

# Robustness of phonolexical representations relates to phonetic flexibility for difficult second language sound contrasts\*

MIQUEL LLOMPART  
*Ludwig Maximilian University Munich*  
EVA REINISCH  
*Ludwig Maximilian University Munich*

(Received: October 09, 2017; final revision received: July 31, 2018; accepted: August 01, 2018)

*Listening to speech entails adapting to vast amounts of variability in the signal. The present study examined the relationship between flexibility for adaptation in a second language (L2) and robustness of L2 phonolexical representations. Phonolexical encoding and phonetic flexibility for German learners of English were assessed by means of a lexical decision task containing nonwords with sound substitutions and a distributional learning task, respectively. Performance was analyzed for an easy (/i/-/ɪ/) and a difficult contrast (/e/-/æ/, where /æ/ does not exist in German). Results showed that for /i/-/ɪ/ listeners were quite accurate in lexical decision, and distributional learning consistently triggered shifts in categorization. For /e/-/æ/, lexical decision performance was poor but individual participants' scores related to performance in distributional learning: the better learners were in their lexical decision, the smaller their categorization shift. This suggests that, for difficult L2 contrasts, rigidity at the phonetic level relates to better lexical performance.*

Keywords: second language learning, speech perception, lexical representations, phonetics, flexibility

Learning a second language (L2) includes building a non-native lexicon, which usually comes with certain difficulties, especially if the L2 is learned later in life. The learning circumstances for the L2 are rather different from how the first, native language (L1) is acquired (Gathercole, Hitch, Service & Martin, 1997; Kuhl, Williams, Lacerda, Stevens & Lindblom, 1992; Werker & Tees, 1984, among others). For many late learners, much of the learning occurs in a formal setting (i.e., classroom) and/or in an environment in which the L2 is rarely spoken (Darcy, Daidone & Kojima, 2013; Díaz, Mitterer, Broersma & Sebastián-Gallés, 2012; Weber & Cutler, 2004). Consequences of this limited experience with the L2 can be observed as deficits at virtually all levels of language processing, including the non-native lexicon. Vocabulary size in an L2 tends to be smaller than in the L1 (Bundgaard-Nielsen, Best & Tyler, 2011), as fewer words reach full integration into

long-term memory (Gollan, Montoya, Cera & Sandoval, 2008). In addition, the impoverished input that learners receive causes them to struggle with building robust phonolexical representations for newly-learned words. L2 phonolexical representations have repeatedly been found to lack relevant phonetic detail: that is, phonetic categories are ‘fuzzily’ or ‘weakly’ encoded into these representations. This is generally reflected in slower and more error-prone word recognition in the L2 compared to the L1 (Cook, 2012; Cook & Gor, 2015; Cook, Pandzda, Lancaster & Gor, 2016; Lancaster & Gor, 2016).

An additional obstacle to establishing accurate phonolexical representations for words in the L2 is that the late learners' perceptual system is already optimized for perceiving the sound contrasts relevant to their L1 (Polka & Werker, 1994; Werker & Tees, 1984). The perception of any later-learned language is therefore modulated by the categories of the L1 phonetic inventory, as detailed, for instance, in L2 sound learning models (Best & Tyler, 2007; Flege, 1995). However, when learning an L2, new phonetic categories need to be established. These new categories may give rise to sound contrasts that are not present in the L1, which often results in perceptual difficulties that are hard to overcome. A case in point is the struggle of German learners of English with the contrast between the open-mid front vowel /e/ and the near open front vowel /æ/. According to the predictions of

\* This project was funded by a grant from the German Research Foundation (DFG; grant nr. RE 3047/1-1) to the second author. This work is part of the first author's Ph.D. project. Parts of this work were presented at the 23rd Annual Conference on Architectures and Mechanisms for Language Processing, 2017, in Lancaster, UK. We would like to thank Rosa Franzke for her help with testing participants and Christopher Carignan for comments on a previous version of the manuscript.

Address for correspondence:

Miquel Llompart, Institute of Phonetics and Speech Processing, Ludwig Maximilian University Munich, Schellingstraße 3, 80799 Munich, Germany.

[M.Llompart@phonetik.uni-muenchen.de](mailto:M.Llompart@phonetik.uni-muenchen.de)

L2 learning models, this should be a ‘difficult’ distinction because German has only one open-mid vowel /ɛ/ (similar to English /ɛ/). /æ/ does not exist in German and German /ɛ/ is its closest L1 category. German learners of English hence tend to associate both English /ɛ/ and /æ/ to their L1 /ɛ/ category. Consequently, at least in the initial stages of learning, they have considerable difficulties in teasing the two vowels apart (Bohn & Flege, 1990; Eger & Reinisch, 2017, 2018; Flege, Bohn & Jang, 1997).

Crucially, to master the /ɛ/-/æ/ distinction, Germans not only need to establish two separate L2 categories corresponding to /ɛ/ and /æ/ (i.e., phonetic encoding) so that they can perceptually identify the two sounds, they also have to accurately encode these two phonetic categories into the representations of words in their L2 lexicon (i.e., phonolexical representations; Cook et al., 2016; Cook & Gor, 2015). For instance, /ɛ/ has to be encoded into the representations of words like *lemon*, and /æ/ has to be encoded into those of words like *dragon* – and not the other way around. This encoding of phonetic categories in phonolexical representations has been shown to be difficult for L2 learners, above and beyond the phonetic identification of L2 sound categories. Learners perform poorly in tasks that involve the recognition of words containing sounds in difficult L2 contrasts even when their phonetic categorization of these contrasts is already within the native range (Amengual, 2016; Darcy et al., 2013; Díaz et al., 2012).

A likely explanation for the difficulties in lexical tasks is that the contrast is still not well encoded in the learners’ L2 phonolexical representations. In particular, evidence from eye-tracking (Escudero, Hayes-Harb & Mitterer, 2008; Llompart & Reinisch, 2017; Weber & Cutler, 2004) suggests that the L2 category that is a worse fit to the native category is less reliably encoded in L2 words than the better fitting counterpart. This results in more errors with words containing the former (Broersma, 2005; Díaz et al., 2012; Simon, Sjerps & Fikkert, 2014). In sum, phonolexical representations containing sounds involved in difficult L2 contrasts not only lack phonetic detail due to limited exposure to the L2, but are also fuzzier than those in the L1 lexicon because of perceptual problems with specific L2 categories (Broersma, 2012; Cutler, Weber & Otake, 2006; Darcy, Dekydtspotter, Sprouse, Glover, Kaden, McGuire & Scott, 2012; see Cutler, 2015, for a review).

In the L2, as well as in the L1, sound category learning and phonolexical encoding needs to be accomplished while being exposed to variability coming from manifold sources, such as physiological differences (Ladefoged & Broadbent, 1957), signal distortions like speech in noise (Cooper, Brouwer & Bradlow, 2015), dialectal variation (Llompart & Simonet, 2018; Sumner & Samuel, 2009), foreign accents (Bradlow & Bent, 2008; Sidaras, Alexander & Nygaard, 2009; Weber, Di Betta &

McQueen, 2014; Witteman, Weber & McQueen, 2013), and personal idiosyncrasies (Norris, McQueen & Cutler, 2003), to name but a few examples. In the native language, listeners are able to make up for such variability by flexibly adapting their perception to the input they receive. For instance, Norris et al. (2003) conducted a lexically-guided perceptual learning study on the /s/-/f/ contrast in Dutch. They presented an ambiguous sound between /s/ and /f/ embedded in either /s/-final or /f/-final words and subsequently examined the listeners’ categorization of that contrast. They found that listeners who heard the ambiguous sound in the /s/ context reported hearing more /s/ during categorization, and the opposite was true for those exposed to the same fricative in an /f/ context. In short, listeners used lexical knowledge to shift the boundary between the two phonetic categories to accommodate the deviant productions of the fricative sound.

Importantly, L2 learners have also been found to be flexible in their perception of at least some sound contrasts in the L2 – specifically sound contrasts that do not pose perceptual difficulties for them (Drozdova, van Hout & Scharenborg, 2016; Reinisch, Weber & Mitterer, 2013; Schuhmann, 2014; but see Bruggeman, 2016). For example, Reinisch et al. (2013) conducted an experiment similar to Norris et al. (2003) on the same Dutch /s/-/f/ contrast, but with German learners of Dutch as well as native Dutch speakers. They found that both groups exhibited similar shifts in categorization after perceptual adaptation. L2 learners were hence as able to adapt their perception of the Dutch /s/-/f/ contrast as L1 listeners. Note that /s/-/f/ is a sound contrast that is also part of the German inventory and German learners of Dutch are expected not to have difficulties with that distinction.

By contrast, not as much is known about how flexible L2 learners are with sounds involved in difficult L2 contrasts (i.e., contrasts in which sounds are perceptually confusable) and whether (and how) such flexibility may be related to the learners’ performance in these distinctions. Critically, some evidence suggests that unlike with easy L2 distinctions, flexibility or adaptability when it comes to difficult non-native categories may be problematic (Sjerps & McQueen, 2010). Sjerps and McQueen found that Dutch learners of English learned to treat the difficult English sound /θ/ (Cutler, Weber, Smits & Cooper, 2004; Hanulikova & Weber, 2011) as a token of /f/ or /s/ in a perceptual learning study. Exposure to /θ/ in contexts where it replaced /f/ or /s/ triggered similar adjustments of /f/-to-/s/ category boundaries as the artificially constructed ambiguous sound between /f/ and /s/ in previous studies (e.g., Norris et al., 2003). That is, the learners’ difficulty with the L2 category /θ/ resulted in ‘unnecessary’ or even ‘unwanted’ adaptation effects in the L1. While Sjerps and McQueen showed how a difficult L2 category can impact the perception of an L1 contrast,

it still remains unclear how flexible L2 learners are in their perception of difficult L2 contrasts. Moreover, it remains to be shown whether flexibility (or lack thereof) for difficult L2 contrasts may relate to the learners' mastery of the contrast, and in particular to how robustly that contrast is encoded into words in the L2 lexicon.

The present study addresses these questions with German learners of English by examining the phonolexical encoding and phonetic categorization of the difficult /ɛ/-/æ/ contrast and a contrast that is expected to be 'easy' for this group of learners, /i/-/ɪ/. German and English have a similar tenseness contrast for high front vowels. Therefore, native German speakers are predicted to straightforwardly associate each of the two English vowels with its German counterpart and treat this L2 contrast much like a native contrast (Best & Tyler, 2007; Flege, 1995). Indeed, experimental evidence has shown that the perceptual match between the English and German categories is very robust (Iverson & Evans, 2007) and that German learners do not have difficulties in distinguishing these two English vowels (Bohn & Flege, 1990, 1992). Therefore this contrast was used as a within-language, within-participant baseline to the difficult /ɛ/-/æ/ contrast. The robustness of phonolexical encoding of the two L2 contrasts was assessed using a lexical decision task containing mispronunciations, while phonetic flexibility was investigated by means of a distributional learning task.

In the lexical decision task, learners were presented with real words that contained sounds from the contrasts /i/-/ɪ/ and /ɛ/-/æ/ as well as nonwords created by replacing the critical sounds by the other member of the contrast (Broersma, 2005; Díaz et al., 2012; Sebastián-Gallés, Echeverría & Bosch, 2005). We took the learners' ability to accept the real words and reject the mispronounced nonwords as an indicator of how well the critical vowels are phonologically represented in their L2 lexicon. Learners were expected to be substantially more accurate with items containing /i/ and /ɪ/ than with items containing /ɛ/ and /æ/ because, for them, the latter two categories are perceptually confusable at the phonetic level. In addition, based on previous findings with similar tasks and Dutch speakers (Broersma, 2005; Díaz et al., 2012; Simon et al., 2014), we also expected that, within the difficult contrast, learners would have more difficulties with nonwords in which /æ/ was mispronounced as [ɛ] (e.g., \*dr[ɛ]gon) than with nonwords in which /ɛ/ was mispronounced as [æ] (e.g., \*l[æ]mon) because of the dominant role of /ɛ/ in the contrast. Crucially, measures derived from the learners' performance in the lexical decision task were compared to performance in the distributional learning task, so as to relate their robustness of phonolexical encoding to their perceptual flexibility at the phonetic level.

Phonetic flexibility was measured by means of a distributional learning task. This task consisted of two

parts in which learners were asked to categorize an /i/-/ɪ/ and an /ɛ/-/æ/ continuum. In each part, categorization phases were always preceded by exposure phases designed to bias perception towards one of the response options during categorization. Listeners were expected to show adaptation to the regularities in continuum step presentation during exposure and shift their category boundaries accordingly (Clayards, Tanenhaus, Aslin & Jacobs, 2008; Escudero, Benders & Wanrooij, 2011; Kleinschmidt, Raizada & Jaeger, 2015; Munson, 2011). By comparing the location of category boundaries following the different types of exposure, a measure was obtained of how flexible the L2 learners' perception of each contrast was (see Munson, 2011, for a detailed discussion of boundary-shifting distributional learning).

Distributional learning was used because it allowed us to examine how L2 listeners shifted phonetic category boundaries in an unsupervised manner. Contrary to other perceptual learning paradigms, we did not provide our listeners with explicit information about where the category boundaries should be located (see Lametti, Krol, Shiller & Ostry, 2014 for an explicit boundary-learning task) or other types of information such as lexical or visual context to anchor phonetic categories (Bertelson, Vroomen & de Gelder, 2003; Norris et al., 2003; Reinisch & Mitterer, 2016). Moreover, given that our population were L2 learners, unsupervised distributional learning seemed more appropriate than, for example, training involving lexical information (e.g., as in Reinisch et al., 2013) because distributional learning mostly relies on phonetic perception. The minimal involvement of the L2 lexicon in this task allowed us to avoid confounds in the observed effects due to differences in general L2 lexical knowledge among learners.

Our main goal was to assess how the phonolexical encoding of sound categories in the L2 relates to phonetic flexibility in easy and difficult L2 sound contrasts, for L2 learners as a group but, critically, also within the individual learners. For the easy /i/-/ɪ/ contrast, we expected learners to show flexibility because both their phonetic perception and phonolexical encoding should be quite robust, and it has already been shown that boundaries between sounds in easy contrasts can be shifted (Reinisch et al., 2013). Moreover, no individual differences due to L2 experience and/or proficiency were expected because these factors do not appear to strongly modulate the learners' perception of easy L2 contrasts (Bohn & Flege, 1990).

Regarding /ɛ/-/æ/, however, two main outcomes can be predicted: if flexibility comes with mastery of the contrast, learners with more robust phonolexical representations of words containing the sounds of the difficult L2 contrast (i.e., better in lexical decision) will also be those who are more flexible at the phonetic category level. Learners with fuzzier phonolexical representations (i.e., worse in

lexical decision) should adapt their boundaries less during distributional learning. The alternative is that learners with more robust phonological representations shift less in the distributional learning task for /ɛ/-/æ/ than those exhibiting poorer phonological encoding. This is because new phonetic categories that are involved in difficult L2 contrasts differ from other categories in that their establishment is always effortful (to name but a few, see Aoyama, Flege, Guion, Akahane-Yamada & Yamada, 2004; Flege et al., 1997; Goto, 1971; Sheldon & Strange, 1982). These difficulties may impact the perception of the /ɛ/-/æ/ contrast so that the accurate distinction of the two sounds comes at the cost of flexibility.

## Methods

### Participants

Forty-two native speakers of German (21 females; mean age = 24.93, *sd* = 4.3), all students at the University of Munich, participated for a small payment. None of them reported hearing problems. Recruitment requirements were that (i) they had not learned any language other than German before starting to learn English (mean age = 9.61, *sd* = 2.21, range 4–12), (ii) had not spent more than 6 months in an English-speaking country, and (iii) were not enrolled in a language-program at the university. Two participants were discarded from all analyses because they did not fulfill these requirements. All participants filled out a language background questionnaire assessing a number of self-estimated measures of English proficiency and experience. From these measures, following Eger and Reinisch (2017), we calculated individual Proficiency/Experience Scores that were the mean of 5 parameters: self-rated comprehension and speaking skills, self-reported frequency of exposure and use of spoken English and self-estimated German accent in English. All measures were provided on a 1-to-7 scale in which 1 indicated good skills, frequent exposure/use and weak foreign accent, and 7 indicated the opposite. Table 1 shows the mean values for these five metrics as well as for the mean Proficiency/Experience Scores.

### Materials

A total of 304 English words was used for the lexical decision task. Words were mono-, di- or trisyllabic nouns, adjectives and verbs, and were selected to be commonly known to learners of English. 104 of these words belonged to the critical sound contrasts (i.e., /i/-/ɪ/ and /ɛ/-/æ/), 26 words per sound. The remaining 200 words were fillers involving the contrasts /p/-/t/, /k/-/m/, /b/-/v/, /ɔ:/-/u/. Filler contrasts were expected to be easy for our participants because they also exist in German.

Table 1. Mean self-estimated English Proficiency/Experience measures for the German learners of English who participated in the study. Values close to 1 indicate good skills, frequent exposure/use and weak foreign accent; values close to 7 indicate poor skills, infrequent exposure/use and strong foreign accent.

Self-reported measure	M	sd
English comprehension skills	2.28	0.86
English speaking skills	3.15	1.11
Exposure to spoken English	2.32	1.2
Use of spoken English	4.03	1.72
German accent in English	4	1.96
Mean Proficiency/Experience Score in English	3.15	0.87

Importantly, half of the words were selected to be used as nonwords. Nonwords were created by exchanging the two sounds in each contrast (e.g., \*l[æ]mon for *lemon* and \*dr[ɛ]gon for *dragon*; the same for /i/-/ɪ/ and the fillers). That is, for each critical sound (/i/, /ɪ/, /ɛ/, /æ/), 13 words appeared as real words and 13 as nonwords containing a single-sound mispronunciation (see Appendix A for a list of words and nonwords for the two contrasts). For the critical contrasts, the mispronunciations always concerned the first stressed vowel of the words. For filler contrasts, the position of the mispronunciation varied such that participants could not adopt the strategy of focusing exclusively on the first two segments of the words. Lexical frequency was controlled within each contrast using the Zipf scale measures provided by Subtlex-UK (van Heuven, Mandera, Keuleers & Brysbaert, 2014, see also Zipf, 1949). Lexical frequency for words of the two critical contrasts was comparable: /i/-/ɪ/ (4.63–4.55), /ɛ/-/æ/ (4.66–4.61).

For the distributional learning task, four English minimal pairs for each critical sound contrast were selected for recording (/i/-/ɪ/: *cheap-chip*, *heat-hit*, *seat-sit*, *sheep-ship*; /ɛ/-/æ/: *bed-bad*, *bet-bat*, *pen-pan*, *said-sad*).

### Recordings

All words were recorded in a sound-attenuated booth by a 26-year-old native speaker of Standard Southern British English who lived in London until moving to Munich at the age of 22. All words for the lexical decision task were recorded in their correct word form and items designated to function as nonwords were additionally recorded with the critical sound substituted by the other member of the contrast. Care was taken that the substitutions sounded natural. Table 2 shows the mean F1, F2 and duration



Table 2. Mean formant values (F1, F2; in Hertz) for Real Words, Base Forms and Mispronounced Nonwords for the /i/-/ɪ/ and /ɛ/-/æ/ contrasts in the lexical decision task. Standard deviations are in parentheses.

Sound	F1			F2		
	Real Words	Base Forms	Mispronounced Nonwords	Real Words	Base Forms	Mispronounced Nonwords
/i/	256 (11)	258 (12)	374 (57)	2357 (53)	2361 (61)	1944 (165)
/ɪ/	430 (80)	438 (131)	272 (30)	1944 (152)	1997 (196)	2375 (75)
/ɛ/	579 (34)	584 (36)	820 (41)	1689 (215)	1664 (228)	1370 (137)
/æ/	815 (41)	795 (50)	573 (37)	1506 (68)	1451 (71)	1760 (67)

values of the real words with /i/, /ɪ/, /ɛ/ and /æ/ (e.g., *record*, *happy*), the real words that were the base for the mispronounced nonwords (i.e., Base Forms; e.g., *lemon* and *dragon*) and the mispronounced nonwords (e.g., \*l[æ]mon, \*dr[ɛ]gon). From these values, it can be seen that the native British speaker produced a large difference between the critical vowels of the two contrasts in both F1 and F2 and that values for mispronounced items were comparable to values for real words of the intended category<sup>1</sup>.

Minimal pairs for the distributional learning task were recorded multiple times. The minimal pairs *sheep-ship* and *bet-bat* were selected for use in the experiment because their consonants are unproblematic for German learners of English, unlike, for instance, the final -d in *bed-bad* due to final devoicing in German (Fourakis & Iverson, 1984; Port & O'Dell, 1985). One recording of each word was chosen so that the two words in each pair were as similar as possible in recording quality, *f0* contour and perceived speech rate. However, since it appeared that the vowels in *bet* and *bat* were consistently produced in creaky voice, we decided to replace the words' onsets (/b/ and vowel) with recordings from tokens of *bed* and *bad* that appeared less creaky. Critically, the vowel duration was shortened to match the original durations of *bet* and *bat* from which the final stops were taken. No splicing was needed for the selected tokens of *sheep* and *ship*.

Two 21-step continua, one between *sheep* and *ship* and the other between *bet* and *bat*, were created through duration manipulation and formant shifting in *Praat* (Boersma & Weenink, 2010). Vowel duration and F1 and F2 values for the endpoints were taken directly from the naturally-produced tokens and continuum steps were set to change linearly in the three dimensions (see Table 3 for endpoint values). The 21-step continua were then

Table 3. Duration and F1 and F2 values of endpoint vowels of the two continua used in the distributional learning task.

Endpoint	Duration (ms)	F1 (Hz)	F2 (Hz)
Sheep (/i/)	235	266	2354
Ship (/ɪ/)	120	440	1944
Bet (/ɛ/)	140	541	1804
Bat (/æ/)	180	841	1477

reduced to 11 steps leaving out every other step (i.e., steps 0, 2, 4, etc.) and pretested on 7 native speakers of German so as to find the perceptually most ambiguous steps. Participants categorized each continuum step 10 times and results showed that step 12 was closest to the category boundary for both contrasts. The final continua were built around this most ambiguous step (see Figure 1). This centering of the continuum was necessary to ensure that, when exposed to the distributions of stimuli, one of the peaks was always located close to the most ambiguous region for all participants (see Procedure below); pretests on native German speakers had shown that only then could distributional learning be observed.

### Procedure

All participants completed the two tasks, lexical decision and distributional learning, in two experimental sessions that were separated by at least one and no more than three weeks. The order was consistent, with lexical decision preceding distributional learning. Participants were tested individually in a sound-attenuated booth. Experiments were run using *Psychopy2* software (v.1.83.01; Peirce, 2007). Auditory stimuli were presented binaurally over headphones at a comfortable listening level.

### Lexical decision

For the lexical decision task, individual presentation lists were built with all items appearing in full random order

<sup>1</sup> A pilot test of the lexical decision task with three native speakers of English indicated that they were around 95% correct on the critical contrasts (/i/-/ɪ/: 96.15%; /ɛ/-/æ/: 95.51%), including nonword rejection (/i/-/ɪ/: 94.87%; /ɛ/-/æ/: 94.87%), replicating thus previous findings with this population in similar tasks (Broersma, 2005; Díaz et al., 2012).

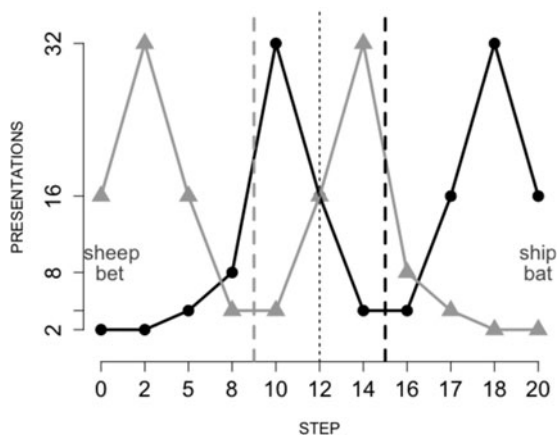


Figure 1. Number of presentations per continuum step in each of the two distributions presented in the distributional learning task. The distribution presented in exposure phases 1 and 3 is in black and the distribution presented in exposure phase 2 is in grey. Dashed lines depict the intended boundary for each distribution. The dotted line signals the most ambiguous step for German listeners as determined by the pretest.

except that no two items of the same contrast could follow one another. Importantly, the sets of items presented as real words and as nonwords with mispronunciations were the same for all participants, so as to facilitate comparisons at the individual level. Participants were told (in English) that they would be hearing real English words and invented words that could sound similar to English words. They were instructed to press the “1” key on the computer keyboard if they considered the presented item to be a real English word and the “0” key if they thought it was not an existing word. They were asked to respond as fast as they could but being as accurate as possible. On every trial, two small boxes appeared on the computer screen: a green one, on the left-hand side, with *word* written on it, and a red one, on the right-hand side, with *not a word* written on it. After the button press, the selected box moved slightly upwards and the other box disappeared for 300ms. This was used to indicate to participants that their response had been stored. There was no time limit for responses and presentation of the next trial started 1.1s after the previous button press. Listeners were given ten practice trials (i.e., 5 words and 5 nonwords) before they started the experiment. None of the practice trials included words or nonwords containing the sounds of the critical contrasts. The lexical decision task took between 15 and 20 minutes to complete.

### Distributional learning

The distributional learning task was divided into two parts. The first part was on the /i/-/ɪ/ contrast and the second on /ɛ/-/æ/ (i.e., order: easy-difficult). Each part consisted

of three exposure-test sequences. During the exposure phases, listeners were asked to listen to randomized orders of tokens sampled from the continuum. The frequency with which each token was sampled was chosen such as to bias listeners’ perception towards one of the endpoint categories. A visual schematization of this sampling is shown in Figure 1. The distributions for exposure phases 1 and 3 were the same (black line), and were expected to trigger more /i/ and /ɛ/ responses, respectively. Exposure phase 2 was expected to result in more /ɪ/ and /æ/ responses (grey line). The third phase was included to test whether listeners would be able to shift their boundaries back to where they were after the first exposure once they had been exposed to the second distribution, which should produce the opposite biasing effect. The two distributions mirrored each other and always contained the same number of tokens at both sides of the intended (shifted) boundary (dashed lines in Figure 1).

Immediately following each exposure phase, listeners were prompted to perform a short categorization task on the selected 11 steps of the continuum they had just heard. Steps (henceforth referred to as going from 1 to 11) were presented 5 times each in fully random order. Participants were instructed to press “1” when the word they heard corresponded to a picture on the left-hand side of the screen (e.g., sheep) and “0” when the word was the one depicted on the right-hand side of the screen (e.g., ship). Orthographic representations of the words (SHEEP-SHIP and BET-BAT) were provided before the first categorization of each continuum. Just as in the lexical decision task, after the button press the selected picture moved slightly upwards and the other picture disappeared for 300ms. There was no time limit for responses and presentation of the next trial started 0.8s after the previous button press. The distributional learning task took approximately 25 minutes.

## Results

### Lexical decision

Three participants had to be excluded from all analyses due to missing files, which reduced the dataset from 40 to 37 participants. In addition, all trials in the lexical decision task that contained words with which listeners indicated to be unfamiliar (as assessed after the experiment) were removed on a by-participant basis. This resulted in an additional loss of 119 trials (0.01%), only 20 of which involved the critical sound contrasts (0.005% of critical trials). Overall performance in the lexical decision task was high (minimum correct = 73.81%;  $M = 85.68\%$ ,  $sd = 35.03$ ).

In order to analyze lexical decision responses to words and nonwords and to take into account any possible bias to report hearing a word more often

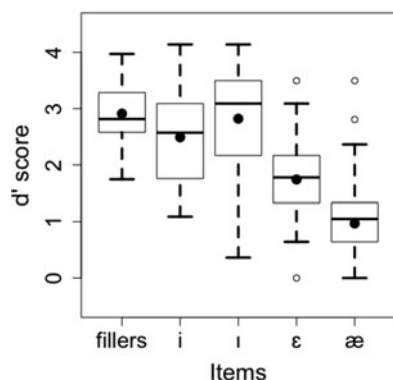


Figure 2. Individual  $d'$  scores for items with the four critical vowels ( $/i/$ ,  $/ɪ/$ ,  $/ɛ/$ ,  $/æ/$ ) and all fillers pooled together. Black dots indicate mean values over all participants.

than a nonword (which is likely if participants do not detect the mispronunciation) all responses were converted into  $d'$  scores per participant per vowel (Macmillan & Creelman, 2005; Noguchi & Kam, 2017; Pallier, 2002).  $d'$  is a measure of sensitivity that takes into account the likelihood that a ‘word’ response was given when the stimulus was a real (correctly pronounced) word (“Hit rate”) and the likelihood of responding ‘word’ when the item was a nonword (i.e., containing a sound substitution; “False alarm rate”). Individual  $d'$  scores per vowel were obtained by subtracting the z-transformed False alarm rate from the z-transformed Hit rate separately for each of the four L2 sounds examined ( $/i/$ ,  $/ɪ/$ ,  $/ɛ/$ ,  $/æ/$ ). The higher the  $d'$  scores, the better participants performed in lexical decision. When referring to  $d'$  scores of a given vowel in the remainder of the paper we will speak of  $d'$  scores for ‘ $/ɛ/$ -items’ when referring to the word set in which  $/ɛ/$  was the intended vowel – that is, words with  $/ɛ/$  and nonwords in which  $/ɛ/$  had been substituted by  $[æ]$  – and analogously for the other vowels.  $d'$  scores for each set of items are shown in Figure 2.

Lexical decision data were first submitted to a linear mixed-effect model (lme4 package 1.1-13, Bates, Mächler, Bolker & Walker, 2015) in R (Version 3.2.2, R Core Team, 2017) with  $d'$  scores as the dependent variable and Sound Contrast as fixed factor. This analysis was performed in order to assess whether there was a difference between  $d'$  scores for the two contrasts. Sound Contrast was contrast coded such that  $/i/-/ɪ/$  was coded as  $-0.5$  and  $/ɛ/-/æ/$  was coded as  $0.5$ . The random-effects structure included a random intercept for participants and a random slope for Sound Contrast over participants. Significance of variables was assessed by means of Satterthwaite’s approximation for degrees of freedom using the lmerTest package (Kuznetsova, Brockhoff & Christensen, 2017). The model revealed a significant effect of Sound Contrast ( $b = -1.30$ ;  $t = -9.47$ ;  $p < .001$ ). As expected,  $d'$  scores were higher – that is,

performance was better – for the items containing the sounds from the  $/i/-/ɪ/$  contrast than for the difficult  $/ɛ/-/æ/$  contrast. Secondly, data were split by contrast in order to examine whether there were differences between the two vowels within each contrast. Two linear mixed-effect models – one for each contrast – were run with  $d'$  scores as the dependent variable and Vowel ( $/i/$  vs.  $/ɪ/$  and  $/ɛ/$  vs.  $/æ/$ , respectively) as fixed factors. We contrast coded the vowels according to their acoustic properties, with the vowel with a lower F1 and a higher F2 in each contrast,  $/i/$  and  $/ɛ/$ , coded as  $-0.5$  and  $/ɪ/$  and  $/æ/$  as  $0.5$ . Note that the effect of Vowel was not assessed together with the effect of Sound Contrast because the levels of the factor Vowel would have been nested within the levels of Sound Contrast (e.g.,  $d'$  scores for  $/i/-$ items and  $/ɪ/-$ items are the  $d'$  scores for the  $/i/-/ɪ/$  contrast). A random intercept for participants was included in each of the two models. A significant effect of Vowel was found in both models ( $/i/-/ɪ/$ :  $b = 0.33$ ;  $t = 2.27$ ;  $p = .029$ ;  $/ɛ/-/æ/$ :  $b = -0.77$ ;  $t = -5.81$ ;  $p < .001$ ). Learners were more accurate with  $/ɪ/-$ items than  $/i/-$ items and with  $/ɛ/-$ items than  $/æ/-$ items.

In order to assess to what extent the results of the lexical decision task relate to the learners’ self-estimated proficiency and experience with English, we conducted additional analyses testing possible correlations between the participants’  $d'$  scores for each of the four sets of items and their Proficiency/Experience Scores from the language background questionnaire (see Participants section). Results showed that only  $d'$  scores for  $/ɛ/-$ items showed a medium-sized correlation with the self-reported measures ( $r(35) = -.36$ ,  $p = .029$ ). The other three correlations were substantially smaller and not significant ( $/æ/-$ items:  $r(35) = -.18$ ,  $p = .29$ ;  $/i/-$ items:  $r(35) = -.07$ ,  $p = .70$ ;  $/ɪ/-$ items:  $r(35) = -.04$ ,  $p = .81$ ). These results will be taken up in the Discussion.

In the following section, the individual  $d'$  scores for each of the four critical vowels were used to examine whether robustness of phonolexical encoding is related to phonetic flexibility.  $d'$  scores for  $/i/-$  and  $/ɪ/-$ items were used as separate predictors in the analysis of the distributional learning data for the  $/i/-/ɪ/$  continuum.  $d'$  scores for  $/ɛ/-$  and  $/æ/-$ items were entered separately into the analysis of the distributional learning data for the  $/ɛ/-/æ/$  continuum.

### Distributional learning

Before we focused on the effects of distributional learning and listeners’ perceptual flexibility, a descriptive analysis of categorization performance was conducted. This was to ascertain that the endpoints of the easy and difficult contrasts were perceived as intended. For the easy contrast, learners correctly identified the two most  $/i/$ -like continuum steps (steps 1 and 2) as  $/i/$  in 98.89% of the cases (minimum = 93.33%;  $sd = 10.49$ ) and the two most

/ɪ/-like continuum steps (steps 10 and 11) as /ɪ/ in 98.33% of the cases (minimum = 86.67%; sd = 12.81). For the difficult contrast, the endpoints and near-endpoint stimuli were correctly identified as /ɛ/ and /æ/ 98.11% (minimum = 76.67%; sd = 13.63) and 97.75% (minimum = 73.33%; sd = 14.84) of the time, respectively. This indicates that participants did not have problems distinguishing prototypical tokens of /ɛ/ and /æ/ at the phonetic level.

Data for the distributional learning task were thereafter analyzed separately for /i/-/ɪ/ and /ɛ/-/æ/ because performance for each contrast in the distributional learning task was to be related to contrast-specific measures of phonolexical encoding from the lexical decision task. For each contrast, data were submitted to a generalized linear mixed-effects model with a logistic linking function with Response (0 = *ship* and *bat*; 1 = *sheep* and *bet*) as the categorical dependent variable. Continuum Step, Exposure Phase and the  $d'$  scores from the lexical decision task were entered as fixed factors. Parameter estimates and significance values for all variables are reported as provided by the lme4 package (Bates et al., 2015), which makes use of maximum likelihood estimation based on the Laplace approximation (Raudenbush, Yang & Yosef, 2000).

Continuum Step was centered on zero (e.g., step 1 = -2.5, step 6 = 0, step 11 = 2.5) and was not allowed to interact with the other factors due to convergence problems (see Mitterer & Reinisch, 2017, for a similar solution). Exposure Phase was re-coded as two linearly independent contrasts to which we will henceforth refer to as Bias and Sequence. “Bias” was coded to represent differences in categorization following the two distributions that should bias perceptual boundaries in opposite directions: that is, categorization patterns after exposure phases 1 and 3 were compared to categorization after exposure phase 2. Exposure phases 1 and 3 were coded as 0.5 and exposure phase 2 as -1. With this coding, a significant positive regression coefficient would indicate that listeners were sensitive to the type of distribution such that there were more “bet” and “sheep” responses following exposure phases 1 and 3 than following exposure phase 2. The second contrast, “Sequence”, assessed differences in the magnitude of shifts from exposure phase 1 to 2 vs. 2 to 3. In other words, it compared categorization following exposure phase 1 and 3, which should bias participants in the same direction. Exposure phase 1 was coded as -0.5, exposure phase 2 as 0 and exposure phase 3 as 0.5. In this case, a significant negative coefficient would mean that, following exposure phase 2, the bias in exposure phase 3 could not re-set listeners’ perceptual boundary to match perception after exposure phase 1.

The  $d'$  scores for each of the two vowels in the given contrast were entered as continuous variables centered around the mean of all participants’  $d'$  scores for that

Table 4. Results of the mixed-effects regression model on the distributional learning data for /i/-/ɪ/.

Predictor	<i>b</i>	<i>z</i>	<i>p</i>
Intercept	-0.13	-0.77	.44
Continuum Step	-3.25	-15.88	< .001
Bias	0.51	6.89	< .001
Sequence	0.05	0.29	.78
/i/ $d'$ scores	-0.19	-0.77	.44
/ɪ/ $d'$ scores	0.16	0.81	.42
Bias * /i/ $d'$ scores	-0.14	-1.26	.21
Bias * /ɪ/ $d'$ scores	0.09	1.01	.32
Sequence * /i/ $d'$ scores	-0.33	-1.46	.14
Sequence * /ɪ/ $d'$ scores	0.06	0.34	.73

vowel. Note that entering  $d'$  scores as a continuous variable allows for stronger conclusions than if participants were grouped based on these scores, a practice common in similar studies (e.g., Díaz, Baus, Escera, Costa & Sebastián-Gallés, 2008; Díaz, Mitterer, Broersma, Escera & Sebastián-Gallés, 2016). This is because significant effects using a continuous variable are harder to find as they critically depend on a (close to) linear effect of the variable. The two variables involving  $d'$  scores were not allowed to interact with each other but were allowed to interact with Bias and Sequence, as such an interaction would suggest a modulation of phonetic flexibility (as measured in distributional learning) by the quality of the learners’ phonolexical representations for a particular type of words (i.e., as measured by the  $d'$  scores). The models’ random-effects structure was the largest that resulted in model convergence: a random intercept for participants with random slopes for Continuum Step and Sequence over participants. The two models were as good a fit to the data as the model with a full random-effects structure in each case, as indicated by log-likelihood ratio tests.

#### /i/-/ɪ/ contrast

Table 4 shows the results of the model for the easy /i/-/ɪ/ contrast. The only significant effects found were those of Continuum Step and Bias. The effect of Continuum Step shows that participants’ responses differed depending on the continuum step to be categorized: the higher the continuum step (the more /ɪ/-like the vowel), the fewer “sheep” responses were given. The significant effect of Bias indicates that categorization after exposure to the /i/-biasing distribution (phases 1 and 3) resulted in more “sheep” responses than exposure to the /ɪ/-biasing distribution (phase 2). This confirms that distributional learning was overall successful at eliciting categorization shifts for the easy contrast. The fact that the effect of Sequence was not significant indicates that categorization



Table 5. Results of the mixed-effects regression model on the distributional learning data for /ɛ/-/æ/.

	b	z	p
Intercept	0.24	1.98	.048
Continuum Step	-2.72	-18.20	<.001
Bias	0.33	4.92	<.001
Sequence	0.13	0.88	.38
/ɛ/ d' scores	0.07	0.35	.73
/æ/ d' scores	-0.02	-0.13	.89
Bias * /ɛ/ d' scores	-0.23	-2.16	.031
Bias * /æ/ d' scores	0.04	0.42	.67
Sequence * /ɛ/ d' scores	0.16	0.68	.49
Sequence * /æ/ d' scores	-0.17	-0.78	.43

responses following exposure phases 1 and 3 – that is, the categorization shifts from exposure phase 1 to 2 and 2 to 3 – did not differ. Finally, neither the  $d'$  scores for /i/-items nor the  $d'$  scores for /ɪ/-items had a significant effect on categorization, nor did they show a significant interaction with Bias or Sequence. This suggests that for the baseline contrast there was no difference whatsoever in the effects of distributional learning as a function of the listeners' performance in the lexical decision task.

#### /ɛ/-/æ/ contrast

Table 5 shows the results of the model for the difficult /ɛ/-/æ/ contrast. As for the easy /i/-/ɪ/ distinction, the model rendered significant effects of Continuum Step and Bias, while the effect of Sequence,  $d'$  scores for /ɛ/-items and  $d'$  scores for /æ/-items were not significant. Crucially, however, the interaction between Bias and  $d'$  scores for /ɛ/-items was significant<sup>2</sup>. This shows that the higher the participants'  $d'$  scores for words containing /ɛ/ and nonwords in which the /ɛ/ had been substituted by [æ] (e.g., \*[æ]mon), the smaller their shift in categorization; that is, the difference in responses following exposure to the biasing distributions in phases 1 and 3 vs. 2 was smaller.

<sup>2</sup> Inspection of individual values in the distributional learning task with the /ɛ/-/æ/ continuum indicates that there is one participant who showed a considerably bigger shift in categorization than the rest of our participants (see Figure 3). Since this outlier could have had an undue influence on the outcome of the analyses, we re-ran the same mixed model with data from this participant excluded. Results mirrored those of the general analysis, showing again a small but significant interaction between Bias and  $d'$  scores for /ɛ/-items ( $b = -0.22$ ;  $z = -1.99$ ;  $p = .046$ ). Since said participant categorized the endpoints of the continuum correctly and showed the expected sigmoid-like categorization functions in distributional learning, and no anomalies were found in the lexical decision data, data from this participant was kept in all analyses reported.

Figure 3 serves as an illustration of how this relationship is instantiated at an individual level. It shows the correlations between individual values for average shift (i.e., distance between 50% crossover points) in distributional learning with /ɛ/-/æ/ ((Exposure 1 – Exposure 2) + (Exposure 3 – Exposure 2) / 2) and  $d'$  scores for /ɛ/-items (left panel) and  $d'$  scores for /æ/-items (right panel). The negative-going regression line for the /ɛ/-items in the left panel mirrors the findings from the mixed-model regression analyses reported above (i.e., the interaction between Bias and  $d'$  scores for /ɛ/-items) that the higher the  $d'$  scores, the smaller the shifts in the distributional learning task. However, unlike the interaction in the mixed-effects regression model, the small-to-medium correlation did not reach significance ( $r(35) = -.26$ ,  $p = .11$ ). This is likely due to differences between the statistical methods.

While the mixed-model reported above is based on all observations from the distributional learning task, the correlation takes into account only one aggregated value per participant, thereby losing information (note that  $d'$  scores were only one value per participant in both analyses). Moreover, mixed-effects modelling allows the specification of random variables (intercept and slopes) such that random variation between participants in their overall performance as well as their responsiveness to certain manipulations are taken into account. For these reasons, among others, mixed-effects regression models are regarded as having a higher sensitivity than analyses with aggregated values (Baayen, Davidson & Bates, 2008; Koerner & Zhang, 2017; Quené & van den Bergh, 2008). This is why, with a sample of 37 participants, it may be not too surprising that the small effect of the interaction reported above (i.e.,  $p = .031$ )<sup>3</sup> did not surface as a significant correlation by  $p < .05$  standards. For /æ/-items, the correlation between  $d'$  scores from the lexical decision task and the average shift in categorization was small and not significant ( $r(35) = -.06$ ,  $p = .73$ ), as was the corresponding interaction in the mixed-effects model reported above.

#### Discussion

The present study explored the relationship between the robustness of L2 phonolexical representations and

<sup>3</sup> Note that if we had used  $d'$  scores to group participants according to their performance with /ɛ/-items (e.g., “poor performers” vs. “good performers”), this would have resulted in a much higher p-value for the critical interaction. To illustrate this point, we ran the exact same mixed-effects regression model for the distributional learning data for the /ɛ/-/æ/ contrast with  $d'$  scores for /ɛ/-items introduced as a categorical predictor. Groups were established by a mean split of  $d'$  scores (below the mean contrast coded as  $-0.5$ , above as  $0.5$ ). In this analysis, the interaction between Group and Bias was significant ( $b = -0.41$ ;  $z = -3.11$ ;  $p = .002$ ) and considerably stronger than the interaction in our main analysis using the continuous variable.

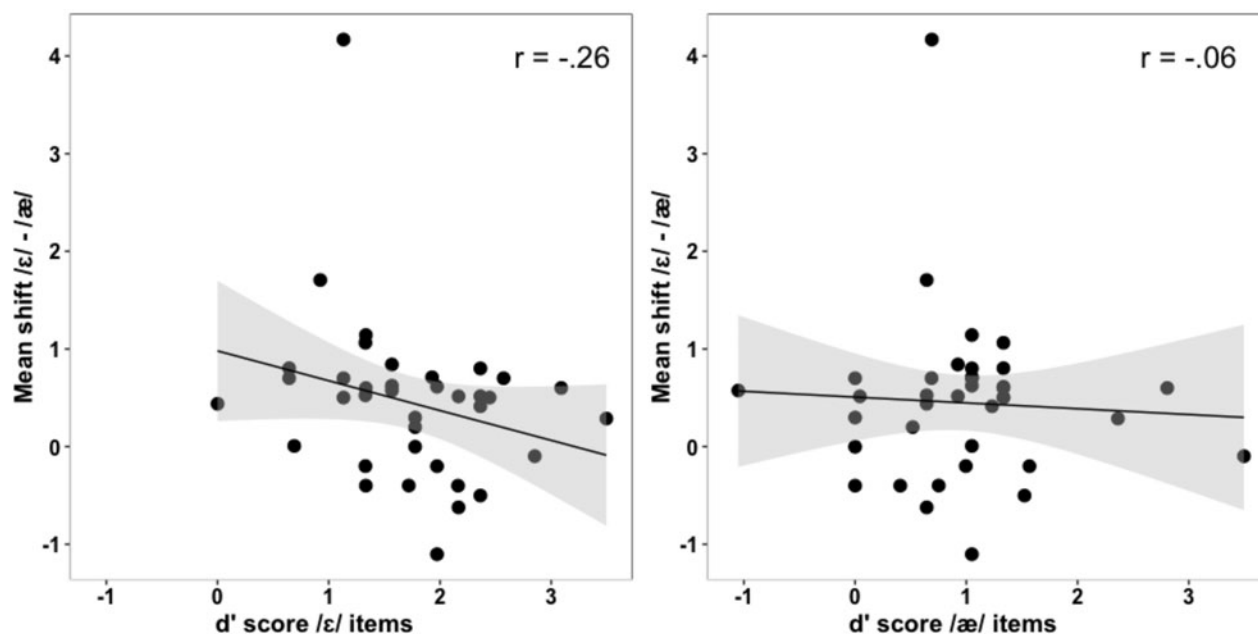


Figure 3. Correlation plots for individual values for the average shift (i.e., distance between 50% crossover points) in distributional learning with  $/\epsilon/-/\text{æ}/$  ((Exposure 1 – Exposure 2) + (Exposure 3 – Exposure 2) / 2) and  $d'$  scores for  $/\epsilon/$  items (left panel) and  $d'$  scores for  $/\text{æ}/$  items (right panel). The shaded areas indicate the 95% confidence intervals of the regression lines.

phonetic flexibility in a non-native language. This relationship was tested with German learners of English for two English vowel contrasts that differ in the perceptual difficulties they cause to these learners. The  $/i/-/ɪ/$  distinction (our L2 control) is easy from a phonetic standpoint because it is shared with our learners' native language (Bohn & Flege, 1992). The  $/\epsilon/-/\text{æ}/$  contrast, however, is difficult to distinguish because  $/\text{æ}/$  does not occur in German and both sounds are often perceived as members of the German category  $/\epsilon/$  (Bohn & Flege, 1990). Participants were tested on two tasks, a lexical decision task containing mispronunciations in which the two sounds of the contrasts were exchanged and a distributional learning task. Lexical decision performance (in  $d'$  scores) was taken as a measure of how robustly learners had encoded each of these sounds in their L2 phonolexical representations. The distributional learning task probed how flexible learners were with the two contrasts at a phonetic level, as assessed through their ability to adjust the boundary between phonetic categories. The central question was whether the robustness of L2 phonolexical representations is related to phonetic flexibility.

Results of the lexical decision task showed that German learners of English were, as expected, more accurate with items corresponding to the easy  $/i/-/ɪ/$  contrast than with items containing  $/\epsilon/$  and  $/\text{æ}/$ . In addition, it was found that there were differences in how accurate learners were with the two vowels within each contrast. For  $/i/-/ɪ/$ , participants

showed higher  $d'$  scores (i.e., were more accurate) for words containing  $/ɪ/$  and nonwords in which this vowel was substituted by  $[i]$  (e.g., \*w[i]nter) than for words with  $/i/$  and the opposite pattern of mispronunciations (e.g., \*n[i]ddle). One possible explanation for this may be that  $/i/$  is longer and more peripheral in the vowel space than  $/ɪ/$  and therefore serves as a better perceptual anchor (Polka & Bohn, 2003, 2011), leading to better detection of an unexpected  $[i]$  instead of  $/ɪ/$  than the other way around.

Nonetheless, the  $d'$  scores for neither of the vowels in the easy contrast related to how much learners shifted their boundary between  $/i/-/ɪ/$  in distributional learning. This replicates previous findings showing that the phonetic boundary between two sounds of an L2 contrast that also occurs in the L1 can easily be shifted (Reinisch et al., 2013) and confirms that  $/i/-/ɪ/$  was an appropriate within-language, within-participant baseline for the difficult L2 distinction. Importantly, the magnitude of shift of the category boundary following distributional learning did not depend on the direction of shift (cf. no effect of Sequence, see Table 4). Distributional learning hence proved a robust measure of overall phonetic flexibility that appears consistent within participants.

While items involving the  $/\epsilon/-/\text{æ}/$  contrast systematically resulted in poorer performance in lexical decision than those involving  $/i/$  and  $/ɪ/$ , there was also a difference between the two vowels within the difficult contrast. Learners were considerably more accurate with  $/\epsilon/$ -items (i.e., words with  $/\epsilon/$  and nonwords of the \*l[æ]mon type)

than for /æ/-items (i.e., words with /æ/ and nonwords of the \*dr[ɛ]gon type). The difference between the *d'* scores for the two vowels in the contrast is likely a reflection of the difficulties that German learners of English have with the new L2 category /æ/ at the lexical level. As mentioned in the introduction, English /ɛ/ appears to be more strongly encoded to words than English /æ/ because it is a better fit to the L1 category. Hence, for German learners of English, nonwords in which /ɛ/ is substituted by [æ] should be relatively easier to reject than nonwords in which /æ/ is substituted by [ɛ]. In this sense, the present results replicate previous findings on an asymmetric perception of this vowel contrast by Dutch (Broersma, 2005; Díaz et al., 2012; Simon et al., 2014) and German learners of English (Llompert & Reinisch, 2017).

Our main finding was, however, that only the individual *d'* scores for /ɛ/-items (and not /æ/-items) predicted the magnitude of the shift in the distributional learning task for the difficult /ɛ/-/æ/ contrast. A factor that likely contributed to this finding is the learners' familiarity with the type of mispronunciations in the /æ/-items (e.g., \*dr[ɛ]gon). These mispronunciations are commonly heard in German-accented English (e.g., Eger & Reinisch, 2017). Therefore, learners may have accepted nonwords like \*dr[ɛ]gon not only because of their fuzzy representations for vowels in these words but also because hearing this type of mispronunciations from fellow native German speakers makes them sound acceptable (see Samuel & Larraza, 2015). Mispronunciations in the /ɛ/-items (e.g., \*[ɛ]mon), in contrast, rarely occur in German-accented English, and thus learners are not likely to consider these nonwords acceptable merely on the basis of familiarity. Rather, if they accept these nonwords as existing, this should be attributed to their deficient phonolexical encoding of the critical vowel. This suggests that, in a task such as ours, *d'* scores for /ɛ/-items may be a more suitable probe to the individual phonolexical encoding of the /ɛ/-/æ/ contrast than *d'* scores for /æ/-items. Note that this argument is further strengthened by the finding that *d'* scores for /ɛ/-items were more strongly related to the learners' self-reported measures of English proficiency and experience than those for /æ/-items.

In the Introduction, we proposed two different outcomes for our main relationship of interest: that is, the one between phonolexical encoding and perceptual flexibility for the /ɛ/-/æ/ contrast. On the one hand, if flexibility developed along general mastery of the contrast, better performance at distinguishing words from nonwords containing the sounds in the difficult contrast should be related to enhanced flexibility for this contrast. On the other hand, it is also possible that a better representation of the difficult sound contrast in the lexicon comes at the cost of flexibility with this contrast. In that case, learners with better scores in the

lexical decision task should be more rigid with their category boundaries between /ɛ/ and /æ/. Our results are mostly in correspondence with this second possibility. The more accurate learners were with /ɛ/-items during lexical decision, the less they shifted their category boundary in the distributional learning task. In general terms, this finding fits well with previous research indicating that flexibility involving difficult non-native categories may be problematic (Sjerps & McQueen, 2010).

We argue that our pattern of results is probably due to the difficulties inherent in the acquisition of a contrast such as /ɛ/-/æ/ for German learners of English. In order to learn such a contrast, German L2 learners have to establish a new phonetic category in their already existing sound category space. Since the L1 acts like a perceptual sieve (Flege, 1995; Trubetzkoy, 1977), it typically takes L2 learners a lot of effort to establish this new L2 category and to accurately separate it from both existing L1 and other L2 categories, with no guarantee of success (e.g., Bohn & Flege, 1990). It may therefore be the case that perceptual properties, such as phonetic flexibility, are developmentally distinct for difficult L2 contrasts. Although the design of our study does not allow us to establish a clear-cut causal or sequential relationship between greater robustness of phonolexical encoding (here for /ɛ/ words and \*[ɛ]mon-type nonwords) and rigidity (i.e., lack of flexibility) in phonetic categorization for /ɛ/-/æ/, two tentative explanations can be put forward for this connection. One account characterizes rigidity of phonetic categories as a CONSEQUENCE of a more robust phonolexical encoding and the other treats it as the CAUSE of the observed enhanced lexical performance with /ɛ/-items.

The first alternative is that as phonolexical representations in the L2 become more robust, the boundary between the two L2 phonetic categories becomes more rigid. At the early stages of L2 learning, perceptual abilities with sounds in difficult L2 contrasts are limited (e.g., Bohn & Flege, 1990) and the encoding of these sounds to L2 phonolexical representations is often faulty. However, over time, learners become better at distinguishing these sounds at the phonetic level, and are able to phonolexically encode them to L2 word representations with a certain success (e.g., Díaz et al., 2012). As phonolexical representations become stronger, learners may become less flexible in their phonetic categorization and less likely to adapt to variation. Phonetic rigidity could therefore be the result of the learners' need to keep their phonetic categories stable once they find themselves able to cope with the difficult non-native contrast with a certain degree of success at both the phonetic and the lexical level.

The second – in our view more likely – alternative is that establishing rigid phonetic categories for sounds in difficult contrasts facilitates their reliable encoding into

the L2 lexicon. Given that the phonolexical encoding of difficult L2 contrasts is extremely challenging, it could be expected that inconsistencies or uncertainty at the phonetic level may have an impact on the degree of success in this encoding. Being excessively flexible or adaptable (i.e., constantly shifting the phonetic boundary between two categories) may be a source of uncertainty compromising the accurate encoding of the phonetic categories into phonolexical representations of words in the L2. By contrast, maintaining rigid phonetic boundaries – once they have been established – would be a way of keeping the phonetic categories stable. Since during L2 acquisition learners are expected to be constantly learning new L2 words and strengthening the representations for already known words (Darcy et al., 2013; Díaz et al., 2012), phonetic rigidity could therefore help improve the encoding of difficult sounds to individual phonolexical representations. This connection would then be apparent, as it is for /ɛ/-items, in a lexical recognition task such as ours. As mentioned above, there are potential reasons why this relationship does not hold for the /æ/-items.

In the present study, we purposefully focused on ‘the average’ German learner of English by targeting a rather homogeneous population whose English abilities were not expected to be close to native-like. This was confirmed in the lexical decision task, where our learners performed poorly on the difficult contrast, especially on /æ/-items but also on /ɛ/-items. A question that remains to be answered, therefore, is whether reaching a better performance in lexical tasks may yet have consequences on flexibility (vs. rigidity) for difficult contrasts. Once mastery of the difficult L2 contrast approaches native-like levels, will phonetic categories stay relatively rigid as those of the best performers with the /ɛ/-items in our lexical decision task, or will they become as flexible as the categories involved in L1 and easy L2 contrasts?

On the one hand, if difficult contrasts like /ɛ/-/æ/ reach a stage in which they are as well established as other less problematic contrasts, then flexibility would be expected to be ultimately attained. Here we showed

that German learners of English consistently shifted the phonetic boundary between /i/-/ɪ/ (an easy L2 contrast) in a distributional learning task. If learners’ mastery of the difficult /ɛ/-/æ/ contrast can approximate that of /i/-/ɪ/ – especially at the lexical level, where a considerable difference between the two contrasts was observed – then we should eventually expect similar results with our experimental paradigm. On the other hand, if rigidity with the newly-established phonetic contrast is indeed central to achieving an accurate encoding of these sounds in the L2 lexicon (as hypothesized above), phonetic categories may still stay relatively rigid in order to avoid compromising phonetic and lexical abilities. However, note that quantifying how much experience with the L2 would be needed for learners to perform within the native range for difficult L2 contrasts in lexical tasks is far from easy, since sometimes not even early and extensive L2 learning guarantees native-like lexical processing (Navarra, Sebastián-Gallés & Soto-Faraco, 2005; Pallier, Colomé & Sebastián-Gallés, 2001; Sebastián-Gallés et al., 2005).

In sum, the present study outlined a relationship between a more successful encoding of a sound involved in a difficult L2 contrast onto L2 phonolexical representations and a reduced flexibility or adaptability for that contrast at the phonetic category level. The German learners of English that were more successful at accepting words with /ɛ/ and rejecting nonwords in which /ɛ/ was substituted by the new L2 category [æ] (e.g., \*[æ]mon) shifted their perceptual boundary between /ɛ/ and /æ/ less in a distributional learning task. While our small effect needs to be interpreted with caution, and further research is needed to confirm and better understand this relationship, our findings suggest that a lack of flexibility at the phonetic level may be advantageous for difficult L2 sound distinctions, reinforcing the idea that the same phonetic flexibility that has been shown to be beneficial in first-language listening (e.g., Bradlow & Bent, 2008; Norris et al., 2003) may be costly when it comes to confusable non-native categories.



**Appendix A.** Words and nonwords used for the /i/-ɪ/ and the /ε/-/æ/ contrasts in the lexical decision task. For nonwords, the base form is also presented.

/i/-ɪ/		/ε/-/æ/	
Words	Nonwords	Words	Nonwords
chief	ch[ɪ]s; cheese	centre	ch[æ]rry; cherry
decent	dr[ɪ]m; dream	check	ch[æ]ss; chess
demon	f[ɪ]ture; feature	chest	d[æ]sk; desk
feeling	f[ɪ]d; feed	death	dr[æ]ss; dress
field	f[ɪ]male; female	dentist	fr[æ]sh; fresh
free	fr[ɪ]dom; freedom	desert	h[æ]lth; health
frequent	l[ɪ]gue; league	helmet	h[æ]lp; help
genius	l[ɪ]gal; legal	helpful	l[æ]gend; legend
leaf	n[ɪ]ddle; needle	lecture	l[æ]mon; lemon
need	r[ɪ]gion; region	record	l[æ]sson; lesson
reader	scr[ɪ]m; scream	rest	s[æ]ntence; sentence
secret	s[ɪ]son; season	smell	w[æ]ther; weather
speaker	sp[ɪ]ch; speech	special	y[æ]llow; yellow
chill	d[i]fference; difference	champion	ch[ε]nnel; channel
discount	d[i]nner; dinner	damage	dr[ε]gon; dragon
dish	d[i]stance; distance	fact	f[ε]ctor; factor
fiction	f[ɪ]gure; figure	fashion	f[ε]ctory; factory
listen	f[ɪ]nger; finger	hammer	fl[ε]g; flag
listener	f[ɪ]sh; fish	hand	g[ε]llery; gallery
ritual	s[ɪ]gnal; signal	happy	h[ε]bit; habit
scissors	spl[i]t; split	language	l[ε]mp; lamp
simple	str[ɪ]ng; string	national	r[ε]mp; ramp
sing	w[ɪ]nd; wind	Saturday	sc[ε]ndal; scandal
singer	w[ɪ]ndow; window	shadow	spl[ε]sh; splash
thing	w[ɪ]nter; winter	smash	st[ε]ndard; standard
wizard	w[ɪ]tch; witch	snack	th[ε]nk; thank

## References

- Amengual, M. (2016). The perception of language-specific phonetic categories does not guarantee accurate phonological representations in the lexicon of early bilinguals. *Applied Psycholinguistics*, 37, 1221–1251. <https://doi.org/10.1017/s0142716415000557>
- Aoyama, K., Flege, J. E., Guion, S. G., Akahane-Yamada, R., & Yamada, T. (2004). Perceived phonetic dissimilarity and L2 speech learning: The case of Japanese /r/ and English /l/ and /r/. *Journal of Phonetics*, 32(2), 233–250. [https://doi.org/10.1016/s0095-4470\(03\)00036-6](https://doi.org/10.1016/s0095-4470(03)00036-6)
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1). doi:10.18637/jss.v067.i01
- Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59(4), 390–412. doi:10.1016/j.jml.2007.12.005
- Bertelson, P., Vroomen, J., & de Gelder, B. (2003). Visual recalibration of auditory speech identification. *Psychological Science*, 14(6), 592–597. <https://doi.org/10.1046/j.0956-7976.2003.psci.1470.x>
- Best, C. T., & Tyler, M. D. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In O.S. Bohn & M. J. Munro (eds.), *Language experience in second language speech learning: In honor of James Emil Flege*, pp. 13–34. Amsterdam: John Benjamins. <https://doi.org/10.1075/llt.17.07bes>
- Boersma, P., & Weenink, D. (2010). Praat: doing phonetics by computer. Version 5.4.22, <http://www.praat.org/>.
- Bohn, O. S., & Flege, J. E. (1990). Interlingual identification and the role of foreign language experience in L2 vowel perception. *Applied Psycholinguistics*, 11(03), 303–328. <https://doi.org/10.1017/s0142716400008912>
- Bohn, O. S., & Flege, J. E. (1992). The production of new and similar vowels by adult German learners of English. *Studies in Second Language Acquisition*, 14(02), 131–158. <https://doi.org/10.1017/s0272263100010792>

- Bradlow, A. R., & Bent, T. (2008). Perceptual adaptation to non-native speech. *Cognition*, 106(2), 707–729. <https://doi.org/10.1016/j.cognition.2007.04.005>
- Broersma, M. (2005). *Phonetic and lexical processing in a second language*. Ph.D. dissertation, Radboud University.
- Broersma, M. (2012). Increased lexical activation and reduced competition in second-language listening. *Language and Cognitive Processes*, 27(7-8), 1205–1224. <https://doi.org/10.1080/01690965.2012.660170>
- Bruggeman, L. (2016). *Nativeness, dominance, and the flexibility of listening to spoken language*. PhD dissertation, Western Sydney University.
- Bundgaard-Nielsen, R. L., Best, C. T., & Tyler, M. D. (2011). Vocabulary size is associated with second-language vowel perception performance in adult learners. *Studies in Second Language Acquisition*, 33(3), 433–461. <https://doi.org/10.1017/S0272263111000040>
- Clayards, M., Tanenhaus, M. K., Aslin, R. N., & Jacobs, R. A. (2008). Perception of speech reflects optimal use of probabilistic speech cues. *Cognition*, 108(3), 804–809. <https://doi.org/10.1016/j.cognition.2008.04.004>
- Cook, S. V. (2012). *Phonological form in L2 lexical access: Friend or foe?* Ph.D. dissertation, University of Maryland.
- Cook, S. V., & Gor, K. (2015). Lexical access in L2: Representational deficit or processing constraint? *The Mental Lexicon*, 10(2), 247–270. <https://doi.org/10.1075/ml.10.2.04coo>
- Cook, S. V., Pandza, N. B., Lancaster, A. K., & Gor, K. (2016). Fuzzy nonnative phonolexical representations lead to fuzzy form-to-meaning mappings. *Frontiers in Psychology*, 7, 1345. <https://doi.org/10.3389/fpsyg.2016.01345>
- Cooper, A., Brouwer, S., & Bradlow, A. R. (2015). Interdependent processing and encoding of speech and concurrent background noise. *Attention, Perception, & Psychophysics*, 77(4), 1342–1357. <https://doi.org/10.3758/s13414-015-0855-z>
- Cutler, A. (2015). Representation of second language phonology. *Applied Psycholinguistics*, 36(01), 115–128. doi:10.1017/s0142716414000459
- Cutler, A., Weber, A., & Otake, T. (2006). Asymmetric mapping from phonetic to lexical representations in second-language listening. *Journal of Phonetics*, 34(2), 269–284. doi:10.1016/j.wocn.2005.06.002
- Cutler, A., Weber, A., Smits, R., & Cooper, N. (2004). Patterns of English phoneme confusions by native and non-native listeners. *The Journal of the Acoustical Society of America*, 116(6), 3668–3678. doi:10.1121/1.1810292
- Darcy, I., Daidone, D., & Kojima, C. (2013). Asymmetric lexical access and fuzzy lexical representations in second language learners. *The Mental Lexicon*, 8(3), 372–420. <https://doi.org/10.1075/ml.8.3.06dar>
- Darcy, I., Dekydtspotter, L., Sprouse, R. A., Glover, J., Kaden, C., McGuire, M., & Scott, J. H. (2012). Direct mapping of acoustics to phonology: On the lexical encoding of front rounded vowels in L1 English–L2 French acquisition. *Second Language Research*, 28(1), 5–40. <https://doi.org/10.1177/0267658311423455>
- Díaz, B., Baus, C., Escera, C., Costa, A., & Sebastián-Gallés, N. (2008). Brain potentials to native phoneme discrimination reveal the origin of individual differences in learning the sounds of a second language. *Proceedings of the National Academy of Sciences*, 105(42), 16083–16088. doi:10.1073/pnas.0805022105
- Díaz, B., Mitterer, H., Broersma, M., Escera, C., & Sebastián-Gallés, N. (2016). Variability in L2 phonemic learning originates from speech-specific capabilities: An MMN study on late bilinguals. *Bilingualism: Language and Cognition*, 19(05), 955–970. doi:10.1017/s1366728915000450
- Díaz, B., Mitterer, H., Broersma, M., & Sebastián-Gallés, N. (2012). Individual differences in late bilinguals' L2 phonological processes: From acoustic-phonetic analysis to lexical access. *Learning and Individual Differences*, 22(6), 680–689. <https://doi.org/10.1016/j.lindif.2012.05.005>
- Drozдова, P., van Hout, R., & Scharenborg, O. (2016). Lexically-guided perceptual learning in non-native listening. *Bilingualism: Language and Cognition*, 19(05), 914–920. <https://doi.org/10.1017/s136672891600002x>
- Eger, N. A., & Reinisch, E. (2017). The role of acoustic cues and listener proficiency in the perception of accent in nonnative sounds. *Studies in Second Language Acquisition*, 1–22. doi:10.1017/s0272263117000377
- Eger, N. A., & Reinisch, E. (2018). The impact of one's own voice and production skills on word recognition in a second language. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. doi:10.1037/xlm0000599
- Escudero, P., Benders, T., & Wanrooij, K. (2011). Enhanced bimodal distributions facilitate the learning of second language vowels. *The Journal of the Acoustical Society of America*, 130(4), EL206–EL212. doi:10.1121/1.3629144
- Escudero, P., Hayes-Harb, R., & Mitterer, H. (2008). Novel second-language words and asymmetric lexical access. *Journal of Phonetics*, 36(2), 345–360. <https://doi.org/10.1016/j.wocn.2007.11.002>
- Flege, J. E. (1995). Second language speech learning: Theory, findings, and problems. In W. Strange (ed.), *Speech perception and linguistic experience: Issues in cross-language research*, pp. 233–277. Timonium, MD: York Press.
- Flege, J. E., Bohn, O. S., & Jang, S. (1997). Effects of experience on non-native speakers' production and perception of English vowels. *Journal of Phonetics*, 25(4), 437–470. <https://doi.org/10.1006/jpho.1997.0052>
- Fourakis, M., & Iverson, G. K. (1984). On the “incomplete neutralization” of German final obstruents. *Phonetica*, 41(3), 140–149. <https://doi.org/10.1159/000261720>
- Gathercole, S. E., Hitch, G. J., Service, E., & Martin, A. J. (1997). Phonological short-term memory and new word learning in children. *Developmental Psychology*, 33(6), 966–979. <https://doi.org/10.1037/0012-1649.33.6.966>
- Gollan, T. H., Montoya, R. I., Cera, C., & Sandoval, T. C. (2008). More use almost always means a smaller frequency effect: Aging, bilingualism, and the weaker links hypothesis. *Journal of Memory and Language*, 58(3), 787–814. <https://doi.org/10.1016/j.jml.2007.07.001>
- Goto, H. (1971). Auditory perception by normal Japanese adults of the sounds “L” and “R.” *Neuropsychologia*, 9(3), 317–323. doi:10.1016/0028-3932(71)90027-3
- Hanulíková, A., & Weber, A. (2011). Sink positive: Linguistic experience with th substitutions influences nonnative word

- recognition. *Attention, Perception, & Psychophysics*, 74(3), 613–629. doi:10.3758/s13414-011-0259-7
- Iverson, P., & Evans, B. G. (2007). English vowel training with different first-language vowel systems. *The Journal of the Acoustical Society of America*, 121(5), 3072–3072. doi:10.1121/1.4781875
- Kleinschmidt, D. F., Raizada, R., & Jaeger, T. F. (2015). Supervised and unsupervised learning in phonetic adaptation. In R. Dale, C. Jennings, P. Maglio, T. Matlock, D. Noelle, A. Warlaumont, & J. Yoshimi (eds.), *Proceedings of the 37th Annual Conference of the Cognitive Science Society*, pp. 1129–1134. Austin, TX: Cognitive Science Society.
- Koerner, T., & Zhang, Y. (2017). Application of linear mixed-effects models in human neuroscience research: A comparison with Pearson correlation in two auditory electrophysiology studies. *Brain Sciences*, 7(12), 26. doi:10.3390/brainsci7030026
- Kuhl, P., Williams, K., Lacerda, F., Stevens, K., & Lindblom, B. (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. *Science*, 255(5044), 606–608. <https://doi.org/10.1126/science.1736364>
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in linear mixed effects models. *Journal of Statistical Software*, 82(13). doi:10.18637/jss.v082.i13
- Ladefoged, P., & Broadbent, D. E. (1957). Information conveyed by vowels. *The Journal of the Acoustical Society of America*, 29(1), 98–104. <https://doi.org/10.1121/1.1908694>
- Lametti, D. R., Krol, S. A., Shiller, D. M., & Ostry, D. J. (2014). Brief periods of auditory perceptual training can determine the sensory targets of speech motor learning. *Psychological Science*, 25(7), 1325–1336. <https://doi.org/10.1177/0956797614529978>
- Lancaster, A., & Gor, K. (2016). Abstraction of phonological representations in adult nonnative speakers. *Proceedings of the Linguistic Society of America*, 1, 24, 1–15 <https://doi.org/10.3765/plsa.v1i0.3725>
- Llompарт, M., & Simonet, M. (2018). Unstressed vowel reduction across Majorcan Catalan dialects: Production and spoken word recognition. *Language and Speech*, 61(3), 430–465. doi:10.1177/0023830917736019
- Llompарт, M., & Reinisch, E. (2017). Articulatory information helps encode lexical contrasts in a second language. *Journal of Experimental Psychology: Human Perception and Performance*, 43(5), 1040–1056. <https://doi.org/10.1037/xhp0000383>
- Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory: A user's guide* (2nd ed.). Mahwah, NJ: Erlbaum.
- Mitterer, H., & Reinisch, E. (2017). Surface forms trump underlying representations in functional generalisations in speech perception: The case of German devoiced stops. *Language, Cognition and Neuroscience*, 32, 1133–1147. <https://doi.org/10.1080/23273798.2017.1286361>
- Munson, C. M. (2011). *Perceptual learning in speech reveals pathways of processing*. Ph.D. dissertation, University of Iowa.
- Navarra, J., Sebastián-Gallés, N., & Soto-Faraco, S. (2005). The perception of second language sounds in early bilinguals: New evidence from an implicit measure. *Journal of Experimental Psychology: Human Perception and Performance*, 31(5), 912–918. <https://doi.org/10.1037/0096-1523.31.5.912>
- Noguchi, M., & Kam, C. L. H. (2017). The emergence of the allophonic perception of unfamiliar speech sounds: The effects of contextual distribution and phonetic naturalness. *Language Learning*, 68(1), 147–176. doi:10.1111/lang.12267
- Norris, D., McQueen, J. M., & Cutler, A. (2003). Perceptual learning in speech. *Cognitive Psychology*, 47(2), 204–238. [https://doi.org/10.1016/S0010-0285\(03\)00006-9](https://doi.org/10.1016/S0010-0285(03)00006-9)
- Pallier, C. (2002). Computing discriminability and bias with the R software. Available at <http://www.pallier.org/pdfs/aprime.pdf>.
- Pallier, C., Colomé, A., & Sebastián-Gallés, N. (2001). The influence of native-language phonology on lexical access: Exemplar-based versus abstract lexical entries. *Psychological Science*, 12(6), 445–449. <https://doi.org/10.1111/1467-9280.00383>
- Peirce, J. W. (2007). PsychoPy—psychophysics software in Python. *Journal of Neuroscience Methods*, 162(1), 8–13. <https://doi.org/10.1016/j.jneumeth.2006.11.017>
- Polka, L., & Bohn, O. S. (2003). Asymmetries in vowel perception. *Speech Communication*, 41(1), 221–231. doi:10.1016/s0167-6393(02)00105-x
- Polka, L., & Bohn, O. S. (2011). Natural Referent Vowel (NRV) framework: An emerging view of early phonetic development. *Journal of Phonetics*, 39(4), 467–478. doi:10.1016/j.wocn.2010.08.007
- Polka, L., & Werker, J. F. (1994). Developmental changes in perception of nonnative vowel contrasts. *Journal of Experimental Psychology: Human Perception and Performance*, 20(2), 421–435. doi:10.1037/0096-1523.20.2.421
- Port, R., & O'Dell, M. L. (1985). Neutralization of syllable-final voicing in German. *Journal of Phonetics*, 13(4), 455–471.
- Quené, H., & van den Bergh, H. (2008). Examples of mixed-effects modeling with crossed random effects and with binomial data. *Journal of Memory and Language*, 59(4), 413–425. doi:10.1016/j.jml.2008.02.002
- R Core Team (2017). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <https://www.r-project.org/>.
- Raudenbush, S. W., Yang, M.-L., & Yosef, M. (2000). Maximum likelihood for generalized linear models with nested random effects via high-order, multivariate Laplace approximation. *Journal of Computational and Graphical Statistics*, 9(1), 141–157. doi:10.1080/10618600.2000.10474870
- Reinisch, E., & Mitterer, H. (2016). Exposure modality, input variability and the categories of perceptual recalibration. *Journal of Phonetics*, 55, 96–108. doi:10.1016/j.wocn.2015.12.004
- Reinisch, E., Weber, A., & Mitterer, H. (2013). Listeners retune phoneme categories across languages. *Journal of Experimental Psychology: Human Perception and Performance*, 39(1), 75–86. <https://doi.org/10.1037/a0027979>
- Samuel, A. G., & Larraza, S. (2015). Does listening to non-native speech impair speech perception?

- Journal of Memory and Language*, 81, 51–71. doi:[10.1016/j.jml.2015.01.003](https://doi.org/10.1016/j.jml.2015.01.003)
- Schuhmann, K. S. (2014). *Perceptual learning in second language learners*. Ph.D. dissertation, State University of New York at Stony Brook.
- Sebastián-Gallés, N., Echeverría, S., & Bosch, L. (2005). The influence of initial exposure on lexical representation: Comparing early and simultaneous bilinguals. *Journal of Memory and Language*, 52(2), 240–255. <https://doi.org/10.1016/j.jml.2004.11.001>
- Sheldon, A., & Strange, W. (1982). The acquisition of /r/ and /l/ by Japanese learners of English: Evidence that speech production can precede speech perception. *Applied Psycholinguistics*, 3(03), 243–261. <https://doi.org/10.1017/s0142716400001417>
- Sidaras, S. K., Alexander, J. E. D., & Nygaard, L. C. (2009). Perceptual learning of systematic variation in Spanish-accented speech. *The Journal of the Acoustical Society of America*, 125(5), 3306. doi:[10.1121/1.3101452](https://doi.org/10.1121/1.3101452)
- Simon, E., Sjerps, M., & Fikkert, P. (2014). Phonological representations in children's native and non-native lexicon. *Bilingualism: Language and Cognition*, 17(1), 3–21. doi:[10.1017/S1366728912000764](https://doi.org/10.1017/S1366728912000764)
- Sjerps, M. J., & McQueen, J. M. (2010). The bounds on flexibility in speech perception. *Journal of Experimental Psychology: Human Perception and Performance*, 36(1), 195–211. doi:[10.1037/a0016803](https://doi.org/10.1037/a0016803)
- Sumner, M., & Samuel, A. G. (2009). The effect of experience on the perception and representation of dialect variants. *Journal of Memory and Language*, 60(4), 487–501. <https://doi.org/10.1016/j.jml.2009.01.001>
- Trubetzkoy, N. S. (1977). *Grundzüge der Phonologie* (6th ed.). Göttingen, Germany: Van den Hoeck & Ruprecht.
- van Heuven, W. J. B., Mandera, P., Keuleers, E., & Brysbaert, M. (2014). SUBTLEX-UK: A new and improved word frequency database for British English. *The Quarterly Journal of Experimental Psychology*, 67(6), 1176–1190. <https://doi.org/10.1080/17470218.2013.850521>
- Weber, A., & Cutler, A. (2004). Lexical competition in non-native spoken-word recognition. *Journal of Memory and Language*, 50(1), 1–25. [https://doi.org/10.1016/s0749-596x\(03\)00105-0](https://doi.org/10.1016/s0749-596x(03)00105-0)
- Weber, A., di Betta, A. M., & McQueen, J. M. (2014). Treack or trit: Adaptation to genuine and arbitrary foreign accents by monolingual and bilingual listeners. *Journal of Phonetics*, 46, 34–51. <https://doi.org/10.1016/j.wocn.2014.05.002>
- Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development*, 7(1), 49–63. [https://doi.org/10.1016/s0163-6383\(84\)80022-3](https://doi.org/10.1016/s0163-6383(84)80022-3)
- Witteman, M. J., Weber, A., & McQueen, J. M. (2013). Foreign accent strength and listener familiarity with an accent codetermine speed of perceptual adaptation. *Attention, Perception, & Psychophysics*, 75(3), 537–556. <https://doi.org/10.3758/s13414-012-0404-y>
- Zipf, G. K. (1949). *Human behavior and the principle of least effort*. Cambridge, MA: Addison-Wesley.