



Resolving ambiguity in familiar and unfamiliar casual speech

Annelie Tuinman^{a,b}, Holger Mitterer^{a,*}, Anne Cutler^{a,b,c}

^a Max Planck Institute for Psycholinguistics, PO Box 310, 6500 AH Nijmegen, The Netherlands

^b Donders Institute for Brain, Cognition and Behaviour, Radboud University Nijmegen, 6500 HE Nijmegen, The Netherlands

^c MARCS Institute, University of Western Sydney, Penrith South DC, NSW 2751, Australia

ARTICLE INFO

Article history:

Received 8 July 2010

revision received 31 January 2012

Available online 3 March 2012

Keywords:

Spoken-word recognition

Casual speech

Insertion

Second-language listening

ABSTRACT

In British English, the phrase *Canada aided* can sound like *Canada raided* if the speaker links the two vowels at the word boundary with an intrusive /r/. There are subtle phonetic differences between an onset /r/ and an intrusive /r/, however. With cross-modal priming and eye-tracking, we examine how native British English listeners and non-native (Dutch) listeners deal with the lexical ambiguity arising from this language-specific connected speech process. Together the results indicate that the presence of /r/ initially activates competing words for both listener groups; however, the native listeners rapidly exploit the phonetic cues and achieve correct lexical selection. In contrast, The Dutch-native advanced L2 listeners to English failed to recover from the /r/-induced competition, and failed to match native performance in either task. The /r/-intrusion process, which adds a phoneme to speech input, thus causes greater difficulty for L2 listeners than connected-speech processes which alter or delete phonemes.

© 2012 Elsevier Inc. All rights reserved.

Introduction

In the Beatles song “A day in the life”, John Lennon sings *I saw a film today* with an intrusive /r/ after *saw*. For a split second, listeners might understand *I soar* or *I sore* or even *eye-sore*, interpretations that are all compatible with an /r/ sound before the following vowel. But given that none of them makes sense in combination with *a film today*, most listeners should ultimately realize that Lennon must have meant *I saw*.

British English dialects in general, and certainly the standard forms such as Received Pronunciation (RP), are “non-rhotic”; /r/ can occur word-initially but it cannot occur word-finally in citation forms. Word-final /r/ can appear in these dialects under two circumstances (Giegerich, 1992; Grimson & Cruttenden, 1994). First, a “linking /r/” surfaces when a word ending with an underlying /r/ precedes a word beginning with a vowel (e.g., *soar up*). Second, “intrusive /r/” may be inserted after a non-high vowel ([ə, a, ɔ] and [ɔ]-final

diphthongs) and before a vowel-initial word, as in *I saw a film today*. Note that a linking /r/ is represented in the spelling, while an intrusive /r/ is not. Intrusive /r/ is a casual speech process, not perhaps found in very formal registers, but highly common in natural conversational British speech.

There is now an extensive body of research on the perception of such connected-speech phonological processes, which distort canonical forms of utterances by altering, deleting, or adding segments. This research has shown that native listeners easily interpret the processes correctly (Connine, Ranbom, & Patterson, 2008; Mitterer, Csépe, Honbolygo, & Blomert, 2006; Mitterer & Ernestus, 2006). In contrast to the case of phoneme inventory differences, which have generated an extensive literature (Bohn & Munro, 2007; Strange, 1995), little attention has been paid to the issue of how listeners who are not familiar with this type of input – second language (L2) listeners, for instance – cope with connected speech processes. How difficult is it to hear and interpret *I saw a film* correctly if one is unfamiliar with the process of intrusive /r/?

In general we know – from experience, even if there were not an extensive literature – that listening to speech

* Corresponding author. Fax: +31 24 3521213.

E-mail address: Holger.Mitterer@mpi.nl (H. Mitterer).

is harder in L2 than in the native language (L1). The most well documented reason for this is phoneme perception difficulties, with L2 contrasts that divide a single L1 category, such as the English /l/-/r/ contrast for Japanese listeners (Goto, 1971), forming the classic example; for reviews see Strange (1995), and on the special case of L2 learning, Best and Tyler (2007). But phoneme perception problems are not the end of the story; L2 word recognition presents further problems. For any listener, spoken-word recognition involves multiple activation of candidate words, with the simultaneously activated forms competing for recognition (for a review see, e.g., McQueen, 2007). Although L2 listeners typically have smaller vocabularies than native listeners, listening to L2 actually involves more such activation and competition between word forms than listening to L1, as studies with the classic word recognition methods of cross-modal priming and eye-tracking have demonstrated (Broersma & Cutler, 2008, 2011; Weber & Cutler, 2004). The increased competition can even persist despite decades of experience with a non-dominant L2 from childhood onwards (Pallier, Bosch, & Sebastian-Gallés, 1997; Sebastián-Gallés, Rodríguez-Fornells, Diego-Balaguer, & Diaz, 2006). Note that the activated competitors in L2 listening may be spurious ones that would not trouble L1 listeners; compare an embedding such as *leg* in *legacy*, that L1 and L2 listeners alike will experience, with *leg* in *regular* that will concern only L2 listeners who cannot distinguish /l/ from /r/ (Cutler, Weber, & Otake, 2006). Crucially, such spurious activation is harder to get rid of than the activation of “truly” embedded words (Broersma & Cutler, 2011).

Although casual speech processes that alter the surface form of words have not so far been widely studied in the L2 context, it is clear that how native listeners deal with them is by efficiently exploiting phonetic detail (Gow, 2002; Kemps, Ernestus, Schreuder, & Baayen, 2004; Mitterer & McQueen, 2009; Quene, 1992). One of the sources of information on which listeners draw is segment duration, particularly the durational difference between word-initial consonants and consonants in other positions. Gow and Gordon (1995) showed that in English ambiguous sequences such as *two lips/tulips*, word-initial phonemes (e.g., the /l/ in *two lips*) were longer than corresponding noninitial phonemes (e.g., the /l/ in *tulips*), and listeners could exploit this durational difference to achieve disambiguation. Shatzman and McQueen (2006) showed that Dutch sequences such as *een spot* ‘a spotlight’ versus *eens pot* ‘once jar’ could be disambiguated in the same way – the /s/ was longer word-initially, and listeners interpreted a longer /s/ as word initial. Spinelli, McQueen, and Cutler (2003) found the same to be true of French phrases involving liaison (in which underlying word-final consonants are pronounced only before a vowel-initial following word); the onset /r/ in *dernier rognon* (‘last kidney’), for example, was longer than the liaison /r/ in *dernier oignon* (‘last onion’), and listeners in a perception study again used this cue.

French liaison has been investigated within the L2 context. With some experience, learners of French can successfully make use of the durational effects and distinguish liaison from word-initial consonants in a forced-choice task (Shoemaker, 2010); they do not reach native levels of

rapidity in distinguishing liaison versus word-initial consonants in word recognition, but they do distinguish them (Tremblay, 2011). However, liaison can also lead L2 listeners to erroneously assume words to be consonant-initial (Dejan de la Batie & Bradley, 1995). Another casual speech process of French is voicing assimilation (whereby, for instance, the final /t/ of *note* ‘note’ becomes voiced in *note grave* ‘low note’). Such consonant-to-consonant voicing assimilation does not occur in English, although assimilation of place is common. Darcy, Peperkamp, and Dupoux (2007) tested judgements about such sequences by native French listeners and English learners of French (beginning or advanced). In a probe detection task, the advanced learners were able to compensate for the assimilation as well as the native listeners did, and even beginning learners showed better than chance performance. Darcy et al. concluded that with learning, compensation for the effects of this assimilation could be native-like. This conclusion is in line with models in which compensating for the effects of assimilation must be learned, such as that proposed by Gaskell (2003).

However, compensation for a connected speech process can appear without relevant experience. Gow and Im (2004) presented native and nonspeakers of Hungarian and Korean with language-specific assimilation phenomena of each language, in a task involving monitoring for segments in assimilated and non-assimilated contexts. Native and non-speakers of Hungarian showed similar context effects related to Hungarian voicing assimilation, and neither native nor non-speakers of Korean showed context effects related to Korean labial-to-velar place assimilation. Similarly, Mitterer, Csépe, and Blomert (2006) presented native speakers of Hungarian and Dutch non-speakers of Hungarian with Hungarian words and nonwords containing a viable versus an unviable liquid assimilation. Viable changed forms were difficult to distinguish from canonical forms, for both listener groups.

This native-like performance by listeners with no experience at all, as well as the native-level performance of advanced learners and even the above-chance performance of Darcy et al.’s (2007) beginning learners, contrasts with the persistently poor performance of L2 listeners in many phoneme and word recognition tasks (e.g., Pallier et al., 1997; Weber & Cutler, 2004). Mitterer, Csépe, and Blomert (2006) argued that assimilation phenomena are perceptually motivated; within the framework of functional/evolutionary phonology (Blevins, 2004; Boersma, 1998), compensation for such phonologically natural processes would in principle be available to non-native listeners and thus would not need to be learned at all.

British English /r/ intrusion has been argued to have arisen historically by a process of analogy (Gimson, 1980). When a dialect of English becomes non-rhotic, words such as *soar* and *saw* become homophones in isolation. But they are not necessarily homophones in connected speech; word-final /r/ will surface in the linking case (*soar up*). Non-rhotic dialect speakers without access to orthographic information (e.g., pre-literate children) have, however, no means to distinguish *soar* and *saw*. Intrusive /r/ occurs in the same contexts as those in which linking /r/ is found, but in words where there is no historic /r/ (and thus no /r/ in rhotic dialects), suggesting that the

linking process of /r/-insertion was overgeneralized to other words in similar contexts (Giegerich, 1992). No perceptual basis has been offered for the intrusive /r/ process.

The case of intrusive /r/ thus forms an interesting test case for L2 listening and connected speech processes. First, it causes ambiguous sequences (*extra ice* with an intrusive /r/ sounds like *extra rice*), and such sequences are known to produce difficulty for L2 listeners. Second, it is an optional process; its use is widespread, but it is sometimes felt to be unsuitable for formal speech and thus avoided (Collins & Mees, 1996; Giegerich, 1992). This could further make its correct perception difficult, especially for inexperienced listeners who could not rely on regularity of where and when an intrusive /r/ might appear. Both these factors suggest that it will cause L2 listening difficulty. Third, its realization incorporates a durational pattern of exactly the type already shown for French liaison and for word-boundary effects across languages: word-initial /r/ is longer than intrusive /r/ (Grimson & Cruttenden, 1994). This latter fact suggests that L2 listeners might be able to acquire the ability to deal with the process, since analogous durational patterns are likely to be available in listeners' native languages. Finally, it is a process without the perceptual basis argued to hold for assimilation and allied casual speech processes. Although some models of compensation for connected speech processes claim that familiarity with a particular process is not crucial for compensation to occur (Gow & Im, 2004; Mitterer, Csépe, Honbolygo et al., 2006), these models have so far only been tested on processes with a potential perceptual basis. Testing this claim on /r/-insertion in British English may hence provide information on the scope of connected speech processes for which such models are valid.

In the present study we examine the recognition of British English words involving intrusive /r/, by native listeners and by Dutch listeners proficient in English. The Dutch language is rhotic, and Dutch speakers produce /r/s word-finally. Dutch displays no pattern resembling British intrusive /r/; specifically, /r/ is never inserted in intervocalic contexts equivalent to *saw* a (Collins & Mees, 1996; Gussenhoven & Broeders, 1997). Dutch listeners to English thus have no native experience with such a phonological pattern.

They should, however, have heard the process many times, since they are exposed to speech in non-rhotic dialects of English. The target English pronunciation taught in Dutch schools and universities is British English (see, e.g., Collins & Mees, 1996; Gussenhoven & Broeders, 1997). British radio and television can be received in all households. Dutch television itself airs both American and British productions, and all foreign-language productions are subtitled, never dubbed. Our listeners are proficient in English, with at least 7 years of English instruction at school.

In a previous study (Tuinman, Mitterer, & Cutler, 2011), we examined the sensitivity of listeners to onset versus intrusive /r/ realisations using a phonetic categorization task. First, we found that, as previously reported in the phonetic literature, intrusive /r/ is acoustically weak compared with onset /r/ – onset tokens are longer in duration and have a larger intensity decrement than intrusive tokens. A power analysis revealed these differences to be

strong enough to be of use to listeners. In a 2AFC categorization experiment, we presented native British, Dutch and American listener groups with synthetically produced sentences, based on a British English model, in which the duration of /r/ was manipulated across a word boundary *saw/more* (r)ice, and an orthographic bias and/or semantic bias favored either the r-initial (e.g. *rice*) or the vowel-initial word (e.g. *ice*). The task was to choose for every token between *ice* and *rice*. British listeners responded categorically, reporting *ice* after short /r/s and *rice* after long /r/s; they were unaffected by the orthographic (*saw/more*) and semantic manipulations. The Dutch group relied less on durational cues than the British listeners, however, and were affected both by orthography, reporting /r/ more often after *saw* than after *more*, and by semantic bias in the sentences. The US listeners' performance fell between the other two groups.

Importantly for the present study, these earlier results suggest that intrusive /r/ does indeed cause perceptual problems for L2 listeners. While native listeners effectively exploit phonetic detail (in this case duration) to categorize a token of /r/, listeners for whom the speech is in a non-native language make much less use of the available phonetic cues.

In the present study, we test this by extending our investigations of /r/ insertion to word recognition in connected speech. We first use a cross-modal priming task with natural speech material, again presented to British and Dutch listeners. Listeners hear spoken sentences containing sequences such as *Canada raided* or *Canada*[intrusive r]aided, and make visual lexical decision to printed words such as *raided* or *aided*. With this paradigm, we thus test in an implicit way how listeners perceive an intrusive /r/ in a natural speech context; the critical question is to what extent intrusive /r/ primes either an r-initial or a vowel-initial word.

Experiment 1

Method

Participants

Seventy-two native speakers of British English (aged 18–35, mean = 23), recruited from the participant pool of the Laboratory of Experimental Psychology, University of Sussex, and 84¹ native speakers of Dutch (aged 18–27, mean = 21), recruited from the participant pool of the Max Planck Institute for Psycholinguistics, took part in the experiment. No participant reported any hearing loss, and all were volunteers and received a small fee or course credits for participation. Dutch participants had a high level of proficiency in English as a second language, based on at least 7 years of English instruction in primary and secondary education.

Stimuli

Twenty-seven pairs of English sentences were constructed (see Appendix). An example sentence is *My*

¹ More Dutch listeners were tested to equate the number of data points in the two groups.

brother likes extra ice/rice when he has dinner. In all sentences, a member of a minimal pair such as *ice/rice* followed a word ending on a low vowel. In such pairs, an onset /r/ will be produced in the r-initial member of the pair, while the vowel-initial member of the pair preceded by the low vowel (in this case, the last vowel of *extra*) creates a context in which an intrusive /r/ can occur. The sentence pairs were ambiguous at least until after the critical word.

All sentences were recorded by a female native speaker of British English who was unaware of the purpose of the study. This speaker typically produces intrusive /r/s in her normal speech. Each sentence was recorded at least twice. We measured the duration of the /r/ and the amplitude difference between the center of the /r/ and the surrounding vowels. In the sentences with an onset /r/, the /r/ was as predicted (Cruttenden & Gimson, 1994; Tuinman

et al., 2011) longer and had a larger intensity decrease than in similar sentences with intrusive /r/s. The full set of measurements may be found in Tuinman et al. (2011). Fig. 1 shows the crucial /r/ portion of the pairs *my sister saw roads/odes* and *Natasha remitted/emitted*, chosen to illustrate the range of /r/ variation in our sentences; in the pair in the left panel, the difference between intrusive /r/ and onset /r/ is particularly strong, while in the pair in the right panel, it is relatively weak.

In addition, 27 control sentences were constructed with target words that were matched for frequency in English to the /r/ minimal pairs (e.g., *My brother likes extra pages...*). A further 108 filler prime sentences were constructed, 54 with unrelated word targets and 54 with nonword targets. Finally, we also chose 27 “semi-experimental” prime sentences which had nonword targets starting with an /r/ in version 1 of the experiment and with a vowel in version

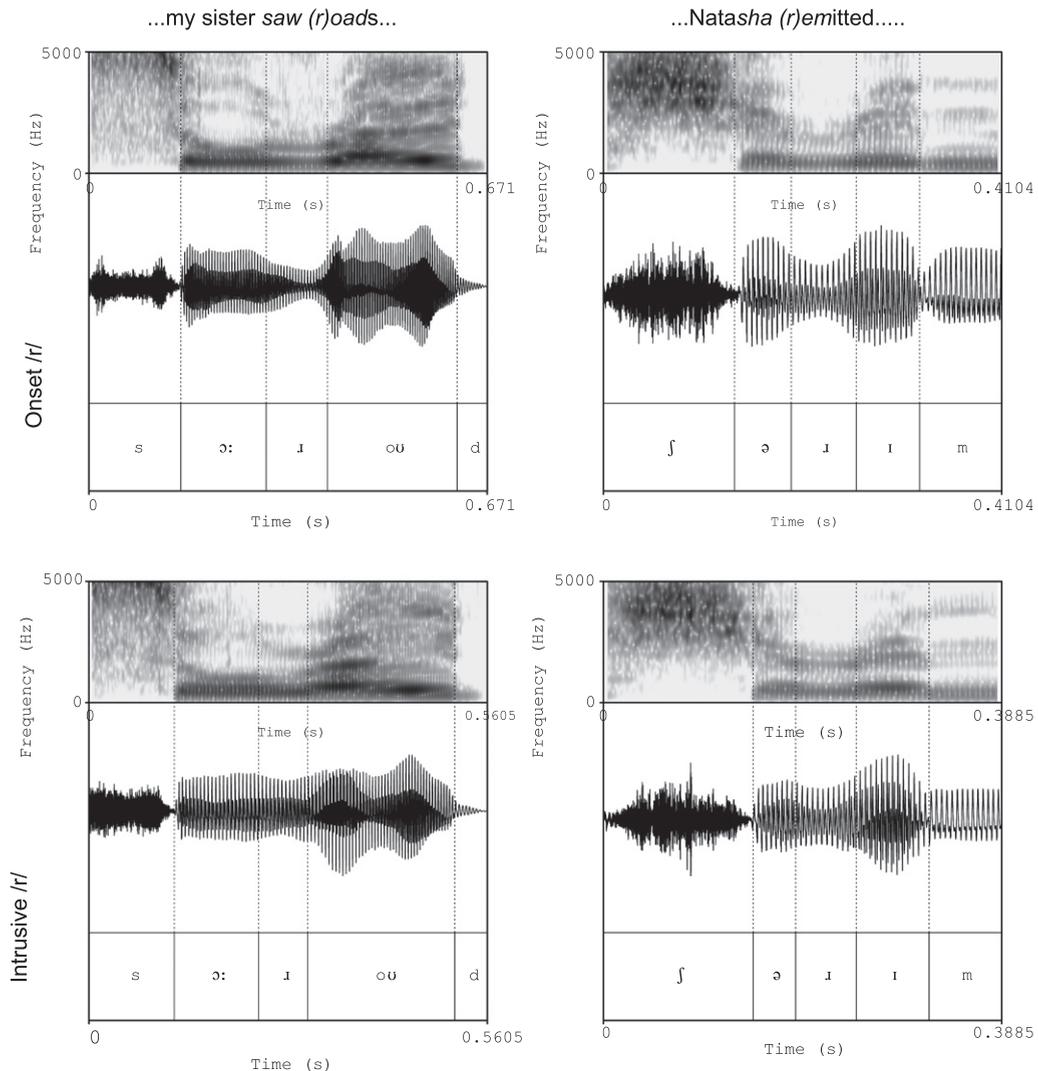


Fig. 1. Spectrograms and oscillograms for two pairs of potentially ambiguous phrases as used in the current experiment. Stimuli in the left panels contain the phrases *saw roads/odes*, stimuli on the right panels the phrases *Natasha remitted/emitted*. The top row shows the tokens with an onset /r/, the bottom row the tokens with an intrusive /r/. The pair on the left has an above-average difference between the two types of /r/, while the pair on the right has a below-average difference. The italicized letters in the column header show the part of the phrase visible in the figure.

2 of the experiment. Nine of these sentences contained a potential linking /r/ (e.g., *I think that your explanation...*) and had nonword targets partially overlapping with the prime words (*explint* or *rexplint*). Another nine sentences contained potential geminates of a linking and an onset /r/ (e.g., *And then my neighbour refused...*) and also had nonword targets partially overlapping with the prime words (e.g., *effint* or *reffint*). These 18 sentences were included to prevent phonological relatedness between prime and target becoming associated with “yes” responses in the lexical decision task. The final nine sentences with unrelated nonword targets (e.g., *ilems* or *rilems*) had no /r/s near the place of truncation (e.g., *I heard on the news that taxes...*).

Finally, 18 question trials were made, on which participants would hear a complete sentence and then receive a yes/no question about it. These trials constituted a check on whether the participants were paying attention to the prime sentences as well as to the visual target words. The filler sentences and question trials were recorded by the same female native speaker of British English who produced the critical items.

Design

The experimental design has the two factors Prime type (within participants, with the three levels R-prime, Vowel-prime, Control) and Target type (between participants, with the two types of target words, r-initial versus vowel-initial). Each participant received each experimental visual target only once, with nine targets in each of the three Prime conditions. In version 1, the targets were the r-initial words (e.g., *rice*), whereby in the R-prime condition targets were preceded by a matching auditory prime with an onset /r/, in the Vowel-prime condition targets were preceded by a vowel-initial prime with an intrusive /r/ (which, in terms of the intended word, mismatched the r-initial target), and in the Control condition targets were preceded by a phonologically and semantically unrelated prime. In version 2, the target words were vowel-initial words (e.g., *ice*), whereby the Vowel-prime condition with the intrusive /r/ matched the vowel-initial target word and the R-prime condition with the onset /r/ mismatched the target word. This between-participants design was chosen to minimize the likelihood of participants becoming aware of the experimental manipulation.

Procedure

Participants were tested one at a time in a sound-attenuated booth. They received English instructions printed on the screen, informing them that on each trial they would hear a part of an English sentence, directly after which an English word or nonword would appear on the screen. Only ambiguous parts of the sentences were played, up to the offset of the critical word. Participants were instructed for each visually presented item to press a green response button labeled “yes” with their dominant hand if they thought that the item was an English word, and a red response button labeled “no” with their other hand if they thought that it was not an English word; they were asked to try to respond as fast and as accurately as possible.

The experiment started with seven practice trials and one practice question trial, after which each participant was presented with (a) the 27 experimental trials, (b) all filler word and nonword trials, and (c) the 18 yes/no questions that could only be answered if attention had been paid to the preceding auditory sentence. Trials were randomized per participant. After every 50 trials participants could take a short break.

Results

Lexical decision reaction times (RTs) were measured from onset of the visual presentation of the target words, with a response window of 1500 ms. For the British listeners, 12.1% of all data was missing (6.2% late or absent responses, 5.9% errors) and for the Dutch listeners, 24.4% (6.3% late or absent responses, 18.1% errors). Fig. 2 shows the priming effects.

For the statistical analysis, RTs were log-transformed to reduce the skew of the RT distribution. A Linear Mixed-Model analysis on the transformed RTs was conducted with Target type, Prime type, and Native Language as fixed factors, and Participants and Target words as random factors. All possible interactions were enabled in the model. The three-way interaction between Target type, Prime Type, and Native Language proved significant $F(2,3208) = 3.7$, $p < .05$, indicating that the pattern of priming effects differed for the two listener groups. To further examine these differences, we conducted separate analyses of the data from each group across Target type and Prime type.

British listeners

In this data set, the interaction of Target type with Prime type was significant [$F(2,1580) = 20.72$, $p < .001$], so the two Target conditions were analyzed separately. For each condition, a Linear Mixed-Model analysis was

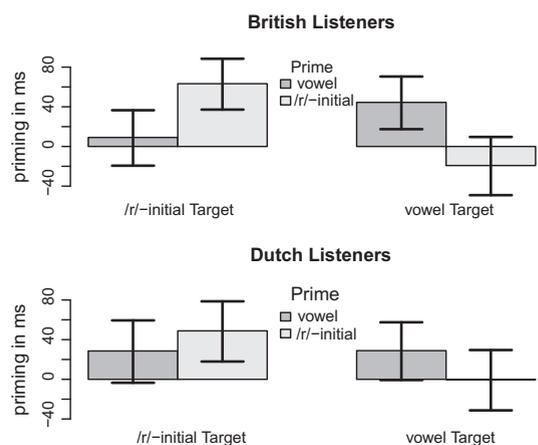


Fig. 2. Priming effects for British English and Dutch listeners in (Exp. 13). The error bars show the confidence intervals based on the regression weights for the two priming conditions from the linear mixed effect model for these data. The reaction times in the baseline conditions were 754 ms and 755 ms for the British Listeners on r-initial target and vowel-initial targets, respectively, and 837 ms and 854 ms for the Dutch listeners.

performed with the same random factors and only trial and Prime type as fixed factors. In such a model, multi-level factors (such as Prime type) are analyzed such that one level is mapped on the intercept and a regression weight is calculated for all other levels, indicating the difference from that intercept condition. We mapped the Control condition to the intercept, giving regression coefficients which showed the difference of the two priming conditions from the Control condition. For the r-initial targets, the Prime effect was significant [$F(2, 796) = 14.33, p < .001$]: RTs were significantly shorter in the (matching) R-prime condition than in the Control condition [$t(796) = -4.94, p < .001$], but RTs in the Vowel-prime and the Control condition did not significantly differ [$t(796) = -.70, p > .1$]. For the vowel-initial targets, Prime type was again significant [$F(2, 783) = 11.7, p < .001$]: RTs were shorter in the (matching) Vowel-prime condition than in the Control condition [$t(783) = -3.48, p < .01$], but RTs in the R-prime and the Control condition did not significantly differ [$t(783) = 1.20, p < 1$]. Thus the native listeners showed significant priming from each type of matching prime, but no priming from mismatching primes.

Dutch listeners

Target type and Prime type [$F(2, 1575) = 3.40, p < .05$] interacted in this data set also, so the two conditions of Target type were again analyzed separately. For r-initial targets, there was a significant main effect of Prime type [$F(2, 797) = 4.66, p < .05$]: RTs were shorter after (matching) /r/ primes than after Control primes [$t(796) = -3.05, p < .01$]. Unlike in the British data, there was also a trend towards faster RTs [$t(796) = -1.71, p = .088$] after the (mismatching) Vowel primes. For the vowel-initial targets there was no main effect of Prime type [$F(2, 776) = 2.63, p = .073$]. In the Dutch data, we thus find selective priming only by the r-initial prime, which leads to priming for /r/ targets but not for vowel-initial targets. The vowel-initial prime generates some priming for both target types, missing significance in each case. This is the source of the three-way interaction in the overall analysis, given that the British listeners show selective priming for both prime types.

Discussion

The native listeners performed as expected: Hearing a vowel-initial word preceded by an intrusive /r/ primed vowel-initial target words but not r-initial target words, while hearing an r-initial word primed r-initial target words but not vowel-initial target words. These results replicate and extend findings by Spinelli et al. (2003) who also found that native listeners are able to use subphonemic cues to resolve apparent ambiguities (in that case liaison in French). The effect that we find is stronger than that found by Spinelli et al., in that the native listeners here show a significant priming effect for the intended phrase only; no hint of a priming effect appears when there is a subphonemic mismatch between the prime and the target (see Fig. 2). This is as predicted from the results of our phonetic categorization study (Tuinman et al., 2011), in which British English listeners based their interpretation of /r/ as onset or intrusive exclusively on the acoustic-phonetic evidence.

The Dutch listeners performed differently, although they too showed some sensitivity to this subphonemic difference. Their priming from onset /r/ versus intrusive /r/ differed, as evident from the Prime type by Target type interaction. However, they were less sensitive than the British English listeners, as shown by the overall three-way interaction. For r-initial targets, the matching primes produced significant priming, but mismatching intrusive /r/ primes also led to a RT reduction (not observed for the native listeners). Moreover, the Dutch listeners showed no significant priming for vowel-initial targets after a matching vowel-initial prime preceded by intrusive /r/ (only a trend in that direction, again). For these non-native listeners, intrusive /r/ was apparently a good match for neither the vowel-initial nor the r-initial targets. Some activation then appeared for both word interpretations (see Fig. 2). These results are also in line with Tuinman et al.'s (2011) findings, where Dutch listeners made almost no use of durational information in interpreting /r/.

However, although these results show that Dutch listeners fail to use the acoustic evidence for /r/ in a native-like manner in recognizing words, the RT patterns were insufficiently conclusive to provide a complete picture of their processing. It is possible, however, to probe listeners' perception of specific phonetic patterns with a method that gives a closer view of the relative activation of alternative word candidates in cases of lexical ambiguity, namely eye-tracking, or the "visual world" paradigm (Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995). This we therefore did in our next study.

Most visual world studies have measured looks to pictures (Tanenhaus & Spivey-Knowlton, 1996), but recent studies have also successfully used printed words in this paradigm. McQueen and Viebahn (2007), for instance, examined the influence of mismatching phonetic evidence on spoken-word recognition in a printed-word version of a study performed by Allopenna, Magnuson, and Tanenhaus (1998) with pictures. Allopenna et al. found that participants looked more at pictures of offset-mismatch competitors (e.g. a beetle, given the spoken target *beaker*) than of onset-mismatch competitors (e.g. a speaker, given the same target), and McQueen and Viebahn similarly found more looks to printed Dutch words which were offset mismatch competitors (e.g., *buffer* 'buffer', given the target *buffel* 'buffalo') than to onset-mismatch competitors (e.g., *lotje* 'lottery ticket', given the target *rotje* 'firecracker'). Subsequent eye-tracking experiments with printed words have examined the effects of connected-speech processes (Mitterer & McQueen, 2009) and adaptation to a speaker of a different dialect (Dahan, Drucker, & Scarborough, 2008).

Eye tracking with printed-word targets is thus able to show subtle effects in speech perception. Thus even though our priming target words do not lend themselves easily to pictorial representation, we can use eye-tracking to gain more information on how the Dutch (and British) listeners perceive intrusive /r/. Eye-tracking is a more powerful method than cross-modal priming, as both target types (r-initial and vowel-initial words) can be tested on a single trial, with listeners asked to choose between them.

Experiment 2

Method

Participants

Twenty-four native Dutch speakers (aged 18–24, $m = 21$ years) meeting the description given in Experiment 1 took part, as did 22 native speakers of British English (aged 18–28, $m = 22$ years). The latter were students and postgraduates at the University of Birmingham. None of either group had participated in Experiment 1.

Stimuli

Participants watched a screen with four printed words while hearing an auditory sentence. There were 108 auditory sentences overall. Fifty-four sentences contained a member of a minimal pair such as *aided/raided* or *ice/rice* – the same 27 pairs as were used in the previous experiment. For each word pair, there was a pronunciation with an onset /r/ (e.g., *raided*) and a pronunciation with an intrusive /r/, where the speaker intended the vowel-initial word (e.g., *aided*). These were in fact the same recordings as used in the earlier experiments, but the sentences were played in full, not truncated after the crucial word. Fifteen of the 27 sentence pairs were disambiguated after the r-initial and vowel-initial words by following context, while the other 12 remained ambiguous. An additional 54 filler sentences were used with filler words inserted in the slot of the r-initial or vowel-initial (e.g., *pages, space*).

For the visual stimuli, 27 quadruplets of words were presented four times. The words were presented on the centers of the four quadrants of the screen. Each of the four words was the target word on one presentation, and a competitor or distractor on the other three occasions. Each critical r-initial and vowel-initial word was hence presented once as a target, once as a competitor (the r-initial words as competitor for the vowel-initial word, and vice versa), and two times as a distractor for other consonant-initial words.

Procedure

Participants were tested individually in a quiet room. They were seated in front of the computer screen at a comfortable reading distance. Eye movements were monitored using a SMI EyeLink eye-tracking system sampling at 250 Hz for the Dutch listeners, and a SR Research EyeLink II eye-tracking system for the British listeners. The experiment was controlled by computer (Compaq 486 for the Dutch listeners, Dell D800 for the British listeners). Audio input was presented over headphones (using NESU software for the Dutch listeners, SR Research ExperimentBuilder software for the British listeners). The eye-tracker was mounted and calibrated for each participant. Both eyes were monitored, but only data from the right eye were analyzed.

Each trial had the following structure. First, a central fixation dot appeared on the screen for 500 ms, followed by a blank screen for 600 ms. Then, a spoken sentence was presented and, at the same time, a two-by-two array with four words appeared on the screen. Participants had

received written instructions to click on the word mentioned in the spoken sentence, using the computer mouse. They were told to click on one of the four words even when they were not sure which word they heard, and informed that they did not have to wait till the end of the sentence to click. After a mouse click the next trial was initiated. The position of the four words in each trial was randomized across the four locations on the screen. After every five trials a fixation point appeared centered on the screen, and participants were instructed to look at it. The experimenter could then correct potential drifts in the calibration of the eye tracker.

A different randomized list was created for each participant, containing the 54 filler and 54 experimental items. Within each list, there were 27 sentences with an r-initial target word, 27 sentences with a vowel-initial target word and 54 filler trials in which these target words formed distractors. Four practice trials began the experiment.

Design

There is one independent variable with two levels; type of /r/ with the levels intrusive /r/ and onset /r/. The two dependent variables are the proportion of correct responses, and the difference in looks to r-initial words and vowel-initial words.

Results

Table 1 shows the percentages of correct responses. Both listener groups made hardly any errors on filler trials. On experimental trials, however, clear differences appeared. The Dutch listeners had a strong preference for r-initial words, even when hearing an intrusive /r/. When the auditory sentence contained an intended onset /r/, participants clicked correctly on the r-initial word in 96.5% of the cases. When the auditory sentence contained an intrusive /r/ and the vowel-initial word was intended, participants still clicked erroneously on the r-initial word in 70.9% of the cases. The British listeners had a moderate bias towards /r/-responses, with 28.5% clicks on r-initial words when the vowel-initial word was intended. Interestingly, the British listeners were slightly less accurate (93.5%) than the Dutch listeners (96.5%) when the r-initial word was the intended target.

These data were analyzed with a linear mixed-effect model with subjects and items as random factors, and four fixed factors. Two of those were factorial: Type of /r/ and Native Language. The other two were covariates: Was the sentence ambiguous throughout or not, and trial number. To allow an estimation of change in results over time, trial number was scaled from -0.5 to 0.5 from the start to the end of the experiment. The analysis began with a full

Table 1
Percentage of correct responses in Experiment 2.

	Condition		
	Vowel targets (%)	r-Initial targets (%)	Fillers (%)
Dutch listeners	29.1	96.5	98.6
English listeners	71.5	93.2	99.9

model with all possible interactions. Table 2 shows the model after pruning of insignificant interactions. The cell with Dutch listeners reacting to r-initial words was mapped on the intercept. The fact that Dutch listeners were almost always correct on r-initial trials (see Table 1) is reflected in this high intercept value of 4 on a logOdds scale (=98% correct responses). The large negative regression weight for “target = vowel” reflects the poor performance of Dutch listeners when the vowel-initial word was a target. Note that the small but significant negative regression weight for nativeLanguage = British is only to be interpreted relative to the intercept: when the speaker intended an r-initial word, British listeners performed slightly worse than the Dutch. The large regression weight for [target = vowel × nativeLanguage = British] reflects the fact that British listeners performed much better than the Dutch when the speaker intended a vowel-initial word. The main effect for the ambiguity covariate indicates that performance was better when the sentence was not ambiguous. The simple effect of trialNumber and its interaction with target = vowel indicates that participants shifted their criterion over the course of the experiment, with more errors on trials with r-initial targets but fewer errors on trials with vowel-initial targets. The numerical similarity of these two effects (−1.36 and 1.44) indicates that overall performance level did not change but rather, participants made fewer errors for vowel targets and more errors for r-initial targets. Note that all these latter effects were independent of native language; the interactions of those effects with native language were not significant and so had been pruned from the model.

Fig. 3 shows the averaged fixation proportions from −200 ms before target onset until 1200 ms after target onset, respectively for onset /r/ and for intrusive /r/ trials. Here the difference between the two groups is very obvious in the vowel target figures; when the speaker utters a vowel-initial word with intrusive /r/, British listeners eventually look at the intended word, but Dutch listeners mostly do not. We analyzed whether participants had a preference for the interpretation of the /r/ as an onset /r/ and to what extent this depended on the stimulus. This was achieved by using a difference measure “fixation to r-initial word – fixation to vowel-initial word” as the dependent variable in a linear mixed-effect model, again with subjects and items as random factors and type of /r/

as a fixed factor. Fixation proportions were transformed into logOdds for this analysis. If, in such model, the intercept is significantly different from zero, this shows a preference for one visual word over the other. Given that the preference measure is defined as “fixation to r-initial word – fixation to vowel-initial word”, positive values indicate more looks to r-initial words, negative values a preference for vowel-initial words. We tested this for time windows with a width of 100 ms, between 100 ms and 1000 ms. Fig. 4 shows the beta weights of the analysis. It takes about 200 ms for eye-movements to reflect an influence of incoming speech (e.g., Matin, Shao, & Boff, 1993), and indeed, in the first two time windows (100–300 ms), neither the type of word nor the type of /r/ had a significant influence for either listener group. For all other time windows there is a significant preference for r-initial words for both groups. A combined analysis shows that the preference for the r-initial word is larger in the British groups from 400 ms to 700 ms, but not in other time windows. There are clear differences driven by the type of /r/ in the stimulus. For British listeners, this effect arises at 300–400 ms, i.e., in the first time window in which significant effects can be observed. For Dutch listeners, it is only significant from 500 to 600 ms onwards. A combined analysis showed that the British listeners were more strongly influenced than the Dutch listeners by the type of /r/ in the stimulus, in all time windows from 300 ms onwards. That is, the effect of /r/ type arises immediately for British listeners, and they immediately differ from the Dutch listeners.

We also tested whether semantic cues had an influence on the fixation proportions. It is important to note here that not all our sentence were “minimal pairs”, such as the example *Emma (r)jected the cassette*. Some sentences diverged after the critical word (e.g., *Sheila elates everybody with her presence* versus *Sheila relates well to other people*). However, whether the sentence remained ambiguous after the target word or not had no measurable influence on the fixation proportions for either group. The preference for the intended word in the late time-windows was therefore signal driven.

Another noteworthy pattern is the difference between the distractors and the competitor for trials with an onset /r/. The Dutch listeners more or less ignore the competitor, the vowel-initial word, giving it hardly more looks than the phonologically unrelated distractors. The British listeners,

Table 2

Regression weights of the analysis for the accuracy of the click responses in Experiment 2. The values are for the final model with all insignificant effects pruned. The regression weights are on a logOdds scale (4 = 98% correct, 3 = 95% correct, 2 = 88% correct, 1 = 77% correct, 0 = 50% correct). See text for further explanation.

	Regression weight	Std. error	z-value
Intercept (Target = r-initial and Native Language = Dutch)	4.00	0.83	4.84***
Target = vowel	−6.62	0.64	−10.28***
nativeLanguage = English	−1.10	0.37	−2.94**
Ambiguity (0 = ambiguous, 1 = unambiguous)	0.86	0.42	2.06*
trialNumber (scaled to [−0.5,0.5])	−1.36	0.60	−2.27*
Target = vowel × nativeLanguage = English	3.61	0.37	9.80***
Target = vowel × trialNumber	1.44	0.67	2.13*

* $p < .05$.

** $p < .01$.

*** $p < .001$.

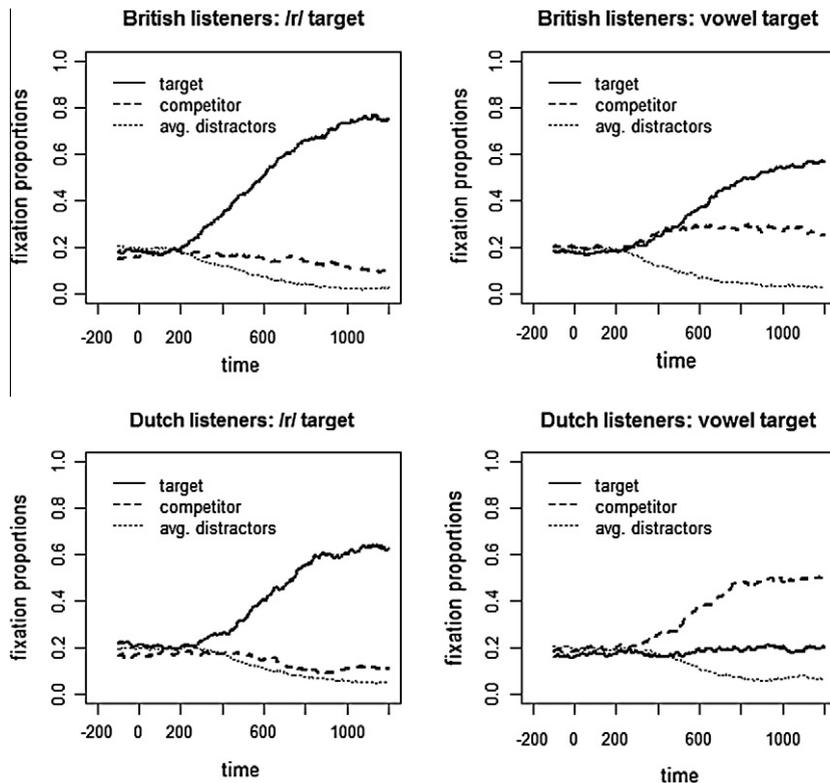


Fig. 3. Fixation proportions over time to the target, the competitor and averaged distractors, for sentences with an intrusive /r/ and sentences with an onset /r/.

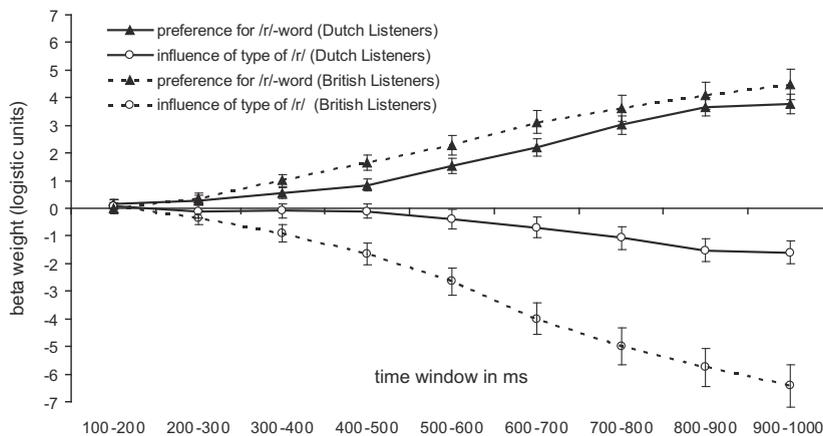


Fig. 4. Time window analysis of the eye-tracking data based on the regression weights. The filled triangles show the extent of the preference for the r-initial word and the open circles the impact of the prime type (intrusive versus onset /r/). The data are based on separate analysis for the two listener groups justified by significant interactions with native language for prime type (from 300 ms onwards) and for overall preference (from 400 to 700 ms).

however, show a clear competition effect. To investigate whether these effects are statistically reliable, we tested whether the proportion of looks to the competitor was larger than the proportion of looks to the distractors, again in consecutive time windows from 100 ms to 1000 ms after target onset. (Proportions were logit-transformed for this analysis as well.) In this analysis, there was no significant competition effect in the Dutch data in any time window ($t_{\max} = 1.8$), while the British listeners had a significant

preference for the competitor over the distractor from 400 ms onwards ($t_{\min} = 2.1$). A combined analysis showed a significantly stronger competition effect for British than for Dutch listeners from 700 ms onward ($t_{\min} = 3.3$).

Discussion

These results broadly confirm those of Experiment 1, in that British listeners efficiently differentiate intrusive from

onset /r/ while Dutch listeners do not. Nevertheless, the present results also amplify what could be learned from Experiment 1, by showing that both groups of listeners are biased towards interpreting /r/ as an onset. Cross-modal priming was able to show that British listeners were very accurate in eventually distinguishing the two types of /r/. Eye-tracking has revealed that British listeners, too, can mistake an intrusive /r/ for an onset /r/. It thus shows that their accuracy is achieved in defiance of this.

As the eye-tracking data show, British participants are much more sensitive than the Dutch listeners to the sub-phonemic difference between an onset and an intrusive /r/. At the first hint of a rhotic in the signal, the Dutch listeners directed their eye-gaze toward the printed r-initial word. That is, in the time window 300–400 ms after word onset, the first time window in which a signal-driven effect can be expected given the latency of eye-movements, we see a preference for r-initial words in the Dutch responses. The difference in realization of an intrusive /r/ and an onset /r/ only influences their eye-movements 300 ms later.

The difference in realization between the two /r/ types is largely in duration, so that to use it listeners might be thought to need to wait for the /r/ to end. However, note that British listeners use this difference in /r/ type in the first time window in which any target-related eye-movement are observed. This suggests that the British listeners are making use of the non-duration difference between onset and intrusive /r/, in the amount of intensity decrement during the /r/. This differentiating information is immediately available for these listeners, but Dutch listeners fail to make early use of it.

One tantalizing aspect of the British performance is that these native listeners make slightly more errors than the Dutch L2 listeners on trials when the speaker intended an onset /r/. That is, they sometimes interpret an onset /r/ as an intrusive /r/, a possibility that does not seem to occur to the Dutch listeners. This interpretation is borne out by an analysis of the strength of the competition effects. British listeners experience stronger competition from vowel-initial words in phrases such as *Canada raided*, apparently because they take into account the possibility that the /r/ may be an intrusive /r/.

General discussion

The /r/-intrusion process of British English has provided a window on substantial differences between the processing applied by native listeners and by proficient L2 listeners. In brief: the intrusion increases, for all listeners, potential lexical competition. The native listeners recover from this very rapidly indeed. The L2 listeners do not.

Our experiments employed the two most widely used methods for examining lexical activation and competition: cross-modal priming (Experiment 1) and eyetracking (Experiment 2). These provide complementary views of the competitor population activated by speech, and how the resulting competition is resolved and the speaker's message understood. The speech examined in our study consisted of British English sentences involving intrusive /r/, a casual speech phenomenon that, like many others

across many languages, creates potentially ambiguous sequences but also offers sub-phonemic cues that listeners can exploit in disambiguation. In each experiment we assessed how such sentences were processed by native British listeners, and by proficient L2 listeners whose native language, Dutch, does not feature this phonological process.

In cross-modal priming, we examined all possible combinations of primes and targets. For example, in a phrase such as *Canada[r]aided*, where /r/ could potentially be an onset (i.e., *raided* is the intended verb) or intrusive (i.e., *aided* is the intended verb), we tested utterances with both these intended verbs as primes, and we tested both *raided* and *aided* as targets. The native British English listeners effectively exploited the sub-phonemic differences between intrusive and onset /r/ to resolve the potential ambiguity: an intended *raided* produced priming for *raided* but not *aided*, and an intended *aided* primed *aided* but not *raided*. The Dutch listeners, however, appeared unable to distinguish effectively between intrusive and onset /r/s. For them, an onset /r/ primed r-initial words but not vowel-initial words, but an intrusive /r/ seemed equivalently relevant to both types of target.

In eye-tracking we presented the same potentially ambiguous sentences, again in both versions, while listeners saw displays of four printed words, including alternatives such as *aided* and *raided*. The task was to click on the word that occurred in the sentence. The difference between native and L2 listeners was here even clearer. British listeners showed a clear preference for the intended word – they looked more at and clicked more often on the r-initial word when the speaker intended that word, and they looked more at and clicked more often on the vowel-initial word when the speaker intended that word and produced an intrusive /r/ before it. Dutch participants, in contrast, interpreted the majority of stimuli (with either onset or intrusive /r/) as an onset /r/, preferring to look at and click on the r-initial word in both versions of our sentences.

Not only were the looking preferences and click choices in the eyetracking study informative; the time course effects were also telling. The British listeners made immediate use of the sub-phonemic differences between intrusive and onset /r/, while the Dutch did not. These L2 listeners looked at first at the r-initial word, irrespective of the stimulus properties of the input. Over time, though, their preference for the r-initial word became stronger in the cases where the speaker actually intended that word, suggesting that they were able to detect the cues that were relevant to eventual disambiguation. Also, the early competition effects made apparent the native listeners' sensitivity to potential /r/ intrusion; on trials in which the speaker intended the r-initial word, although L2 listeners hardly looked more at the vowel-initial word than at unrelated distractors, the native listeners did look at the vowel-initial option, clearly entertaining the possibility that the /r/ might be an intrusive one. These listeners thus know that evidence for /r/ in the signal does not necessarily constitute evidence for an intended utterance of /r/, and so they show a stronger competition effect from vowel-initial competitors when hearing an onset /r/.

These results are informative in several different ways. They highlight L1–L2 listening differences, and further specify the view of spoken-word processing to be gained from the cross-modal priming and the eyetracking techniques; they shed light on what sort of information can be effectively imported from L1 to L2; and, most generally, they expand our knowledge of how listeners deal with casual speech phenomena.

With respect to the L1–L2 comparison, first, it is very clear that our two groups performed differently, and that this was due primarily to the L1 group's greater familiarity with this particular phonological process, and with how information in the signal disambiguates it. Both groups could hear that /r/ was present, from the earliest moment that the signal allowed (see Fig. 4). This translated to a significantly greater proportion of looks to r-initial words, for all listeners overall. But for the British listeners, and only for them, this preference was modulated by the knowledge that /r/ could also be an intruder preceding a vowel-initial word, and that durational and other cues needed to be consulted to sort this out. This can be seen in Fig. 4 as the greater influence they show of the acoustic cues to type of /r/. However, the L2 listeners can also hear the difference between /r/ types; their beta weight for this value in Fig. 4 is also non-null. The effect of this sensitivity for them can be seen in the lower panels of Fig. 3, by comparing their looks to the dispreferred items in the display. These L2 listeners are by preference looking at and clicking on the r-initial words in both panels, meaning that in the left lower panel in this figure, they are looking at the correct target, but in the right lower panel (where the target is vowel-initial), they are looking at the competitor. But their dispreferred choices differ in the two panels; on the left the competitor and distractors are looked at equally, but on the right, the vowel-initial (target) word receives more looks than the distractors through to the last measured time window. They too are thus sensitive to differences in the realization of /r/ in the signal; but they do not adequately act upon this sensitivity.

In our phonetic categorization study (Tuinman et al., 2011), where an artificial durational manipulation was built into our materials, we also found that L1 and L2 listeners alike could perceive this manipulation; the native listeners based their *ice-rice* categorizations exclusively on the durational factor, and ignored all else, but the L2 listeners allowed semantic and orthographic factors to outweigh it. The present cross-modal priming results have reinforced the conclusion that native listeners can act upon their processing of the precise realization of /r/; they showed fully appropriate priming patterns for the intended word only, whichever it was. The L2 listeners' cross-modal priming results further emphasized that they cannot muster such lexical selectivity.

Note that on a methodological level, our eye-tracking and priming results, especially for the British listeners, confirm the complementary nature of these two tasks. The priming data alone, with totally selective priming patterns for each word, indicate that British listeners are perfectly able to differentiate between intrusive and onset /r/. These data alone would not tell us at what level this differentiation was achieved. The eye-tracking data, in turn,

allow us to see that the British listeners do consider the alternative interpretations; alone they would not show that recovery from this competition process is both rapid and complete. Together the two tasks provide the full picture of initial multiple evaluation of word candidates and exploitation of information in the signal that culminates in completely secure decision-making at the level of final lexical selection.

This pattern of results thus extends and confirms the picture that has arisen from L2 listening research across decades: L1–L2 differences are found less in perceptual sensitivity to speech input than in the linguistic processing of it. Just as the inability to distinguish unfamiliar foreign speech sounds arises from phonological interpretation of the signal rather than perceptual insensitivity (Best, McRoberts, & Sithole, 1988), so do non-native listening experiments repeatedly reveal perceptual sensitivity in conjunction with inability to exploit its results at higher levels of processing. Thus Japanese listeners' goodness ratings on tokens of English /r/ and /l/ pattern differently for the two categories (Iverson et al., 2003), but at the lexical level this difference is not exploited to enable, for instance, correct early choice between *rocket* versus *locker* in an eyetracking task (Cutler et al., 2006). The same pattern appears with Dutch listeners' processing both of [æ] versus [ɛ] (vowels that divide a single Dutch category) and of word-final voicing distinctions (which are neutralized in Dutch); in both cases Dutch listeners perform quite well in a low-level choice task (Broersma, 2005), but fail to distinguish lexical minimal pairs such as *cattle-kettle* or *robe-rop* in cross-modal priming (Broersma & Cutler, 2008, 2011) or in eyetracking (Escudero, Hayes-Harb, & Mitterer, 2008; Weber & Cutler, 2004). In fact one of the most well-known L2 effects, the disproportionate difficulty of listening to an L2 in a noisy environment, shows the same pattern; a review of four decades of literature on this topic (Lecumberri, Cooke, & Cutler, 2010) motivated the conclusion that noise impinges upon the initial uptake of speech by L1 and L2 listeners to an equivalent degree, but L1 listeners recover better from its effects. We see here a similar pattern in the processing of casual speech phenomena.

What is striking about the L2 listeners' performance in the present case is how far it falls short of the reports of L2 casual speech processing that were described in the introduction. Both French liaison (Shoemaker, 2010) and French voicing assimilation (Darcy et al., 2007) were reported to be dealt with in a native-like manner by advanced listeners to French. Yet our advanced Dutch listeners to English were far from equaling native levels of performance in dealing with the British English intrusive /r/ process. This is despite the fact that the type of processing required – exploiting sub-phonemic distinctions such as differences in segment duration – is certainly familiar to Dutch listeners from their L1. They successfully distinguish ambiguous sequences such as *een spot* versus *eens pot* on the basis of the duration of /s/ (Shatzman & McQueen, 2006), and they distinguish *ham* 'ham' from the first syllable of *hamster* 'hamster' on the basis of vowel duration (Salverda, Dahan, & McQueen, 2003). Dutch has as many casual speech processes as any of its linguistic relatives, and native Dutch listeners deal with these efficiently,

including with segment insertion processes such as vowel epenthesis (indeed, vowel epenthesis, as when *film* /film/ becomes /fɪləm/, can make word recognition easier: van Donselaar, Kuijpers, & Cutler, 1999). Furthermore, our participants would surely have been exposed to instances of /r/ intrusion simply because they will have been exposed to speech in British English. Why then do they show such poor ability to exploit the cues that allow onset versus intrusive /r/ to be distinguished?

As noted in the introduction, /r/ intrusion is likely to be difficult because it is optional (but so are many assimilation processes), and because it causes ambiguity (but so do other processes). We suggest that the crucial reason for the difficulty of /r/ intrusion for L2 listeners is that the way it causes ambiguity is by adding a new segment to the speech stream. Although processing can be enhanced for words with an added segment arising from vowel epenthesis in a consonant cluster (van Donselaar et al., 1999), this latter result was traced to its effects of enhancing the clarity of consonants, which have long been known to be easier to perceive when adjacent to a vowel rather than to another consonant (Liberman, Delattre, Cooper, & Gerstman, 1954). Further, of course, this finding is from L1 listening and no information is available on the perception of such epenthised words by L2 listeners whose L1 shows no equivalent process. Addition of a consonant into a sequence of vowels in the manner of /r/-intrusion is not a process known in Dutch.

The effect of an added phoneme is to enlarge the lexical competitor population with spurious competitors supported by the intruding phoneme. Native listeners can discard these extra competitors by correctly ascribing the added phoneme to its external source, but for listeners without the experience to motivate such a solution, competitors of this kind are difficult to get rid of. Making the correct selection in a phrase such as *Canada aided* with added /r/ in fact requires leaving a vowelless residue of the input unaccounted for (the /r/), in violation of the Possible Word Constraint (Norris, McQueen, Cutler, & Butterfield, 1997). This constraint on speech segmentation and competition is operative across languages (Cutler, Demuth, & McQueen, 2002; Hanulíková, McQueen, & Mitterer, 2010; McQueen, Otake, & Cutler, 2001; Yip, 2004) and very effectively removes competition from embedded words – such as, for instance, *aid* in (intended) *raid*.

In contrast, the processes with which L2 listeners have been shown to cope well do not add an extra phoneme. Altering or deleting phonemes may well weaken support for an intended word, but no spurious competitors are added. Assimilation processes, for instance, change a single feature of a phoneme, to the value of an adjacent phoneme; the number of segments stays the same. Liaison causes an underlying phoneme to surface; but note that French liaison is an obligatory process (not doing it counts as mispronunciation), and as such it is taught in school French lessons. Further, the surfacing phoneme is present in the orthography and in other non-liaison forms (consider French *petit*, the masculine form of ‘small’, with a final *t* that is pronounced before a vowel-initial word, e.g., in *petit agneau* ‘little lamb’; in *petite*, the feminine form, the second *t* is always pronounced).

Also, the best L2 performance in these cases has appeared in rather lower-level tasks than the word recognition procedures employed here. Shoemaker (2010) used two-alternative forced choice, and Darcy et al. (2007) asked listeners to decide whether a word in a sentence was the same as a prior pronunciation of the word in isolation (probe detection that was effectively same-different judgement). As Broersma (2005; Broersma & Cutler, 2008, 2011) showed with phoneme contrasts, low-level tasks may reveal discrimination that does not carry through to lexical processing. We do not in fact know whether Darcy et al.’s participants would continue to demonstrate native-level performance in cross-modal priming or eye-tracking. Dejean de la Batie and Bradley’s (1995) finding that L2 listeners respond to liaison consonants as onset consonants in phoneme-monitoring suggests that Shoemaker’s discrimination findings do not translate simply to the lexical processing of liaison. Tremblay’s (2011) eye-tracking study of liaison sequences confirms this; less advanced L2 listeners showed early looks to consonant-initial competitors where native listeners did not.

Finally, and crucially, then, our results also illuminate the processing of casual speech phenomena in general. From our L1 results, we now have further strong evidence for native listeners’ use of phonetic detail to resolve ambiguity arising from inter-word phonetic processes in natural speech. From our L2 results, we derive new evidence that may constrain the models that have been proposed in this domain.

As described in the introduction, models in this area fall into two general classes; those that explain successful compensation in terms of natural processes based in speaking and listening (Gow, 2003; Mitterer, Csépe, & Blomert, 2006; Mitterer, Csépe, Honbolygo et al., 2006), and those that explain it in terms of learning accumulated from linguistic experience (e.g., Darcy et al., 2007; Gaskell, 2003). Supporting the former kind of model is native-like performance by listeners without any prior experience of a process (Gow & Im, 2004; Mitterer, Csépe, & Blomert, 2006), while supporting the latter is performance improvement that parallels language experience (Darcy et al., 2007; Shoemaker, 2010; Tremblay, 2011). Our present results suggest that compensation for /r/ intrusion does not fall within the scope of the former class of models. However, our results also do not support the latter class of models, since our L2 listeners were experienced but still failed to compensate for the /r/ intrusion.

We suggest that one crucial dimension in accounting for how a connected-speech process is perceived is the naturalness of the process. Many assimilation processes are common across languages and grounded in processes of articulation and perception. The place and voicing assimilations investigated by Darcy et al. (2007), for instance, stem from the spreading of articulatory features to neighboring segments. The Hungarian assimilation investigated by Mitterer, Csépe, and Blomert (2006) and Mitterer, Csépe, Honbolygo et al. (2006), involved liquids, for which the acoustic consequences of assimilation are not highly salient, and listeners without experience of the rule in question behaved similarly to native listeners. The two cases investigated by Gow and Im (2004) differed

somewhat in naturalness, with the voicing assimilation being less marked than the labial-to-velar place assimilation, and indeed, Gow and Im's results indicated that the voicing assimilation was easier for their listeners to process.

In contrast, the /r/ insertion process does not make claim to such naturalness. We are aware of one claim for an articulatory source of /r/-insertion – Broadbent's (1991) interpretation in terms of 'glide formation', the process which gives rise to intervocalic /j/ or /w/ in *kiosk* or *hoeing*. This claim however has little basis. Unlike these two cases, the articulatory target for /r/ does not follow naturally from a gestural transition from a mid or low vowel to another vowel. Further, the intrusive /r/ is not dependent on the place or height of the following vowel, only on the preceding one; this also indicates that it does not naturally arise from a gestural transition. It is far from universal; other languages have developed other strategies to realize a transition between a vowel-final and a vowel-initial word, such as glottal stops, often reduced to creaky voice, in German (Kohler, 1994). This strategy is also employed by some British English speakers to avoid the intrusive /r/, and, as noted, the intrusive /r/ is indeed avoided in more formal speech (Brown, 1988).

Natural processes such as place and voice assimilation, then, will be relatively easy to compensate for at an early processing level, and this compensation should carry through to higher processing levels such as word recognition. The more natural a process, the less necessity for learning. Learning will, however, assist in compensation at higher levels in the sense that advanced learners will have larger vocabularies and more accurate knowledge of the range of variant realisations of words. In other words, both classes of models can usefully add to an account of natural connected-speech phenomena.

For less natural processes, both classes of models predict that learning is needed, and here listener ability to compensate will depend on the precise effect of the process. Phoneme alterations (e.g., Gow and Im's (2004) labial-to-velar assimilation) or deletions (e.g., of /t/, see Tuinman, 2006) will be easier to cope with than additions; among the latter, those that enhance signal clarity (e.g., vowel epenthesis) or have an otherwise obvious source (e.g., liaison) will be easier than those for which neither advantage holds. British English /r/ intrusion thus has several strikes against it; L1 listeners can invoke their experience and deal with it successfully, but L2 listeners can only cope with John Lennon's *I saw a film* by realizing, at some level, that *I soar a film* makes no sense at all.

Acknowledgments

This research was supported by a doctoral scholarship from the Max Planck society to the first author. Partial reports of the results were presented to the 16th International Congress of Phonetic Sciences, Saarbrücken, August 2007, and to the 5th Workshop on Disfluency in Spontaneous Speech, Tokyo, September 2010. We thank Alan Garnham for enabling the testing of Experiment 1 at the University of Sussex, and Sotaro Kita for enabling the test-

ing of Experiment 2 at the University of Birmingham. The first author is now at the Centraal Bureau voor de Statistiek, Den Haag.

Appendix

Sentence pairs used in the experiments.

A few days ago, I saw aces when I looked at my cards
A few days ago, I saw races from the British Superbike Championship

I heard that Canada aided the area of Lesotho in Africa
I heard that Canada raided the area of Lesotho in Africa

I really thought that I saw air burning
I really thought that I saw rare animals

Did you know that the terracotta ear of the statue was broken?
Did you know that the terracotta rear of the statue was broken?

My brother likes extra ice when he has dinner
My brother likes extra rice when he has dinner

I think that Rebecca owes a lot to London
I think that Rebecca rows a lot to London

In north Malaysia itches are number one on the list of annoyances
In north Malaysia riches are frowned upon

And then Emma ejected the cassette
And then Emma rejected the cassette

I read that people from China etch whenever they feel like doing so
I read that people from China retch whenever they feel like doing so

My youngest sister saw odes to Rome made by many different people
My youngest sister saw roads to Rome made by many different people

I saw on Discovery Channel that people in Panama ate cows' eyes in former times
I saw on Discovery Channel that people in Panama rate David as the nicest city

My favourite grandma aged because her dog suddenly died
My favourite grandma raged because her dog suddenly died

Do you think it was sepia ash on her body?
Do you think it was sepia rash on her body?

We had not expected it, but suddenly Natasha emitted a cry of pain
We had not expected it, but suddenly Natasha remitted the money to us

Usually, Sheila elates everybody with her presence
Usually, Sheila relates well to other people

The Broadway musical “Clarissa” evolves into a deeply moving metaphor for the struggles of an entire generation

The Broadway musical “Clarissa” revolves around a young American girl in London

I read that the president of Algeria evokes the idea of using nuclear weapons

I read that the president of Algeria revokes his decision to cut taxes

I heard that the Australia Alley is a nice street to live in

I heard that the Australia rally lasts two weeks

For Buddhists in India enunciation is of great importance

For Buddhists in India renunciation is part of their daily lives

We asked the children to draw apt presents for Mother’s Day

We asked the children to draw wrapped presents for Mother’s Day

Obviously, the bourgeois eye expert was bragging about his salary to his Harvard friends

Obviously, the bourgeois rye expert was bragging about his salary to his Harvard friends

Davidson and Brooks claim to be thorough ale-manufacturers

Davidson and Brooks claim to be thorough rail-manufacturers

The president of Russia eagerly awaits his caviar

The president of Russia regally decorated his new office

In this area anglers are still ice fishing

In this area wranglers are experienced horse back tour guides

In postwar America aches and pains are usually ignored by people without health insurance

In postwar America rakes and planes are sold in hardware shops

You should use extra old cheese in this recipe

You should use extra rolled oats in this recipe

I think that the extra ink cartridges were too expensive

I think that the extra rink did not result in more ice skaters visiting it

References

- Allopenna, 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. doi:10.1006/jmla.1997.2558.
- Best, C. T., McRoberts, G. W., & Sithole, N. M. (1988). Examination of perceptual reorganization for nonnative speech contrasts – Zulu click discrimination by English-speaking adults and infants. *Journal of Experimental Psychology: Human Perception and Performance*, 14, 345–360. doi:10.1037/0096-1523.14.3.345.
- Best, C. T., & Tyler, M. D. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In O.-S. Bohn & M. J. Munro (Eds.), *Language experience in second language speech learning: In honor of James Emil Flege* (pp. 13–34). Amsterdam and Philadelphia: John Benjamins.
- Blevins, J. (2004). *Evolutionary phonology: The emergence of sound patterns*. Cambridge: Cambridge University Press.
- Boersma, P. (1998). *Functional phonology: Formalizing the interactions between articulatory and perceptual drives*. The Hague: Holland Academic Graphics.
- Bohn, O.-S., & Munro, M. J. (Eds.). (2007). *Language experience in second language speech learning: In honor of James Emil Flege*. Amsterdam and Philadelphia: John Benjamins.
- Broadbent, J. (1991). Linking and intrusive r in English. *UCL Working Papers in Linguistics*, 3, 281–302.
- Broersma, M. (2005). Perception of familiar contrasts in unfamiliar positions. *Journal of the Acoustical Society of America*, 117, 3890–3901. doi:10.1121/1.1906060.
- Broersma, M., & Cutler, A. (2008). Phantom word activation in L2. *System*, 36, 22–34. doi:10.1016/j.system.2007.11.003.
- Broersma, M., & Cutler, A. (2011). Competition dynamics of second-language listening. *Quarterly Journal of Experimental Psychology*, 64, 74–95. doi:10.1080/17470218.2010.499174.
- Brown, A. (1988). Linking, intrusive, and rhotic /r/ in pronunciation models. *Journal of the International Phonetic Association*, 18, 144–151. doi:10.1017/S0025100300003765.
- Collins, B., & Mees, I. M. (1996). *The phonetics of English and Dutch*. Leiden, The Netherlands: Brill.
- Connine, C. M., Ranbom, L. J., & Patterson, D. J. (2008). Processing variant forms in spoken word recognition: The role of variant frequency. *Perception & Psychophysics*, 70, 403–411.
- Cruttenden, A., & Gimson, A. C. (1994). *Gimson’s pronunciation of English* (5th ed.). London: Arnold.
- Cutler, A., Demuth, K., & McQueen, J. M. (2002). Universality versus language-specificity in listening to running speech. *Psychological Science*, 13, 258–262. doi:10.1111/1467-9280.00447.
- Cutler, A., Weber, A., & Otake, T. (2006). Asymmetric mapping from phonetic to lexical representations in second-language listening. *Journal of Phonetics*, 34, 269–284. doi:10.1016/j.wocn.2005.06.002.
- Dahan, D., Drucker, S. J., & Scarborough, R. A. (2008). Talker adaptation in speech perception: Adjusting the signal or the representations? *Cognition*, 108, 710–718.
- Darcy, I., Peperkamp, S., & Dupoux, E. (2007). Bilinguals play by the rules: Perceptual compensation for assimilation in late L2-learners. In J. Cole & J. I. Hualde (Eds.), *Laboratory Phonology 9* (pp. 411–442). Berlin: Mouton de Gruyter.
- Dejan de la Batie, B., & Bradley, D. C. (1995). Resolving word boundaries in spoken French – Native and nonnative strategies. *Applied Psycholinguistics*, 16, 59–81. doi:10.1017/s0142716400006421.
- Escudero, P., Hayes-Harb, R., & Mitterer, H. (2008). Novel second-language words and asymmetric lexical access. *Journal of Phonetics*, 36, 345–360. doi:10.1016/j.wocn.2007.11.002.
- Gaskell, M. G. (2003). Modeling regressive and progressive effects of assimilation in speech perception. *Journal of Phonetics*, 31, 447–463. doi:10.1016/S0095-4470(03)00012-3.
- Giegerich, H. J. (1992). *English phonology: An introduction*. Cambridge, UK: Cambridge University Press.
- Gimson, A. C. (1980). *An introduction to the pronunciation of English* (3rd ed.). London: Edward Arnold.
- Goto, H. (1971). Auditory perception by normal Japanese adults of the sounds “l” and “r”. *Neuropsychologia*, 9, 317–323.
- Gow, D. W. (2002). Does English coronal place assimilation create lexical ambiguity. *Journal of Experimental Psychology: Human Perception and Performance*, 28, 163–179. doi:10.1037/0096-1523.28.1.163.
- Gow, D. W. (2003). Feature parsing: Feature cue mapping in spoken word recognition. *Perception & Psychophysics*, 65, 575–590. doi:10.3758/BF03194584.
- Gow, D. W., & Gordon, P. C. (1995). Lexical and prelexical influences on word segmentation – Evidence from priming. *Journal of Experimental Psychology: Human Perception and Performance*, 21, 344–359. doi:10.1037/0096-1523.21.2.344.
- Gow, D. W., & Im, A. M. (2004). A cross-linguistic examination of assimilation context effects. *Journal of Memory and Language*, 51, 279–296. doi:10.1016/j.jml.2004.05.004.
- Grimson, A., & Cruttenden, A. (1994). *Gimson’s pronunciation of English* (5th ed.). London: Edward Arnold.

- Gussenhoven, C., & Broeders, A. (1997). *English pronunciation for student teachers* (2nd ed.). Groningen: Wolters-Noordhoff-Longman.
- Hanulíková, A., McQueen, J. M., & Mitterer, H. (2010). Possible words and fixed stress in the segmentation of Slovak speech. *Quarterly Journal of Experimental Psychology*, 63, 555–579. doi:10.1080/17470210903038958.
- Iverson, P., Kuhl, P. K., Akahane-Yamada, R., Diesch, E., Tohkura, Y., Kettermann, A., et al. (2003). A perceptual interference account of acquisition difficulties for non-native phonemes. *Cognition*, 87, B47–B57. doi:10.1016/S0010-0277(02)00198-1.
- Kemps, R., Ernestus, M., Schreuder, R., & Baayen, H. R. (2004). Processing reduced word forms: The suffix restoration effect. *Brain and Language*, 90, 17–127. doi:10.1016/S0093-934X(03)00425-5.
- Kohler, K. J. (1994). Glottal stops and glottalization in German data and theory of connected speech processes. *Phonetica*, 51, 38–51.
- Lecumberri, M. L. G., Cooke, M., & Cutler, A. (2010). Non-native speech perception in adverse conditions: A review. *Speech Communication*, 52, 864–886. doi:10.1016/j.specom.2010.08.014.
- Liberman, A. M., Delattre, P. C., Cooper, F. S., & Gerstman, L. J. (1954). The role of consonant-vowel transitions in the perception of the stop and nasal consonants. *Psychological Monographs*, 68, 1–13.
- Martin, E., Shao, K. C., & Boff, K. R. (1993). Saccadic overhead: Information processing time with and without saccades. *Perception & Psychophysics*, 53, 372–380.
- McQueen, J. M. (2007). Eight questions about spoken-word recognition. In M. G. Gaskell (Ed.), *The Oxford handbook of psycholinguistics* (pp. 37–53). Oxford: Oxford University Press.
- McQueen, J. M., Otake, T., & Cutler, A. (2001). Rhythmic cues and possible-word constraints in Japanese speech segmentation. *Journal of Memory and Language*, 45, 103–132. doi:10.1006/jmla.2000.2763.
- McQueen, J. M., & Viebahn, M. (2007). Tracking recognition of spoken words by tracking looks to printed words. *Quarterly Journal of Experimental Psychology*, 60, 661–671. doi:10.1121/1.419865.
- Mitterer, H., Csépe, V., & Blomert, L. (2006). The role of perceptual integration in the recognition of assimilated word forms. *Quarterly Journal of Experimental Psychology*, 59, 1395–1424. doi:10.1080/17470210500198726.
- Mitterer, H., Csépe, V., Honbolygo, F., & Blomert, L. (2006). The recognition of phonologically assimilated words does not depend on specific language experience. *Cognitive Science*, 30, 451–479. doi:10.1207/s15516709cog0000_57.
- Mitterer, H., & Ernestus, M. (2006). Listeners recover /t/s that speakers lenite: Evidence from /t/-lenition in Dutch. *Journal of Phonetics*, 34, 73–103. doi:10.1016/j.wocn.2005.03.003.
- Mitterer, H., & McQueen, J. M. (2009). Processing reduced word-forms in speech perception using probabilistic knowledge about speech production. *Journal of Experimental Psychology: Human Perception and Performance*, 35, 244–263. doi:10.1037/a0012730.
- Norris, D., McQueen, J. M., Cutler, A., & Butterfield, S. (1997). The possible-word constraint in the segmentation of continuous speech. *Cognitive Psychology*, 34, 191–243. doi:10.1006/cogp.1997.0671.
- Pallier, C., Bosch, L., & Sebastian-Gallés, N. (1997). A limit on behavioral plasticity in speech perception. *Cognition*, 64, B9–B17. doi:10.1016/S0010-0277(97)00030-9.
- Quene, H. (1992). Durational cues for word segmentation in Dutch. *Journal of Phonetics*, 20, 331–350.
- 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. doi:10.1016/S0010-0277(03)00139-2.
- Sebastián-Gallés, N., Rodríguez-Fornells, A., Diego-Balaguer, R. d., & Diaz, B. (2006). First and second language phonological representations in the mental lexicon. *Journal of Cognitive Neuroscience*, 18, 1277–1291.
- Shatzman, K. B., & McQueen, J. M. (2006). Prosodic knowledge affects the recognition of newly-acquired words. *Psychological Science*, 17, 372–377. doi:10.1111/j.1467-9280.2006.01714.x.
- Shoemaker, E. M. (2010). Nativelike attainment in L2 listening: The segmentation of spoken French. Paper presented at the New Sounds 2010: Sixth international symposium on the acquisition of second language speech, Poznan.
- Spinelli, E., McQueen, J. M., & Cutler, A. (2003). Processing resyllabified words in French. *Journal of Memory and Language*, 48, 233–254. doi:10.1016/S0749-596X(02)00513-2.
- Strange, W. (1995). *Speech perception and linguistic experience. Theoretical and methodological issues in cross-language speech research*. Timonium, MD: York Press.
- Tanenhaus, M. K., & Spivey-Knowlton, M. J. (1996). Eye-tracking. *Language and Cognitive Processes*, 11, 583–588. doi:10.1080/016909696386971.
- Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, 268, 1632–1634. doi:10.1126/science.7777863.
- Tremblay, A. (2011). Learning to parse liaison-initial words: An eye-tracking study. *Bilingualism-Language and Cognition*, 14, 257–279. doi:10.1017/S1366728910000271.
- Tuinman, A. (2006). Overcompensation of /t/-reduction in Dutch by German-Dutch bilinguals. In C. Fougerson (Ed.), *Abstracts of laboratory phonology 10* (pp. 101–102), Paris.
- Tuinman, A., Mitterer, H., & Cutler, A. (2011). Perception of intrusive /r/ in English by native, cross-language and cross-dialect listeners. *Journal of the Acoustical Society of America*, 130, 1643–1652. doi:10.1121/1.3619793.
- van Donselaar, W., Kuijpers, C., & Cutler, A. (1999). Facilitatory effects of vowel epenthesis on word processing in Dutch. *Journal of Memory and Language*, 41, 59–77.
- Weber, A., & Cutler, A. (2004). Lexical competition in non-native spoken-word recognition. *Journal of Memory and Language*, 50, 1–25. doi:10.1016/S0749-596X(03)00105-0.
- Yip, M. C. W. (2004). Possible-Word Constraints in Cantonese speech segmentation. *Journal of Psycholinguistic Research*, 33, 165–173. doi:10.1023/B:JOPR.0000017225.65288.4b.