Journal of Memory and Language 66 (2012) 530–544

Contents lists available at SciVerse ScienceDirect

Journal of Memory and Language

journal homepage: www.elsevier.com/locate/jml

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.

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 [ə]-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).

0749-596X/$ - see front matter © 2012 Elsevier Inc. All rights reserved.
doi:10.1016/j.jml.2012.02.001

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. Viably 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