Available online at www.sciencedirect.com

SCIENCE DIRECT®

Journal of Phonetics 34 (2006) 269–284

Journal of Phonetics

www.elsevier.com/locate/phonetics

Asymmetric mapping from phonetic to lexical representations in second-language listening

Anne Cutler[a,*], Andrea Weber[b], Takashi Otake[a]

[a] *Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands*
[b] *Saarland University, Saarbrücken, Germany*

Received 16 December 2004; received in revised form 14 June 2005; accepted 23 June 2005

Abstract

The mapping of phonetic information to lexical representations in second-language (L2) listening was examined using an eyetracking paradigm. Japanese listeners followed instructions in English to click on pictures in a display. When instructed to click on a picture of a rocket, they experienced interference when a picture of a locker was present, that is, they tended to look at the locker instead. However, when instructed to click on the locker, they were unlikely to look at the rocket. This asymmetry is consistent with a similar asymmetry previously observed in Dutch listeners' mapping of English vowel contrasts to lexical representations. The results suggest that L2 listeners may maintain a distinction between two phonetic categories of the L2 in their lexical representations, even though their phonetic processing is incapable of delivering the perceptual discrimination required for correct mapping to the lexical distinction. At the phonetic processing level, one of the L2 categories is dominant; the present results suggest that dominance is determined by acoustic–phonetic proximity to the nearest L1 category. At the lexical processing level, representations containing this dominant category are more likely than representations containing the non-dominant category to be correctly contacted by the phonetic input.

[*] Corresponding author. Tel.: +31 24 352 1377; fax: +31 24 352 1213.
E-mail address: anne.cutler@mpi.nl (A. Cutler).

0095-4470/$ - see front matter © 2005 Elsevier Ltd. All rights reserved.
doi:10.1016/j.wocn.2005.06.002

1. Introduction

A Google search on *flied lice* (May 15, 2005) yields 856 hits. As this suggests, such forms are part of our general cultural knowledge: we are all too aware that second-language users are often unable to distinguish certain phonetic contrasts. Second-language phoneme perception problems have received widespread attention in speech research over several decades (see, for instance, the papers in Strange, 1995). Less attention, however, has been paid to the question of what these problems actually mean for communication in a second language (L2). For instance, just how is L2 word recognition affected by phoneme discrimination difficulty?

The *flied lice* example evokes the phenomenon that many listeners to English (e.g. from Japan or other Asian countries) have difficulty distinguishing utterances containing /r/ from utterances containing /l/. Dutch and German listeners to English similarly have difficulty distinguishing utterances containing /æ/ from utterances containing /ɛ/. Does this imply that minimal pairs of English words such as *write* and *light* are effectively homophones for Japanese listeners, and pairs such as *cattle* and *kettle* are effectively homophones for Dutch and German listeners? Certainly there is evidence from lexical decision experiments that is consistent with this interpretation. In the (auditory) lexical decision task, listeners hear spoken forms and decide as rapidly as possible whether or not each form is a real word. In this task, repetition effects are extremely robust (i.e., response time is always faster to an item which occurs for a second time), and there is evidence that minimal pairs behave like repetitions for listeners who cannot discriminate the contrasts which distinguish them. Thus Dutch listeners' responses to *kettle* are faster in the sequence *bromp cattle magazine top revare kettle* than in the sequence *bromp window magazine top revare kettle*, and Japanese listeners' responses to *light* are faster in the sequence *pin write plorve napkin light* than in the sequence *pin pause plorve napkin light* (Cutler & Otake, 2004). The same kind of repetition priming for minimal pairs is observed in Spanish-dominant Spanish–Catalan bilinguals who have difficulty with some contrasts which are distinctive in Catalan but not in Spanish (Pallier, Colomé, & Sebastián-Gallés, 2001).

However, L2 users are often painfully aware that there should be a difference between such minimal pairs, and aware for instance that they fail to convey the necessary distinctions when they utter the words. Moreover, at least in the case of the English distinctions described above, the pairs are distinguished by spelling, and whatever proficient L2 users may perpetrate in pronunciation, as far as we know they rarely write about *kettle-rustlers* or *letter-lighting*. Thus it is reasonable to propose that L2 users' stored lexical representations of minimal-pair words encode the phonetic distinction in some form, even if speech perception cannot reliably apprehend it.

Indeed, in a recent study (Weber & Cutler, 2004) we observed evidence consistent with the suggestion that a difference between two phonemes which collapse to a single category for L2 listeners may be encoded by those same listeners at the lexical level.

The task we used was one which is sensitive to spoken-word recognition as it proceeds across time. Whereas lexical decision tasks (as used in the repetition priming experiments of Cutler & Otake, 2004 and Pallier et al., 2001) provide evidence of the results of lexical recognition, they do not offer a window into the spoken-word recognition process as it unfolds. The eyetracking paradigm does provide such a window. In this paradigm, listeners wear a head-mounted camera which tracks the movements of their eyes as they carry out simple spoken instructions, e.g. to click on one of several objects in a display. The task was introduced to spoken-word recognition research by Tanenhaus, Spivey-Knowlton, Eberhard, and Sedivy (1995), who were concerned with changes in the effective competitor population of words as a function of unfolding phonetic input to listeners. They observed that a display containing, for example, both a candy and a candle attracts looks to both these objects as listeners hear *Click on the cand-*. The task is very simple, and listeners will of course look at the specified target object (say, the candle) as later incoming phonetic information distinguishes between the two competitors. The interesting aspect of the eyetracking methodology is that it provides a window into the listeners' processing before it is fully certain which word is being heard, when alternative word candidates still compete for recognition. It has been shown that such competition, defined as fixation proportions to pictures, closely maps to activation levels of word candidates as simulated in models of spoken-word recognition such as TRACE (Allopenna, Magnuson, & Tanenhaus, 1998; Dahan, Magnuson, & Tanenhaus, 2001).

If two words such as *cattle* and *kettle* or *write* and *light* are homophonous, i.e., phonetically indistinguishable, both should compete with each other. Of course, homophones are not generally used in eyetracking experiments; confronted with a display containing a fir and a fur, and instructed to click on one of them, subjects might rightly object that their task was unclear. But parts of words can be potentially homophonous too. Thus the beginning portions of *panda* and *pencil* or *rocket* and *locker* could be effectively indistinguishable for Dutch and for Japanese listeners, respectively, and only the later parts of the words would disambiguate the input (just as we saw with *candy* and *candle*).

Our recent eyetracking study suggested, however, that simple homophony was not what L2 listeners experienced. We presented Dutch listeners (and also native British–English listeners) with spoken English instructions to click on items in a display containing, for example, a panda and a pencil, or a palace and a pelican. We found that when instructed to click on the panda, the Dutch listeners were likely to look at the pencil during the first syllable of the target word in the instruction. English listeners did not do this. However, when instructed to click on the pencil, the Dutch listeners did not experience such competition from a picture of a panda, and their performance more closely resembled that of the native English listeners. In other words, the interference was asymmetric.

Weber and Cutler (2004) interpreted this result as indicating that the lexicon encoded a distinction which perceptual processing at an earlier level could not satisfy. All input, they suggested, seemed to be interpreted as belonging to only one of the two L2 categories, not as ambiguous between the two categories (as might occur with true homophony). Thus one category was dominant—in this case, whether /æ/ or /ɛ/ was heard, what was passed on to the lexicon was apparently always /ɛ/. The lexicon, however, accepted /ɛ/ as a match only to words really containing /ɛ/, such as *pencil* or *pelican*. This implied that words really containing an /æ/, such as *panda* and *palace*, were represented in a different way than words really containing an /ɛ/—i.e., the distinction was indeed encoded in the lexicon even though it was not accurately reflected in mapping from the input to the lexicon. One effect of this combination of dominance in the input but distinction in the lexicon was, Weber and Cutler pointed out, that words with the non-dominant category would not necessarily be activated by input matching either of the two phoneme categories in question. Indeed, target activation patterns in their experiment confirmed this suggestion: looks to targets with the dominant vowel (e.g. the pencil, the pelican, etc.) began to rise some 200 ms earlier than looks to targets with the non-dominant vowel (the panda, the palace, etc.), suggesting that resolution of competition effectively occurred earlier (e.g. with a contribution from the first vowel) for words like *pencil* and *pelican*, but later (only with second-syllable input, i.e. after the arrival of at least one definitively disambiguating phoneme) for words like *panda* and *palace*.

Dutch listeners are highly proficient in English (Dutch students are expected, for instance, to be able to follow lectures in English, and generally have extensive reading experience in the language). Is the asymmetry which Weber and Cutler observed only to be expected in such highly proficient L2 listeners with all the consequences this might imply for specification in lexical entries? Or might it, indeed, be specific to the relationship between these two languages, or even to the vowel phonemes tested? For these reasons (and also because the result was an unexpected outcome of a study actually designed to test several aspects of nonnative listening), we undertook the search for similar asymmetry in other listener groups.

Additionally, in the case of Dutch listeners to English there is more than one plausible explanation for why the one L2 category should be dominant instead of the two being indistinguishable. On the one hand, it could be that /ɛ/ is perceived as—and indeed may phonetically be—a closer match to the Dutch category /ɛ/ than /æ/ is (although Dutch /ɛ/ is typically lower than either Southern British English /ɛ/ or General American /ɛ/; Gussenhoven, 1999). Thus English /æ/ and /ɛ/ are perceived as /ɛ/ because /ɛ/ is closer to the Dutch phonetic category. Alternatively, there could be an influence from orthography. Words written with *e* are pronounced similarly in Dutch and English, but words written with *a* are pronounced very differently—*a* represents a back vowel in the first syllable of the Dutch words *panda* and *paleis*, for example. In the construction of the lexical representation, only words containing orthographic *e* would be labeled as containing a front central vowel. Thus if both /æ/ and /ɛ/ in the input were then correctly perceived as a front central vowel, the differences in the lexicon would allow only words containing orthographic *e* to be matched by either in lexical access.

To address both these sets of questions, we carried out a further eyetracking study, this time with Japanese listeners to English. The /r/-/l/ contrast in English for Japanese listeners is perhaps the most widely studied case of L2 phonetic category conflation (since Goto, 1971; see Yamada, 1995, for a review). If we observe asymmetric mapping of this contrast to lexical representations of these listeners, then such asymmetry is clearly not confined to the listener population Weber and Cutler tested, nor to the particular contrast they tested, or indeed to vocalic contrasts rather than consonantal. Moreover, the Japanese case lends itself well to testing between the alternative explanations we posed for the dominance of one category at the input level. The two explanations predict different results in the Japanese case. The single Japanese phonemic category closest to the English voiced postalveolar approximant /r/ and voiced alveolar lateral approximant /l/ is the voiced alveolar flap /ɾ/, and it is phonetically closer to the English /l/ than /r/ in most English dialects. The closeness is primarily articulatory—place of articulation is closer for /l/ and /ɾ/ than for /r/ and /ɾ/—but importantly, it has perceptual effect: Japanese listeners rate prevocalic /l/ as having a higher goodness of fit than prevocalic /r/ to their native prevocalic /ɾ/ (Takagi, 1995; Iverson et al., 2003), and they identify /l/ more often than /r/ as /ɾ/ (Iverson et al., 2003). The difference in effective similarity has been shown to correlate with success in acquiring the contrast in perception and production, in that acquisition of the more dissimilar sound /r/ proceeds more rapidly than acquisition of the more similar sound /l/ (Aoyama, Flege, Guion, Akahane-Yamada & Yamada, 2004).

If phonetic closeness to the L1 category is what determines dominance in the case of an asymmetry, then both /r/ and /l/ should be perceived as /l/ because /l/ is closer to /ɾ/. The phonetic explanation thus predicts that in the case of an asymmetry, /l/ should be the dominant category. The transliteration of Japanese words into alphabetic orthography, however, uses the letter *r* to represent /ɾ/—consider *Narita, Kirin, tempura*. If orthography plays an important role in dominance, then Japanese-speakers' lexical representations for English words might label only words spelled with *r* as containing a voiced (post)alveolar approximant. In the input, both /r/ and /l/ would be perceived as such an approximant, but in lexical access, the difference in the representations would detemine that only words spelled with *r* would be matched. The orthographic explanation thus predicts that in the case of an asymmetry, /r/ should be the dominant category.

2. Method

2.1. Participants

Twenty-four native speakers of Japanese, mostly visitors and students at Saarland University (mean age of 27, ranging from 20 to 36), took part in the experiment for monetary compensation. They had normal or corrected-to-normal vision and normal hearing. All had been born in Japan and all spoke Japanese in the home. Seven had never lived outside Japan for longer than a year; the remaining 17 had been away for longer periods, and for six of them this included a period resident in an English-speaking country. On average, they had studied English as a foreign language for 8 years in secondary and tertiary education, starting at a mean age of 12. They underwent an English multiple choice test after completing the eyetracking experiment to confirm their proficiency in the language. For 20 nouns, similar in frequency to target nouns in the eyetracking experiment, they had to choose the correct English definition (from three) taken from the Longman Dictionary of Contemporary English (1987). Most false definitions described nouns that were phonologically or semantically related to the target noun. Their average score was 81% correct.

2.2. Materials

We selected 20 pairs of English words which all were the names of picturable objects. All words were listed in the Genius English–Japanese Dictionary (2001), an English dictionary widely used in school and college English classes in Japan. Since there are no available familiarity ratings for English words by English learners of Japanese, it was not possible to match words as might be desirable for native-language experiments. English dictionaries in Japan do mark words as to whether they should be known by students at junior high school, at high school, or at college level; our materials contained a mix of these three categories, but, given the picturability constraints plus the necessity to construct phonologically similar pairs, it was impossible to select, for example, only from the first two of these categories.

Within each pair of words, both names began with an onset (singleton or cluster) containing in one member an /l/ and in the other member an /r/; the vowel in the first syllable was the same in both members, and the vowel was followed by a consonant which was either the same in both members of the pair (11 cases) or shared nearly all features (nine cases). Eleven of the 20 pairs were bisyllabic, five paired a bisyllable with a trisyllable, and the remaining four were monosyllabic. The full list of materials is: *lady/radio, ladder/radish, laptop/wrapper, leopard/reptile, lighthouse/writer, lobster/robber, laser/racing-car, lifeguard/rifle, lighter/rider, lipstick/ribbon, livingroom/river, lizard/wristwatch, locker/rocket, clockface/crocodile, glasses/grasshopper, crown/cloud, blazer/bracelet, blinds/bride, loaf/rose, legs/wreck*.

The overall lexical frequency of /r/-initial and /l/-initial lemmas, as computed with CELEX (Baayen, Piepenbrock, & Van Rijn, 1993), did not differ significantly ($F<1$). It is known that the frequency of target and competitor names can affect the likelihood of fixating corresponding pictures, but does not influence fixation probabilities of phonetically unrelated distractors (Dahan et al., 2001).

Pictures were selected for each of these 40 target items, and we further selected pictures to serve as non-competitor and filler pictures. All pictures were black and white line drawings, taken from the Snodgrass and Vanderwart (1980) and the Cycowicz, Friedman, Rothstein, and Snodgrass (1997) picture sets, as well as the IMSI MasterClips Image Collection (1990). For all pictures used in experimental trials, we then collected rating and naming data from native English speakers. Ten native speakers (five from America or Canada, five from the UK or Australia) named the pictures; agreement between their responses and the names intended for use in the experiment was 90.3% (correct names included synonyms, for example *equestrian* for *rider*, and short forms, for example *clock* for *clockface*). An additional 10 native speakers (again five from America or Canada, five from the UK or Australia) rated the goodness of the pictures as depictions of the intended object on a scale from 0 (poor) to 6 (excellent); the mean rated goodness was 5.1.

Each of the 20 experimental pairs formed the basis for three four-picture displays. Each display contained two distractor pictures. One of the three displays further contained both members of the experimental pair, while the other two contained one member plus a non-competitor picture. For the *locker/rocket* pair, for example, the distractor pictures were of a fountain and a waiter (see Fig. 1), while the non-competitor picture was an apple. Thus the three displays contained: a locker, a rocket, a waiter and a fountain; a locker, an apple, a waiter and a fountain; a rocket, an apple, a waiter and a fountain. Similarly, for the *lady/radio* pair, the distractors were socks and stairs, while the non-competitor was a door. Since either member of the experimental pair could serve as the target, the first of these three displays appeared in two conditions, once with the target being the word containing an /l/ and once with the target being the word containing an /r/. Further, each experimental display had two versions, with position of the target and the competitor counterbalanced. Targets occurred equally often in each cell position.

There were 30 four-picture displays which were used in filler trials, and a further two such displays used in two preliminary practice trials. Some of these displays contained pairs of pictures with phonologically similar names (e.g. a balloon and a banana, or a piano and a pirate). Others contained distractor items beginning with /l/ or /r/ (e.g. a ring or a lamb or a robot). The target item never began with /l/ or /r/ in a filler trial. No pictures were used in more than one trial.

The spoken sentences to accompany the displays contained only the target noun; that is, competitor and distractors were not heard during the experiment. Target nouns were embedded in the carrier *Click on the*…. These spoken sentences were recorded onto minidisc in a sound-attenuated room by a female native speaker of American English with a standard educated

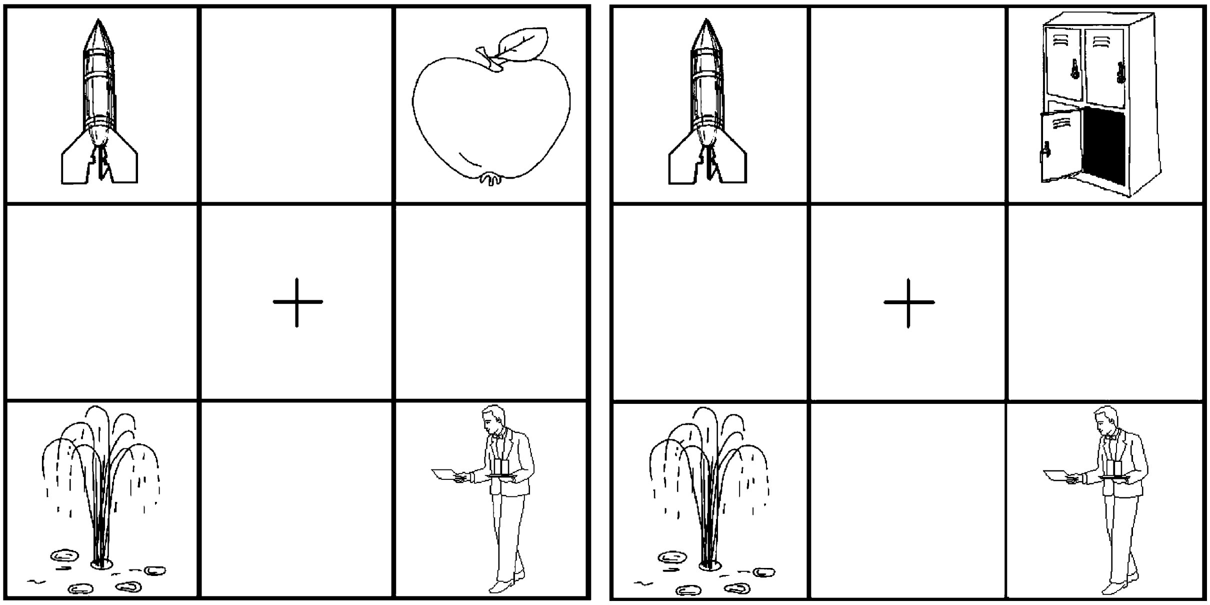

Fig. 1. Sample target displays for the target *rocket*. The right display includes the competitor picture (a locker), the left display includes the non-competitor picture (an apple).

midwestern accent, sampling at 44.1 kHz (later down-sampled to 22 kHz). The recordings were measured and target word onsets ascertained by visual and auditory checking. The constant preceding context facilitated this process, in that there was always a drop in F_1 and a decrease in intensity of F_2 following the vowel of *the*. Word onset markers were placed at a zero crossing corresponding to the lower value of F_1 after the drop. The average duration of putative overlap between the target noun and its competitor was 285 ms for targets with initial /r/ (e.g., duration of /rak/ in *rocket*) and 306 ms for targets with initial /l/ (e.g., duration of /lak/ in *locker*).

To check whether significant information in the signal cued the nature of the following phoneme prior to the point which we had determined as word onset, truncated versions of the target instructions ("Click on the -") were presented to 14 Japanese listeners (students at Daito Bunka University, Tokyo) who were asked to decide on the actual continuation, with four alternatives to choose from. The speech materials were truncated at exactly the point determined as word onset and used for results calculation. To disguise the recurrence of r- and l-alternatives, the two further options always began with /s/ and /n/ (e.g. *socket* and *knocker* were paired with *rocket* and *locker*). The results (for singleton-r trials: 40% correct, 34% judged as /l/, 26% other choice; for singleton-l trials: 30% correct, 27% judged as /r/, 43% other; for clusters: 39% correct, 37% judged as minimal pair, 24% other) revealed that the preceding signal did contain some information about the upcoming sound, but it was insufficient for these listeners to identify the sound precisely. The proportion of "other" judgments was highest for singleton-l instructions, suggesting that an upcoming /l/ provided these listeners with even less usable information than an upcoming /r/ or stop consonant.

Four versions of the experiment were constructed. Each version began with the two practice trials and further contained 20 experimental trials and all 30 filler trials, in pseudo-random order such that before each experimental trial there was at least one filler trial. Experimental trials appeared once in a given list, and target-initial phoneme (/r/-initial versus /l/-initial targets) and competitor presence (competitor or non-competitor displayed) were counterbalanced across versions. Six participants were presented with each version of the experiment; thus for the *locker/rocket* pair, 12 listeners were instructed to click on the locker and 12 to click on the rocket, and within each of these sets of 12, half saw a display which also contained a picture of the competitor item, while the other half saw a display which contained, instead, a picture of an apple. All comparisons were thus within-subject, with each participant receiving five trials in each possible combination of target-initial phoneme and competitor presence.

2.3. Procedure

At the beginning of a session, participants received written instructions in English, telling them to click with a computer mouse on the object mentioned in the spoken sentences. A list with the randomized set of pictures and their names was shown to the participants before the experiment.

Our experiment used the SMI EyeLink eye tracker, via which a detailed record can be made of what the eye is doing across time. There are three possibilities: fixation on a scene (i.e. the eye rests in one position, and information can be taken in), saccade (the eye moves from one position to another, and no information can be taken in), or blink (the eye is closed, and no information can be taken in; see Matin, Shao, & Buff, 1993 for further detail). After the eye tracker had been individually calibrated, each participant was presented with the trials from one of the four lists. Sentences were presented auditorily over headphones and started 500 ms after the appearance of the pictures on the screen. A camera on the participants' dominant eye provided the input to the tracker. Onset and offset times and spatial coordinates of the participants' fixations were recorded (250 Hz sampling rate). Along with the eye movements, the position of the mouse click was stored. Grid cells in which the pictures were displayed measured 7.5×7.5 cm, corresponding to a visual angle of approximately 7°, which is well within the resolution of the eye tracker (better than 1°).

3. Results

For the analysis, custom-made graphical software was used to display the locations of participants' fixations as dots superimposed on trial displays. Fixations were coded as pertaining to the cell of the target picture, the competitor, or one of the two unrelated distractors. Fixations that lay outside the cell of any picture were coded as looks to the background. Blinks were added to previous fixations; saccade times were not added to fixation times. Note that it takes typically about 200 ms before a programmed eye movement is launched (Matin et al., 1993). Thus fixations that are triggered by the first 100 ms of acoustic information are observable around 300 ms after target noun onset. Time windows for the analysis therefore usually begin between 200 and 300 ms after target noun onset.

The target *wrapper–laptop* was removed from the analysis because more than 40% clicking errors were made on that particular item. In addition, 27 trials were excluded because participants had clicked on an object other than the target object (5.9% of all trials). Proportion of fixations were calculated by adding the number of trials for each participant and each item that each of the

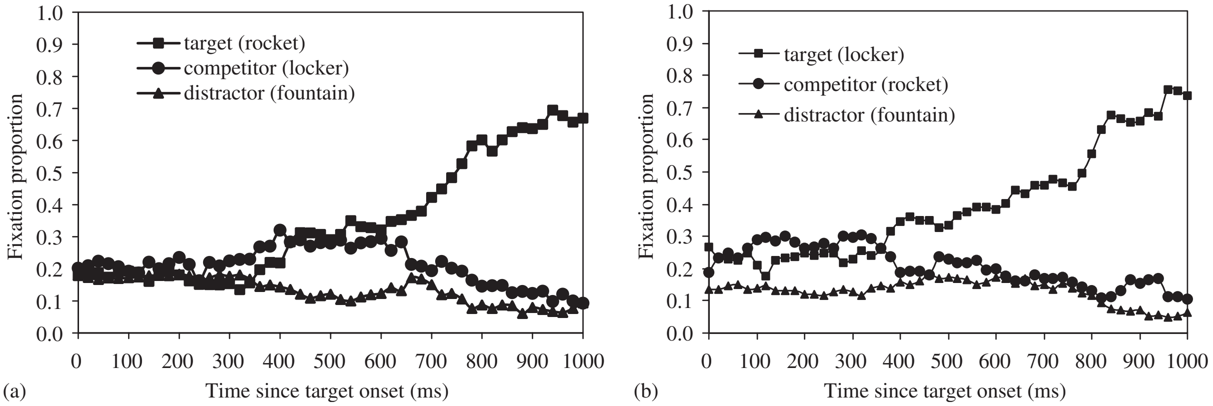

Fig. 2. (a) Fixation proportions over time from word onset for /r/-initial target items (e.g. the rocket), the competitor (the locker) and the averaged distractor items (e.g. the fountain, the waiter). (b) Fixation proportions over time from word onset for /l/-initial target items (e.g. the locker), the competitor (the rocket) and the averaged distractor items (e.g. the fountain, the waiter).

above four alternatives, respectively, was fixated during successive 10 ms time frames. The sum for each alternative picture type was then divided by the total sum of all fixations during the interval. Fig. 2 presents the resulting averaged proportions of fixations over time for trials in which a competitor was displayed, respectively for targets with /r/-initial names (Fig. 2a) and for targets with /l/-initial names (Fig. 2b). To simplify the figures, fixation proportions for the two unrelated distractors are averaged.

Separate one-factor ANOVAs with picture (with the two levels 'competitor' and 'unrelated distractors') as the within-participants and within-items factor were conducted on the fixation proportions averaged over participants and over items. For /r/-initial targets (Fig. 2a), the listeners fixated competitor objects (e.g. the locker) more often than distractor objects. Between 300 and 700 ms after target noun onset, the proportion of fixations was on average 27% for the competitor and 16% for the unrelated distractors, a significant difference ($F1[1, 23] = 6.94$, $p < .05$; $F2[1, 18] = 10.26$, $p < .01$). This suggests that during the presentation of *rocket* the competitor *locker* was considered as a potential candidate for recognition.

When the target word was /l/-initial, however (Fig. 2b), no significant activation of /r/-initial competitors was observed between 300 and 700 ms. That is, the difference between looks to the competitor and to the unrelated distractors was not significant ($F1[1, 23] = 2.54$, $p > .1$; $F2 < 1$). Note that in Fig. 2b between 0 and 300 ms the competitor (e.g. the rocket) received more looks than the unrelated distractors ($F1[1, 23] = 10.86$, $p < .01$; $F2[1, 18] = 8.31$, $p < .05$). This was apparently an effect of the competitor picture, not of the spoken input, since there was no difference among alternatives in competitor-absent trials in this time range (both $F\text{s} < 1$). Thus prior to the point that fixations could be directed by acoustic information from the target noun, the competitor picture was preferred. Crucially, however, after 300 ms the number of looks to the competitor only decreased over time. For example, between 300 and 400 ms, fixation proportions for the competitor decreased from 30% to 19% (compared to an increase from 15% to 22% for the competitor in Fig. 2a). The lack of a significant difference between looks to the competitor and to the distractors in the 300–700 ms time window suggests that during the presentation of *locker* the competitor *rocket* was not considered as a potential candidate for recognition.

The explanation of the early looks to the /r/-initial pictures in /l/-target competitor-present but not competitor-absent trials in terms of picture attractiveness should predict a similar effect in /r/-target trials, since the pictures were the same across trials. However, there was no such effect in /r/-target trials. In /r/-target trials, none of the alternatives (target, competitor, distractor) received an above-average proportion of looks in the early (<300 ms) range. We can only speculate as to why this was so, but recall that the listeners who heard the truncated materials were slightly better at finding useful information in the instructions from /r/-target trials. It may be that on those trials there was just enough early information available to counteract the effect of the intrinsic picture attractiveness (although insufficient to induce a significant proportion of early looks to the target or the competitor, since this did not occur).

In addition to the above analyses, we compared the effect of displayed competitor objects on target activation. For this analysis we compared the half of our items in which competitor objects were displayed with the half in which they were replaced with non-competitor objects whose names did not overlap with the target. For the latter trials, target activation should not be hindered by potential competitor activation. Fig. 3 shows the average fixation times to the target items in these two trial types, separately for /r/-initial and /l/-initial targets. It can be seen that between 300 and 700 ms, /r/-initial targets received 45% of fixations when no competitor was displayed, but only 29% when an /l/-initial competitor was shown. This difference was significant ($F1[1, 23] = 9.39$, $p < .01$; $F2[1, 18] = 13.07$, $p < .01$). In contrast, fixations for /l/-initial targets between 300 and 700 ms hardly differed as a function of the presence or absence of /r/-initial competitors (35% target fixations with competitor, 40% without), and indeed, the difference between the trial types for these targets was not significant (both $F\text{s} < 1$). Thus competitor activation of /l/-initial nouns significantly slowed recognition of /r/-initial targets, but the reverse was not the case: /r/-initial competitors did not slow recognition of /l/-initial targets. Not only the pattern of looks to non-target items (Fig. 2) showed a significant difference between trials with

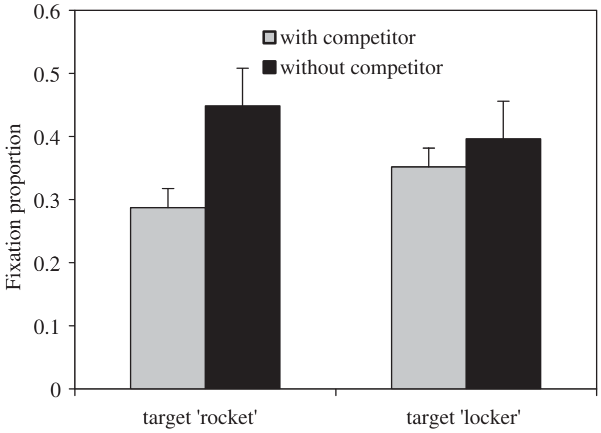

Fig. 3. Average fixation proportions in the time window 300–700 ms for /r/-initial targets (e.g. the rocket) and /l/-initial targets (e.g. the locker) as a function of presence versus absence of a competitor item in the display. The vertical bars represent standard deviations.

/l/-initial and /r/-initial targets, therefore, but the speed with which these two types of targets were themselves recognised (Fig. 3) was likewise significantly different.

4. Discussion

The results are very clear: for Japanese listeners to English, mapping from phonetic representations to lexical representations is asymmetric. Instructions to click on a picture of a rocket induce looks to a picture of a locker, but instructions to click on a picture of a locker do not induce looks to a picture of a rocket. Precisely the same sort of asymmetry which Weber and Cutler (2004) had observed in the case of Dutch listeners' processing of vowel contrasts in English has here been observed also, in the already well-studied case of Japanese listeners' difficulties with English /r/ and /l/.

We have not established in the present study that native English listeners would be able to distinguish *rocket* and *locker* with equal ease from competitor words, but we have no reason to suspect that they would have any difficulty doing so. Of course, the contrastive function in the English vocabulary of the phonetic contrasts we have examined (and the fact that jokes about *flied lice* exist!) is the existence proof that English listeners can easily distinguish the contrasting sounds. But Weber and Cutler (2004) did test native English listeners with their vowel-contrast pairs, and found, as expected, that words like *panda* and words like *pencil* were equally effectively distinguished from their competitors via information in the first vowel. We would predict similar processing efficiency for the consonantal onset contrast at issue in the present case. Studies of phonetic confusability do not show asymmetry in English listeners' identifications of /r/ and /l/; the most recent such study, by Cutler, Weber, Smits and Cooper (2004), which examined American English listeners' identifications of syllables spoken by a female American-English speaker at three signal-to-noise ratios (SNRs), reported that even at the worst SNR, /l/ in onset position was identified as /r/ in only 2.1% of cases, and /r/ was never identified as /l/. Asymmetric mapping plays no major role in L1 listening.

Our present finding suggests, however, that such asymmetry may be quite general in L2 listening; certainly it is not specific to the case studied by Weber and Cutler (2004). That is, in the first instance, the effect is not specific to vowel contrasts, since it appeared here with a consonantal distinction, implemented exclusively as contrasts between syllable onsets. And second, it is not specific to the Dutch listeners tested in the previous study, with their arguably very high competence in the L2. Of course, it is difficult to compare across groups of L2 users (not least because any such group tends to range over extensive individual differences), but we are under the impression that the English competence of the present Japanese listener group was not as high as that of the Dutch listener group tested by Weber and Cutler. Our choice of participants in the present case was severely constrained, primarily by the location of the eyetracking equipment; the latter does not, alas, lend itself to flexible testing wherever the most extensive populations of participants may be found. Our listeners in this study were certainly competent in the language—any listener group who can pass a word recognition test and successfully carry out an experiment such as ours must be so; but we would argue that the appearance of the asymmetry in the present listener group at least suggests that the effect is not restricted to L2 users who are (as Dutch users of English generally are) very skilled indeed.

The reason why a distinction which cannot be reliably perceived in speech input has been established nonetheless in the lexicon is not directly addressed by our results, but the most obvious candidate is, of course, explicit language instruction, and especially instruction in writing the L2. We suggest that L2 listeners have been taught that *write* and *light* (or *cattle* and *kettle*) are supposed to sound different, and that they have accordingly incorporated this distinction in some form in the phonological representations which they maintain of words containing the sounds in question in the lexicon. It is an open question (and one which seems to us certainly worthy of further research) whether there would be any such asymmetry in processing by listeners who have acquired an L2 without explicit teaching.

The wider generality of an asymmetry in mapping from phonetic to lexical processing in L2 listening has, however, important implications. If the failure to discriminate in phonetic processing propagated through the language processing system in the form of indiscriminability at all levels, then we would expect pairs like *write* and *light* or *rock* and *lock* to be functionally homophonous in all respects. But this appears not to be an accurate description of the representation of such pairs in L2 listeners' lexicons. Instead, the pattern of the present results is consistent with the type of account which Weber and Cutler (2004) proposed for Dutch listeners' mapping of English speech to lexical representations. That is, for the listeners in the present study, the inputs *lock*- and *rock*- both appear to have been interpreted as *lock*-. Whether listeners heard the beginning of *locker* or of *rocket*, they were inclined to look at the locker. Their representation of the input thus obviously did not contact both lexical entries beginning with *lock*- and with *rock*- equally effectively (as would have been the case with simple homophony), because if it had done so, a competition effect would have been observed whatever the input. Rather, their representation contacted lexical entries with /l/ to a greater degree than it contacted lexical entries with /r/. Thus in the present case, just as in the preceding study, a distinction appears to be maintained at the lexical level between lexical representations containing the two separate English categories—even though the input from phonetic processing to the lexicon fails to satisfy this distinction. L2 listeners can thus exhibit lexical competence even though this competence is in practice betrayed by inadequacies in phonetic perception.

A further important implication of this result is that indiscriminability of two L2 phonetic categories does not imply that the two categories are essentially equivalent. In the case of the contrasts under study, one is clearly dominant. In the Weber and Cutler (2004) study, the dominant category for Dutch listeners' perception of English /æ/ versus /ɛ/ was /ɛ/, and in the present study, the dominant category for Japanese listeners' perception of English /r/ versus /l/ is /l/. Note that it is still not the case that the dominant category is capable of contacting the lexical representations of all L2 words containing either phoneme. Lexical competence entails that the L2 distinction is in fact quite accurately represented at the lexical level, and perceptual representations of the dominant category contact preferentially lexical representations containing that category.

What is the case for representations containing the non-dominant category? The picture we have sketched seems to be compatible with two possibilities. One is that such representations never receive a relevant match from the input—the input is always interpreted as the dominant category and always contacts lexical representations containing that category. In the present case, lexical representations containing /r/ would be matched neither by /r/ nor by /l/ in the input. It may be that they are also not mismatched by either input—for instance, it may be that, as suggested by Broersma (2002), L2 listeners have learned to switch off inhibitory connections between lexical representations where these correspond to categories which they know to be different but cannot reliably discriminate—but they receive no positive activation from either category in the input. This was the proposal put forward by Weber and Cutler (2004), and in support of it, they pointed to the slower rate of rise in looks to targets with names containing the non-dominant category than to targets with the dominant category in their results. Only when later parts of the input were consistent with the word containing the non-dominant category but inconsistent with the word containing the dominant category, they argued, was the former able to become the most activated word candidate, with a consistent rise in looks to the correct target as a result; that is, *panda* could only win in competition with *pencil* once the input presented a /d/. Our results show a similar effect: looks to the target given the input *rocket* only rose above 35% and continued to rise steadily from about 550 ms (Fig. 2a), whereas the 35% point followed by continuous rise was reached at around 400 ms given the input *locker* (Fig. 2b).

Another possibility, however, is that the asymmetry is more graded. We have not so far addressed the form in which the inter-category distinction is coded in the phonological representations in the lexicon, but a likely possibility is that the coding makes reference to the L1 category. The dominant category is coded as a reasonable match to the single L1 category, while the non-dominant category is coded as a much poorer match to the same L1 category. Whatever category occurs in the input, the L1 category will be activated and an estimate will be made of the degree of match between the input and that category. This estimate will in fact be unreliable, because "perceptual magnet" effects (Kuhl & Iverson, 1995) will ensure that the degree of match will always seem closer than it actually is, for either input. Thus input containing the dominant category will be rated as a better match to the L1 category than it actually is, and will be even more likely to be routed to lexical representations containing the dominant category; input containing the non-dominant category will be rated as a less poor match to the L1 category than it actually is, and will also therefore attain a higher probability of being erroneously routed to lexical representations containing the dominant category. This still leaves some possibility of contact for representations containing the non-dominant category, when a poor match to the L1 category is indeed registered. This view too is supported by evidence in our data, in that the early looks to the competitor in Fig. 2a competed with, but did not in fact dominate, looks to the target.

The evidence to date does not force acceptance of either of these explanations, and it is not inconceivable that each of them may, for different cases, hold true.

Current models of spoken-word recognition assume that speech input activates multiple word candidates in parallel, with active competition between these candidates leading eventually to a single winner. In models such as TRACE (McClelland & Elman, 1986) or Shortlist (Norris, 1994), in which the competition process is implemented via inhibitory connections between alternative candidates for any given portion of the input, the more one candidate is activated by matching input, the more it is able to inhibit its rivals. However, this does not mean that a candidate which is the most activated alternative at a given point is necessarily destined to emerge as winner. The competition can effectively seesaw as later input forces an alternative interpretation. Norris (1994) presents a Shortlist simulation for the input *ship inquiry*; from the second syllable onwards, the input favors *shipping* over *ship*, then *shipping choir* over *ship inquiry*, and only when the input of the final vowel /i/ arrives does the pattern of competition radically change and the string *ship inquiry* emerge as the correct winner.

Experimental evidence that partially mismatched candidate words remain in the competition process has been provided by Connine, Blasko, and Titone (1993), Connine, Titone, Deelman, and Blasko (1997), Frauenfelder, Scholten, and Content (2001) and Marslen-Wilson, Moss, and Van Halen (1996); in all of these studies, non-words which differed from real words in the initial phoneme exercised lexical effects on responses in speech perception tasks. The word which these non-words most resembled thus remained available despite the fact that it was partially mismatched by the input. Most relevantly for the present study, evidence from eyetracking supports the same conclusion. Allopenna et al. (1998) observed competition not only from pictures with names overlapping in onset with that of a target picture (e.g., *candy–candle*), but also from pictures with names rhyming with that of a target (e.g., *speaker–beaker*). The competition in the rhyming case was weaker than that in the onset case, but was still noticeable. Thus as long as a word candidate is supported by evidence from the input, it remains in the competition process, even though it may not at all times be in a winning position.

All of this evidence is consistent with the picture of L2 listening which the present study suggests. Lexical candidates which contain the non-dominant phoneme receive less input support, and as a consequence may be temporarily less active than candidates which contain the dominant phoneme. However, they do not drop out of the competition, with the result that they can win after all when they eventually receive sufficient support from later matching input—perhaps especially when that input clearly mismatches an alternative candidate containing the dominant phoneme. Note that L2 listening must thus necessarily involve, as indeed Weber and Cutler argued, more competition than besets native-language listening.

It is also consistent with the previous studies of L2 lexical processing of words containing confusable phonemes. Both members of a minimal pair will be activated by the input which matches them. For L1 speakers, a single phoneme mismatch will count against the mismatched competitor. But for L2 listeners, the mismatch will be less effective; both competitors will remain active, leading to repetition priming in lexical decision (Cutler & Otake, 2004; Pallier et al., 2001), and also allowing false-alarm recognition of non-words such as *dask* or *retter* as words (Broersma, 2002; Sebastián-Gallés, Echeverría, & Bosch, 2005).

The current study further advances our understanding of the mapping from L2 phonetic to lexical processing in that it clarifies the mechanism underlying dominance of a single L2 category in phonetic processing. As we described in the introduction, the Weber and Cutler (2004) result could admit of at least two explanations for the dominance of /ε/ over /æ/ in Dutch listeners' perception of English: It could have been due to /ε/ being phonetically closer to the single native Dutch vowel in that portion of the vowel space, or it could have been the result of orthographic influence in the construction of lexical representations, since only the letter *e* can represent a front central vowel in Dutch and thus only lexical representations containing this letter would be contactable by such a vowel. The present results from Japanese resolve this uncertainty. In our study, English /l/ and not /r/ was the dominant category for our Japanese listeners. This cannot be an orthographic effect, since the single Japanese liquid consonant is always written in Romanji with the letter *r*, never with the letter *l*. The Japanese consonant in question is, however, phonetically closer to English /l/ than to English /r/ (Aoyama et al., 2004). Thus our results clearly suggest that when two L2 categories activate a single L1 category in speech perception, it is the perceived acoustic–phonetic proximity to the L1 category which is the deciding factor in determining which of the two L2 categories is dominant.

Finally, we call attention to the methodological innovation which our study represents for the field of speech perception. The eyetracking paradigm has allowed a more sensitive look at the way in which incoming phonetic information is mapped to lexical entries over time. Failure to discriminate between L2 phonetic categories, especially where the two are subsumed by a single L1 category, is one of the most well-established findings in speech perception research, demonstrated via numerous empirical techniques in many laboratories around the world. But only this new window has allowed us to observe that such perceptual indiscriminability can paradoxically co-exist with accurate implementation of discrimination between the same two categories at the level of lexical processing.

Acknowledgments

This research was supported by the SPINOZA project "Native and Non-native Listening" of the Nederlandse Organisatie voor Wetenschappelijk Onderzoek (NWO) and by the SFB378 project "ALPHA" of the Deutsche Forschungsgemeinschaft. We thank Katja Pfahler and Magdalena Mitova for their help in collecting the eyetracking data, Sally Butterfield, Laurence Bruggeman and Kiyoko Yoneyama for assistance with the collection of additional listening data, and Núria Sebastián-Gallés and an anonymous reviewer for helpful commentary.

References

Allopenna, P., Magnuson, J., & Tanenhaus, M. (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.

Baayen, H., Piepenbrock, R., & Van Rijn, H. (1993). *The CELEX lexical database* (CDROM). Philadelphia: Linguistic Data Consortium, University of Pennsylvania.

Broersma, M. (2002). Comprehension of non-native speech: Inaccurate phoneme processing and activation of lexical competitors. *Proceedings of the 7th international conference on spoken language processing* pp. 261–264.

Connine, C. M., Blasko, D. G., & Titone, D. A. (1993). Do the beginnings of spoken words have a special status in auditory word recognition? *Journal of Memory and Language*, *32*, 193–210.

Connine, C. M., Titone, D., Deelman, T., & Blasko, D. G. (1997). Similarity mapping in spoken word recognition. *Journal of Memory and Language*, *37*, 463–480.

Cutler, A., & Otake, T. (2004). *Pseudo-homophony in non-native listening*. Paper presented to the 75th meeting. New York: Acoustical Society of America.

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.

Cycowicz, Y., Friedman, D., Rothstein, M., & Snodgrass, J. (1997). Picture naming by young children: Norms for name agreement, familiarity, and visual complexity. *Journal of Experimental Child Psychology*, *65*, 171–237.

Dahan, D., Magnuson, J., & Tanenhaus, M. (2001). Time course of frequency effects in spoken-word recognition: Evidence from eye movements. *Cognitive Psychology*, *42*, 317–367.

Frauenfelder, U. H., Scholten, M., & Content, A. (2001). Bottom–up inhibition in lexical selection: Phonological mismatch effects in spoken word recognition. *Language and Cognitive Processes*, *16*, 583–607.

Genius English–Japanese Dictionary. (2001). 3rd edition. Tokyo: Taishukan Shoten.

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

Gussenhoven, C. (1999). Dutch. In *Handbook of the international phonetic association (pp. 74-77)*. Cambridge, UK: Cambridge University Press.

IMSI Masterclips Image Collection (1990). Premium image collection 303,000 [Image database]. (http://www.imsisoft.com)

Iverson, P., Kuhl, P. K., Akahane-Yamada, R., Diesch, E., Tohkura, Y., Ketterman, A., & Siebert, C. (2003). A perceptual interference account of acquisition difficulties for non-native phonemes. *Cognition*, *87*, B47–B57.

Kuhl, P. K., & Iverson, P. (1995). Linguistic experience and the "perceptual magnet effect". In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 121–154). Baltimore: York Press.

Longman Dictionary of Contemporary English (2nd ed.). (1987). London, Longman Group UK Limited.

Marslen-Wilson, W. D., Moss, H. E., & Van Halen, S. (1996). Perceptual distance and competition in lexical access. *Journal of Experimental Psychology: Human Perception and Performance*, *22*, 1376–1392.

Matin, E., Shao, K., & Boff, K. (1993). Saccadic overhead: information processing time with and without saccades. *Perception & Psychophysics*, *53*, 372–380.

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, *18*, 1–86.

Norris, D. (1994). Shortlist: A connectionist model of continuous speech recognition. *Cognition*, *52*, 189–234.

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.

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

Snodgrass, J., & Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory*, *6*, 174–215.

Strange, W. (Ed.) (1995). *Speech perception and linguistic experience: Issues in cross-language research*. Baltimore: York Press.

Takagi, N. (1995). Signal detection modeling of Japanese listeners' /r/-/l/ labeling behavior in a one-interval identification task. *Journal of the Acoustical Society of America*, *97*, 563–574.

Tanenhaus, M., Spivey-Knowlton, M., Eberhard, K., & Sedivy, J. (1995). Integration of visual and linguistic information during spoken language comprehension. *Science*, *268*, 1632–1634.

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

Yamada, R. (1995). Age and acquisition of second language speech sounds: Perception of American English /ɹ/ and /l/ by native speakers of Japanese. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 305–320). Baltimore: York Press.