non-words
% correct* (SD)	84.6 (7.1)	86.1 (8.3)
RT (in ms)** (SD)	355.7 (124.0)	541.3 (210.2)

\* Mean accuracy (i.e., "yes" responses to critical items)

\*\* Mean RT for correct items (i.e., "yes" responses to critical items) measured from word offset

Table 4. L1 German listeners: Results for all stimuli in the LDT

Table 5 illustrates the L1 German–L2 English listeners' response time and acceptance rates of the critical stimuli during the lexical decision task. Overall, the RT and acceptance data suggest that the critical stimuli also sounded acceptable to the non-native listeners, although non-native listeners accepted slightly fewer of the ambiguous critical stimuli and responded slightly more slowly to the ambiguous critical stimuli in comparison to the natural critical stimuli. Participants accepted fewer (12.86%) of the ambiguous [?f] compared to natural [f] items (75.6% compared to 88.46%, respectively), and responded about 76.9 ms more slowly to ambiguous [?f] items. Participants also accepted slightly fewer (2.28%) of the ambiguous [?s] items compared to natural [s] items (82.12% compared to 84.40%, respectively), and responded slightly more slowly to ambiguous [?s] items (38.9 ms more slowly). (Note again, however, that each participant heard either only ambiguous [?f] and natural [s] words, *or* ambiguous [?s] words and natural [f] words.)

	Natural Stimuli		Ambiguous Stimuli	
	[f]	[s]	[?f]	[?s]
% correct*	88.46%	84.40%	75.60%	82.12%
(SD)	(10.08)	(8.94)	(11.30)	(11.59)
RT (in ms)**	347.8	336.7	424.7	375.5
(SD)	(151.9)	(141.0)	(160.6)	(129.3)

\* Mean accuracy (i.e., "yes" responses to critical items)

\*\* Mean RT for correct items (i.e., "yes" responses to critical items) measured from word offset

Table 5. L1 German listeners: Results for critical stimuli in the LDT

3.2. RESULTS: CATEGORIZATION PHASE. The responses to the phoneme categorization tasks were analyzed to see whether exposure to atypical phonetic realizations of the /f/ or /s/ phoneme in the exposure phase (conducted entirely in English), would lead to changes in the /f-s/ phoneme category boundary. The analysis is based on the average percent labial (%f) responses calculated for each continuum for each participant.

Starting with native English speakers, the following figure shows that atypical [?f] sounds and atypical [?s] sounds in the exposure phase change how the perceptual system interprets cues to the /f-s/ place features in English. Participants in condition B, who were exposed to atypical [?f] sounds and unmodified [s]-stimuli during the auditory lexical decision task, categorized more of the phones on the /f-s/ continuum in English as /f/ than participants in condition A, who were exposed to atypical [?s] sounds and unmodified [f]-stimuli ( $F(1,53)=9.807, p=.003$ ). The effect size – the difference in /f/ responses to the English /f-s/ continuum between the two participant groups – is 7.4%. This perceptual learning effect is illustrated in Figure 1 below.

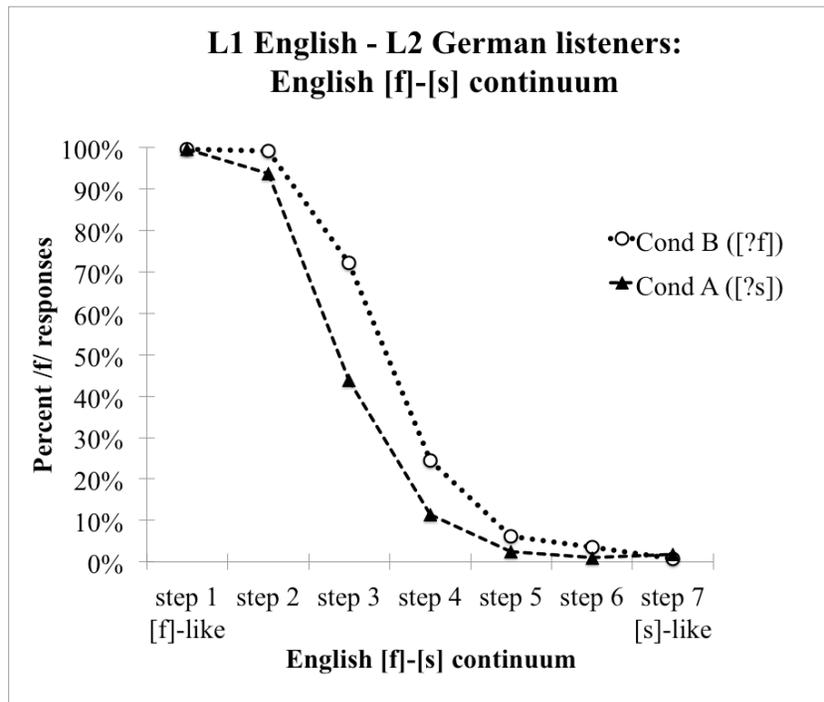


Figure 1. L1 English–L2 German listeners: English [f]-[s] continuum

The same native English speakers also categorized phones on an /f-s/ continuum in German, their second language. Note that the listeners did not hear any German in the exposure phase; i.e., they were not exposed to atypical fricatives in German in the lexical decision task. As Figure 2 depicts, atypical sounds in their native language English change how the perceptual system interprets cues to the /f-s/ place features in their L2 German. Participants in condition B, who were exposed to atypical [ʔf] sounds and unmodified [s]-stimuli in English during the lexical decision task, categorized more of the phones on the /f-s/ continuum in German as /f/ than participants in condition A, who were exposed to atypical [ʔs] sounds and unmodified [f]-stimuli in English ( $F(1,53)=4.226, p=.045$ ). The effect size – the difference in /f/ responses to the German /f-s/ continuum between the two participant groups – is 6.1%.

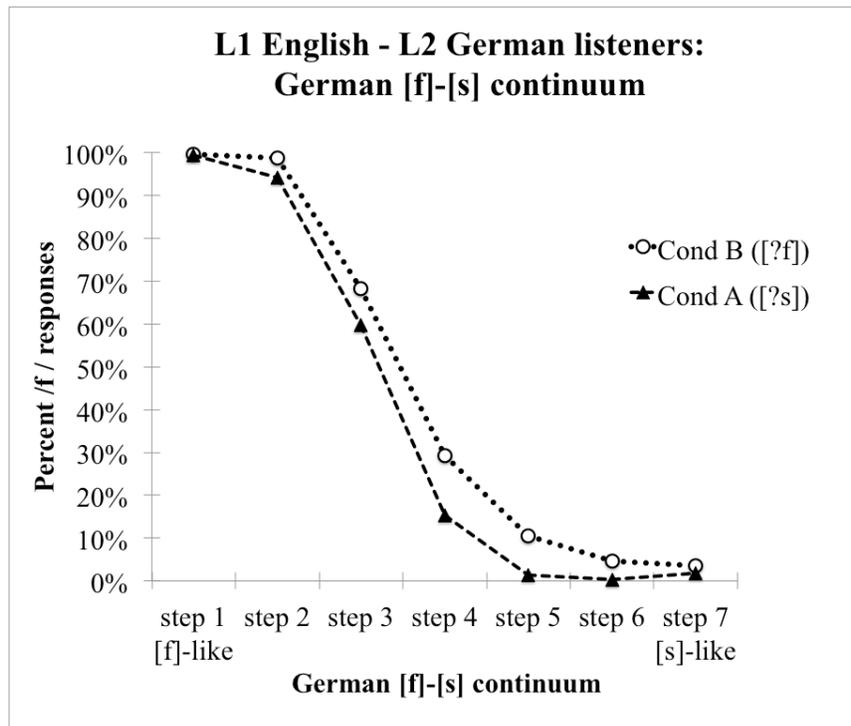


Figure 2. L1 English–L2 German listeners: German [f]-[s] continuum

Next, we present results for the second study, which was conducted with non-native English speakers. As Figure 3 shows, for German learners of English, atypical [ʔf] sounds and atypical [ʔs] sounds in the exposure phase change how the perceptual system interprets cues to the /f-s/ place features in English. Participants in condition B, who were exposed to atypical [ʔf] sounds and unmodified [s]-stimuli in English during the auditory lexical decision task, categorized more of the phones on the /f-s/ continuum in English as /f/ than participants in condition A, who were exposed to atypical [ʔs] sounds and unmodified [f]-stimuli ( $F(1,49)=9.472, p=.003$ ). The effect size – the difference in /f/ responses to the English /f-s/ continuum between the two native German participant groups – is 9.7%.

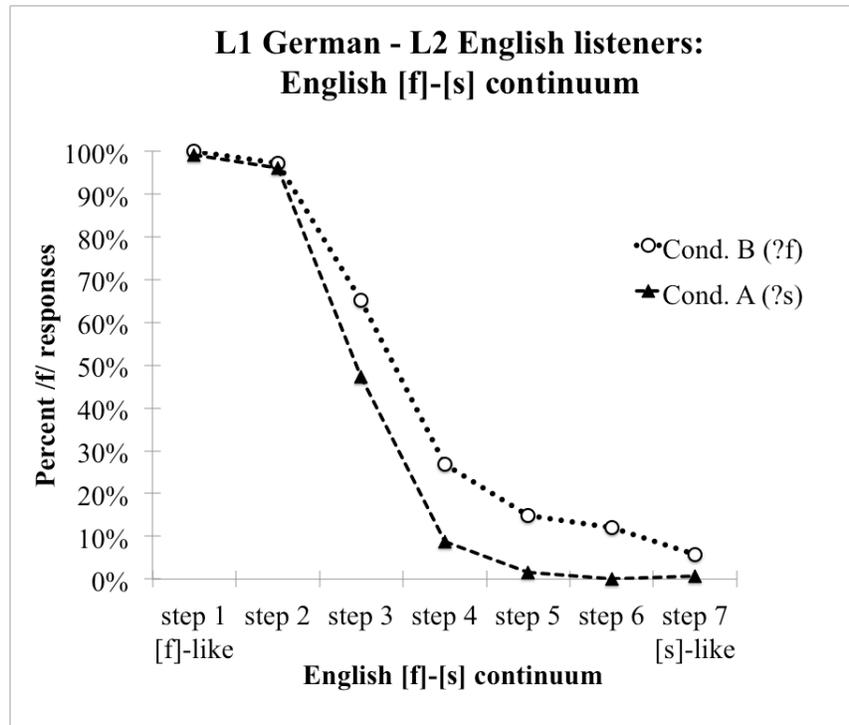


Figure 3. L1 German–L2 English listeners: English [f]-[s] continuum

The same German learners of English also categorized phones on an /f-s/ continuum in German, their first language. Note that the listeners did not hear any German in the exposure phase; i.e., they were not exposed to atypical fricatives in German in the lexical decision task. As Figure 4 depicts, atypical sounds in their non-native language English also change how the perceptual system interprets cues to the /f-s/ place features in German. Participants in condition B, who were exposed to atypical [?f] sounds and unmodified [s]-stimuli in English during the lexical decision task, categorized more of the phones on the /f-s/ continuum in German as /f/ than participants in condition A, who were exposed to atypical [?s] sounds and unmodified [f]-stimuli in English ( $F(1,49)=4.687, p=.035$ ). The effect size – the difference in /f/ responses to the German /f-s/ continuum between the two participant groups – is 6.4%.

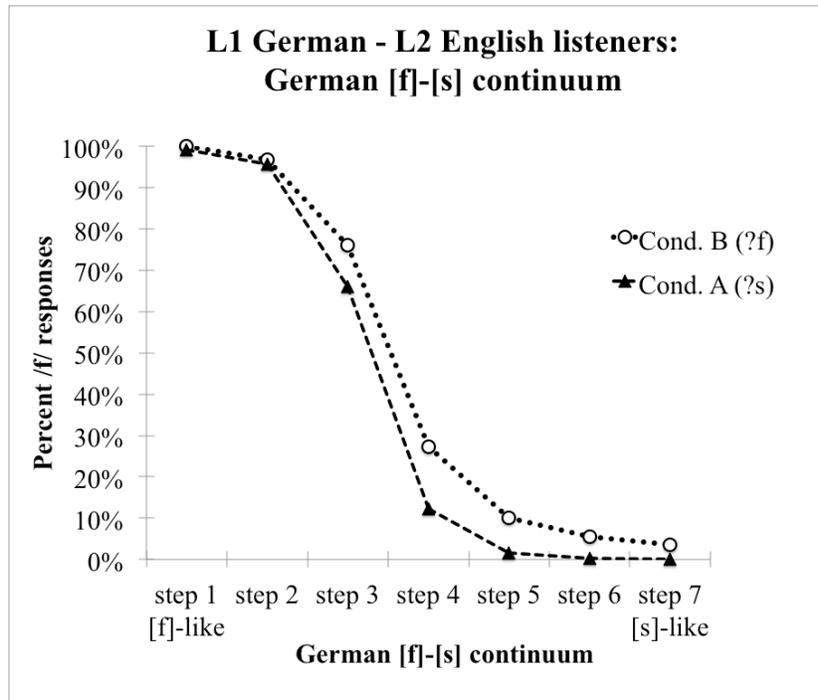


Figure 4. L1 German–L2 English listeners: German [f]-[s] continuum

To summarize, both native and non-native English listeners show shifts in their perceptual category between English /f/ and /s/, having been exposed to atypical sounds in English. Moreover, both groups of listeners also show shifts in their perceptual category between German /f/ and /s/, although they were not exposed to atypical sounds in German.

**4. Discussion and Conclusion.** We set out to determine whether phonetically similar sounds in two languages are represented separately in the minds of bilinguals and advanced second language learners. In this work, we use the perceptual learning paradigm to test whether changes in the representation of sound contrasts in one language (English) also affect listeners’ other language (German). Our two studies with English–speaking learners of German and German–speaking learners of English revealed perceptual learning effects in English as well as in German in both groups of participants, even though participants were exposed to atypical pronunciations only in English.

The results of our two studies show cross-linguistic perceptual learning effects in both groups of participants. Native English listeners show the perceptual learning effect on the English /f-s/ contrast, which was expected based on growing evidence that adult native listeners continue to fine-tune their perceptual representations, thus allowing for perceptual learning effects to take place (e.g., Norris et al. 2003, Eisner & McQueen 2005, Kraljic & Samuel 2006, *inter alia*). At the same time, native English listeners also adjust the way they perceive the /f-s/ contrast in their L2 German. Together with Schuhmann (2014), this appears to be the first time L1-to-L2 generalizations have been shown for cross-linguistic perceptual learning effects.

Similarly, native German listeners show perceptual learning effects in their L2 English after exposure to unusual pronunciations in English. This provides evidence that relatively advanced L2 listeners can fine-tune their speech perception in a non-native language, even when

they live in an L1 context where their native language is clearly dominant.<sup>4</sup> This suggests that advanced non-native listeners are also sensitive to non-canonical, atypical pronunciations, which in this case change how the perceptual system interprets cues to the /f-s/ place features. In the results reported here, the malleability of the /f-s/ contrast in English is even larger in non-native listeners than native listeners (9.7 % vs. 7.4%, respectively). Most importantly, the non-native English listeners in our study generalize changes in how the perceptual system interprets cues to the /f-s/ place features from English to their native language German.

All listeners heard a native English speaker during the auditory lexical decision task, which was used to expose the participants to atypical [ʔfs] phones within the critical stimulus items. Thus, the cross-linguistic generalization of perceptual learning effects from English to German was not the result of a German accent in the English stimuli during the exposure phase (unlike the L1 Dutch–L2 English situation in Reinisch et al. 2013).<sup>5</sup>

Taken together, the bidirectional cross-linguistic perceptual learning effects in our studies indicate that the perceptual representations of the phoneme contrast /f-s/ in English and in German cannot be fully separate or independent in our intermediate to advanced non-native listeners: exposure to atypical sounds in English also affects the untrained German /f-s/ phoneme boundary, both for English-speaking learners of German and for German-speaking learners of English. At the same time, the different perceptual learning effect sizes in German and English suggest that the representations of the /f-s/ phoneme contrast in English and the representation of the /f-s/ phoneme contrast in German are not fully identical. The effect sizes and the exact nature of the perceptual categorization of the stimulus continua are different in the two languages, rendering the notion of shared representations for the /f-s/ contrast between German and English unlikely. A summary of the perceptual learning effect sizes in the two groups of participants is provided in Table 6 below.

	English /f-s/	German /f-s/
L1 English-L2 German	Effect size: 7.4%	Effect size: 6.1% =82.4% of English effect size
L1 German-L2 English	Effect size: 9.7%	Effect size: 6.4% =66.0% of English effect size

Table 6. Summary of perceptual learning effect sizes

Based on these data, we propose that the relationship between phonetically highly similar sounds in listeners' L1 and L2 likely consists of separate yet interconnected mental representations for speech perception. Under this account, phoneme contrast representations between two languages are not statically either separate or merged, but interrelated and dynamic in nature. The strengths of these interconnections between phonemes, in turn, depend on the language background and language use of the bilingual listener: The connection between the

<sup>4</sup> Future research is necessary to determine whether perceptual learning in non-native listeners also takes place for a L2 phoneme contrast that does not exist in L1.

<sup>5</sup> The speaker who recorded all our stimuli was also fluent in German and recorded the German non-word stimulus items for the categorization phase. At the same time, the speaker was clearly an L1-dominant speaker and currently resided in the US. Moreover, any potential cross-linguistic influences from his German onto his English are unlikely to have led to realizations of the /f-s/ contrast in English that might be atypical or non-canonical for English, though more typical for German.

representations of phonetically highly similar L1 and L2 sounds is ‘stronger’, i.e., allowing for more ‘transfer’ or generalization effects, in the direction of L1 to L2, rather than in the opposite direction, from L2 to L1. This is in line with both anecdotal evidence and an abundance of L1-to-L2 transfer effects, but generally fewer or smaller L2-to-L1 transfer effects. Similarly, it captures the fact that the generalized perceptual learning effect from English to German is larger in the L1 English–L2 German group than in the L1 German–L2 English group, namely 82.4% vs. 66.0%, respectively (cf. Table 2 above). Finally, the strength of the interconnections seems to be dynamically modulated by language use, at least in novel L2 learners (cf. Schuhmann 2014).

The findings presented here suggest that phonologically equivalent phonemes that are phonetically highly similar do not necessarily have one shared or merged perceptual representation in the minds of intermediate to advanced L2 listeners, although this might seem uneconomical. Additional studies are warranted to test whether similar evidence from other phonetically highly similar phonemes across two languages will corroborate this conclusion. Further research could also help determine to which specific phonetic-acoustic cues native and non-native listeners attend when discriminating phoneme contrasts such as /f-s/ in English and in German. Finally, additional research with less proficient L2 listeners, novice L2 listeners, and near-native L2 listeners would allow us to establish how the differentiation and the interaction of the representations unfolds throughout L2 development.

## References

- Best, Catherine T. & Michael D. Tyler. 2007. Nonnative and second language speech perception. Commonalities and complementarities. In M. J. Munro & Ocke-Schwen Bohn (eds.), *Second language speech learning: The role of language experience in speech perception and production*. 13-34. Amsterdam: John Benjamins.
- Boersma, Paul & David Weenink. 2011. Praat: Doing Phonetics by Computer. [Computer program]. Retrieved <http://www.praat.org/>.
- Caramazza, Alfonso, Grace H. Yeni-Komshian, Edgar B. Zurif & Ettore Carbone. 1973. The acquisition of a new phonological contrast: The case of stop consonants in French-English bilinguals. *The Journal of the Acoustical Society of America* 54. 421-428. <http://dx.doi.org/10.1121/1.1913594>.
- Cutler, Anne, Laurence Bruggeman & Mark Antoniou. 2015. Variation in perceptual flexibility across languages within listeners. Abstract for *AMLaP (Architectures and Mechanisms for Language Processing)* 2015. <https://drive.google.com/file/d/0B0a477xEnvzXVWEyNkNMVUtdEk/view> (29 May, 2016.)
- Eisner, Frank & James M. McQueen. 2005. The specificity of perceptual learning in speech processing. *Perception & Psychophysics* 67. 224-38. <https://doi.org/10.3758/BF03206487>.
- Escudero, Paola. 2009. Linguistic perception of "similar" L2 sounds. In Paul Boersma & Silke Hamann (eds.), *Phonology in perception*. 151-190. Berlin: Mouton de Gruyter.
- Flege, James E. 1995. Second language speech learning. Theory, findings, and problems. In Winfred Strange (ed.), *Speech perception and linguistic experience: Issues in cross-language research*. 233-277. Timonium, MD: York Press.
- Flege, James E. 2003. Assessing constraints on second-language segmental production and perception. In A. Meyer & N. Schiller (eds.), *Phonetics and phonology in language comprehension and production. Differences and similarities*. 319-355. Berlin: Mouton de Gruyter.

- Flege, James E. 2007. Language contact in bilingualism: Phonetic system interactions. In J. Cole & J. I. Hualde (eds.), *Laboratory phonology*. 353-382. Berlin: Mouton de Gruyter.
- Flege, James E. & Wieke Eefting. 1987. Cross-language switching in stop consonant perception and production by Dutch speakers of English. *Speech Communication* 6. 185-202. [http://dx.doi.org/10.1016/0167-6393\(87\)90025-2](http://dx.doi.org/10.1016/0167-6393(87)90025-2).
- Garcia-Sierra, Adrian, Randy L. Diehl & Craig Champlin. 2009. Testing the double phonemic boundary in bilinguals. *Speech Communication* 51(4). 369-378. <http://dx.doi.org/10.1016/j.specom.2008.11.005>.
- Hazan, Valerie, Souhila Messaoud-Golusi, Stuart Rosen, Suzan Nouwens & Bethanie Shakespeare. 2009. Speech perception abilities of adults with dyslexia: Is there any evidence for a true deficit? *Journal of Speech, Language, and Hearing Research* 52(6). 1510–1529. [http://dx.doi.org/10.1044/1092-4388\(2009/08-0220\)](http://dx.doi.org/10.1044/1092-4388(2009/08-0220)).
- International Phonetic Association. 1999. *Handbook of the International Phonetic Association: A guide to the use of the International Phonetic Alphabet*. Cambridge University Press.
- Kraljic, Tanya & Arthur G. Samuel. 2006. Generalization in perceptual learning for speech. *Psychonomic Bulletin & Review* 13. 262-268. <http://dx.doi.org/10.3758/BF03193841>.
- Ladefoged, Peter. 1990. Some reflections on the IPA. *UCLA Working Papers in Phonetics* 74. 61-76. <http://escholarship.org/uc/item/71s1701m>. (29 May, 2016.)
- Lisker, Leigh & Arthur Abramson. 1964. A cross-language study of voicing in initial stops: Acoustical measurements. *Word* 20. 384–442. <http://dx.doi.org/10.1080/00437956.1964.11659830>.
- Mitterer, Holger. 2011 (accessed). Interpolation Script. [Computer program]. Retrieved from [http://www.holgermitterer.eu/HM/interpolate between voiced sounds.praat](http://www.holgermitterer.eu/HM/interpolate%20between%20voiced%20sounds.praat)
- Norris, Dennis, James M. McQueen & Anne Cutler. 2003. Perceptual learning in speech. *Cognitive Psychology* 47. 204-38. [http://dx.doi.org/10.1016/S0010-0285\(03\)00006-9](http://dx.doi.org/10.1016/S0010-0285(03)00006-9).
- Reinisch, Eva & Lori L. Holt. 2014. Lexically guided phonetic retuning of foreign-accented speech and its generalization. *Journal of Experimental Psychology: Human Perception and Performance* 40. 539-55. <http://dx.doi.org/10.1037/a0034409>.
- Reinisch, Eva, Andrea Weber & Holger Mitterer. 2013. Listeners retune phoneme categories across languages. *Journal of Experimental Psychology: Human Perception and Performance* 39. 75-86. <http://dx.doi.org/10.1037/a0027979>.
- Schuhmann, Katharina S. 2014. *Perceptual learning in second language learners*. Stony Brook, NY: State University of New York dissertation.
- Williams, Lee. 1977. The perception of consonant voicing by Spanish–English bilinguals. *Perception & Psychophysics* 21. 289-297. <http://dx.doi.org/10.3758/BF03199477>.
- Williams, Lee. 1979. The modification of speech perception and production in second-language learning. *Perception & Psychophysics* 26. 95-104. <http://dx.doi.org/10.3758/BF03208301>.
- Zeno, Susan M., Stephen H. Ivens, Robert T. Millard & Raj Duvvury. 1995. *The educator's word frequency guide*. Brewster, NY: Touchstone Applied Science Associates.

## Appendices

### Appendix 1: List of critical stimuli (Frequency based on Zeno et al. 1995)

	/s/-stimuli	syllables	frequency	/f/-stimuli	syllables	frequency
1	accuracy	4	50.9	amphibian	4	38.8

2	aerosol	3	13.2	beneficial	4	47.3
3	Arkansas	3	47.1	cacophony	4	25.1
4	chromosome	3	47.2	calligraphy	4	29
5	coliseum	4	32.4	chlorophyll	3	45.6
6	condensation	4	44.6	clarification	5	41.7
7	condescend	3	29.3	daffodil	3	36.3
8	connoisseur	3	32.4	endorphin	3	31.3
9	cul-de-sac	3	32.3	gleeful	2	36.4
10	delicacy	4	42.6	glorification	5	27.4
11	democracy	4	53.1	manufacture	4	51.7
12	dinosaur	3	49.9	meaningful	3	49.9
13	embassy	3	41.9	microphone	3	46.8
14	eraser	3	42.9	modification	5	45.2
15	indecision	4	39	Newfoundland	3	46.6
16	Johnson	2	54	orthography	4	22.1
17	legacy	3	40.7	perform	2	56.2
18	medicine	3	57.1	qualification	5	39.8
19	reconcile	3	40.9	qualify	3	46.3
20	rehearsal	3	43.2	unofficial	4	37.9
	<i>Average</i>	<i>3.25</i>	<i>41.74</i>	<i>Average</i>	<i>3.65</i>	<i>40.07</i>

## Appendix 2: Creation of [ʔfs] mixtures (simplified)

All stimuli for the Lexical Decision Task (LDT) and the fricative continua were recorded in a sound-treated room, using a Marantz Portable Solid State Recorder, a Shure microphone, a bit depth of 16 bit, and a sampling rate of 44.1 kHz. For the critical stimuli, the speaker produced each of the 40 critical words in a natural way with the correct /f/ or /s/ phoneme, respectively. In addition, the speaker also produced a second version for each critical stimulus in which the crucial /f/ fricative had been replaced with an /s/ phoneme, and vice versa. In effect, for critical /s/-stimuli like *legacy*, the speaker produced both one version containing [s], [lɛgəsi], and one version containing [f], [lɛgəfi]. For critical /f/-stimuli like *microphone*, the speaker produced both one version containing [f], [maɪkɹəfoʊn], and one version containing [s], [maɪkɹəsoʊn].

For the LDT, the critical stimuli with the ambiguous fricatives were created by segmenting and mixing the [f] and [s] fricatives obtained from the two types of recording for each stimulus word. For example, for the /f/-stimulus *microphone*, the [f] segment from [maɪkɹəfoʊn], and the [s] segment from [maɪkɹəsoʊn] were segmented using the software Praat (Boersma & Weenink 2011). For each stimulus word, these two [f] and [s] segments were then mixed in Praat by means of interpolation, using a (modified) script by Mitterer (Mitterer 2011), which “uses PSOLA to equate duration and pitch contour, and then interpolates between the manipulated sounds” (Mitterer 2011, introduction to script).

The script was used to create 21 [f/s] fricative mixtures for each of the 40 stimulus words. The mixtures were then spliced into the original recordings with the labial fricative pronunciation (the “[f]-frame”) for both the /f/-stimuli and the /s/-stimuli; e.g., [ʔfs] fricative mixtures were inserted into the [f]-frame [maɪkɹəfoʊn] for /f/-stimuli like *microphone*, and into the [f]-frame [lɛgəfi] for /s/-stimuli like *legacy*. The mixture was inserted in place of the original labial fricative. In order to make the /f/-stimuli sound more [f]-like, the intensity of the mixed fricatives was lowered for 17 of the 20 /f/-stimuli. The author and one to two native English speakers judged, for each stimulus item, which of the 21 mixed fricatives spliced into the [f]-frame sounded most ambiguous while also still sounding natural rather than machine-generated. For each stimulus item, the mixture that met the criteria of sounding most ambiguous yet natural was chosen.

The fricative continua for the categorization phase were constructed in a similar manner. Each continuum was created based on the recordings of two German or two English non-words, for example German [rifi] and German [risi]: The [f] and [s] fricatives were segmented from the disyllabic non-word endpoints and mixed with the same Mitterer script as above for each continuum. The script created 21 fricative mixtures, which were spliced one at a time into the [f]-frame of the disyllabic non-word, replacing the [f] segment.

The same seven steps between the two endpoints served as the seven steps of the continuum for all continua used during the categorization phase. This means that the same number of mixture steps (three steps) was skipped between the chosen steps in both continua. In other words, the chosen mixtures were always three steps apart. The diagram in (1) below illustrates which of the 21 mixtures were chosen to form an evenly spaced seven-step continuum. (Note that the numerous mixtures generated by the script are referred to as “mix#”, whereas the seven steps chosen for the continua are referred to as “step#”.)

(1) The seven mixtures chosen as steps (inserted into disyllables) for the continua

