

Novel second-language words and asymmetric lexical access

Paola Escudero

Institute of Phonetic Sciences, University of Amsterdam

Rachel Hayes-Harb

Department of Linguistics, University of Utah

Holger Mitterer

Max Planck Institute for Psycholinguistics, Nijmegen

The lexical and phonetic mapping of auditorily confusable L2 nonwords was examined by teaching L2 learners novel words and by later examining their word recognition using an eyetracking paradigm. During word learning, two groups of highly proficient Dutch learners of English learned 20 English nonwords of which 10 contained the English contrast / ϵ /-/ æ / (a confusable contrast for native Dutch speakers). One group of subjects learned the words by matching their auditory forms to pictured meanings, while a second group additionally saw the spelled forms of the words. We found that the group who received only auditory forms confused words containing / æ / and / ϵ / symmetrically, *i.e.*, both / æ / and / ϵ / auditory tokens triggered looks to pictures containing both / æ / and / ϵ /. In contrast, the group who also had access to spelled forms showed the same asymmetric word recognition pattern found by previous studies, *i.e.*, they only looked at pictures of words containing / ϵ / when presented with / ϵ / target tokens but looked at pictures of words containing both / æ / and / ϵ / when presented with / æ / target tokens. The results demonstrate that L2 learners can form lexical contrasts for auditorily confusable novel L2 words. However and most importantly, this study suggests that explicit information over the contrastive nature of two new sounds may be needed to build separate lexical representations for similar-sounding L2 words.

1. Introduction

Speakers of different languages have specific difficulties distinguishing similar sounding words in a new language. For instance, native speakers of Japanese have difficulty distinguishing the English words *rocket* and *locket* (Aoyama, Flege, Guion, Akahane-Yamada & Yamada 2004), while native speakers of Spanish (Flege, Bohn & Jang 1997) and Portuguese (Bion, Escudero, Rauber and Baptista 2006; Rauber, Escudero, Bion & Baptista 2005) have difficulty distinguishing the English words *beat* and *bit*, and native speakers of Dutch have difficulty with the English words *bet* and *bat* (Broersma 2005; Cutler & Broersma 2005). A common characteristic among these pairs of words is that they are differentiated by L2 contrasts which do not have immediate counterparts in the learners' native languages. A critical issue in the

literature on L2 lexical representations and their phonological structure is whether the difficulties that learners have distinguishing similar-sounding words has to do with perceptual abilities or with the phonological content of the words' representation in the L2 lexicon.

Many studies have suggested that this difficulty has a lexical origin; that is, that listeners do not have distinct lexical representations for words distinguished minimally by novel L2 contrasts and therefore treat them as homophones. For example, Pallier, Colomé and Sebastián-Gallés (2001) found that Spanish-dominant Spanish-Catalan bilinguals exhibited repetition priming effects for Catalan minimal pairs that were contrasted by the /o/ and /ɔ/ vowels, a contrast that is not present in Spanish. They interpreted this finding as evidence for the homophonous lexical representation of these minimal pair—when presented with either auditory form (e.g., *dóna* “s/he gives” and *dona* “woman”), the Spanish-dominant bilinguals activated both lexical entries. Cutler and Otake (2004) found similar results for native Dutch speakers and English /æ/-/ɛ/ minimal pairs. However, such findings may also (or instead) reflect perceptual confusion of the minimal pairs, namely that the learners could not perceive the contrasts distinguishing the words in these online listening tasks and therefore accessed the same lexical representations regardless of which word was present in the auditory input.

Curtin, Goad and Pater (1998) attempted to tease apart the contributions of online perception and lexical storage of novel L2 contrasts to L2 word identification by examining monolingual English speakers' ability to contrast the Thai three-way stop voicing distinction among voiced, voiceless unaspirated, and voiceless aspirated consonants in their memory of Thai minimal triplets. They found that subjects' ability to do so differed depending on whether the contrast tested was voiced vs. voiceless-unaspirated or voiceless-unaspirated vs. voiceless-aspirated. Despite subjects' relatively better ability to discriminate voiceless-unaspirated vs. voiceless-aspirated stop pairs than voiced vs. voiceless-unaspirated pairs in a (perceptual) discrimination task, when the task was lexical, subjects were better at distinguishing voiced vs. voiceless-unaspirated word pairs. Curtin *et al.* (1998) claim that because English has a voicing distinction, represented lexically by the feature [±voi], which distinguishes voiced and voiceless-unaspirated stops, but not an aspiration distinction (e.g., [±asp]), they are better able to lexically encode the voiced vs. voiceless-unaspirated contrast (because it is distinguished by the familiar feature [±voi]) than the voiceless-unaspirated vs. voiceless-aspirated contrast (which is distinguished by the novel feature [±asp]). Additionally, Hayes-Harb and Masuda (to appear)

found that native English speakers have difficulty storing the contrast between Japanese singleton and geminate consonants lexically, despite the fact that native English speakers are able to discriminate the contrasts relatively well in a perceptual task (Hayes-Harb 2005).

Despite this evidence in favor of lexical homophony, it should be noted that Pater (2003), in a follow-up to Curtin *et al.* (1998), examined the role of task effects on English speakers' performance on lexical and (perceptual) discrimination tasks with respect to the features [±voi] and [±asp], and found that when the lexical and non-lexical tasks were made to be more similar to each other, performance on the two tasks was comparable for words containing the features [±voi] and [±asp]. That the Curtin *et al.* (1998) and Pater (2003) studies found such different results with slightly different experimental methods indicates that the Curtin *et al.* (1998) finding is not robust across experimental conditions.

Using eye-tracking technology, a more recent approach to the study of L2 lexical representation and their phonological content has provided a new way of distinguishing between the contributions of perceptual and lexical components in L2 word recognition. Eye-tracking technology provides a continuous measure of how listeners evaluate incoming auditory input over time and therefore can provide information about both speech perception and lexical processing as they unfold during an auditory word recognition task (Salverda, Dahan & McQueen 2003). Weber and Cutler (2004) exploited this technology to observe L2 learners' auditory word processing. In their study, native Dutch-speaking subjects were asked to click on pictures presented on a computer screen that matched auditorily-presented picturable target words. They found that these listeners fixated their visual attention longer and more frequently on distractor pictures of words which are auditorily confusable than on distractor pictures of words which are auditorily dissimilar. For example, they looked longer and more frequently at a picture of a *pencil* when the target word was *panda* than on a less confusable distractor (*e.g.*, *beetle* when the target word was *bottle*). Note that Dutch listeners have problems to distinguish these words based on the first syllable alone (/pæn/ in *panda* and /pɛn/ in *pencil*), because they differ only in the non-native vowel contrast /æ/-/ɛ/. This finding suggests that at the point where their subjects heard the first syllable of the word *panda*, they also inappropriately activated the word *pencil*. However, Weber and Cutler (2004) also demonstrated that the situation is more complex than it appears at first, as the listeners' pattern of inappropriate L2 lexical activation was asymmetric, with the target *panda* causing activation of *pencil* but not the reverse.

The authors explain this asymmetric pattern of word activation by arguing that the learners have indeed encoded the /ɛ/-/æ/ contrast lexically, because they do not look at the picture of a *panda* when they hear a token of the word *pencil*, but that they have difficulty perceiving this contrast in the on-line auditory word identification task. Thus, the authors suggest that the learners perceptually neutralize the contrast, which leads them to activate lexical entries containing /ɛ/ regardless of which of the two vowels is present in the auditory input. This perceptual neutralization hypothesis can explain why native Dutch listeners inappropriately activate the word *pencil* when they hear the word *panda*. On the other hand, learners are prevented from searching for lexical entries containing /æ/ without the support of additional auditory information (*i.e.*, the disambiguating information contained in rest of the word), which explains why they were able to ultimately identify the picture of the panda as matching the target word *panda*. Weber and Cutler conclude that this asymmetry in lexical activation provides evidence that these Dutch listeners who are highly proficient in English have established the /ɛ/-/æ/ contrast in their lexical representations of English words but not in their perceptual representations or in the on-line perceptual processing of words containing the two vowels.

A question that emerges from the results of this study is why the Dutch listeners have a bias for the English category /ɛ/ in their lexical decisions. To answer this question, Weber and Cutler (2004) appeal first to the Dutch native vowel inventory, which contains the phonetic category /ɛ/ but not /æ/. The authors suggest that the native status of the category /ɛ/, although it has a different phonetic realization in the two languages, is responsible for its use when perceiving the non-native contrast. In addition, the authors appeal to previous perception studies to suggest that Dutch /ɛ/ is the closest phonemic match for both English vowels. That is, a number of perceptual identification studies (Cutler, Weber, Smits & Cooper 2004, Broersma 2005) show a perceptual bias towards English /ɛ/ when Dutch listeners classify tokens containing the two English vowels. However, this /ɛ/ bias may also have another explanation, which is developed and further tested in Cutler, Weber, and Otake (2006).

Cutler *et al.* (2006) set out to answer two questions raised by the Weber and Cutler (2004) study: First, can the asymmetry reflecting contrastive lexical representations be replicated in different listener populations, and second, what explains the lexical “dominance” of one of the members of the L2 contrast over the other? Cutler *et al.* (2006) performed a replication of the Weber and Cutler (2004) study, this time with native Japanese speakers listening to words

contrasted by English /r/ and /l/, and found a similar asymmetric pattern of lexical activation, with auditory /r/ words inducing looks to /l/ pictures, but not the reverse, suggesting that /l/ was the dominant category for lexical activation. In an effort to understand the direction of the asymmetry, they first noted that the Japanese liquid consonant is phonetically closer to English /l/ than it is to English /r/ (Aoyama *et al.*, 2004). Cutler *et al.* (2006) thus concluded that the L2 category that is phonetically *closer* to the L1 category emerges as the dominant category for the purpose of auditory word recognition.

Does this phonetic closeness argument explain the Weber and Cutler (2004) findings with respect to the dominance of the /ɛ/ category? The argument that Dutch /ɛ/ is phonetically closer to English /ɛ/ than to English /æ/ does not in fact seem to hold when comparing vowel acoustic data from both General American and Standard Southern British English (Peterson & Barney 1952 and Deterding 1997 respectively) with Dutch data (Pols, Tromp & Plomp 1973). These acoustic data (in the form of F1 and F2 values) show that Dutch /ɛ/ seems to lie between the two English vowels. Weber and Cutler (2004) suggested that the phonetic closeness may be perceptual in that English /ɛ/ may be a better perceptual match to Dutch /ɛ/ than English /æ/. However, Cutler *et al.* (2004: 273) show that Dutch listeners classified both English /ɛ/ and /æ/ with the same level of accuracy (*i.e.*, below 60%), which suggests that having a common native category does not necessarily result in more accurate perceptual performance.

Cutler *et al.* (2006) proposed a second possible explanation for the direction of the asymmetry, based on the influence of orthographic information, which they then rejected in the case of their Japanese learners of English /l/ and /r/. The authors suggest that the orthographic hypothesis would predict English /r/ as the dominant category because the transliteration of Japanese words containing /r/ into alphabetic orthography (*i.e.*, a writing system called *romanji*) uses the letter 'r' (as in, *e.g.*, *tempura*). Thus, if the native Japanese speakers are using previous knowledge of grapheme-phoneme correspondences from their native Japanese (via *romanji*), this would predict that the asymmetry would be in the opposite direction (*i.e.*, with /r/ as the dominant category) from what they actually found (*i.e.*, /l/ was the dominant category). However, it is still an open question whether the influence of orthography can explain the direction of asymmetric lexical activation in the case of Dutch learners and the English contrast between /æ/ and /ɛ/, or even whether orthography plays a role in the development of contrastive lexical representations. Do the differential spellings of /æ/- and /ɛ/- words in English result in a

differential storage of English words containing these letters by Dutch learners of English? In Dutch, the letters ‘e’ and ‘a’ map to separate phonemes—‘e’ maps to a Dutch front central vowel and ‘a’ maps to a low back vowel. If English words containing orthographic ‘e’ are lexically stored by learners with a Dutch front central vowel, while words containing orthographic ‘a’ are stored with a Dutch back vowel, this would lead to the contrastive lexical representation of English words spelled with ‘a’ and ‘e’. Consequently, even though Dutch listeners correctly perceive both /æ/ and /ɛ/ in English as front vowels, they only access entries containing orthographic ‘e’ because they constitute their only lexical match for their perception (as words containing orthographic ‘a’ can only be accessed if a back vowel is perceived). Some evidence suggests that this hypothesis is plausible: Perception studies that have indicated a bias towards /ɛ/ in Dutch learners’ perception of English (Cutler *et al.* 2004, Broersma 2005) have used materials that included orthographic representations of words, such as ‘bet’ or ‘bat’, to measure Dutch listeners’ classification of English syllables or words containing /æ/ and /ɛ/.

A conclusion that can be made from both the Weber and Cutler (2004) and the Cutler *et al.* (2006) studies is that L2 learners are able to lexically encode a novel contrast that they cannot reliably use in the process of auditory word comprehension. This is in striking contrast with most of the studies reviewed in the first paragraphs of this introduction, namely Curtin *et al.* 1998, Pallier *et al.* 2001 and Hayes-Harb and Masuda (to appear), all of which provide evidence that is consistent with homophonous lexical entries for L2. The present study aims to answer the question of how L2 learners can establish contrastive lexical representations for phonemes that they do not reliably discriminate in listening tasks. Although the Cutler *et al.* (2006) orthographic hypothesis for the direction of the asymmetry was not supported for Japanese learners of English /l/ and /r/, it is possible that it may still hold for the Dutch case. Thus, in addition to the hypothesis that knowledge of contrastive spellings for L2 words can lead learners to encode novel phoneme contrasts in the lexicon, we also hypothesize that Dutch learners’ orthographic representations influence the direction of the asymmetry found by Weber and Cutler (2004).

The bias in the asymmetric pattern of recognition found in Weber and Cutler (2004) as well as the orthographic explanation presented by Cutler *et al.* (2006) is explored in the present study, which constitutes a further investigation of the recognition of English words by highly proficient Dutch listeners. We were particularly interested in determining whether the

asymmetric patterns of assimilation would be found with novel L2 words, where subjects' knowledge of the spelled forms of such novel words could be experimentally manipulated. This goal was supported by a recent study by Shatzman and McQueen (2006), who reported that native Dutch listeners transferred the abstract prosodic patterns of their language to novel Dutch words. In normal Dutch speech, the duration of a syllable (*e.g.* [ham]) is longer in a monosyllabic word (*e.g.* [ham] 'ham') than in a bisyllabic word (*e.g.* [hamster] 'hamster'). Shatzman & McQueen (2006) tested whether listeners apply this knowledge to newly learned words. In their study, listeners were taught to associate nonwords with nonobjects. The nonwords were pairs of the form, *e.g.*, [bap] and [baptu], where the syllable [bap] appeared both as a monosyllabic word and as the first syllable of a bisyllabic word. During training, all [bap] syllables, regardless of whether they appeared in monosyllabic or bisyllabic words, had the same duration. At test, subjects were asked to "Click on the picture of the ___", where some auditory forms contained shorter [bap] syllables and some longer [bap] syllables than the trained forms. Results indicate that subjects expected longer syllables to belong to monosyllabic words and shorter syllables to belong to bisyllabic words *regardless of trained durations*. For example, during the word recognition task, a [bap] token with a duration of 316 msec was more likely to activate the monosyllabic word /bap/ than a [bap] token of 292 msec, and a [bap] of 266 msec was more likely to activate the bisyllabic word /baptu/ than a [bap] of 292 msec. Shatzman and McQueen (2006) interpret this finding as indicating that newly-learned nonwords are treated similarly to existing lexical items and that native phonological knowledge is transferred to novel words.

In addition to examining the possible transfer of Weber and Cutler's (2004) asymmetric pattern to novel L2 words, in the present study we directly address the role of orthography in learners' ability to establish contrastive lexical representations for novel L2 words. To this effect, we conducted a word-learning experiment similar to that conducted by Shatzman and McQueen (2006), but which manipulated the availability of the words' orthographic forms to learners. In the present study, native Dutch speaking subjects with the same linguistic and social background as Weber and Cutler's (2004) subjects were taught a set of English nonwords containing /ɛ/ and /æ/ as labels for pictures of nonobjects (*e.g.*, [tɛnzə] and [tændək]). In one training condition, subjects were provided only with the auditory forms and the pictures, and in the other, they were additionally shown the spelled forms of the nonwords (*e.g.*, 'tenzer' and 'tandek'). Importantly,

the letters ‘e’ and ‘a’ provide a visual cue to the /ɛ/-/æ/ contrast. After this word learning phase, subjects were presented with the same word recognition test described in Weber and Cutler (2004), but with the novel words and pictures that they were exposed to during the word learning phase.

Thus the main question driving the present study is: Does the availability of the spelled forms of the nonwords affect the phonological content of learners’ lexical representations, such that the Weber and Cutler (2004) results with respect to asymmetric lexical access be replicated with novel English words learned together with their spelling? We hypothesize that learners’ ability to encode the /æ/-/ɛ/ contrast lexically (and thus whether or not they exhibit asymmetric lexical access) will be determined by the availability of non-auditory and abstract knowledge of a difference between the sounds, which in this study is instantiated as the availability of a difference in word spelling during the word learning phase.

2. Method

2.1. Participants

Fifty Dutch-English bilingual volunteers from the subject pool at the Max Planck Institute for Psycholinguistics were paid to participate in the study. Listeners filled out a background questionnaire prior to performing the study. From these questionnaires, the mean age was computed as 21.1 years, and their mean self-estimated English comprehension as 5.46 (standard deviation = 0.8) on a scale between 0 (no understanding) and 7 (understanding as a native speaker), which suggests that they considered themselves to be highly proficient in their understanding of the English language. It should be noted that these listeners were drawn from a comparable pool of subjects as those recruited by Weber and Cutler (2004).

2.2. Auditory materials

A male native speaker of Standard Southern British English produced twenty sentences of the form “Click on the [target word] and then on the [geometric form]”. The target words were bisyllabic English nonwords; all conformed to English phonotactic restrictions and were parallel in form to the bisyllabic words presented in Weber and Cutler (2004; see Table 1 for examples). Five target-word pairs were created, where the words in each pair overlap in the first syllable with the exception of whether the vowel is /ɛ/ or /æ/ (e.g., [tɛnzə], [tændək]). Each of the ten

target words was also paired with a control word that was identical to the target with the exception of the vowel in the first syllable, which was always /u/ (e.g., [tɛnzə]-[tunzə]; [tændək]-[tundək]). These control words were included for two reasons: (1) to encourage subjects to pay attention to the quality of the vowels in the first syllables of the words and not just to the consonants, and (2) to provide a set of control data using two relatively perceptible vowel contrasts (*i.e.*, /ɛ/-/u/ and /æ/-/u/) against which to compare subjects' performance on the /æ/-/ɛ/ contrast. Note that Dutch listeners can easily perceive the difference between English /ɛ/ and /u/ and between English /æ/ and /u/, as evidenced by the finding that they almost never confused these two vowel pairs in Cutler *et al.*'s (2004) identification task (p. 3675). The geometric forms were “triangle”, “square”, “circle”, and “diamond”, and were included following Weber and Cutler (2004) and Shatzman and McQueen (2006).

	/ɛ/ words		/æ/ words	
	Column A (/ɛ/ target words)	Column B (/ɛ/ control words)	Column C (/æ/ target words)	Column D (/æ/ control words)
Row 1	[tɛnzə] 'tenzer'	[tunzə] 'toonzer'	[tændək] 'tandek'	[tundək] 'toondek'
Row 2	[mɛskəl] 'meskle'	[muskəl] 'mooskle'	[mæstək] 'mastik'	[mustək] 'moostik'
Row 3	[gɛbət] 'gebbet'	[gubət] 'goobet'	[gæbəl] 'gabble'	[gubəl] 'gooble'
Row 4	[hɛstəl] 'hestel'	[hustəl] 'hoostel'	[hæskəm] 'haskum'	[huskəm] 'hooskum'
Row 5	[bɛstət] 'bestet'	[bustət] 'boostet'	[bæskəl] 'baskle'	[buskəl] 'booskle'

Table 1. *Target and control words and their spelled forms*

2.3. Visual materials

Twenty line drawings of nonobjects were selected from the same data base of non-objects used by Shatzman and McQueen (2006). These pictures were randomly associated with the twenty target and control word stimuli discussed above. Two versions of each picture were created: One of the pictures alone, and one of the pictures with its spelled form below it (see Figure 1 for examples). There were two word learning conditions. In the Auditory Forms Only word learning condition, subjects were provided only with the auditory forms and the pictures

(see Figure 1A). In the Auditory + Spelled Forms word learning condition, subjects were additionally shown the spelled forms of the nonwords (e.g., ‘tenzer’; see Figure 1B). A complete list of pictures is presented in the Appendix.

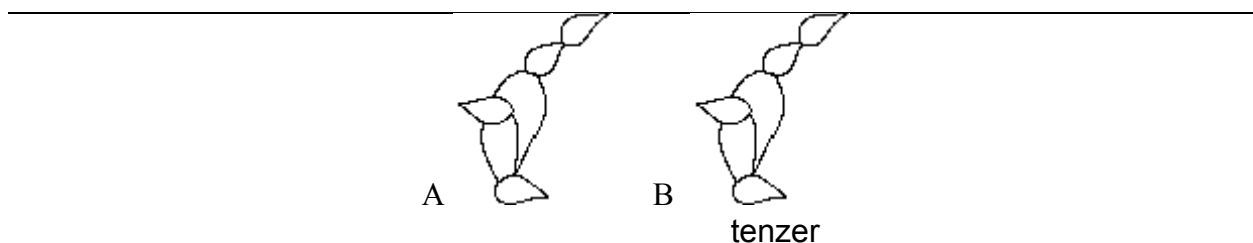


FIGURE 1. The nonobject associated with the auditory form [tɛnzə], in both conditions. In (A) the Auditory Forms Only condition, and in (B) Auditory + Spelled Forms condition.

2.4. Procedure

The entire experiment session lasted approximately one hour. During the experiment, participants were addressed only in English, were told that they were going to learn English words, and were instructed that they would be subsequently tested on how well they had learned these new English words. Additionally, the stimuli they heard were produced by a native speaker of Standard Southern British English and, as mentioned above, consisted of the novel words to be learned embedded in an English carrier sentence. The stimuli contained a number of sounds which can not be found in Dutch and are always associated with English by these highly proficient English learners, namely [t^h] in ‘tenzer’ and [g] in ‘goobet’. This ensured that the participants remained in an English language mode throughout the experiment. Importantly, sound perception and word recognition studies have shown that when L2 learners (especially very advanced ones) are tested entirely in their L2 they have much less interference from their L1 than when they are tested in both languages (Escudero & Boersma 2002, Kroll & Sunderman 2003, and Marian & Spivey 2003). Therefore, we believe that the subjects treated the words they learned as English novel words and that they therefore would employ a similar learning mechanism to the present word learning task as they usually employ in any L2 word learning situation.

2.4.1. Word Learning Phase

A challenge associated with studies of novel word learning is determining a criterion for “lexicalization,” or at the point at which the episodic memories associated with new words become actual lexical items. There is some debate in the literature on word learning over precisely when lexicalization takes place. For example, Dumay and Gaskell (2007) present evidence that a period of sleep intervening between word learning and testing is necessary for newly-learned native language words to participate in lexical competition in word recognition tasks. However, the Shatzman and McQueen (2006) results discussed above indicate that even within a relatively short experiment session (~one hour), with no period of sleep intervening between word learning and testing, subjects transfer the abstract prosodic patterns associated with their native language to newly-learned native language words. In addition, in a study of novel L2 word learning, Hayes-Harb and Masuda (to appear) demonstrated that native English-speaking learners of Japanese were able to associate Japanese nonword labels and pictures during an hour-long experiment session (although their performance was more accurate for nonwords containing only familiar phonemes than for nonwords containing novel Japanese phonemes). Together these studies indicate that even though it is not clear exactly when lexicalization takes place, and that the time course of lexicalization depends on what criterion is used, some aspects of lexicalization can occur during a relatively short experiment session. In the present study, the word learning phase is modeled after that of Shatzman and McQueen (2006), but on the basis of findings from pilot experiments, had four extra training blocks of 60 trials each, thus an extra 240 trials (600 instead of 360 trials), in order to compensate for the fact that the words in the present study were not native to the subjects.

Subjects were tested individually. The experiment was controlled by a Compaq 486 computer; pictures were presented on a ViewSonic 17PS screen, and the auditory stimuli were presented over headphones using the NESU software (Wittenburg, Nagengast, & Baumann 1998; see <http://www.mpi.nl/tools/nesu.html>). In the word learning phase, each of the twenty words listed in Table 1 appeared three times per block. There were ten word learning blocks in total consisting of 60 trials each. In each trial, the subject first heard, for instance, “Click on the [tɛnzə] and then on the triangle”. Thus, participants were asked to click on the picture that they thought represented the word they had heard and then to click on a geometric form. As soon as they clicked on the geometric form, only the chosen picture remained on the screen (the rest of

the display, including all other picture options and geometric forms, disappeared) and feedback was provided. Participants received two types of feedback. The first type of feedback was determined by whether the subject clicked on the correct picture (*i.e.*, “that was right” or “that was wrong”). The second type of feedback was “Here you can see the X again”, where X is the word provided in the instructions, which was presented together with a display of the correct picture. During this second type of feedback, in the Auditory Forms Only condition, only the picture remained on the screen; in the Auditory and Spelled Forms condition, the picture and spelled form were on the screen when this feedback was played.

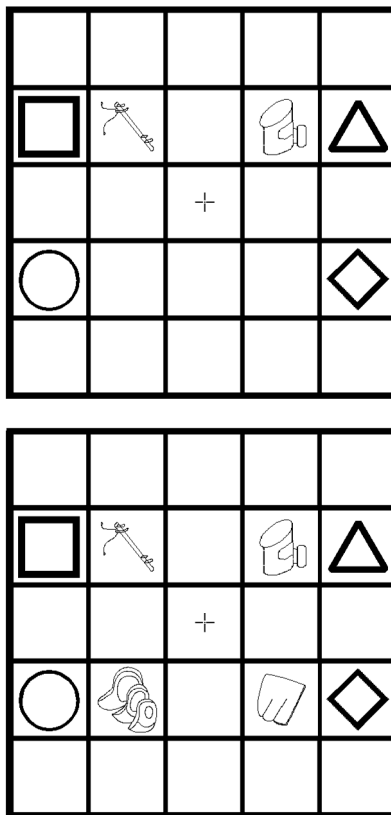


FIGURE 2. Task display in the learning task. The upper panel shows a display from the two-way forced-choice learning task, the lower panel a display from the four-way forced-choice learning task.

In each word learning trial, two (or four, depending on block number; see below) pictures appeared on the screen. In the first five word learning blocks (1-5), the picture for the target word was accompanied by only one distractor picture, selected randomly from a different row in Table 1 (*e.g.*, if the word to click on was [tɛnzə], the distractor was selected randomly from rows

2-5), as demonstrated in the upper panel of Figure 2. The pictures were presented on a 5x5 grid with equal distances between each cell in the grid. The pictures presented during these first five blocks could have one of the two positions shown in the figure, while the position of the geometric forms was fixed throughout the word learning and testing phases.

In word learning blocks 6-10, to increase the difficulty of the task as subjects became better able to recognize the words, subjects performed the same task as in blocks 1-5, only now four pictures appeared on the screen (the target plus three distractor pictures, as shown in the lower panel of Figure 2). Just as in the first five training blocks, the distractors were taken from any of the columns in Table 1 but from different rows than the row of the stimulus in order to avoid having the similar-sounding competitors in the learning phase. For these word learning blocks, the same 5x5 grid was used for display but this time the location of the pictures could vary between the four cells shown in the figure.

Thus the only difference between the two word learning conditions is that in the Auditory + Spelled Forms condition, when all but the correct picture disappeared from the screen and the feedback sentences were played, the written form for the pictured object appeared underneath the picture.

2.4.2. Testing Phase

Immediately following completion of the word learning phase, subjects were fitted with the eye tracker (SMI EYELINK 1 software version 2.1, SensoMotoric Instruments GmbH, Teltow, Germany; sampling at 250 Hz). This head mounted eye-tracker has one camera each to track the right and left pupil. The position of these cameras relative to the screen is monitored with four tracking points, so that the position of the fixation on the screen can be estimated. After a brief calibration phase, the test trials started. The test phase examined subjects' ability to identify the correct picture among not only distractor pictures, but also to distinguish it from the picture of its competitor. A target's competitor is the other target from the same row. For example, [tɛnzə] and [tændək] are each other's competitors, as they have first syllables that are identical except for the vowel.

Of the twenty test trials in each block, ten were experimental trials. In five of the experimental trials, the picture which the subjects were instructed to click on came from column A in Table 1 (containing /ɛ/) and the competitor from column C (containing /æ/) in the same

row; in the other five experimental trials, the pictures to click on came from column C and the competitor from column A of the same row. A first distractor was randomly chosen from another row. If the first distractor was a member of a /æ/-/ɛ/ pair (*i.e.*, from either column A or C), the other member of the pair was used as the second distractor. If the first distractor had a /u/ vowel (*i.e.*, from either column B or D), the other member of the pair served as the second distractor. In this way, there would always be two "pairs" appearing on one screen display, and the participants could not predict which pair was more likely to contain the eventual target. The other ten test trials were filler trials. In five of the filler trials, the picture to click on came from column B (containing /u/) and the competitor from column C in the same row, and in the other five, the picture to click on came from column D (containing /u/) and the competitor came from column A in the same row. Again, two distractors were chosen randomly from other rows. The entire block of test trials was presented three times, with different positions on the screen for target and competitor pictures and in a different random order for each repetition of a block of 20 items. Subjects responded to the same stimuli multiple times so that an average across repetitions could be computed in order to avoid that any one response determined the whole of the data for each item. It is worth mentioning that previous research has not found negative influence of repetition in an eye-tracking paradigm (Allopenna, Magnuson, & Tanenhaus 1998).

The procedure for the test trials is identical to that of the training phase blocks 6-10 (where four pictures appear on the screen which could vary between the same four cell positions).

2.5. Analysis

2.5.1. Word Learning Phase

First we discuss the word learning accuracy data. The data were aggregated as percentages correct over blocks for subjects and items separately. For statistical analysis, the percentages were arcsine-transformed, with word learning condition (two levels: Auditory Forms Only and Auditory + Spelled Forms) and block number as independent variables.

2.5.2. Testing Phase

In addition to percent correct data as described above, in the testing phase, eye-position data is also considered. Eye-positions were sampled at 250 Hz and time-locked to the onset of the word that subjects were instructed to click on. The fixation proportions to the different objects on the screen were analyzed using two independent variables: word learning condition (two levels: Auditory Forms Only and Auditory + Spelled Forms) and target vowel (two levels: /æ/ and /ε/). The impact of these factors was evaluated for looks to the targets and competitor effects (comparing competitors and distractors) in subject- and item-analysis (see, *e.g.*, Dahan & Tanenhaus 2004; Huettig & McQueen, in press). Note that word learning condition is a between-subjects variable in the subject-analysis, but a within-subjects variable in the item-analysis, while target vowel is a between-items and a within-subjects variable. The dependent variable was the fixation proportions over a window determined by the overall data pattern. Similar as in the data of Weber and Cutler (2004) and Cutler *et al.* (2006), the strongest competitor effects are visible in the time window from 300 ms to 700 ms (indicated by vertical lines in all figures). Given that programming a saccade takes roughly 200 ms (*e.g.*, Fischer, 1992), 300 ms after target onset is approximately the point at which fixations driven by the first 100 ms of acoustic information from the target word can be seen. Fixation proportions were calculated with blinks and saccades as missing data, but include fixation to no object and hence do not necessarily add up to one. A fixation was counted to be "on" an object if it fell within the margins of the grid containing this object.

2.6. Results

2.6.1. Word Learning Phase

Figure 3A shows the mean percentages correct for all subjects for the first part of the word learning phase, where subjects had two alternatives. The figure suggests unsurprisingly that the participants improve block-by-block and that learning is easier for subjects in the Auditory + Spelled Forms condition. These observations are borne out by statistical analysis: There is a significant main effect of block number ($F(4, 192) = 110.0, p < 0.001$; $F(4, 76) = 138.0, p < 0.001$) and a significant main effect of word learning condition ($F(1, 48) = 5.1, p < 0.05$; $F(1, 19) = 29.9, p < 0.001$), but no interaction of the two by either subjects or items ($F_s < 1$). In the second part of the word learning phase (with four alternatives; see Figure 2-Panel B for a

stimulus and Figure 3-Panel B for the results), there is a significant main effect of block number ($F1(4, 192) = 15.7, p < 0.001$; $F2(4, 76) = 16.2, p < 0.001$), while the effect of word learning condition is marginally significant by subjects ($F1(1, 48) = 3.6, p = 0.06$; $F2(1, 19) = 40.5, p < 0.001$). The interaction is significant neither by subjects nor by items ($F1(4, 192) = 2.2, p = 0.07$; $F2(4, 76) = 1.0, p > 0.1$). Figure 3B seems to suggest the presence of an interaction, but the arcsine-transform extends the range near the 100% correct ceiling, so that the interaction is not significant. These findings indicate that, though learning was clearly facilitated by the availability of the spelled forms of words for subjects in the Auditory + Spelled Forms condition in the two-alternative task, this advantage was only marginally significant in this second task after the participants started to be familiar with the items. We also checked whether performance just before testing differed between items containing /æ/ and /ɛ/ (92.8% correct) and fillers (91.2% correct); this was not the case ($|t|s < 1$).

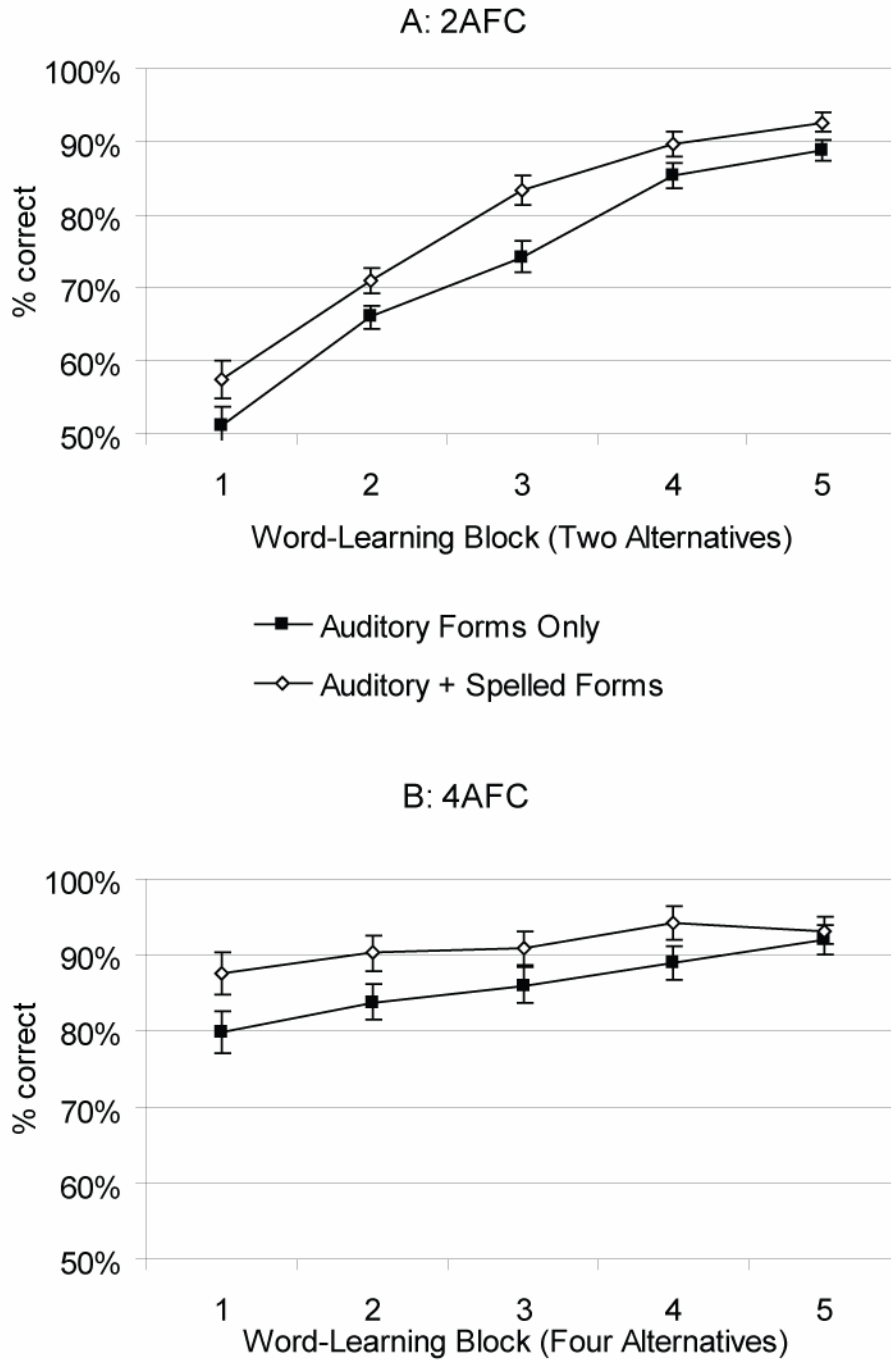


FIGURE 3. Word learning phase data percent correct data: (A) shows the two-alternative word learning blocks and (B) the four-alternative word learning blocks. The error bars indicate the standard error of the mean.

2.6.2. Testing Phase

Two participants were excluded from further analysis because their average percentage correct did not exceed 70% in the final word learning block. This leaves 48 subjects (24 in each word learning condition) for analysis. Performance in the testing phase showed no decrement in accuracy, which remained in the vicinity of 90% correct (see Figure 4). The accuracy is slightly higher in the Auditory + Spelled Forms Learning condition than in the Auditory Forms Only learning condition (92% versus 89%). This 3% accuracy advantage for the Auditory + Spelled Forms Condition in Figure 4 was significant by items ($F(2, 19) = 7.75, p < 0.05$), but not by subjects ($F(1, 48) = 1.92, p = 0.17$).

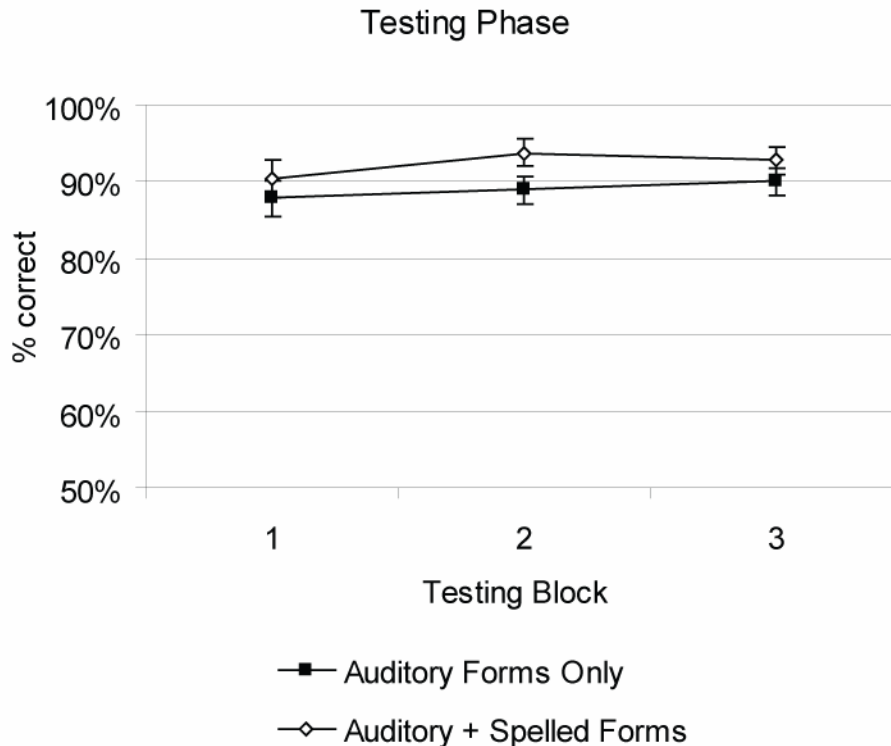


FIGURE 4. Testing phase percent correct data by block number and word learning condition. The error bars indicate the standard error of the mean.

Figure 5 shows the results from the eye-tracking data from the remaining 48 subjects during the test session for the critical experimental trials in which the target was a novel word with either an /æ/- or an /ε/-vowel. The fixation proportions were averaged with a sliding average window with a width of 20 ms (= 5 data points). The function for the looks to the distractors is

the average over both distractor pictures (e.g., [tunzə] or [tundək]). Figure 6 shows the same data but now directly comparing targets and competitors from all four combinations of learning condition and target vowel. In both figures, vertical lines at 300 ms and 700 ms indicate the window used for the statistical analysis below (cf. Cutler *et al.*, 2006, Weber & Cutler 2004).

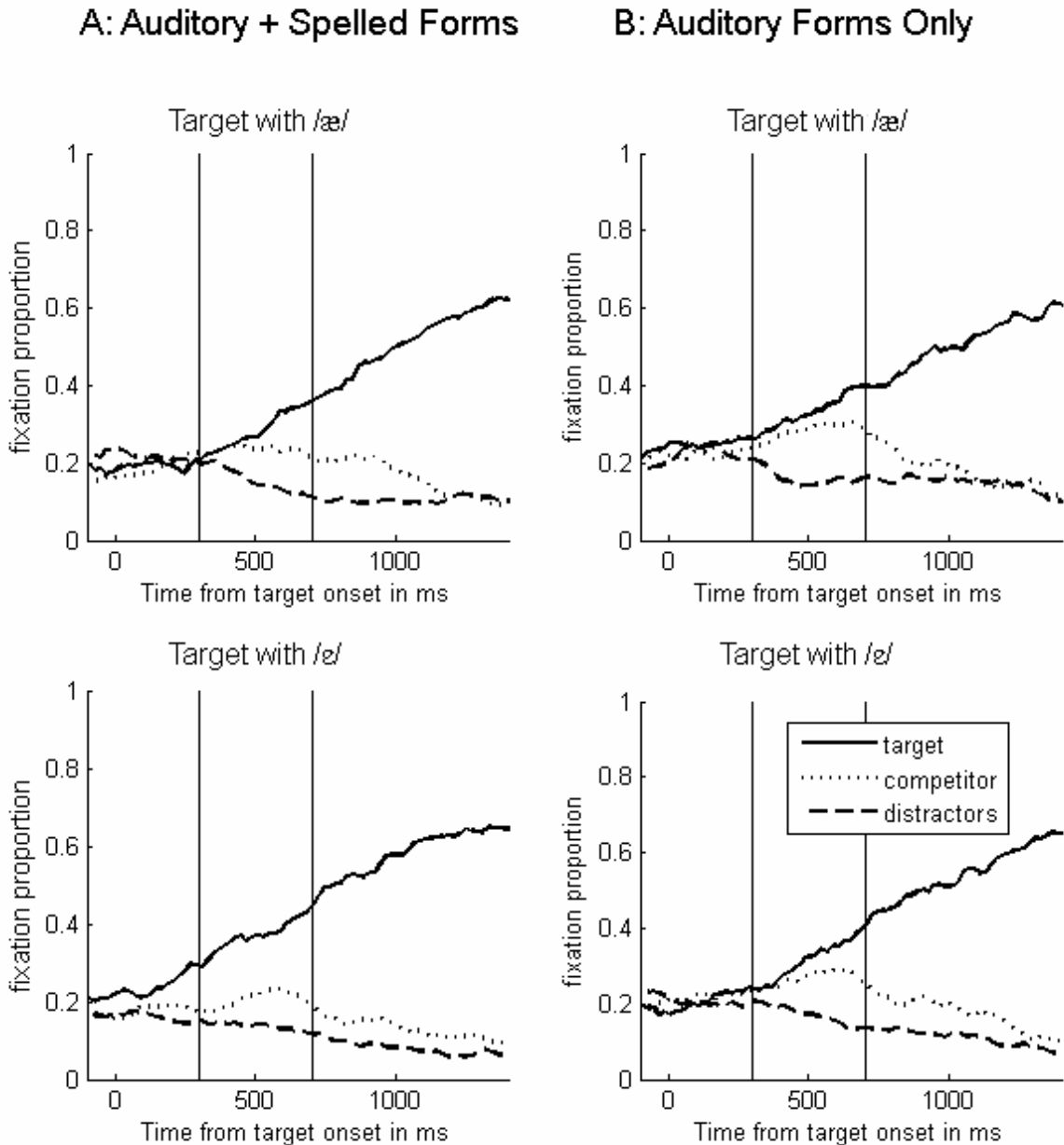


FIGURE 5. Fixation proportions to targets, competitors, and distractors (averaged for the two distractors in each display) with the different learning conditions in different panels. The two vertical lines indicate the 300-700 ms time-window used for the statistical analysis.

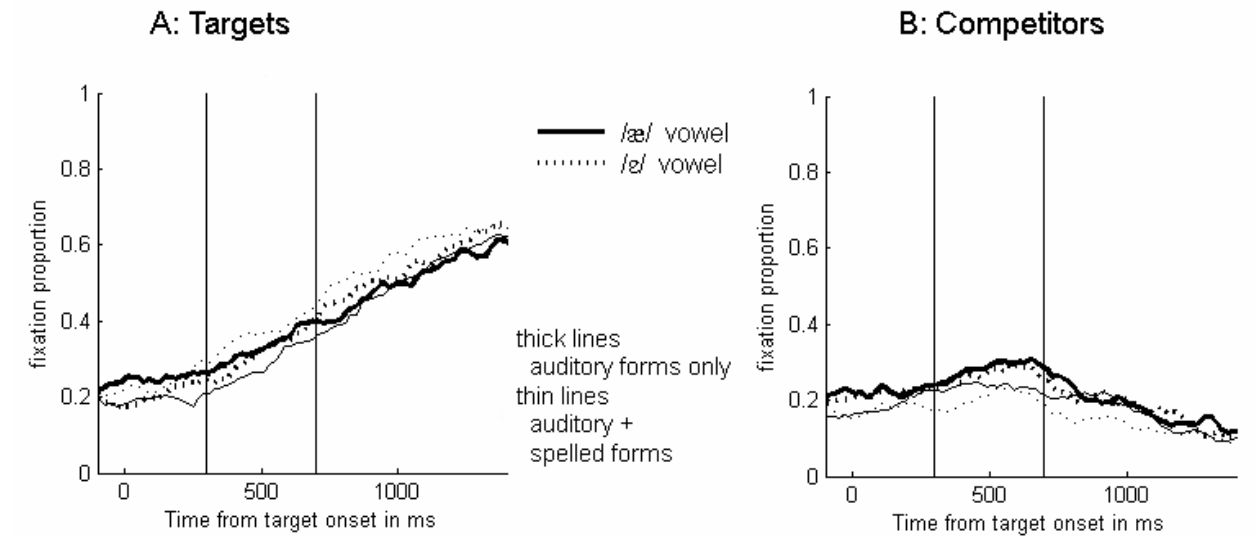


FIGURE 6. Fixation proportions to targets (A) and competitors (B) for the two learning conditions.

As Figure 6 shows, both the /æ/-targets and /ɛ/-targets (e.g., [tændək] and [tɛnzə]) received a similar amount of looks if learned with an auditory form only. The functions for both targets and competitors are virtually on top of each other in the analysis window, as they both rise from a fixation proportion of about 25% between 0 and 200 ms after target onset (before the speech signal can influence the looks) to 60% after 1200 ms after word onset. Adding the spelled forms during the learning condition had a different effect on the trials with /æ/-targets (e.g., [tændək]), and on the trials with /ɛ/-targets (e.g., [tɛnzə]). The /ɛ/-targets became more attractive and received more fixations in the analysis window if learned with auditory and spelled forms, while adding the spelled forms reduced the number of fixations to the /æ/-targets. Adding the spelled forms overall reduced the attractiveness of the competitors as well, but this reduction is more pronounced for the /æ/-competitors, e.g., [tændək], for /ɛ/-targets, e.g., [tɛnzə], than for /ɛ/-competitors for /æ/-targets.

	Learning Condition			
	Auditory Forms Only		Auditory + Spelled Forms	
	Target Vowel		Target Vowel	
Object	/æ/	/ɛ/	/æ/	/ɛ/
Target	26.4%	25.2%	21.6%	29.8%
Competitor	24.3%	22.9%	21.5%	18.7%
Distractors	12.5%	13.7%	13.1%	10.7%

Table 2. Mean fixation proportions on targets, competitors, and distractors in the time window 300-700 ms depending on Learning Condition and Target Vowel.

We calculated the mean of the early fixations in a 300-700 ms window (*cf.* Cutler *et al.*, 2006, Weber & Cutler, 2004) over subjects and items. Table 2 shows the fixation proportions to the targets, competitors, and distractors in this window, that is, the mean "height" of the fixation functions in the 300-700 ms interval. An ANOVA on the target fixations revealed an interaction between Learning Condition and Target Vowel ($F(1, 46) = 4.79, p < 0.05$; $F(1, 8) = 7.48, p < 0.05$), while the main effects of Learning Condition ($F_s < 1$) and Target Vowel ($F(1, 46) = 2.63, p > 0.1$; $F(1, 8) = 2.40, p > 0.1$) were not significant. To investigate the nature of the interaction, we computed *t*-tests for the effect of Target Vowel for both Learning Conditions (*cf.* Keppel 1991). These tests showed that there were significantly more looks to the /ɛ/-target than to the /æ/-target in the Auditory + Spelled Forms condition ($t(1, 23) = 2.91, p < 0.01$; $t(1, 8) = 2.6, p < 0.05$), while there was no preference in the Auditory Forms Only condition ($|t|s < 1$). That is, participants have a preference for the /ɛ/-target (*e.g.*, [tɛnzə]) only if orthographic information indicated that the first vowel in the /æ/-targets (*e.g.*, 'tandik') is different than the vowel in the /ɛ/-targets (*e.g.*, 'tenzer').

To test whether a similar effect occurs with the competitors, we compared the looks of competitors (*e.g.*, [tɛnzə] or [tændək]) and distractors (*e.g.*, [tunzə] or [tundək]) in a 3-way ANOVA with Role (two levels: Competitor vs. Distractor), Learning Condition, and Target

Vowel. This analysis revealed a significant effect of Role ($F1(1, 46) = 64.2, p < 0.001; F2(1,8) = 9.45, p < 0.05$), with more looks to competitors (21.8%) than to distractors (12.5%). The effect of Condition was significant only by items ($F1(1, 46) = 2.31, p > 0.1; F2(1,8) = 41.65, p < 0.001$), and all other effects and interactions failed to reach significance in both subject- and item-analyses. This indicates that the advantage for the /ɛ/-competitors over /æ/-competitors visible in Figure 6 above was not significant.

One possible explanation for the stronger attractiveness of the /ɛ/-targets (e.g., [tɛnzə]), if learned with spelled forms, could be that these pictures were somehow more attractive if learned with auditory and spelled forms. Therefore, we compared the looks to /æ/-objects (e.g., the [tændək]), and /ɛ/-objects (e.g., the [tɛnzə]), when they served as distractors on trials in which one of the fillers was the target. Figure 7 shows that the /ɛ/-objects were less attractive in the Auditory + Spelled Forms learning condition (15%) than in the Auditory Forms Only condition (19%), while the opposite was true for /æ/-objects (15% vs. 13%). Although these effects are not significant ($F1(1,16) = 3.6, p > 0.05, F2(1,8) = 3.0, p > 0.1$), they are in the opposite direction of the effects on targets, and hence they show that the effects on targets discussed above cannot be explained by an unexpected effect of learning condition on the baseline attractiveness of the pictures.

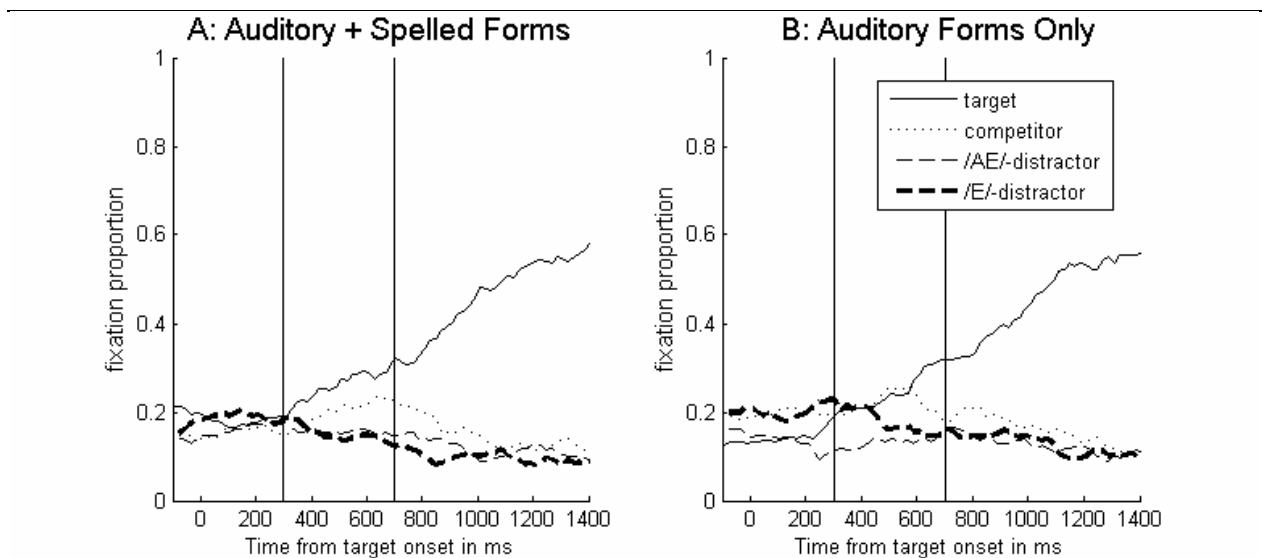


FIGURE 7. Fixation proportions in the filler trials for targets, competitors and the two distractors containing an /æ/- and /ɛ/-vowel separately, with the different learning conditions in different panels.

3. Discussion

In this study, we set out to test whether the Weber and Cutler (2004) and Cutler *et al.* (2006) asymmetric word recognition findings could be replicated with novel L2 words. To this end, we taught two groups of native Dutch speakers who are highly proficient in English to learn 20 English nonwords. Given that the aim of the study was to measure their word recognition abilities, we first confirmed that the listeners had indeed learned the 20 words. The results from the final block of the word learning phase indicate that L2 learners can build a vocabulary of 20 similar-sounding L2 words in a short period of time, *i.e.*, less than 30 minutes. That is, they can achieve a percentage correct of 90% or above when matching auditorily presented words with their correct picture meanings. Furthermore, this accuracy also transferred to a subsequent word recognition task (the testing phase); this can be taken to mean that L2 learners are able to distinguish among newly-learned words after a very short learning period. Thus, these results suggest that the L2 learners had constructed traces of the newly-learned words in their lexicons, which they could later retrieve during the word recognition task.

Perhaps more interestingly, our study shows that auditorily confusable novel words can also lead to asymmetric confusion in word recognition. Recall that Weber and Cutler (2004) found that advanced Dutch speakers of English exhibited an asymmetric pattern of recognition between words containing the contrast such that words containing /æ/ (*e.g.*, [tændək]) trigger looks to pictures of /æ/- and /ɛ/-words (*e.g.*, [tændək] and [tɛnzə]), while words with /ɛ/ only trigger looks to pictures of /ɛ/-words such as [tɛnzə]). Weber and Cutler (2004) explain this finding by suggesting that these listeners lexically contrast /æ/ and /ɛ/ despite an inability to differentiate the vowels auditorily. The asymmetric pattern reported by Weber and Cutler (2004) was replicated with novel L2 words when words were learned on the basis of their auditory and spelled forms. In this learning condition, the /ɛ/-words were more attractive than the /æ/-words, both as targets and competitors. Statistically, this was only significant for the analysis of fixation to targets, but not in the analysis of fixations to competitors. The lack of a statistical effect for the competitors may be due to the fact that lexical competition effects only arise after learning has consolidated, probably during sleep: Gaskell and Dumay (2003, see also Dumay & Gaskell, 2007) showed that a newly learned word (such as *cathedruke*) only compete for recognition with existing words (in this case, *cathedral*) after a period of sleep.

The present findings also indicate that if listeners are only presented with auditory tokens of novel L2 words (as in the Auditory Forms Only condition), they show symmetric confusion between words containing the /æ/-/ɛ/ vowel contrast. That is, listeners had equal number of fixations to pictures of words containing /æ/ or /ɛ/ when the words to be recognized contained either of the two vowels. This finding provides evidence (in response to, *e.g.*, Cutler *et al.*'s 2006: 280 open question regarding the source of the asymmetry) that knowledge of spelled forms of words can trigger asymmetric lexical activation that is not present for learners who do not have knowledge of spelled forms of words. Thus, Weber and Cutler (2004) and Cutler *et al.*'s (2006) asymmetric pattern of word recognition is not replicated when auditory forms are the sole information available to the listeners. Moreover, the present findings suggest that auditorily confusable L2 words which are learned only on the basis of their auditory forms will not result in a lexical contrast.

We thus conclude that the availability of some source of abstract knowledge about a novel phonemic contrast (in this case, the mapping between spelled forms and phonological forms) can result in establishing of a lexical contrast which can then be used in auditory word recognition. When building an orthographic/graphemic representation, learners must learn to associate a number of different sounding tokens to the same visual representation. They must also learn that there are different styles of writing for a particular grapheme: printed, manually written, etc., which look different but must be associated with the same graphemic representation. Thus, the mapping of auditory tokens to graphemes is an instance of discrete, abstract and arbitrary knowledge, because there is no obvious link between, for instance, the visual form 'p' and the closing gestures of the lips, which is necessary to produce the corresponding phoneme /p/. This is because different visual and auditory tokens in the world are arbitrarily and consistently classified as the same form, a grapheme or a phoneme respectively. Therefore, we conclude that our finding reflects the use of abstract knowledge in establishing the *phonological* lexical representation for novel L2 words.

Additionally, we have provided evidence that an orthographic/graphemic difference (reflecting a phonemic contrast) which exists in the learners' native language can be used or transferred to the learning of L2 words after a short training period. This suggests a tight link between phonological and orthographic representations, a link that has previously been

emphasized from a different angle—for gaining lexical access from a written input—in the field of visual-word recognition (see Van Orden 1987 for a classic example).

4. References

- Alloppenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language*, 38, 419-439.
- 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, 233-250.
- Bion, R. A. H., Escudero, P., Rauber, A. S., & Baptista, B.O. (2006). Category formation and the role of spectral quality in the perception and production of English front vowels. *Proceedings of Interspeech 2006*, 1363-1366.
- Broersma, M. (2005). Perception of familiar contrasts in unfamiliar position. *Journal of the Acoustical Society of America*, 117, 3890-3901.
- Curtin, S., Goad, H., & Pater, J.V. (1998). Phonological transfer and levels of representation: The perceptual acquisition of Thai voice and aspiration by English and French speakers. *Second Language Research*, 14, 4, 389-405.
- Cutler, A., & Broersma, M. (2005). Phonetic precision in listening. In W. J. Hardcastle and J. Mackenzie Beck (Eds.), *A Figure of Speech: a Festschrift for John Laver* (pp. 63-91). London: Lawrence Earlbaum.
- Cutler, A., & Otake, T. (2004). Pseudo-homophony in non-native listening. *Journal of the Acoustical Society of America*, 115, 2392.
- Cutler, A., Weber, A., & Otake, T. (2006). Asymmetric mapping from phonetic to lexical representations in second-language listening. *Journal of Phonetics*, 34, 269-284.
- Cutler, A., Weber, A., Smits, R., & Cooper, N. (2004). Patterns of English phoneme confusions by native and non-native listeners. *Journal of the Acoustical Society of America*, 116, 3668-3678.
- Dahan, D., & Tanenhaus, M. K. (2004). Continuous mapping from sound to meaning in spoken-language comprehension: Immediate effects of verb-based thematic constraints. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30, 498-513.
- Deterding, D. (1997). The Formants of Monophthong Vowels in Standard Southern British English Pronunciation. *Journal of the International Phonetic Association*, 27, 47-55.
- Dumay, N., & Gaskell, M.G. (2007). Sleep-associated changes in the mental representation of spoken words. *Psychological Science*, 18, 35-39.
- Escudero, P., & Boersma, P. (2002). The subset problem in L2 perceptual development: Multiple category assimilation of Dutch learners of Spanish. In B. Skarabela, S. Fish, A. H.-J. Doh

- (eds.), *Proceedings of the 26th Boston University Conference on Language Development* (pp. 208-219). Mahwah, MA: Cascadilla.
- Fischer, B. (1992). Saccadic reaction time: Implications for reading, dyslexia and visual cognition. In K. Rayner (Ed.), *Eye Movements and Visual Cognition: Scene Perception and Reading* (pp. 31–45). New York: Springer Verlag.
- Flège, J. E., O.-S. Bohn, & Jang, S. (1997). Effects of experience on non-native speakers' production and perception of English vowels. *Journal of Phonetics*, 25, 437–470.
- Gaskell, G. M., & Dumay, N. (2003). Lexical competition and the acquisition of novel words. *Cognition*, 89, 105-132
- Hayes-Harb, R. (2005). Optimal L2 speech perception: Native speakers of English and Japanese consonant length contrasts. *Journal of Language & Linguistics* 4, 1, 1-29.
- Hayes-Harb, R., & Masuda, K. (to appear). Development of the ability to lexically encode novel L2 phonemic contrast. *Second Language Research*.
- Huetig, F., & McQueen, J. M. (in press). The tug of war between phonological, semantic, and shape information in language-mediated visual search. *Journal of Memory and Language*
- Keppel, G. (1991). *Design and analysis: a researcher's handbook* (3rd edition). Upper Saddle River, NJ: Prentice Hall.
- Kroll, J. F., & Sunderman, G. (2003). Cognitive processes in second language acquisition: The development of lexical and conceptual representations. In C. Doughty & M. Long (Eds.), *Handbook of second language acquisition* (pp. 104-129). Cambridge, MA: Blackwell Publishers.
- Marian, V., & Spivey, M. (2003). Competing activation in bilingual language processing: Within- and between-language competition. *Bilingualism*, 6, 2, 97-115.
- Pallier, C., Colomé, A., & Sebastián-Gallés, N. (2001). The influence of native-language phonology on lexical access: Exemplar-based versus abstract lexical entries. *Psychological Science*, 12, 445-449.
- Pater, J. (2003). The perceptual acquisition of Thai phonology by English speakers: task and stimulus effects. *Second Language Research*, 19; 209–223.
- Peterson, G.E., & Barney, H.L. (1952). Control methods used in a study of the vowels, *Journal of the Acoustical Society of America*, 24, 175-184.
- Pols, Louis C.W., Tromp, H.R.C., & Plomp, R. (1973). Frequency analysis of Dutch vowels from 50 male speakers. *Journal of the Acoustical Society of America*, 53, 1093-1101.
- Rauber, A. S., Escudero, P., Bion, R. A. H., & Baptista, B. O. (2005). The interrelation between the perception and production of English vowels by native speakers of Brazilian Portuguese. *Proceedings of Interspeech 2005*, 2913-2916.
- Salverda, A.P., Dahan, D., & McQueen, J.M. (2003). The role of prosodic boundaries in the resolution of lexical embedding in speech comprehension. *Cognition*, 90, 51–89.
- Shatzman, K.B., & J.M. McQueen. (2006). Prosodic knowledge affects the recognition of newly acquired words. *Psychological Science*, 17, 5, 372-377.

Van Orden, G. C. (1987). A ROWS is a ROSE: spelling, sound, and reading. *Memory and Cognition*, 15, 181–198.

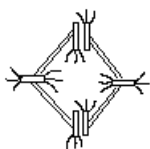
Weber, A., & Cutler, A. (2004). Lexical competition in non-native spoken-word recognition. *Journal of Memory and Language*, 50, 1-25.

Wittenburg, P., Nagengast, J., & Baumann, H. (1998). NESU-the Nijmegen experiment setup. In A. Trapp, N. Hammond & C. Manning (Eds.), *CIP98 conference proceedings* (pp. 92-93). York, England: CTI Centre for Psychology.

5. APPENDIX



baskle



bestet



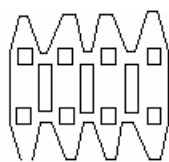
booskle



gabble



gebbet



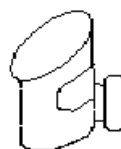
goobet



gooble



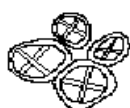
haskum



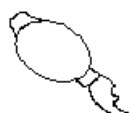
hestel



hooskum



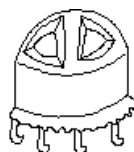
hoostel



mastik



meskle



mooskle



moostik



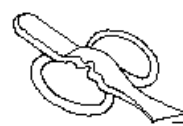
tandek



tenzer



toondek



toonzer