Bilingualism: Language and Cognition

cambridge.org/bil

No L1 privilege in talker adaptation

Laurence Bruggeman[1] and Anne Cutler[2]

[1]The MARCS Institute for Brain, Behaviour and Development, Western Sydney University, ARC Centre of Excellence for the Dynamics of Language and [2]The MARCS Institute for Brain, Behaviour and Development, Western Sydney University, ARC Centre of Excellence for the Dynamics of Language, Max Planck Institute for Psycholinguistics

Research Article

Cite this article: Bruggeman L, Cutler A (2020). No L1 privilege in talker adaptation. *Bilingualism: Language and Cognition* **23**, 681–693. https://doi.org/10.1017/S1366728919000646

Received: 7 January 2019
Revised: 29 July 2019
Accepted: 27 August 2019
First published online: 30 October 2019

Keywords:
perceptual learning; native language; second language; immersion; emigrants

Author for correspondence:
Laurence Bruggeman,
E-mail: l.bruggeman@westernsydney.edu.au

Abstract

As a rule, listening is easier in first (L1) than second languages (L2); difficult L2 listening can challenge even highly proficient users. We here examine one particular listening function, adaptation to novel talkers, in such a high-proficiency population: Dutch emigrants to Australia, predominantly using English outside the family, but all also retaining L1 proficiency. Using lexically-guided perceptual learning (Norris, McQueen & Cutler, 2003), we investigated these listeners' adaptation to an ambiguous speech sound, in parallel experiments in both their L1 and their L2. A control study established that perceptual learning outcomes were unaffected by the procedural measures required for this double comparison. The emigrants showed equivalent proficiency in tests in both languages, robust perceptual adaptation in their L2, English, but no adaptation in L1. We propose that adaptation to novel talkers is a language-specific skill requiring regular novel practice; a limited set of known (family) interlocutors cannot meet this requirement.

Introduction

The way we listen to speech is shaped by the properties of the language we first acquire: our native tongue (L1). The phonology of the L1 constrains the set of sounds we listen for, and the L1 vocabulary furnishes the set of words that compete for our attention. L1 listening thereby becomes generally efficient and effortless.

By comparison, listening to a non-native language acquired later than in childhood (L2) is typically less efficient and often much harder, in part because our listening strategies are attuned to the L1 and may not be appropriate strategies for the L2. This is, unsurprisingly, especially true of L1-dominant participants, the typical population on which most L2 listening research has been carried out. Even highly proficient L2 listeners with regular L2 exposure appear unable to draw on the full range of skills associated with efficient L1 listening (e.g., Broersma & Cutler, 2011; Sebastián-Gallés, Echeverría & Bosch, 2005; Weber & Cutler, 2004).

However, there are many scales of variation on which L2 skills may be rated, each in turn having subordinate domains. Proficiency is one (whereby proficiency in pronunciation, in word and sentence production, and in comprehension may vary to different degrees). But exposure and usage are at least as important (and here relative amount of listening and talking may vary, as may relative usage of different registers, or relative exposure to lesser known varieties or unfamiliar talkers). Consider, for example, the emigrant experience. Many emigrants find themselves predominantly using the L2 in daily life, and the range of different exposures that comes their way in the L2 is likely to far outstrip that now offered in the L1. Under these circumstances, does L1 listening still retain its advantage over L2 listening? Are the resulting effects for these listeners uniform across the domains and processes of listening, or may some processes be selectively susceptible to the change in experience, and even eventually become more efficient in L2 than L1, while others retain an L1 advantage?

As an initial approach to this question, in this study we assessed a group of Dutch–English bilingual emigrants in Australia and examined how they adapt to a new talker's speech. Crucially, we tested such adaptation both in their L1 and in their L2. In contrast to emigrants from many other countries, Dutch emigrants have a tendency to abandon their L1 in favour of the language of their new environment (Clyne & Pauwels, 1997). Dutch emigrants in Australia indeed predominantly use their L2, English, in daily life.

Our study focused on just one type of adaptation required of listeners. Listeners are adept at many types of adjustment to changing circumstances, such as noisy environments (e.g., Warren, 1970), or speech signals that have been deformed, degraded or speeded (e.g., Remez, Rubin, Pisoni & Carrell, 1981; Shannon, Zeng, Kamath, Wygonski & Ekelid, 1995). They can also adapt quickly to foreign accents (e.g., Bradlow & Bent, 2008; Clarke & Garrett, 2004; Witteman, Weber & McQueen, 2013), although note that such speech may still cause added competitor activation (Trude, Tremblay & Brown-Schmidt, 2013). Adaptation to dialects of the L1 (Dahan, Drucker & Scarborough, 2008; Evans & Iverson, 2004), novel native-like accents (Maye, Aslin & Tanenhaus, 2008), or speech impediments

© Cambridge University Press 2019

CAMBRIDGE UNIVERSITY PRESS

(e.g., Borrie et al., 2012) can likewise be quickly achieved. Such rapid adaptation is frequently absent in the L2; for instance, listening to the L2 in noise is a well-attested problem (Garcia Lecumberri, Cooke & Cutler, 2010), and listeners likewise find dialectal variation in their L2 difficult to cope with (e.g., Mitterer & McQueen, 2009; Ockey & French, 2016).

L1 listeners also adapt very rapidly to novel talkers, and this type of adaptation formed the focus of our study. The perceptual learning process underlying such adaptation was charted by Norris et al. (2003) in a study in which participants were first exposed to an ambiguous sound between /s/ and /f/ in the context of a lexical decision task. For one group of listeners, this ambiguous sound consistently occurred in a lexical context that favoured an interpretation as /s/ (e.g., *carcass*), while another group of listeners heard the same ambiguous sound, but in other lexical contexts that led them to interpret it as /f/ (e.g., *carafe*). Listeners in a control group were exposed to the same ambiguous sound but only in non-word contexts.

This initial exposure phase was immediately followed by a test phase which was the same for all listener groups and consisted of a phonetic categorisation task. Listeners heard stimuli from an [ɛs]-[ɛf] continuum and were asked to categorise them as /s/ or /f/. Listeners who had heard the ambiguous sound in lexical contexts favouring an /f/-interpretation categorised more stimuli from the [ɛs]-[ɛf] continuum as /f/ than listeners from the other, /s/-interpretation group. Listeners in the control group showed no such bias; their categorisation responses fell between those of the two other listener groups. This showed that listeners were able to shift the boundaries of their phoneme categories to allow for the correct interpretation of ambiguous sounds, and that the lexical context in which the ambiguous sounds occurred supplied the information that guided this category boundary readjustment.

Since Norris et al.'s (2003) study, which was conducted in Dutch, this type of learning has been shown to be maximally rapid (Mitterer & Reinisch, 2013), long-lasting (Eisner & McQueen, 2006), equivalent in listeners both younger than the typical undergraduate participant population (McQueen, Tyler & Cutler, 2012) and older (Scharenborg & Janse, 2013; Scharenborg, Weber & Janse, 2015), and generalisable across the lexicon (McQueen, Cutler & Norris, 2006). Further, the findings have been extended to other languages (e.g., English: Kraljic & Samuel, 2005; Mandarin: Burchfield, Luk, Antoniou & Cutler, 2017), to other phonemes than fricatives (e.g., stops: Kraljic & Samuel, 2007; laterals and rhotics: Scharenborg & Janse, 2013) and to non-phonemic speech sounds (e.g., lexical tone: Mitterer, Chen & Zhou, 2011). Importantly for our study, talker adaptation of this kind contrasts with listening in noise or adaptation to dialectal variation, in that it has been attested in L2 listening, both when L1 and L2 were related and phonologically similar (English/German/Dutch: Drozdova, Van Hout & Scharenborg, 2016; Reinisch, Weber & Mitterer, 2013; Schuhmann, 2014) and when L1 and L2 were unrelated, with significant differences of phonological structure (English-Mandarin: Cutler, Burchfield & Antoniou, 2018).

This is unsurprising, given that the learning underlying talker adjustment rests on apparently general processes, in which perceptual experience is exploited to alter subsequent processing. Consider the studies on foreign accent cited above. In Bradlow and Bent's (2008) study, the accuracy of English speakers transcribing sentences in Chinese-accented English increased significantly over the course of the experiment. Likewise, in Clarke and Garrett's (2004) study on foreign-accented speech, English listeners hearing Chinese-accented and Spanish-accented speech were initially noticeably slower at a word verification task than a control group hearing American-English (unaccented) speech, but after a brief period (just one minute) of exposure, all groups were responding equally rapidly. Again, in a cross-modal priming study with Dutch listeners who were relatively unfamiliar with German-accented Dutch (Witteman et al., 2013), strongly accented spoken prime words exercised a priming effect only upon those listeners who had been exposed to a 4-minute story by the same speaker prior to the cross-modal task. Thus, in each case, just a short period of exposure enabled listeners to adapt to strongly accented speech. Adjustment to native-like accentual variation in vowel pronunciations, as studied by Maye et al. (2008), is similarly rapid.

All these studies thus demonstrate listeners' flexibility in adjusting to different L1 listening situations, a flexibility that in turn supports the robustness of L1 listening. However, as noted above, this is not always the case for listening to speech in L2. Even highly competent L2 listeners may still struggle with tasks that require listening flexibility and that L1 listeners accomplish with relatively little effort (e.g., listening in noise; Garcia Lecumberri et al., 2010). In that light, the fact that a comparable process of lexically-guided perceptual learning appears to occur in L2 (Drozdova et al., 2016; Reinisch et al., 2013; Schuhmann, 2014) may be indicative of differences between one type of adaptation and another: adaptation to noise, to accent, to dialects and to talkers may each involve different processes or depend differently on the contribution of exposure.

The first talker adaptation study in L2 using the Norris et al. (2003) paradigm was conducted by Reinisch et al. (2013). The study compared Dutch L1 listeners' and German L2 listeners' perceptual learning of a Dutch ambiguous fricative between /s/ and /f/. All participants lived in the Netherlands at the time of testing, which constituted an L2 immersion environment for the German-speaking participants. Following Norris et al. (2003), a lexical decision task served as the exposure phase. In the test phase, listeners categorised the ambiguous final phoneme of Dutch minimal pairs (e.g., *doos* – *doof*, "box – deaf"). Results showed that retuning of the Dutch phoneme categories /s/ and /f/ was just as strong for the L2 listeners as it was for the L1 listeners, establishing for the first time that perceptual learning can occur in L2.

Reinisch et al.'s L2 listeners were in an L2 immersion situation, but this factor does not appear to have been a necessary condition for the appearance of perceptual learning in L2 given that an experiment with L2 listeners in a NON-immersion environment subsequently also found learning (Schuhmann, 2014). Participants in this study were German listeners to L2 English, living in Germany. The experiment again included a lexical decision task and a phoneme categorisation task, and used an American-English /s/-/f/ ambiguous fricative; these German L2 listeners succeeded in retuning their phoneme categories in English.

Additional evidence for perceptual learning by non-immersed L2 listeners comes from the remaining two L2 studies. Drozdova et al. (2016) exposed Dutch L2 listeners in the Netherlands to a British-English story containing an ambiguous sound between /l/ and /r/, and subsequently performed a phonetic categorisation task in which they categorised the ambiguous sound in English minimal pairs, such as *arrive* – *alive*, and *correct* – *collect*. Cutler et al. (2018) used the Norris et al. (2003) design of lexical decision followed by phonetic categorisation, and tested listeners

https://doi.org/10.1017/S1366728919000646 Published online by Cambridge University Press

whose L1 was English on ambiguous fricatives in their L2, Mandarin; the experiment was conducted in the L1 listeners' English-speaking environment. In each of these studies, the listeners successfully adapted to the novel pronunciation and adjusted the boundary between the phonetic categories in question.

These studies all suggest that listeners may be flexible when listening to their L2, at least when talker adaptation is concerned, and that the robustness of speech perception may thus to a significant degree be comparable in L1 and in L2 listening. However, no direct comparisons have been made of the perceptual learning process itself in L1 and L2, as studies to date (outside our laboratory) have involved only comparisons between separate listener groups. It is not known to what degree perceptual learning does in fact vary, either in L1 or L2, and to what extent such variation might then stem from differences on listener-, language- or situation-specific measures. Keeping these factors constant, as far as possible, at the very least allows a less confounded approach to the L1 versus L2 comparison.

Our study thus tests both L1 (Dutch) and L2 (English) perceptual learning in a group of bilingual emigrants living in Australia (i.e., an L2 immersion environment). Perceptual learning has been repeatedly demonstrated both in L1 Dutch (e.g., Norris et al., 2003; Scharenborg & Janse, 2013) and in L1 Australian English (e.g., Cutler et al., 2018; McQueen et al., 2012). Here participants' perceptual learning was investigated in both L1 and L2, with the experiments being administered a few weeks apart.

General method

Paradigm adjustments

In the lexically-guided perceptual learning paradigm as developed by Norris et al. (2003), listeners adapt the category boundary between (for example) /s/ and /f/ upon exposure to an ambiguously pronounced sound. The measure of perceptual learning is the degree of shift in the /s/-/f/ category boundary as a function of training to interpret this ambiguous sound as /s/ versus /f/. Ambiguous sounds used in such studies are typically chosen from a continuum formed by gradually merging the two endpoint sounds (e.g., /s/, /f/). The ambiguous sound presented in the exposure and test phase is selected based on a categorisation pretest involving a separate group of participants.

In this study, however, we could not rely on the availability of such a closely comparable pre-test group. We conducted instead individual pre-tests for each participant, which then allowed personalised tailoring of the exposure and test phase. Besides making a separate pre-test group unnecessary, this procedural change also deals with the greater-than-usual variability within our participant group. Although perceptual learning has proven to be consistent across the lifespan (McQueen et al., 2012; Scharenborg & Janse, 2013), age variation (with its concomitant variation in high-frequency hearing ability) is usually smaller within participant groups than we could achieve here. In addition, length of residence in Australia varied, potentially affecting the total amount of exposure participants had received to both Dutch and English throughout their lives. Although we tested our participants' proficiency (in each language), the length of residence factor could also influence such factors as surety in perceiving and categorising fricatives. Personalised tailoring of test materials addresses each of these variability points.

Such tailoring should make the paradigm highly sensitive to evidence of perceptual learning. However, it raises other potential issues. First, the pre-test with an [s]-[f] continuum could draw attention to the /s/ or /f/ manipulation in the exposure phase. To mask this relationship, a further phonetic judgement task was added along with the fricative task. Second, the proposed experiment would comprise at least five parts for each language instantiation: the pre-test to establish the individual category boundary, the filler phonetic task, the exposure phase to invoke learning, the post-test to assess the degree of learning achieved, and the language proficiency test also. This compares with the two-part procedure of prior studies of this type. Third, the experimenter would need to compute the individual choice for the most ambiguous fricative sound (henceforth referred to as [?]) and create the resulting set of individual exposure materials, in the time between pre-test and exposure (previously non-existent, but now devoted to the language proficiency test). Fourth, it would also mean that participants would not all be exposed to the same step of the [s]-[f] continuum, a factor of uncertain relevance. In short, the procedural alterations were significant enough to require pre-testing in their own right. Experiment I below compared the novel procedure with the traditional procedure; this comparison was carried out in Dutch. Experiment II then tested Dutch–English bilinguals' perceptual learning in both Dutch and English. The overall design and the materials used in the experiments are described first.

Design, materials and procedure

The perceptual learning tasks were closely based on those of Norris et al. (2003), which examined listeners' adaptation to an ambiguously realised fricative sound between /s/ and /f/, by using a lexical decision task for exposure and a phonetic categorisation task for test. In these respects we exactly replicated their procedure. However, we selected the ambiguous sound differently, via a pre-test at the experiment outset. Further, the extra filler task was inserted at two points: after the fricative pre-test but before the exposure phase, and (since participants still had to complete the second session of the experiment a few weeks later) after the fricative categorisation test phase of the first testing session. In this filler task, participants categorised vowels on a [tɔ]-[tɑ] (Dutch) or [tɔ]-[tɐ] (English) continuum. The filler task stimuli were spoken by the same speaker who recorded the stimuli for the rest of the experiment.

Two versions of all parts of the experiment were created, one in Dutch and one in English. All stimulus materials for the Dutch version were recorded by a female native speaker of Dutch living in the Netherlands; all English materials were recorded by a female native speaker of Australian English living in Australia. Recordings were made digitally (sampling rate 44.1 kHz, sampling resolution 16-bit) in a sound-attenuated booth.

For the pre-test, the syllables /ɛs/, /ɛf/, and /ɛθ/ were recorded and from these recordings the [s] and [f] sounds were isolated using PRAAT (Boersma & Weenink, 2013). Following Norris et al. (2003), fricatives were excised from a zero-crossing at frication onset to a zero-crossing near the end of frication. To facilitate the creation of a continuum, the duration of both fricatives within a language was kept the same (at 345 ms for the Dutch fricatives, and at 218 ms for the English fricatives, determined by the speaker's mean duration for the shorter phoneme, which was /f/ in both languages). For each language, the isolated [s] and [f] sounds were then merged to create a 41-step [s]-[f] continuum, using a sample-averaging method (Repp, 1981) as in the predecessor perceptual learning studies. At each step of the continuum, different

https://doi.org/10.1017/S1366728919000646 Published online by Cambridge University Press

proportions of each sound were used, with all steps being equidistant. Thus, step 1 consisted of 100% [s] and 0% [f], step 21 of 50% [s] and 50% [f], and step 41 of 0% [s] and 100% [f]. For use in the pre- and post-test, each step of this continuum was then spliced onto the vowel [ɛ] (Dutch: 151 ms, English: 173 ms) of the [ɛθ] recording of the same language. This created a 41-step [ɛs]-[ɛf] continuum. Using [ɛ] from [ɛθ], rather than from [ɛs] or [ɛf], ensured that acoustic cues in test token vowels would not bias listeners' fricative categorisation towards either /s/ or /f/. Pilot phonetic categorisation experiments with native listeners of the respective languages confirmed that both the Dutch and English continua were perceived categorically (see Appendices A and B).

In each language, the filler vowel categorisation continuum ([tɔ]-[tɑ] for Dutch; [tɔ]-[tɐ] for English) was developed by a parallel procedure.

The ambiguous [?] was selected individually per participant (by the experimenter, while the participant undertook another task). Appropriate tokens were chosen based on each participant's pre-test result on the relevant continuum, using the same decision metric applied groupwise in prior studies. The step of the [s]-[f] continuum that received closest to 50 percent /f/-responses in a participant's pre-test became [?] in that participant's exposure phase. Again individually for each participant, four additional sounds from the continuum, being the steps that received closest to 90, 70, 30 and 10 percent pre-test /f/-responses, were chosen for use in the post-test.

In each language version of the experiment, the exposure phase stimuli comprised 40 critical words, 60 filler words and 100 non-words. Appendices C and D list the critical words for the Dutch and English version, respectively. Twenty of the critical words ended in /s/, and twenty in /f/. No further [f], [v], [s], [z], [ʧ]or [ʤ] sounds occurred in any stimuli. In addition, to avoid co-activation of lexical items, no critical fricative-final items were cognates in Dutch and English. The English lists of critical /s/-final and /f/-final words each contained six monosyllables (e.g., *dress*, *rough*), 11 disyllables (e.g., *embrace*, *midwife*) and three trisyllables (e.g., *hideous*, *autograph*). The Dutch lists contained two monosyllables (e.g., *krijs* "scream", *braaf* "honest"), six disyllables (e.g., *matroos* "sailor", *doolhof* "maze"), six trisyllables (e.g., *ananas* "pineapple", *ongeloof* "disbelief") and six four-syllable words (e.g., *bekentenis* "confession", *rentetarief* "interest rate"); these frequencies accurately reflect average word length distributions across the two languages.

Mean word form frequency in the CELEX lexical database (Baayen, Piepenbrock & Gulikers, 1995) was 33 and 37 per million for the English /s/-final and /f/-final words respectively; for the Dutch critical stimuli this was 3 per million for each of the /s/-final and /f/-final sets. The item frequencies were lower in Dutch than in English since the exclusion of cognates left insufficient high-frequency Dutch /s/-final and /f/-final words available. Many higher-frequency words used in previous Dutch perceptual learning studies had English cognates and could therefore not be used here. (Despite their lower frequency, the Dutch stimulus words were well-known by participants, as evidenced by the results of the lexical decision task; see Results sections). All critical words were recorded both naturally and with a word-final [θ] (e.g., *embrace* was recorded both as [ɛm'breɪs] and as [ɛm'breɪθ]). Ambiguous versions of the critical words were created by removing the final sound from the [θ]-final recording and replacing it by [?].

Further, in each experimental version, two stimulus lists were created for the lexical decision task used as exposure. Both contained all 40 test words, all 60 filler words and all 100 non-words. One list – the s-bias list – contained natural versions of all /f/-final words, and ambiguous versions of all /s/-final words. The other list – the f-bias list – had ambiguous versions of all /f/-final words, and natural versions of all /s/-final words. Stimulus order was randomised per participant, with the exception of the first 12 trials, in which the same six filler words and six non-words occurred in the same order for all participants.

The new testing procedure was the same for both experiments and for both sessions of Experiment II. Participants were tested individually and were seated in front of a computer screen in a sound-attenuated booth. The Lexical Test for Advanced Learners of English (LexTALE; Lemhöfer & Broersma, 2012) was presented in MATLAB (R2013b; The MathWorks, Inc.); all other tasks were conducted using DMDX (version 4.3.0.0; Forster & Forster, 2003). Auditory stimuli were presented over headphones at a comfortable sound level, constant for all participants. For all tasks, written instructions were provided (displayed on screen or printed on paper) in the language of the experiment and were also spoken by the experimenter, to ensure participants fully understood the tasks. To equalise interaction with the experimenter (i.e., the amount and type of language used) across participants, participants were asked to remain inside the sound-attenuated booth between exposure and post-test, though they could leave the booth for a short break between other parts of the experiment.

Each testing session started with two phoneme categorisation tasks (pre-test and filler). The pre-test consisted of six blocks, each containing all 41 steps of the [ɛs]-[ɛf] continuum in random order, for a total of 246 trials. Participants were asked to categorise the final sound of each token as either /s/ or /f/ using the left and right shift buttons on a keyboard, and to give their response both quickly and accurately. The letters S and F appeared on either side of the computer screen, above the corresponding shift key. For half of the participants, the /s/-response was assigned to their dominant hand, for the other half the /f/-response was assigned to their dominant hand. There was no time limit for participants to respond to a trial, and stimuli were always presented 1.5 s after the response to the previous trial. After each block a self-paced break was provided, during which participants remained in the experiment booth. Immediately after the pre-test, participants completed the filler task, which had two blocks of 30 trials each (three repetitions of ten selected steps from the [tɔ]-[tɑ] or [tɔ]-[tɐ] continuum) in randomised order. Participants categorised the final sound of each token using the left and right shift buttons on a keyboard. They were once again instructed to respond both quickly and accurately.

Next, participants completed the LexTALE (Lemhöfer & Broersma, 2012) in the language of the session, while, unbeknownst to participants, the experimenter analysed the pre-test results, so that the ambiguous fricatives for the exposure phase could be selected. Participants then moved on to the exposure phase, an auditory lexical decision task in which they had to indicate, using the left and right shift buttons, whether the items they heard were real words or nonsense words. Half of the participants were exposed to the f-bias list, while the other half heard the s-bias list.

Response instructions emphasised both speed and accuracy. There were 200 trials, and "word"-responses were always given with the participant's dominant hand. The texts *woord* ("word") and *geen woord* ("not a word"; for the L1 task), or *word* and *not a word* (for the L2 task) were displayed on either side of the

https://doi.org/10.1017/S1366728919000646 Published online by Cambridge University Press

computer screen, above the corresponding shift key. There was no time limit for participants' responses. The next word was presented 1 s after participants had responded to the previous trial.

Immediately after the exposure phase, participants continued with the phoneme categorisation post-test. Instructions for the post-test were presented on the screen and not further explained orally. The procedure of the post-test was identical to that of the pre-test. The same response button was assigned to a participant's dominant hand as during the pre-test, so that for half of the participants, the /s/-response was assigned to their dominant hand, and for the other half of the participants the /f/-response was assigned to their dominant hand. The post-test contained twelve repetitions of five steps of the [ɛs]-[ɛf] continuum, selected from the participant's pre-test as described above. In total, there were 60 trials in the post-test, presented in a random order. In the first experimental session of Experiment II only, the post-test was followed by the filler vowel categorisation task, containing 12 repetitions of five steps of the [tɔ]-[tɑ] or [tɔ]-[tɐ] continuum (a total of 60 trials).

Experiment I: Testing the personalised procedure

Participants

Fifty–four participants (aged 18–36 years, $M = 22.6$, $SD = 4.2$; 40 females) were recruited from the participant pool of the Centre for Language Studies at the Radboud University Nijmegen, the Netherlands. They received a gift voucher in return for participating. Thirty participants took part in a Personalised condition, 24 in a Standard condition. All were native speakers of Dutch and none reported any hearing problems. A further two participants' results were excluded from analysis as they had grown up speaking languages other than Dutch. On the Dutch version of LexTALE (Lemhöfer & Broersma, 2012), participants achieved a mean score of 92.2% ($SD = 5.7$, range: 75-100%). Written informed consent was obtained from each participant before the experiment.

Procedure

The experiment consisted of a single testing session, in Dutch. Two conditions were created: a Personalised condition and a Standard condition. Personalised condition participants completed the pre-test and the filler task, followed by the LexTALE, the exposure phase and the post-test; Standard condition participants completed the exposure phase and post-test only, and subsequently took the LexTALE. For participants in this latter condition, without individual pre-test results, the selection of [?] for exposure and post-test used the averaged pre-test results of the participants in the Personalised condition. Thus the Standard condition post-test used the five steps of the 41-step [s]-[f] continuum that on average received nearest to 10%, 30%, 50%, 70% and 90% /f/-responses from the Personalised participants (respectively steps 20, 24, 26, 28, and 30), while [?] was the step most nearly receiving 50% /f/-responses (step 26). Stimuli were presented over Sennheiser HD 215 headphones.

Results and discussion

Exposure

Following Norris et al. (2003), all participants who accepted less than 50% of [?]-final items as words were excluded from further analysis. This was the case for two Personalised participants and one Standard participant, all in /f/-bias groups. In consequence, the analysis included data from 28 Personalised condition participants (14 per bias condition) and 23 Standard condition participants (13 s-bias, 10 f-bias). Overall, f-bias and s-bias participants respectively responded correctly to 93.4% and 94.0% of the word fillers and 97.1% and 95.5% of the non-word fillers. Table 1 shows listeners' responses to the fricative-final experimental items. The ambiguous words presented to the f-bias group were rejected most often, with "yes" responses to only 84.6% of these items.

Table 1. Response times and percentage of correct responses to experimental items in the exposure phase of Experiment I. Response times are measured from target word onset.

	f-bias group		s-bias group	
	natural fricatives	ambiguous fricatives	natural fricatives	ambiguous fricatives
Mean % "yes"	95.2	84.6	95.9	93.1
Mean RT "yes" (in ms)	1085	1194	1157	1187

Table 2. Fixed-effect estimates of listeners' accuracy in the exposure phase of Experiment I.

Fixed effect	β	SE	z	p
Intercept	−4.294	0.401	−10.706	< .001
Exposure bias	−1.041	0.610	−1.706	.088
Pronunciation	1.466	0.642	2.283	.023
Exposure bias * Pronunciation	0.055	1.197	0.046	.964

We fitted a generalised linear mixed-effects model to the data in R (R Core Team, 2018), using family 'binomial' and the logit-link function from the lme4 package (Bates, Maechler, Bolker & Walker, 2015). Exposure bias (with s-bias coded as 0.5, and f-bias as -0.5) and Pronunciation (Ambiguous coded as 0.5, Natural coded as -0.5) were added to the model as deviation coded fixed factors. A maximal random effects structure was used (Barr, Levy, Scheepers & Tily, 2013), with random intercepts for participants and items, as well as random slopes for Pronunciation by participant. Results of the model fit are displayed in Table 2. The model confirmed a significant effect of Pronunciation: words ending in a naturally pronounced fricative were accepted as existing words more often than words ending in an ambiguous fricative. No other significant effects or interactions were found.

Post-test

Figure 1 shows the mean percentage of /f/-responses for the 51 participants included in the analysis (separate figures by Condition are provided in Appendices E and F). To compare the responses in the Personalised condition to those in the Standard condition, we fitted another generalised linear mixed-effects model with family 'binomial' and the logit-link function. All fixed factors were again deviation coded. Exposure bias (with s-bias coded as 0.5, and f-bias as -0.5) and condition

https://doi.org/10.1017/S1366728919000646 Published online by Cambridge University Press

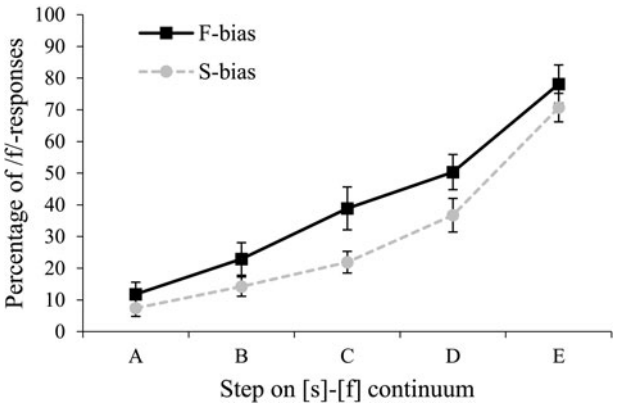


Fig. 1. Mean percentage of /f/-responses for the post-test of Experiment I for listeners in the f-bias and s-bias groups. Error bars represent standard errors. Step A is the most /s/-like of the tested ambiguous stimuli, while step E is the most /f/-like.

(Personalised coded as 0.5, Standard coded as -0.5) were included in the model as fixed categorical predictors, while fricative step (A-E) was included as a fixed continuous predictor and centred on step C (step A coded as -2, step B as -1, step C as 0, step D as 1, and step E as 2). Random intercepts were added for participants and items, as well as random slopes for fricative step by participant, and for exposure bias and condition by item. Results of the model fit are displayed in Table 3. The model confirmed a significant effect of step and of exposure bias; listeners in the f-bias group categorised significantly more tokens as /f/ than listeners in the s-bias group, indicating perceptual learning. No interactions were found involving step and exposure bias. Moreover, the model showed no effect of condition, nor any interactions involving condition, indicating that exposure phase outcomes were unaffected by whether or not listeners had completed a pre-test.

The results of Experiment I confirm that the personalised testing procedure induces perceptual learning in the same way as the standard procedure used in prior studies. Further, it appears that none of our procedural alterations – addition of a pre-test, use of individually selected ambiguous tokens, multiple components of the experimental session – exercises any adverse effect on perceptual learning. Lastly, the set of Dutch stimuli used in Experiment I (selected to avoid Dutch words with English cognates) has also been proven to replicate previous Dutch materials sets in guiding the retuning of listeners' phonetic categories.

Experiment II: Testing perceptual learning in an individual's two languages

Participants

Thirty-two participants (age range 24-73 years, $M = 47.8$, $SD = 16.8$; 21 females) were recruited from the Dutch immigrant community in the wider Sydney area, and paid for their participation. Data of two additional participants were excluded from analysis, in one case because it was revealed at a late point that the participant had lived in Australia for part of childhood, and in the other case because the participant did not complete the testing sessions. All filled out a background questionnaire, which showed that all included participants had been born and raised in the Netherlands, were native speakers of Dutch and had migrated to Australia as adults (mean age at migration = 28.5 years, $SD = 7.4$, range: 18-52). The mean length of residence in Australia was 19.4 years ($SD = 16.7$, range: 9 months-55 years). Participants' mean score on the English version of the LexTALE (Lemhöfer & Broersma, 2012) was 91.4% ($SD = 8.8$, range: 62.5-100%), indicating a high proficiency in English. On the Dutch version of LexTALE their mean score was 92.2% ($SD = 5.08$, range: 80-100%), showing that all participants still maintained high proficiency in their L1 despite migration. No participant reported any hearing problems. Written informed consent was obtained from each participant before starting the initial session.

Table 3. Fixed-effect estimates of listeners' performance in the post-test of Experiment I.

Fixed effect	β	SE	z	p
Intercept	−1.094	0.225	−4.863	< .001
Fricative step	1.172	0.127	9.214	< .001
Exposure bias	−0.823	0.391	−2.106	.035
Condition	0.268	0.414	0.649	.516
Fricative step * Exposure bias	0.039	0.200	0.197	.844
Fricative step * Condition	0.299	0.222	1.348	.178
Exposure bias * Condition	0.590	0.776	0.760	.447
Fricative step * Exposure bias * Condition	−0.509	0.395	−1.288	.198

Procedure

Experiment II comprised two perceptual learning sessions – one for each language – completed approximately three weeks apart ($M = 21.6$ days, $SD = 4.18$, range: 14-28). In the L1 session, all stimuli and tasks were in Dutch; in the L2 session they were in English. Each session included administration of the respective version of LexTALE. Order of sessions was counterbalanced across participants. Materials were presented over Sennheiser HD 650 headphones.

Results and discussion

Exposure: L1

As in Experiment I, participants who accepted less than 50% of [?]-final items as words during the exposure phase were excluded from further analysis. This was the case for three participants, who had all been exposed to natural /s/-final words and ambiguous /f/-final words (the f-bias group). Thus, the analyses included results of 29 participants: 13 in the f-bias group and 16 in the s-bias group (natural /f/-final, ambiguous /s/-final words).

Both exposure groups were presented with the same filler items and the percent correct for each group was high (f-bias group: 96.6% "yes" to filler words, 95.7% "no" to filler non-words; s-bias group: 95.4% "yes" to words, 96.2% "no" to non-words). Responses to the fricative-final experimental items are shown in Table 4. As in Experiment I, the ambiguous words presented to the f-bias group were more likely to be rejected; only 82.3% of these items were accepted as existing words.

https://doi.org/10.1017/S1366728919000646 Published online by Cambridge University Press

Table 4. Response times and percentage of correct responses to experimental items in the exposure phase of the L1 task of Experiment II. Response times are measured from target word onset.

	f-bias group		s-bias group	
	natural fricatives	ambiguous fricatives	natural fricatives	ambiguous fricatives
Mean % "yes"	97.3	82.3	96.9	92.5
Mean RT "yes" (in ms)	1268	1479	1320	1325

We fitted a generalised linear mixed-effects model to the data, with family 'binomial' and the logit-link function, using the same fixed and random structure as used for the exposure phase of Experiment I. Results of the model fit are shown in Table 5. A significant effect of Pronunciation indicated that words ending in naturally pronounced fricatives were accepted as existing words more often than words ending in ambiguous fricatives. No other effects or interactions were significant.

Exposure: L2

Using the same 50%-acceptance criterion as for the L1 exposure task, three participants, all in the f-bias group, were excluded from further analysis. An additional three participants (two in the f-bias, one in the s-bias group) were excluded because they showed no sign of categorical perception in the pre-test phase. Due to a technical error, results from one further participant were not saved and therefore could not be analysed. Thus, the results of 25 participants were included in the analyses reported below: 11 in the f-bias and 14 in the s-bias group.

Again, filler performance was high: f-bias group 96.1% "yes" to words and 83.8% "no" to non-words, s-bias group 94.3% "yes" to words, 88.8% "no" to non-words. Responses to the fricative-final experimental items are shown in Table 6. The ambiguous words were again most often rejected as existing words by listeners in the f-bias group; only 84.1% of these items received a "yes" response.

Responses were analysed using the same generalised linear mixed-effects model as described previously for the exposure phase. Results of the model fit (see Table 7) revealed no significant effects or interactions.

Post-test: L1

Figure 2 shows the mean percentage of /f/-responses to all five ambiguous stimuli in the L1 phonetic categorisation post-test. As for Experiment I, we fitted a generalised linear mixed-effects model to the data, with family 'binomial' and the logit-link function. Included as deviation coded fixed factors were fricative step (centred on step C; step A: -2, step B: -1, step C: 0, step D: 1, step E: 2) and exposure bias (s-bias: 0.5, f-bias: -0.5). Random intercepts were added for participants and items, and random slopes for fricative step by participant, and for exposure bias by item. Results of the model fit are displayed in Table 8. As expected, they showed a significant effect of step, indicating appropriate sensitivity to the continuum progression. However, they did not show any effect of exposure bias; in fact, the very small difference between the two curves is actually in the wrong direction. Nor was there any interaction between exposure bias and step. Listeners in the two exposure conditions thus did not significantly differ in their categorisation of the ambiguous fricatives, i.e., there is no evidence of perceptual learning.

Table 5. Fixed-effect estimates of listeners' accuracy in the exposure phase of the L1 task of Experiment II.

Fixed effect	β	*SE*	*z*	*p*
Intercept	−4.125	0.443	−9.311	< .001
Exposure bias	−0.788	0.657	−1.199	0.231
Pronunciation	1.644	0.692	2.374	0.018
Exposure bias * Pronunciation	−1.334	1.179	−1.132	0.258

Table 6. Response times and percentage of correct responses to experimental items in the exposure phase of the L2 task of Experiment II. Response times are measured from target word onset.

	f-bias group		s-bias group	
	natural fricatives	ambiguous fricatives	natural fricatives	ambiguous fricatives
Mean % "yes"	88.6	84.1	93.2	90.0
Mean RT "yes" (in ms)	1300	1399	1165	1262

Table 7. Fixed-effect estimates of listeners' accuracy in the exposure phase of the L2 task of Experiment II.

Fixed effect	β	*SE*	*z*	*p*
Intercept	−4.804	0.704	−6.828	< .001
Exposure bias	−1.532	1.027	−1.491	0.136
Pronunciation	2.144	1.100	1.950	0.051
Exposure bias * Pronunciation	−0.058	1.868	−0.031	0.975

As some listeners had completed the L1 task in session 1 and others in session 2 (i.e., after they had already completed the L2 task), session (first vs second) was included as a deviation coded fixed factor (session 1 coded as 0.5, session 2 coded as -0.5) in a separate model. This model suggested that our finding does not appear to be related to the order of testing, as no effect of session ($\beta = 1.322$, $p = .070$) was found, nor any significant interactions between session and step ($\beta = 0.141$, $p = .581$), session and exposure bias ($\beta = 1.936$, $p = .177$), or session, exposure bias and step ($\beta = -0.436$, $p = .376$).

Post-test: L2

Although one participant's lexical decision results had been lost and hence could not be analysed, this same participant's phonetic categorisation data WERE available and were included in the analyses. The results reported below are thus from 26 participants, 11 in the f-bias and 15 in the s-bias group.

Figure 3 shows the mean percentage of /f/-responses to all five ambiguous stimuli (steps A-E) in the phonetic categorisation task. The perceptual learning effect is represented by the difference between the two lines. A model of the same type as used for the analysis of the L1 post-test, and including the same fixed factors

https://doi.org/10.1017/S1366728919000646 Published online by Cambridge University Press

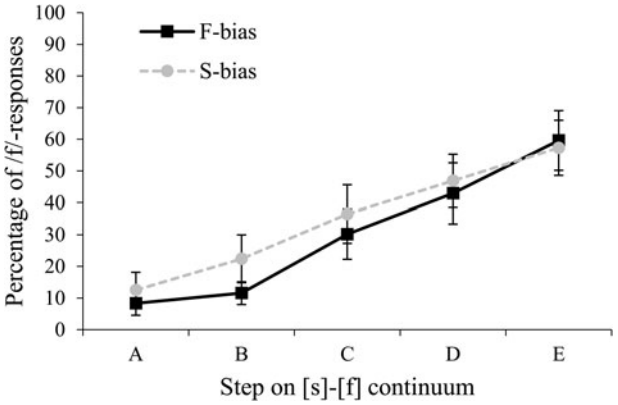


Fig. 2. Mean percentage of /f/-responses in the fricative post-test of the L1 task of Experiment II for listeners in the f-bias and s-bias groups. Error bars represent standard errors. Step A is the most /s/-like of the tested ambiguous stimuli, while step E is the most /f/-like.

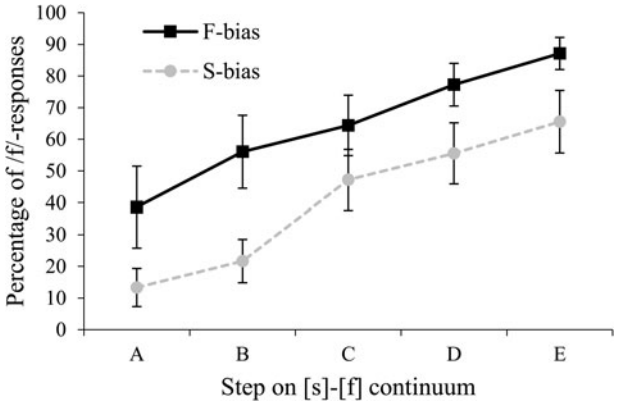


Fig. 3. Mean percentage of /f/-responses in the fricative post-test of the L2 task of Experiment II for listeners in the f-bias and s-bias groups. Error bars represent standard errors. Step A is the most /s/-like of the tested ambiguous stimuli, while step E is the most /f/-like.

Table 8. Fixed-effect estimates of listeners' performance in the post-test of the L1 task of Experiment II.

Fixed effect	*β*	*SE*	*z*	*p*
Intercept	−1.410	0.386	−3.657	< .001
Fricative step	1.029	0.135	7.634	< .001
Exposure bias	0.420	0.755	0.557	.578
Fricative step * Exposure bias	−0.128	0.249	−0.514	.608

Table 9. Fixed-effect estimates of listeners' performance in the post-test of the L2 task of Experiment II.

Fixed effect	*β*	*SE*	*z*	*p*
Intercept	0.300	0.593	0.506	.613
Fricative step	1.067	0.124	8.590	< .001
Exposure bias	−2.779	1.192	−2.331	< .001
Fricative step * Exposure bias	0.079	0.287	0.277	.782

and random structure, was fit to the data (see Table 9 for results of this fit). The model showed significant effects of step and exposure bias. Listeners in the f-bias group, trained to interpret the ambiguous fricative sound [?] as /f/, categorised significantly more tokens as /f/ than listeners in the s-bias group (trained to interpret the same sound as /s/). This difference is an indication of perceptual learning; listeners have adjusted the boundaries of their phoneme categories to accommodate for the ambiguous pronunciation of the fricative sound [?] and have done so based on the lexical context in which this ambiguous fricative occurred.

Again, a separate analysis with session (first vs second) as an additional fixed factor was conducted to examine whether the perceptual learning may have been affected by this factor, and again this was not the case: no effect of session ($\beta = -1.205$, $p = .342$) appeared, nor any significant interaction of session and step ($\beta = 0.399$, $p = .109$), session and exposure bias ($\beta = -4.165$, $p = .101$), or session, exposure bias and step ($\beta = -0.199$, $p = .732$).

Finally, we directly compared responses in the L1 and L2 phonetic categorisation tasks. To this end, we once more constructed a generalised linear mixed-effects model with family 'binomial' and the logit-link function. Included as deviation coded fixed factors were fricative step (centred on step C; step A: -2, step B: -1, step C: 0, step D: 1, step E: 2), exposure bias (s-bias: 0.5, f-bias: -0.5), and language (L1: 0.5, L2: 0.5). Random intercepts were added for participants and items, and random slopes for fricative step and language by participant, and for exposure bias by item. Results of the model fit (see Table 10) showed a significant effect of language, and a significant interaction between language and exposure bias, confirming that listeners responded differently to the exposure bias in L1 and L2.[1]

In sum, then, our examination of L1 and L2 talker adaptation by a single group of listeners (native Dutch-speaking emigrants immersed in an L2 English environment in Australia) has produced an asymmetric outcome. When these listeners are exposed to stimuli containing ambiguously pronounced fricatives, lexically-guided perceptual learning indeed occurs in L2. But this contrasts starkly with an absence of equivalent learning in L1. The result cannot be ascribed to our personalised procedure as that produced significant learning in Experiment I and also in the English half of Experiment II. Nor can it be ascribed to the particular materials we used, as these showed learning for the Experiment I participants. The L2 results are in line with previous demonstrations of lexically-guided perceptual learning in Dutch listeners to L2 English; but the L1 findings contrast with the substantial previous evidence of perceptual learning in L1 Dutch. Most strikingly, the L1 results do not replicate the L2 perceptual learning success BY THE SAME LISTENERS.

[1]Since the perceptual learning paradigm depends on a comparison at a group level (i.e., it is not possible to determine the perceptual learning effect for participants individually), we were unable to explore the relationship between individual participants' perceptual learning in L1 and L2. However, we did calculate – for each step of the continuum and for each participant – the percentage of 'learning-consistent' responses (i.e., the percentage of /f/-responses for listeners in the f-bias group, and the percentage of /s/-responses for listeners in the s-bias group; Scharenborg & Janse, 2013). While these do not constitute perceptual learning per se (as they are not compared to a baseline established before exposure), they may provide some indication of how many responses a participant gave in line with the bias to which they were exposed. No correlations were found between participants' percentage of learning-consistent responses in L1 and L2, for any of the five steps.

https://doi.org/10.1017/S1366728919000646 Published online by Cambridge University Press

Table 10. Fixed-effect estimates of listeners' performance in both L1 and L2 post-tests of Experiment II.

Fixed effect	*β*	*SE*	*z*	*p*
Intercept	−0.591	0.426	−1.386	.166
Fricative step	1.029	0.112	9.191	< .001
Exposure bias	−0.958	0.729	−1.314	.189
Language	−1.746	0.653	−2.673	.008
Fricative step * Exposure bias	0.043	0.133	0.325	.746
Fricative step * Language	−0.017	0.102	−0.165	.869
Exposure bias * Language	2.927	1.395	2.098	.036
Fricative step * Exposure bias * Language	−0.284	0.254	−1.119	.263

General discussion

An individual's perceptual learning abilities are not necessarily constant across languages. Here we have seen a remarkable asymmetry in the ability to adjust to the speech of a novel interlocutor in the L1 versus in the L2, in our test population of long-term emigrants. When listening to an unfamiliar speaker of Australian English (i.e., of their L2), these emigrants used lexical knowledge to disambiguate ambiguously pronounced fricatives and retuned their phoneme categories, as expected, to accommodate to the specific characteristics of the new talker's speech. When the same listeners heard ambiguously pronounced fricatives from an unfamiliar speaker of their L1, Dutch, however, no category retuning occurred. Thus with respect to talker adaptation, we have demonstrated that no automatic privilege accrues as a result of listening to speech in the L1 rather than an L2.

Methodologically, we have further shown that the addition of a pre-test in no way alters the pattern of results produced in the traditional perceptual learning paradigm. This finding is important in that it may allow future investigations of perceptual learning in otherwise scarce or difficult-to-study populations, where a pre-test with a separate group of pilot participants is undesirable or simply not feasible.

Our L2 results confirm previous findings for perceptual learning in L2 speech perception (Cutler et al., 2018; Drozdova et al., 2016; Reinisch et al., 2013; Schuhmann, 2014) and indicate that the flexibility required for perceptual learning is indeed also available during L2 listening. Although the robustness of speech perception may under some circumstances be significantly worse in L2 than in L1 listening, for instance in noisy environments (Garcia Lecumberri et al., 2010), perceptual abilities in L2 appear not to lag behind in the case of talker adaptation. Reviewing the literature on L2 listening in noise, Garcia Lecumberri et al. proposed that L1 listeners may outperform L2 listeners in this situation because they have at their disposal more higher-level resources of the kind that support the process of recovering from the inevitable impact of noise or signal degradation on low-level perception (e.g., a larger vocabulary, a wider knowledge of syntactic patterns). The findings from the present study indicate that whatever resources are required to enable flexible adaptation to novel talkers, L2 listeners have enough of them at their disposal. Thus it seems likely that the resources involved in coping with noise are the linguistic resources required for comprehension, and thus not the same as those involved in coping with unfamiliar interlocutors, which, as argued in the introduction, principally involve general processes.

Nativeness in a language is thus not NECESSARY for perceptual learning; but neither is it in itself SUFFICIENT. We assume that there is no selective factor at play here in the sort of person who becomes an emigrant, or in the nature of the migration experience. There are also several other potential factors that our analyses suggest may be ruled out.

For instance, consider first the acoustic properties of the Dutch and English fricative phonemes used in our study. While /f/ is spectrally similar in both languages, English /s/ typically contains more energy in the higher frequency regions than Dutch /s/ (Collins & Mees, 2003). After extended immersion in the L2 environment, could Dutch emigrants have adapted to the acoustic characteristics of English /s/, leading to an adjusted perception of that phoneme in their L1? If so, this could have prevented potential perceptual learning effects from being revealed by our experimental paradigm. The emigrants' categorisation of the L1 [ɛs]-[ɛf] continuum during their pre-test was therefore compared to the results of the pilot categorisation test with L1 listeners in the Netherlands (Appendix A). This comparison revealed no significant differences between the response distributions of the two participant groups ($F(1, 47) = 0.785$, $p = .380$, $\eta_p^2 = .02$). It thus seems very unlikely that acoustic differences between Dutch and English /s/ have resulted in unusual L1 phonetic categories for the emigrants which in some way interfered with the appearance of perceptual learning effects in the L1 half of Experiment II.

Second, consider the age of the emigrants, who, with a mean age of 47.75 years, were older both than typical participants in prior perceptual learning studies, and than the control participants of Experiment I (mean age 22.6 years). This potential issue (with its implications concerning hearing ability) is easy to rule out given that our participants did show perceptual learning in L2 and we know of no reason why the effects of age and/or auditory acuity should differ across an individual's languages, especially in such a way as to disfavour the L1. Our present findings thus add further support to earlier studies that have found that perceptual learning is not restricted to younger adults but occurs in older adults as well (Scharenborg & Janse, 2013; Scharenborg et al., 2015), with individual hearing sensitivity being uncorrelated with the strength of perceptual learning (Scharenborg & Janse, 2013).

A third factor is language proficiency. The experimental paradigm used here relies on lexical information to induce perceptual learning. If a word is not present in a listener's mental lexicon, it cannot contribute towards disambiguation of the ambiguous fricative, so that listeners with less well-stocked lexicons would have fewer opportunities to learn during the exposure phase. Although the emigrants of Experiment II left their L1 environment, on average, in their 20s and the participants in Experiment I were, on average, in their 20s, it is nonetheless possible that the two groups may have differed in proficiency. Although vocabulary size in fact tends to increase with age (Ramscar, Hendrix, Shaoul, Milin & Baayen, 2014), it is still possible that a proficiency difference in the opposite direction occurred such that the older emigrant group had a smaller Dutch vocabulary than the younger control group (who were students), and thus had insufficient opportunities to learn from exposure to the ambiguous fricative. There are several ways to compare proficiency and vocabulary size in our data set. The first way is to compare the accuracy of responses by the participants of Experiment I and II to both words and non-words in the lexical decision task. Compared to listeners with a large vocabulary, listeners with a smaller vocabulary might recognise

https://doi.org/10.1017/S1366728919000646 Published online by Cambridge University Press

fewer words as such and/or erroneously accept more non-words as existing words.

The results presented above suggest that such an asymmetry does not hold; when presented in their natural form, 97.3% of /s/-final and 96.9% of /f/-final Dutch words were correctly identified by the emigrants as existing words. Recognition of the Dutch words was better, in fact, than that of the English words, which received correct lexical decisions in 88.6% of /s/-final and 93.2% of /f/-final cases (see Tables 4 and 6, Results section of Experiment II). Nonetheless, we fitted a generalised linear mixed-effects model to the lexical decision data, using family 'binomial' and the logit-link function. Responses to experimental items (words) as well as fillers (words and non-words) were included. Fixed predictors were Lexical status (words coded as -0.5, non-words coded as 0.5) and Listener group (emigrants coded as -0.5, controls as 0.5). Random intercepts for participants and items were included, as well as random slopes for Lexical status by participant, and for Listener group by item. Results of the model fit showed no significant difference between listener groups ($\beta = 0.071$, $p = .768$), nor any interaction between Listener group and Lexical status ($\beta = -0.411$, $p = .197$).

Another way to assess proficiency is of course via the LexTALE (Lemhöfer & Broersma, 2012) proficiency scores that all participants provided. These too provide no reason to suspect a proficiency gap. On average, the emigrants whose data were included in the analysis of the post-test received a proficiency score of 92.2%, compared to a mean score of 92.0% for all included control participants (an insignificant difference on an independent-samples t-test: $t(74) = 0.15$, $p = .883$). Thus emigrants and control participants seem to have had a similar amount of lexical information available to them to disambiguate the ambiguously pronounced fricative-final words they were exposed to.

Thus imperfect phonetic category match across languages, participant age and hearing ability, and participant proficiency levels offer no account of the lack of phonetic flexibility displayed by the Dutch emigrants in the present experiment. Note that we certainly assume our results to reflect absence of a flexibility that these participants once had. No perceptual learning study in any language has suggested that talker adaptation skills are absent in any population as a whole. Considering their age at migration, the emigrant participants had presumably been exposed to at least as many Dutch talkers as the control participants in Experiment I. The flexibility they presumably once held is however not on show in this study, and this is surely a result of the move out of the L1 environment and into an L2 environment.

But moving is not the whole story. Immersion in a particular language environment is, as noted in the Introduction, not a prerequisite for perceptual learning of this kind; category retuning has been found for the English L2 of Germans living in Germany (Schuhmann, 2014) and of Dutch residents of the Netherlands (Drozdova et al., 2016), and for the Mandarin L2 of English-speaking Australians (Cutler et al., 2018). Rather than immersion as such, we suggest that the ability to retune phonetic categories as a means of adapting to novel talkers actually requires a regular supply of novel talkers. And this is a language-specific requirement; regularly available novel English-talkers will not help maintain adaptability in Dutch. The Dutch emigrants in our study live in Australia and are not only likely to be exposed to large QUANTITIES of speech input in the immersion language, English, they are also likely to be exposed to a great VARIABILITY in that speech input, by conversing with both familiar and unfamiliar talkers on a daily basis. But as it is not a matter of where one lives, but of who one converses with, they need to keep up the supply of novel Dutch-talkers to maintain the readiness to retune the categories of that language.

Without exception, all the emigrant participants claimed to be fluent in Dutch (as well as English) and to use Dutch (as well as English) regularly; as described above, their Dutch proficiency indeed remained high. But with whom do they speak Dutch?

Evidence on this question was, in fact, available, since the background questionnaire filled in by the participants prior to embarking on the present series of experiments had included a set of seven questions concerning situations of use of the two languages. The tally of answers can be seen in Appendix G. It is immediately clear: all participants without exception reported talking Dutch more often with relatives, but English more often with friends. For those who had a partner, the language spoken with the partner was more often English than Dutch. For those who had a workplace, English was the language there, and never Dutch. Social situations outside the home likewise involved English and rarely Dutch. In short, these participants accrued greater regular variation in English, in many different situations and with much larger likely sets of interlocutors. When they spoke Dutch, in contrast, it was likely to be with family, which we presume to be a very much smaller set. Thus novel interlocutors were available in plenty in L2, while L1 interlocutors were not novel at all, but long known.

This asymmetry, we propose, is the underlying source of the remarkable absence of perceptually learned talker adaptation in these listeners' L1. The phonetic flexibility that supports such learning concerns the adjustment of phoneme category boundaries, and as phoneme categories themselves are language-specific, so, unsurprisingly, is the set of skills needed to adapt them. Thus keeping the adaptation skills ready for use requires practice for each language in which the skills are to be exercised. The emigrants in our study receive such practice in their L2, but not in their L1. With the L1 being principally spoken with familiar interlocutors, the skills in question are neither needed nor missed.

The previous findings of successful perceptual learning in L2 despite no immersion in an L2 environment suggest the availability of sufficient exposure to novel talkers in those cases also. For the Dutch undergraduates in Drozdova et al.'s (2016) study of L2 English, this is indeed likely; although English is not one of the Netherlands' official languages, it is omnipresent in Dutch society, especially in the media. English-spoken television programmes are not dubbed but subtitled, and thus provide a great amount of exposure to English, as do a number of English language television channels that are available to nearly all households, such as CNN and BBC. In addition, many university students in the Netherlands receive lectures in English. The participants tested by Schuhmann (2014) and by Cutler et al. (2018) must be assumed to have been similarly exposed to sufficient variability in their L2; note that in both cases, a proportion of the participants were taking classes in the L2.

Talker adaptation allows us to deal with inter-talker variability. Without variability, adaptation is unnecessary and not practised and, in consequence, the mechanisms of adaptation may decline, or be temporarily suspended, regardless of whether the language involved is the L1 or the L2. The precise amount of necessary practice to keep adaptation available, or the lower limits of it, cannot yet be known (and may even vary across individuals). However, if our hypotheses are correct then they suggest several predictions, as well as several interesting further avenues to explore.

First, we predict that the absence of adaptation flexibility due to lack of practice with novel talkers is likely to be reversed by making such practice available. We would assume that our

https://doi.org/10.1017/S1366728919000646 Published online by Cambridge University Press

emigrants would recover their L1 flexibility on every visit to their homeland, for instance, assuming that their visits involved sufficient talking in their L1 outside the family.

Second, we predict that relative maintenance of adaptation skills will apply to an individual's languages however many they are, with availability or non-availability holding for every language independently. Proficiency will not determine adaptation success.

Third, we predict that emigration is not the only situation that could trigger disuse of the ability to adapt in the L1. So-called heritage users, who typically also use their heritage language in family situations only and the language of the environment in every other situation, should also show a similarly asymmetric pattern. It should be even possible to suspend adaptation skills in one's L1 when no other language is known; consider (admittedly rare) situations such as those of hermits or troglodytes or members of religious communities or residents in isolated and distant small villages.

Further research should not only examine these other populations, where possible, but could also address the timing of skill acquisition and abandonment, the requisite quantity of practice and the parameters of adaptation skill restoration. Also of interest would be within-listener comparisons, such as we have reported here, in bilingual listeners whose L1 and L2 share fewer characteristics than do the closely related Dutch versus English; we do not yet know whether adaptation of phonetic category boundaries is sensitive to the number of (similar) categories in a language, nor do we know whether transfer of adaptation skills from one phonological system to another could be affected by dissimilarity between the L1 and L2 in number of contrasts, type of contrast, or syllabic positions in which contrasts may occur. Then, if a bilingual's two languages happened to have similar phoneme repertoires but were from quite different language families, would this allow some relative protection of adaptation skills against the effects of interlocutor scarcity?

Acknowledgements. Financial support was provided by a MARCS Institute doctoral fellowship to the first author. We thank Vincent Hoofs for assistance with data collection, and Sammie Tarenskeen and Anne Dwyer for recording the stimuli.

References

Baayen RH, Piepenbrock R and Gulikers L (1995) *The Celex lexical database (CD-rom)*.

Barr DJ, Levy R, Scheepers C and Tily HJ (2013) Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language* **68**, 255–278. doi:10.1016/j.jml.2012.11.001

Bates D, Maechler M, Bolker B and Walker S (2015) Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* **67**, 1–48. doi:10.18637/jss.v067.i01

Boersma P and Weenink D (2013) Praat: doing phonetics by computer (Version 5.3.42) [Computer software]. Retrieved from http://www.praat.org

Borrie SA, McAuliffe MJ, Liss JM, Kirk C, O'Beirne GA and Anderson T (2012) Familiarisation conditions and the mechanisms that underlie improved recognition of dysarthric speech. *Language and Cognitive Processes* **27**, 1039–1055. doi:10.1080/01690965.2011.610596

Bradlow AR and Bent T (2008) Perceptual adaptation to non-native speech. *Cognition* **106**, 707–729.

Broersma M and Cutler A (2011) Competition dynamics of second-language listening. *The Quarterly Journal of Experimental Psychology* **64**, 74–95. doi:10.1080/17470218.2010.499174

Burchfield LA, Luk SK, Antoniou M and Cutler A (2017) Lexically guided perceptual learning in Mandarin Chinese. *Proceedings of the 18th Annual Conference of the International Speech Communication Association (Interspeech 2017)*, pp. 576–580. Stockholm, Sweden.

Clarke CM and Garrett MF (2004) Rapid adaptation to foreign-accented English. *The Journal of the Acoustical Society of America* **116**, 3647–3658. doi:10.1121/1.1815131

Clyne M and Pauwels A (1997) Use, maintenance, structures, and future of Dutch in Australia. In Klatter-Folmer J & Kroon S (eds.), *Dutch Overseas: Studies in maintenance and loss of Dutch as an immigrant language*. Tilburg, the Netherlands: Tilburg University Press, pp. 34–49.

Collins B and Mees IM (2003) *The phonetics of English and Dutch*. Leiden, the Netherlands: Brill.

Cutler A, Burchfield LA and Antoniou M (2018) Factors affecting talker adaptation in a second language. In Epps J, Wolfe J, Smith J, & Jones C (eds.), *Proceedings of the 17th Australasian International Conference on Speech Science and Technology*, pp. 33–36. Sydney, Australia.

Dahan D, Drucker SJ and Scarborough RA (2008) Talker adaptation in speech perception: Adjusting the signal or the representations? *Cognition* **108**, 710–718. doi:10.1016/j.cognition.2008.06.003

Drozdova P, Van Hout R and Scharenborg O (2016) Lexically-guided perceptual learning in non-native listening. *Bilingualism: Language and Cognition* **19**, 914–920. doi:10.1017/S136672891600002X

Eisner F and McQueen JM (2006) Perceptual learning in speech: Stability over time. *The Journal of the Acoustical Society of America* **119**, 1950–1953. doi:10.1121/1.2178721

Evans BG and Iverson P (2004) Vowel normalization for accent: An investigation of best exemplar locations in northern and southern British English sentences. *The Journal of the Acoustical Society of America* **115**, 352–361. doi:10.1121/1.1635413

Forster KI and Forster JC (2003) DMDX: A Windows display program with millisecond accuracy. *Behavior Research Methods, Instruments, & Computers* **35**, 116–124. doi:10.3758/BF03195503

Garcia Lecumberri ML, Cooke M and Cutler A (2010) Non-native speech perception in adverse conditions: A review. *Speech Communication* **52**, 864–886. doi:10.1016/j.specom.2010.08.014

Kraljic T and Samuel AG (2005) Perceptual learning for speech: Is there a return to normal? *Cognitive Psychology* **51**, 141–178. doi:10.1016/j.cogpsych.2005.05.001

Kraljic T and Samuel AG (2007) Perceptual adjustments to multiple speakers. *Journal of Memory and Language* **56**, 1–15. doi:10.1016/j.jml.2006.07.010

Lemhöfer K and Broersma M (2012) Introducing LexTALE: A quick and valid Lexical Test for Advanced Learners of English. *Behavior Research Methods* **44**, 325–343. doi:10.3758/s13428-011-0146-0

MATLAB (Version R2013b). [Computer software]. Natick, MA, United States: The MathWorks, Inc.

Maye J, Aslin RN and Tanenhaus MK (2008) The weckud wetch of the wast: Lexical adaptation to a novel accent. *Cognitive Science* **32**, 543–562. doi:10.1080/03640210802035357

McQueen JM, Cutler A and Norris D (2006) Phonological abstraction in the mental lexicon. *Cognitive Science* **30**, 1113–1126. doi:10.1207/s15516709cog0000_79

McQueen JM, Tyler MD and Cutler A (2012) Lexical retuning of children's speech perception: Evidence for knowledge about words' component sounds. *Language Learning and Development* **8**, 317–339. doi:10.1080/15475441.2011.641887

Mitterer H, Chen Y and Zhou X (2011) Phonological abstraction in processing lexical-tone variation: Evidence from a learning paradigm. *Cognitive Science* **35**, 184–197. doi:10.1111/j.1551-6709.2010.01140.x

Mitterer H and McQueen JM (2009) Foreign subtitles help but native-language subtitles harm foreign speech perception. *PLoS ONE* **4**, e7785. doi:10.1371/journal.pone.0007785

Mitterer H and Reinisch E (2013) No delays in application of perceptual learning in speech recognition: Evidence from eye tracking. *Journal of Memory and Language* **69**, 527–545. doi:10.1016/j.jml.2013.07.002

Norris D, McQueen JM and Cutler A (2003) Perceptual learning in speech. *Cognitive Psychology* **47**, 204–238. doi:10.1016/S0010-0285(03)00006-9

Ockey GJ and French R (2016) From one to multiple accents on a test of L2 listening comprehension. *Applied Linguistics* **37**, 693–715. doi:10.1093/applin/amu060

R Core Team. (2018) R: A language and environment for statistical computing (Version 3.5.1.) [Computer Software]. Vienna, Austria: R

https://doi.org/10.1017/S1366728919000646 Published online by Cambridge University Press

Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

Ramscar M, Hendrix P, Shaoul C, Milin P and Baayen H (2014) The myth of cognitive decline: Non-linear dynamics of lifelong learning. *Topics in Cognitive Science* **6**, 5–42. doi:10.1111/tops.12078

Reinisch E, Weber A and Mitterer H (2013) Listeners retune phoneme categories across languages. *Journal of Experimental Psychology: Human Perception and Performance* **39**, 75–86. doi:10.1037/a0027979

Remez RE, Rubin PE, Pisoni DB and Carrell TD (1981) Speech perception without traditional speech cues. *Science* **212**, 947–950.

Repp BH (1981) Perceptual equivalence of two kinds of ambiguous speech stimuli. *Bulletin of the Psychonomic Society* **18**, 12–14. doi:10.3758/bf03333556

Scharenborg O and Janse E (2013) Comparing lexically guided perceptual learning in younger and older listeners. *Attention, Perception, & Psychophysics*, 1–12. doi:10.3758/s13414-013-0422-4

Scharenborg O, Weber A and Janse E (2015) The role of attentional abilities in lexically guided perceptual learning by older listeners. *Attention, Perception, & Psychophysics* **77**, 493–507. doi:10.3758/s13414-014-0792-2

Schuhmann KS (2014) Perceptual learning in second language learners. Unpublished doctoral dissertation, Stony Brook University.

Sebastián-Gallés N, Echeverría S and Bosch L (2005) The influence of initial exposure on lexical representation: Comparing early and simultaneous bilinguals. *Journal of Memory and Language* **52**, 240–255. doi:10.1016/j.jml.2004.11.001

Shannon RV, Zeng F.-G., Kamath V, Wygonski J and Ekelid M (1995) Speech recognition with primarily temporal cues. *Science* **270**, 303–304. doi:10.1126/science.270.5234.303

Trude AM, Tremblay A and Brown-Schmidt S (2013) Limitations on adaptation to foreign accents. *Journal of Memory and Language* **69**, 349–367. doi:10.1016/j.jml.2013.05.002

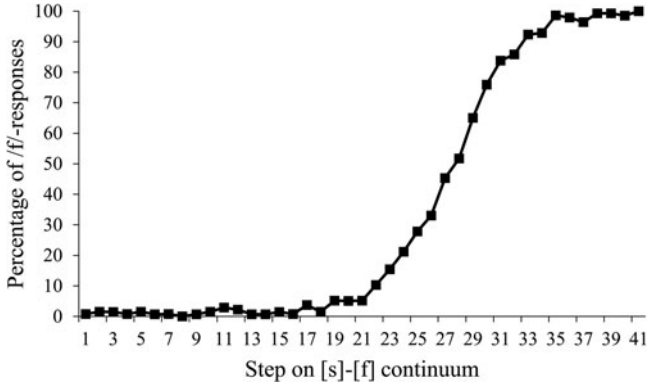
Warren RM (1970) Perceptual restoration of missing speech sounds. *Science* **167**, 392–393.

Weber A and Cutler A (2004) Lexical competition in non-native spoken-word recognition. *Journal of Memory and Language* **50**, 1–25. doi:10.1016/S0749-596X(03)00105-0

Witteman MJ, Weber A and McQueen JM (2013) Foreign accent strength and listener familiarity with an accent codetermine speed of perceptual adaptation. *Attention, Perception, & Psychophysics*, 1–20. doi:10.3758/s13414-012-0404-y

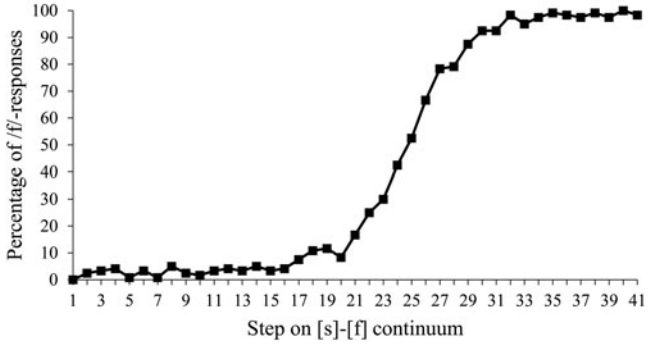
Appendix A

Mean percentage of /f/-responses for the Dutch [s]-[f] continuum in a pilot with 20 native listeners of Dutch.



Appendix B

Mean percentage of /f/-responses for the English [s]-[f] continuum in a pilot with 20 native listeners of Australian English.



Appendix C

Dutch fricative-final words used in Experiments I and II.

/f/-final	/s/-final
aanhef (beginning)	ananas (pineapple)
achterneef (great nephew)	anijs (aniseed)
autokerkhof (wrecking yard)	appelmoes (apple sauce)
braaf (honest)	bekentenis (confession)
doolhof (maze)	bordes (steps)
dreigbrief (threatening letter)	collectebus (collecting box)
educatief (educational)	doctorandus (doctoral candidate)
hartedief (darling)	gedachteloos (thoughtless)
landbouwbedrijf (farm)	geitenkaas (goat's cheese)
loopgraaf (trench)	gemeentehuis (town hall)
middenrif (diaphragm)	hagedis (lizard)
onderlijf (lower body)	hakmes (machete)
ongeloof (disbelief)	krijs (scream)
operatief (surgically)	lerares (teacher)
ophef (fuss)	matroos (sailor)
praalgraf (mausoleum)	naaldbos (pine forest)
proef (test)	nis (niche)
recreatief (recreational)	pedagogisch (pedagogical)
rentetarief (interest rate)	pimpelmees (blue tit)
tortelduif (turtle dove)	relaas (account)

https://doi.org/10.1017/S1366728919000646 Published online by Cambridge University Press

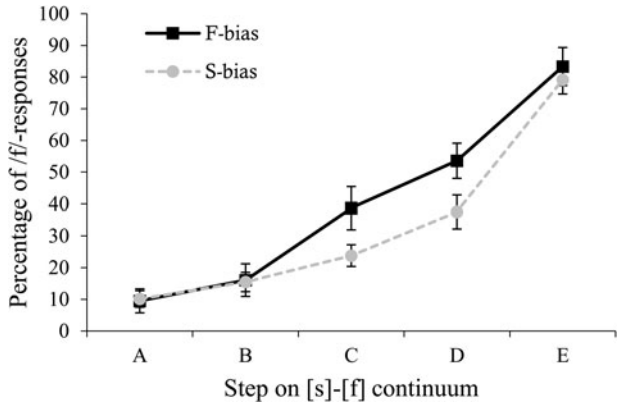
Appendix D

English fricative-final words used in Experiment II.

/f/-final	/s/-final
aloof	across
autograph	awareness
beef	bias
behalf	bliss
belief	crease
bullet-proof	dress
dandruff	embrace
debrief	entice
dwarf	eyewitness
earmuff	harass
enough	hideous
handcuff	immerse
laugh	necklace
loaf	noose
midwife	pierce
plaintiff	promise
proof	remorse
rough	replace
whooping-cough	twice
wildlife	unless

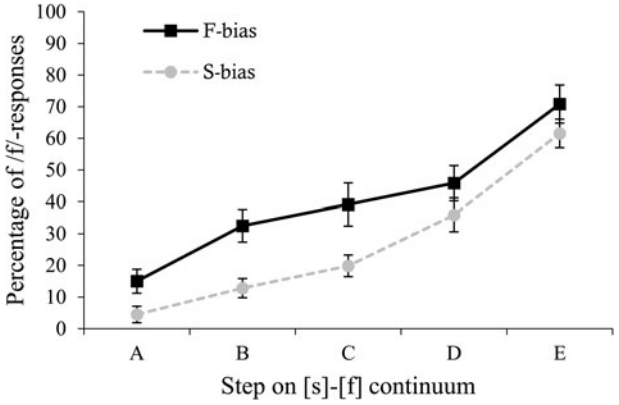
Appendix E

Mean percentage of /f/-responses for the post-test of Experiment I for listeners in the f-bias and s-bias groups of the Personalised condition. Error bars represent standard errors. Step A is the most /s/-like of the tested ambiguous stimuli, while step E is the most /f/-like.



Appendix F

Mean percentage of /f/-responses for the post-test of Experiment I for listeners in the f-bias and s-bias groups of the Standard condition. Error bars represent standard errors. Step A is the most /s/-like of the tested ambiguous stimuli, while step E is the most /f/-like.



Appendix G

The tally of answers to the question "Please indicate to what extent you use Dutch and English in the situations listed" by the 30 participants whose data were included in the analyses of Experiment II. Data of 25 participants were included in both language versions. Data of a further four participants were included in the L1 version only, whereas data for one other participant were included in the L2 version only.

I speak Dutch	always	often	regularly	every now and then	never
With my partner	6	1	3	8	10
With relatives	17	2	3	6	2
With friends	4	2	6	12	6
At work	0	0	4	5	16
In church	0	0	0	1	12
In shops	1	0	0	3	25
At clubs/ organisations	0	1	2	5	18
I speak English	**always**	**often**	**regularly**	**every now and then**	**never**
With my partner	14	5	5	2	2
With relatives	7	4	4	5	10
With friends	11	10	7	1	1
At work	21	3	2	0	0
In church	7	0	0	1	6
In shops	28	2	0	0	0
At clubs/ organisations	20	2	3	0	1

https://doi.org/10.1017/S1366728919000646 Published online by Cambridge University Press