Processing casual speech in native and non-native language

ISBN: 978-90-76203-39-3

Printed and bound by Ipskamp Drukkers, Nijmegen
Cover design by Harald Pieper, In Zicht Grafisch Ontwerp

Processing casual speech in native and non-native language

Een wetenschappelijke proeve op het gebied van de

Sociale Wetenschappen

Proefschrift

ter verkrijging van de graad van doctor

aan de Radboud Universiteit Nijmegen

op gezag van de rector magnificus prof. mr. S.C.J.J. Kortmann,

volgens besluit van het college van decanen

in het openbaar te verdedigen op woensdag 6 april 2011

om 15.30 uur precies

door

Annechien Elina Tuinman

geboren op 16 januari 1981

te Groningen

Promotor:	Prof. dr. A. Cutler
Copromoter:	Dr. H. Mitterer (MPI)

Manuscriptcommissie:

Prof. dr. A.G. van Hell

Prof. dr. W.A. van Dommelen (Norwegian University of Science and Technology)

Dr. E. Spinelli (Pierre Mendès France University of Grenoble)

The research reported in this thesis was supported by a grant from the Max-Planck-Gesellshaft zur Förderung der Wissenshaften, München, Germany.

Voor Elina Birza

Dankwoord

In mijn hoofd had ik dit dankwoord al vele malen geschreven. Maanden keek ik uit naar het moment waarop mijn proefschrift af was en ik eindelijk dit stuk tekst mocht schrijven.

Allereerst wil ik Anne Cutler, mijn promotor, hartelijk bedanken. Zonder haar enthousiasme, vertrouwen in mij en mijn werk en haar vele bemoedigende woorden was dit proefschrift er vermoedelijk nooit gekomen. Met haar verhalen over Winston Churchill en zijn "big black dog" stak ze mij een hart onder de riem en overtuigde zij mij ervan dat ik na al die tijd mijn proefschrift toch kon en moest afmaken. Anne, bedankt dat je me gestimuleerd hebt en altijd tijd nam om stukken te lezen, ook ver weg in Australië toen de deadline naderde.

Ten tweede gaat mijn dank uit naar Holger Mitterer, mijn begeleider en co-promoter. Als zijn eerste PhD-student heb ik er ongetwijfeld soms voor gezorgd dat hij met de handen in het zat. Dit heeft hij echter nooit laten merken. Holger, bedankt dat je altijd rustig bent gebleven, en voor de hulp die je hebt geboden bij het overdoen van analyses—want na al die tijd waren die oude ANOVA's natuurlijk hopeloos achterhaald. Je bracht structuur aan in de laatste maanden en nam alle tijd als ik op het MPI was.

Mijn taakgroepchefs op het Centraal Bureau voor de Statistiek, Hilda Folkerts en Douwe Kuurstra, wil ik bedanken voor de tijd die ze me hebben gegeven om dit proefschrift af te ronden. Het bleek de steun in de rug die ik nodig had. Gerrit Zijlmans en Elisabeth Eenkhoorn, bedankt voor jullie interesse in mijn 'vage' onderzoek.

Ook wil ik iedereen in de Comprehension Group van het MPI, oude en nieuwe mensen, bedanken voor de goede sfeer. Zelfs toen ik na een lange radiostilte weer op het instituut verscheen, waren jullie vriendelijk en behulpzaam. In het bijzonder wil ik Marijt Wittemans en Patrick van der Zande noemen omdat ze mij me weer thuis lieten voelen. Ook denk ik met genoegen terug aan alle group meetings, gezellige lunches en potluck dinners met of zonder Holgers chocolademousse.

Warme herinneringen bewaar ik aan Claudia Kuzla met wie ik 's zomers zat te puffen in vermoedelijk de warmste kamer van het instituut. Met Susanne Brouwer ('the other tall, blond Dutch girl') zat ik een tijd op de 'koele' kamer, allebei druk met

het afronden van onze proefschriften. Susanne, bedankt dat je een stimulans was om hard door te werken, en dat het toch ook heel gezellig was.

In mijn 'cohort' zaten ook Heidrun Bien, Laura Herbst en Federico Rossano. Dankzij Heidrun voelde ik me al snel thuis op het MPI en binnen een paar weken trainden we voor de Vierdaagse van Nijmegen, die we in de zomer van 2005 glorieus uitliepen. Met Laura deelde ik tijdens een congres in Saarbrücken zelfs een hotelbed, omdat er geen kamers met twee eenpersoonsbedden meer beschikbaar waren. Ik denk niet dat ik dat snel vergeet. Federico 'a little less conversation' Rossano, your presence on the first floor meant a lot to me.

Mijn reisgenote Mirjam Broersma wil ik bedanken voor een leuke tijd in Brighton waar we in korte tijd een groot aantal proefpersonen testten. Overdag hard aan het werk en 's avonds heerlijk uit eten in een organisch-biodynamisch-vegetarisch restaurant (met verse gemberthee).

Ook zonder de mensen van de technichal group, de administratie, de bibliotheek en de receptie was het werken aan dit proefschrift een stuk moeilijker en minder leuk geweest. Agnes de Boer, bedankt voor de leuke gesprekken aan de balie, Pim Folten en Thea Schopman bedankt voor alle lekkere broodjes en salades. Rian Zondervan wil ik graag bedanken dat ze er altijd was als ik haar hulp nodig had.

I would like to thank Alan Garham of the Laboratory of Experimental Psychology at the University of Sussex for the opportunity to test the British participants I needed for my research. Thanks to Delphine Dahan at the University of Pennsylvania for having me in her lab for my Fulbright research for six months, where I could also test the American participants for this PhD research. It was a wonderful and intense time that I would not have wanted to miss.

Evelien, je was tijdens mijn stageperiode op het NICI een geweldige huisgenoot en werd een goede vriendin. Ik kan je niet genoeg bedanken voor alle momenten dat ik bij je terecht kon, letterlijk en figuurlijk. Zonder jou was de laatste periode dat ik aan mijn proefschrift werkte nooit zo goed verlopen. Sarah, het MPI was niet meer hetzelfde zonder jou een verdieping lager. Bedankt voor de Vierdaagses, je luisterend oor, je scherpe analyses en je vertrouwen in mij. Ook nu we allebei ouder, wijzer én moeder zijn, ben ik blij dat jij er bent voor mij. Ik ben heel gelukkig dat Evelien Akker en Sarah Schimke mijn paranimfen willen zijn.

Ten slotte wil ik mijn familie en vrienden bedanken. Het werken aan mijn proefschrift was niet altijd leuk, maar jullie hebben me er doorheen gesleept en met rust gelaten wanneer dat nodig was. Lieve pap en mam, bedankt voor alles.

Lieve Senne, jouw komst was de mooiste deadline die ik had kunnen wensen. Lieve lief, zonder jou was ik helemaal nergens.

Contents

Introduction

Chapter 1

Learning a second language

As most people know from their own experience, learning another language can be quite a challenge. Learning a new language not only involves learning new words and expressions, but also new grammatical structures and possibly even a different script. Dutch people are often admired because they speak several foreign languages, but what is forgotten is that they mainly master languages spoken in neighbouring countries that are very similar to the Dutch language: German and English. For Dutch people it is relatively easy to acquire German and English because the script, the vocabulary and the structure of the languages are very much alike. Spanish and Italian are already harder to learn for Dutch speakers because of larger differences in vocabulary and grammar. Japanese and Arabic prove to be even a bigger challenge, because in these cases Dutch speakers also have to learn a completely different script. In other words, learning another language is easiest when it is similar to your native language; the new language is then already somewhat familiar. The aim of this thesis was to investigate the role of familiarity in the perception of casual, connected speech. Familiarity here means that the listeners are familiar with a casual speech process in the second language from their first language experience.

Many studies have investigated the role of familiarity in second language learning. Rhythmic similarity between the native and non-native language, for example, appears to be helpful in non-native listening. Listeners rely on native-language rhythm in segmenting the speech stream into words; in different languages stress-, syllable- or mora-based rhythm is exploited. This language-specificity affects listening to a non-native language if native strategies are applied even though they are inefficient in segmenting the non-native language. However, speakers of two languages with similar rhythmic patterns such as English and Dutch effectively segment both languages in a similar fashion. Even speakers of the unrelated languages Telugu and Japanese effectively use similar procedures in segmenting speech in the other language, as Telugu and Japanese are rhythmically comparable (Cutler, Murty, & Otake, 2003).

Similarity in speech sounds between the native language and the second language also makes it easier to learn the second language. If speech sounds in the second language are very difficult to distinguish for the non-native listeners this logically hinders second language learning. A notorious example of confusable sounds is that of /r/ and /l/ for Japanese listeners. The sounds are not only difficult to pronounce for native speakers of Japanese (Flege, Takagi, & Mann, 1995), but it is also difficult for them to hear the difference (Best & Strange, 1992). This phoneme contrast is difficult because in Japanese there is no /r/ or /l/, but the two sounds both strongly resemble one Japanese sound which is somewhere between /r/ and /l/. This explains why mistakes such as *Eric Crapton* instead of *Eric Clapton* are made on Japanese CD covers—most Japanese speakers find it hard to distinguish /r/ and /l/ (Fleming, 2002).

Languages can also have similar or different phonotactic contraints. Phonotactic constraints determine which speech sounds and which combinations of sounds can occur where. The language-specificity of phonotactic contraints has consequences for second language listening. First, second language listeners have less experience with the phonotactic contraints of their second language so they cannot optimally make use of this kind of information. A second consequence of the language-specificity of phonotactics is that second language listeners also use native language phonotactic contraints which are not helpful in their second language. It is, for example, an English phonotactic constraint that the cluster /pf/ cannot occur in a pre- or postvocalic position. In German, however, the cluster /pf/ does exist (e.g., *Pfanne*, 'pan' and *Kopf*, 'head'). Because of the different phonotactic contraints in English and German, native speakers of these languages will possibly parse the English phrase *lope for* differently; for English listeners the word boundary can only be between *lope* and *for*, whereas for German listeners *low pfor* is also a possible word segmentation. And indeed, even though proficient second language listeners can acquire the phonotactic probabilities of their second language and use them effectively in segmenting speech, at the same time they are not completely able to prevent interference for phonotactic constraints in their native language while listening to their second language (Weber & Cutler, 2006).

Casual speech processes

There is another dimension on which languages can differ: casual speech processes. Casual speech processes are processes that occur in spontaneously spoken speech which lead to phonetic forms which deviate from the canonical pronunciation of the words as can be heard in read speech in for example news broadcasts. And most of the speech that listeners hear is spoken spontaneously and abounds with casual speech processes such as assimilation, reduction, deletion and intrusion. Due to the process of *assimilation*, for example, the phrase *right berries*, where the final [t] of *right* is assimilated to a [p], can sound much like *ripe berries* (Gow, 2002). Speech variation due to the *extreme reduction* of word forms or phonemes occurs frequently and the word *natuurlijk* /naˈtyrlək/ ("of course"), for example, can be reduced to [nˈtyk] (Ernestus, 2000: 137). The *epenthesis*, or insertion, of phonemes is another process that diverges speech from the canonical form. In Dutch, for example, the insertion of an optional epenthetic vowel in word-final consonant clusters is a widespread phenomenon and can turn *film* into *filəm*, making it a two-syllable word (van Donselaar, Kuijpers, & Cutler, 1999).

In recent years, psycholinguistics has turned increasingly to the investigation of casual speech, and how native listeners deal with the non-canonical forms it presents. A grossly over-simplified summary of the accrued results to date is that while listeners are extremely good at exploiting the fine phonetic detail of utterances and identifying intended words even when casual speech processes have altered them from their canonical form, the alterations can often (temporarily) mislead listeners, and can often result in word recognition being harder than it would have been for the canonically pronounced versions. The fine differences between intended phonemes and phonemes resulting from a casual speech process have been shown to be exploited by listeners, for example in the case of place of articulation assimilation (e.g., to distinguish the /p/ of English *ripe* in *ripe berries* from the assimilated final phoneme of *right* in *right berries*; Gow, 2002), and in liaison (e.g., to distinguish the word-initial /p/ in French *trop partisan* from the liaison realization of a word-final /p/ in *trop artisan*; Spinelli, McQueen & Cutler, 2003). Listeners are successful at identifying word forms despite assimilation of place (Gaskell & Marslen-Wilson, 1996; Gow, 2001) or of voice (Snoeren, Segui, & Hallé, 2008) and despite reduction

(Ernestus, Baayen, & Schreuder, 2002) or other non-canonical realizations (e.g., van Alphen & McQueen, 2006; Sumner & Samuel, 2005).

Despite all this success at dealing with real-speech forms, however, native listeners are also often misled. Thus in a phoneme detection task they respond to phonemes which are not actually in the input at all, because they have been deleted in a casual pronunciation (Kemps, Ernestus, Schreuder, & Baayen, 2004), and they respond to phonemes which are accidentally inserted there, such as a medial /p/ in a casual pronunciation of *something* (Warner & Weber, 2001). Their word recognition response times are slowed by many different types of casual speech forms (Andruski, Blumstein, & Burton, 1994; LoCasto & Connine, 2002; Racine & Grosjean, 2000), and they can be seriously misled, at least temporarily, into assuming that a quite different word is being heard (Brouwer, Mitterer, & Ernestus, 2008).

Higher-level information in spoken-word recognition

Several studies have shown that native listeners use higher-level knowledge to compensate for bad or ambiguous signals. Already more than three decades ago it was shown that listeners tend to label a perceptually ambiguous sound on a voicing continuum (e.g., a continuum between [d] and [t]) such that the utterance they heard formed an existing word (as [d] in *dice-tice*, but as [t] in *dype-type*) (Ganong, 1980). In other words, listeners used their word knowledge to interpret the ambiguous sound. This effect was later replicated in word-medial position (e.g., [d] and [g] in *cradle-cragle* and *badel-bagel*, Connine, 1990) and word-final position (e.g., [s] and [ʃ] in *kiss-kish* and *fiss-fish*; McQueen, 1991).

Not only lexical knowledge has been shown to be used by native listeners, but also sentential knowledge. Listeners are more likely to identify the word that makes most sense in the sentence they have heard compared with words that are less sensible (Borsky, Shapiro, & Tuller, 2000; Borsky, Tuller, & Shapiro, 1998; Connine, 1987; Connine, Blasko, & Hall, 1991; Gaskell & Marslen-Wilson, 2001). When listeners, for example, heard stimuli from a *goat-coat* continuum embedded in sentences that biased interpretation to either "goat" (*The busy farmer hurried to milk the* [?]oat *in the drafty barn*) or "coat" (*The careful laundress had to dry-clean the* [?]oat *in the cluttered attic*), a sentential context effect was found in that listeners gave more *goat* identifications in sentences with a *goat* interpretation bias and more *coat* responses for sentences with a *coat* bias.

But how does the use of higher-level knowledge work when listeners hear an ambiguous sound? Models of spoken-word recognition account for lexical context effects (e.g., [d] and [t] in *dice-tice* and *dype-type*) by assuming that the word recognition system has two basic levels. The first is the prelexical level, where an abstract representation of the speech input is made. The second is the lexical level at which all words that match the prelexical (phonemic) representation of the input to a certain extent are activated and compete for recognition. However, the models have different assumptions about the nature and possible directions of the information flow, and thus different explanations for lexical context effects. One explanation for lexical effects, incorporated in interactive models such as TRACE (McClelland & Elman, 1986), is that information can flow from the lexicon back to the prelexical level and directly influence perception. The ambiguous signal [?ype], for example, would activate the lexical representation of *type* on the lexical level, which feeds back support to the representation of the phoneme [t] at the prelexical level. As a result, more [t] responses are expected in the [?ype] context than in the [?ice] context (in which feedback would produce a bias toward [d]). Autonomous models of speech perception (Cutler & Norris, 1979; Forster, 1976), on the other hand, argue that context effects are not due to feedback in speech processing, but represent integration of independent information sources. In the Merge model, for example, information flows from the prelexical level to the lexicon without feedback; in other words, information can only flow from the bottom up (McQueen, Norris, & Cutler, 1999; Norris, McQueen, & Cutler, 2000). Phonemic decisions are made at dedicated decision units that continuously receive input from both the prelexical and lexical levels of processing. At the decision stage, information from the two levels is merged and they both influence the final decision. The Merge model assumes that a signal like [?ype] activates the decision units for both [t] and [d] through connections from the prelexical level to the phoneme decision units. At the same time, the signal also activates the word representation for *type*, which feeds support to the phoneme decision unit [t], creating a lexical bias toward [t]. Accordingly, more [t] responses are expected in a [?ype] context than in a [?ice] context. Sentential context effects (as in the *coat-goat* example given above) can be explained in a similar fashion by both the feedback account and the account based on a decision bias.

This dissertation

Most research on how listeners cope with casual speech processes has been carried out with native listeners. These studies showed that even native listeners are sometimes misled, so what is going to happen when second language listeners hear the same sort of ambiguous speech? Hear it they will, because all languages have casual speech processes, and second language listeners cannot always confine themselves to speech situations in which the input is as close to canonical perfection as it is in the classroom or on language tapes. Research on non-native speech comprehension, on the other hand, has mainly dealt with the perception of segments (vowels and consonants) which are present in the second language but not in the first. Therefore this dissertation takes a different approach and further examines the perceptual effects of casual speech processes in second language listening, focusing on the perception of connected speech in the second language with the use of segments (consonants) that *do* exist in both the first and the second language. However, what was varied in the different chapters is whether the connected speech processes that these segments undergo exist in both the first and the second language.

First, it was investigated how native listeners and proficient non-native listeners cope with word-final /t/-reduction in Dutch. The casual speech process of /t/-reduction is found in many languages (Guy, 1980). In the Germanic languages English, German and Dutch, the process patterns very similarly. For instance, /t/ is very likely to be deleted after /s/ or before a bilabial, so that most utterances of English *postman*, German *Postbeamter* 'postal worker' or Dutch *postbode* 'postman' are equally unlikely to contain much of a detectable trace of /t/. The perception of /t/ in various degrees of reduction has been intensively studied in Dutch (e.g., Janse, Nooteboom, & Quené, 2007; Mitterer & Ernestus, 2006; Mitterer & McQueen, 2009); from this body of work we know that Dutch listeners are highly sensitive to the patterning of this reduction process in their native language. The fact that reduction is more likely before labial than before other coronal segments, for example, drives perceptual interpretations, including the choice between word forms (e.g., deciding whether one has heard *kas* 'greenhouse' or *kast* 'cupboard'; Janse et al., 2007; Mitterer & McQueen, 2009). If native listeners find it useful to draw on knowledge of the production patterning in perception, the same knowledge would surely be of use to second language listeners as well. Therefore, in this dissertation, Dutch speech was presented to native speakers of Dutch and to native speakers of German who are highly proficient in Dutch. It was investigated how well German listeners could cope with /t/-reduction in their second language Dutch, and whether they had more trouble with it than the native Dutch listeners. As described above, it is easier to learn a second language if it is very similar to the native language. Thus, the expectation was that the German listeners should not have major problems with /t/-reduction in Dutch, as their native language knowledge of the process should be similar to that of the Dutch listeners since both languages have /t/-reduction and it patterns similarly in each.

It was also investigated how well second language listeners cope with a casual speech process that does not exist in their native language. Therefore, the process of /t/-reduction, which is frequent of occurrence, was compared with a far less widespread process, namely the insertion of /r/ between words beginning and ending with vowels, in British English sequences such as *idea of*. This process is unknown in many other languages, for instance in Dutch (Collins & Mees, 1999). British English speech was presented to native listeners of British English and to speakers of Dutch who are highly proficient in English as a second language to test the assumption that /r/-intrusion is problematic for Dutch listeners as they are unfamiliar with this casual speech process from their native language. Additionally, it was investigated whether native speakers of the same language, but a different variety, could distinguish an inserted /r/ from a regular /r/. Hence, native speakers of American English also listened to British English sentences with or without an intrusive /r/. Because British English has the /r/-insertion process but Dutch and most American dialects do not, it could be tested what is the influence of the native language; native speakers of American might find /r/-insertion easier to cope with than the Dutch native speakers because this process at least occurs in their native language albeit not in their native variety.

One likely possibility is that second language listeners might not be able to make use of the bottom-up signal to the same extent as native listeners when confronted with casual speech processes. To compensate for this, they might rely more strongly on higher-level knowledge such as lexical and sentence context information to make sense of the, for them, ambiguous bottom-up input. As reviewed above, native listeners also make use of higher-level knowledge when confronted with ambiguous input. The range of input that is ambiguous may, however, be larger for second language listeners than for native listeners. Consequently, it was also investigated what the influence was of higher-level knowledge on the perception of both /t/-reduction and intrusive /r/. As described above, higher-level knowledge can be used by listeners to resolve ambiguities in the speech input, and second language listeners might (need to) use it more than native listeners to compensate for casual speech processes.

As reviewed above, it is still a matter of debate how higher-level knowledge is integrated in the process of spoken-word recognition. Research focusing on this issue has usually employed artificially generated ambiguous speech input. As it turns out, casual speech processes are a natural source of ambiguous input, and may therefore be a more appropriate test bed for the controversy between bottom-up and interactive models of spoken-word recognition. Therefore, it was investigated how the integration of information works in the case of a casual speech process. One type of model assumes that information can flow from the lexicon back to the prelexical level and directly influence perception. According to the other type of model information flows from the prelexical level to the lexicon without feedback; in other words, information can only flow from the bottom up. Different studies have found evidence for both types of model, but these studies never used "real" ambiguous speech as can be the result of a casual speech process. It might be the case that when listeners have to compensate for ambiguous speech due to a casual speech process they do this differently than when they have to compensate for arbitrary ambiguous signals. Therefore, it was investigated whether feedback of information is necessary to compensate for /t/-reduction in Dutch.

To summarize, this thesis presents three lines of empirical work. In the first line, the perception of casual speech processes that was familiar to native and second language listeners was investigated. The process under study was /t/-reduction that occurs in both Dutch and German, and the empirical work tested how Dutch listeners and German learners of Dutch perceive /t/-reduction in Dutch. The second line of research made use of the process of /r/-insertion in British English. It was tested how three different listener groups deal with this process: British listeners, who are very familiar with this process, American English listeners, who are not familiar with /r/-insertion but still listen to their native language, and Dutch learners of English, who are not familiar with /r/-insertion given their native language background. The third line of research investigates how higher-level information is integrated in spoken-word recognition when ambiguity arises due to a casual speech process.

References

Alphen, P. M. van., & McQueen, J. M. (2006). The effect of voice onset time differences on lexical access. *Journal of Experimental Psychology: Human Perception and Performance, 32*, 187-196.

Andruski, J. E., Blumstein, S. E., & Burton, M. (1994). The effect of subphonetic differences on lexical access *Cognition, 52*, 163-187.

Best, C. T., & Strange, W. (1992). Effects of phonological and phonetic factors on cross-language perception of approximants. *Journal of Phonetics, 20*, 305-330.

Borsky, S., Shapiro, L. P., & Tuller, B. (2000). The temporal unfolding of local acoustic information and sentence context. *Journal of Psycholinguistic Research, 29*, 155-168.

Borsky, S., Tuller, B., & Shapiro, L. P. (1998). "How to milk a coat": The effects of semantic and acoustic information on phoneme categorization. *Journal of the Acoustical Society of America, 103*, 2670-2676.

Brouwer, S., Mitterer, H., & Ernestus, M. (2008). How acoustically reduced forms activate the lexicon: Evidence from eye-tracking *Laboratory Phonology 11*. Wellington, New Zealand.

Collins, B., & Mees, I. M. (1999). *The phonetics of English and Dutch* (4th ed.). Leiden: Brill.

Connine, C. M. (1987). Constraints on interactive processes in auditory word recognition: The role of sentence context. *Journal of Memory and Language, 26*, 527-538.

Connine, C. M. (1990). Effects of sentence context and lexical knowledge in speech processing. In G. T. M. Altmann (Ed.), *Cognitive models of speech processing: Psycholinguistic and computational perspectives* (pp. 281-294). Cambridge, MA: MIT Press.

Connine, C. M., Blasko, D. G., & Hall, M. (1991). Effects of subsequent sentence context in auditory word recognition: Temporal and linguistic constraints. *Journal of Memory and Language, 30*, 234-250.

Cutler, A., Murty, L., & Otake, T. (2003). Rhythmic similarity effects in non-native listening? *Proceedings of the 15th International Congress of Phonetic Sciences (ICPhS 2003)* (pp. 329-332). Adelaide: Causal Productions.

Cutler, A., & Norris, D. (1979). Monitoring sentence comprehension. In W. E. Cooper & E. C. T. Walker (Eds.), *Sentence processing: Psycholinguistic studies presented to Merrill Garrett*. Hillsdale, NJ: Erlbaum.

Ernestus, M. (2000). *Voice assimilation and segment reduction in Dutch [dissertation]*. Utrecht, The Netherlands: LOT.

Ernestus, M., Baayen, H., & Schreuder, R. (2002). The recognition of reduced word forms. *Brain and Language, 81*, 162-173.

Flege, J. E., Takagi, N., & Mann, V. (1995). Japanese adults can learn to produce English /r/ and /l/ accurately. *Language and Speech, 38*, 25-55.

Fleming, B. (2002). Eric Crapton. Retrieved August, 24, 2010, from http://www.engrish.com/2002/02/eric-crapton/.

Forster, K. I. (1976). Accessing the mental lexicon. In R. J. Wales & E. Walker (Eds.), *New approaches to language mechanisms* (pp. 257-287). Amsterdam: North-Holland.

Ganong, W. F. (1980). Phonetic categorization in auditory word perception. *Journal of Experimental Psychology: Human Perception and Performance, 6*, 110-125.

Gaskell, M. G., & Marslen-Wilson, W. D. (1996). Phonological variation and inference in lexical access. *Journal of Experimental Psychology: Human Perception and Performance, 22*, 144-158.

Gaskell, M. G., & Marslen-Wilson, W. D. (2001). Lexical ambiguity resolution and spoken word recognition: Bridging the gap. *Journal of Memory and Language, 44*, 325-349.

Gow, D. W. (2001). Assimilation and anticipation in continuous spoken word recognition. *Journal of Memory and Language, 45*, 133-159.

Gow, D. W. (2002). Does English coronal place assimilation create lexical ambiguity? *Journal of Experimental Psychology: Human Perception and Performance, 28*, 163-179.

Guy, G. R. (1980). Variation in the group and the individual. In W. Labov (Ed.), *Locating language in time and space* (pp. 1-36). New York: Academic Press.

Janse, E., Nooteboom, S. G., & Quené, H. (2007). Coping with gradient forms of /t/-deletion and lexical ambiguity in spoken word recognition. *Language and Cognitive Processes, 22*, 161-200.

Kemps, R. J., Ernestus, M., Schreuder, R., & Baayen, H. (2004). Processing reduced word forms: The suffix restoration effect. *Brain and Language, 90*, 117-127.

LoCasto, P., & Connine, C. M. (2002). Rule-governed missing information in spoken word recognition: Schwa vowel deletion. *Perception and Psychophysics, 64*, 208-219.

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

McQueen, J. M. (1991). The influence of the lexicon on phonetic categorization: Stimulus quality in word-final ambiguity. *Journal of Experimental Psychology: Human Perception and Performance, 17*, 433-443.

McQueen, J. M., Norris, D., & Cutler, A. (1999). Lexical influence in phonetic decision-making: Evidence from subcategorical mismatches. *Journal of Experimental Psychology: Human Perception and Performance, 25*, 1363-1389.

Mitterer, H., & Ernestus, M. (2006). Listeners recover /t/s that speakers reduce: Evidence from /t/-lenition in Dutch. *Journal of Phonetics, 34*, 73-103.

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.

Norris, D., McQueen, J. M., & Cutler, A. (2000). Merging information in speech recognition: Feedback is never necessary. *Behavioral and Brain Sciences, 23*, 299-325.

Racine, I., & Grosjean, F. (2000). Influence de l'effacement du schwa sur la reconnaissance des mots en parole continue. *L'Année Psychologique, 100*, 393-417.

Snoeren, N., Segui, J., & Hallé, P. (2008). Perceptual processing of partially and fully assimilated words in French. *Journal of Experimental Psychology: Human Perception and Performance, 34*, 193-204.

Spinelli, E., McQueen, J. M., & Cutler, A. (2003). Processing resyllabified words in French. *Journal of Memory and Language, 48*, 233-254.

Sumner, M., & Samuel, A. G. (2005). Perception and representation of regular variation: The case of final /t/. *Journal of Memory and Language, 52*, 322-338.

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.

Warner, N., & Weber, A. (2001). Perception of epenthetic stops. *Journal of Phonetics, 29*, 53-87.

Weber, A., & Cutler, A. (2006). First-language phonotactics in second-language listening. *Journal of the Acoustical Society of America, 119*(1), 597-607.

Transfer of L1 knowledge helps in doubly adverse conditions: Perception of casual speech in L2

Chapter 2

Tuinman, A., & Mitterer, H. (under revision). Transfer of L1 knowledge helps in doubly adverse conditions: Perception of casual speech in L2. *Language and Cognitive Processes.*

Abstract

Casual speech processes, such as /t/-reduction, make word recognition harder. Additionally, word recognition is also harder in a second language. Combining these two adverse conditions, we investigated whether second language learners have recourse to knowledge from their native language when dealing with casual speech processes in their second language. In three experiments, production and perception of /t/-reduction was investigated. The first experiment compared incidence of /t/-reduction in Dutch and German. Reduction occurred in both languages and patterned similarly in proper nouns. Where /t/ was a verbal inflection, however, the languages differed: Dutch speakers reduced inflectional /t/ more often than German speakers did. Two perception experiments compared the performance of German learners of Dutch with that of native speakers. Mirroring the production patterns, German learners' performance strongly resembled that of native Dutch listeners when the reduced /t/ was part of a word stem, but deviated from the native where /t/ was a morphological marker. These results suggest that the casual speech processes show the same tight perceptual coupling on first and second language as the rest of speech perception.

Introduction

Speech perception studies are often performed under ideal circumstances. A listener shielded from environmental noises by a sound-proof booth listens to his or her native language. The played sound files have been recorded under similarly optimal circumstances by a speaker, who is carefully reading out loud. Outside the laboratory, the situation is often less ideal. Environmental noises are (too) common, speakers are less careful than during reading, and the language we listen to might not be our native language. All of these influences make speech perception harder.

To start, speech perception in a second language is notoriously difficult. Non-native listeners have the disadvantage that speech in their second language (L2) often contains sounds that they are unfamiliar with given their native language (L1). A large body of research (see, e.g., Strange, 1995 and Bohn and Munro, 2007 for an overview) has shown that so-called mismatches between the phoneme repertoires of the L1 and L2 cause perceptual difficulties, which can make distinguishing speech sounds extremely difficult for L2 listeners.

The segmental contrast between /r/ and /l/ in English, for example, is difficult for Japanese listeners as these speech sounds match a single native category equally well (e.g., Bradlow, Pisoni, Akahane-Yamada, & Tohkura, 1997; Ingram & Park, 1998; Underbakke, Polka, Gottfried, & Strange, 1988). In an eye-tracking study, Cutler, Weber, and Otake (2006) observed that Japanese listeners who were instructed to click on a picture of a *rocket*, experienced interference when a picture of a *locker* was present; they tended to look at the *locker* before eventually focusing on the *rocket*. Ultimately, the Japanese listeners comprehend what the speaker meant by the use of disambiguating information that follows (i.e., ...*ocker* vs. …*ocket*), but the initial consonant is not enough for the listeners to quickly deactivate the competitor word *locker*. This perceptual confusion is also evident when other methods are used. With auditory repetition priming, Cutler and Otake (2004) found that English minimal pairs differing in the /r/-/l/ contrast (e.g., *write-light*) activated one another for Japanese listeners. For Dutch listeners, on the other hand, the /r/-/l/ contrast is not problematic, but they do have difficulties with the English vowel contrast /æ/-/ɛ/, and minimal pairs such as *flash-flesh* activated each other for Dutch listeners. Other studies also provide evidence that minimal pairs activate each other more for non-native than for native listeners (Broersma & Cutler, 2008; Pallier, Colomé, & Sebastián-Gallés, 2001).

These perceptual problems of L2 listeners are not limited to possibly difficult minimal pairs, such as *cattle-kettle*. This stems from the fact that all spoken word recognition, native and non-native, involves activation of multiple word candidates which compete for recognition. For example, when listeners hear the word *captain*, not only the word *captain* is activated, but also words such as *cap* and *capitol* are temporarily activated (Davis, Marslen-Wilson, & Gaskell, 2002). During this process, hard to disambiguate contrasts create additional competitors in non-native listeners and as a result, more word candidates are activated in L2 listening than in L1 listening. Broersma and Cutler (2008) showed that L2 listening can involve the "phantom activation" of words which are not actually in the input. In a lexical decision study, Dutch listeners with L2 English accepted spoken nonwords such as *groof* or *flide* as real English words, whereas L1 listeners did not. Furthermore, a priming study showed that hearing these same nonwords facilitated recognition of the printed words *groove* and *flight* for the non-native listeners, but not for the native listeners. This finding suggests that, for the non-native listeners only, the real words had been activated by nonword input, causing additional complication in non-native speech recognition. Further research showed that in cross-modal priming, so-called near-words extracted from word or phrase contexts (*daf* from *DAFfodil*, *lemp* from *eviL EMPire*) caused activation of corresponding real words (*deaf*, *lamp*) for Dutch non-native listeners of English, but not for native listeners (Broersma & Cutler, 2011).

As this shows, listening to a non-native language is, compared to L1 listening, an adverse condition by itself. Non-native listeners are confronted with difficult segmental contrasts leading to spurious lexical activation. To make matters worse, L2 perception seems to be more strongly affected by additional challenges. For example, several studies have shown that, in the presence of background noise, speech perception by non-native listeners is more strongly affected than that of native listeners (see Lecumberri, Cooke, & Cutler, 2010, for an overview).

However, as indicated above, noise-masking is not the only possible challenge for speech perception. As we will review below, spontaneous speech is also much harder to comprehend than carefully-read speech. Little research, however, has focused on how non-native listeners deal with the adverse conditions created by casual speech. Is casual speech, like background noise, especially difficult to overcome for non-native listeners?

This is an important question. After all, casual speech is what, as language users, we mostly hear and produce. This speech is extremely variable as is evident from casual speech corpora. Examples are the American English Buckeye Corpus (Dilley & Pitt, 2007; Pitt, et al., 2007), the Dutch Corpus Gesproken Nederlands (Oostdijk, 2000; Pluymaekers, Ernestus, & Baayen, 2005), and the Nijmegen Corpus of Casual French (Torreira, Adda-Decker, & Ernestus, 2010). Processes such as assimilation, epenthesis and extreme reduction occur frequently in these corpora and result in pronunciation variations and create ambiguities. Consonant reduction such as the deletion of /t/ in *lost* confronts listeners with unintended words, in this case *loss*. Recently, connected speech processes, and how listeners deal with the resulting non-canonical forms, have received widespread attention in speech perception research. Alterations can often (temporarily) mislead listeners, and results in word recognition being harder than it would have been for the canonically pronounced versions (Ernestus, Baayen, & Schreuder, 2002; Sumner & Samuel, 2005, 2009; Tucker & Warner, 2007; Warner, Fountain, & Tucker, 2009).

Still, native listeners are somehow able to cope with the extreme variabilities that occur in spontaneous speech. They prove to be extremely good at exploiting the fine phonetic detail of utterances and identifying intended words even when casual speech processes have altered them from their canonical form. In the case of place of articulation assimilation, for example, listeners are able to distinguish the /p/ of English *ripe* in *ripe berries* from the assimilated final /t/ of *right berries* based on the formant offset frequencies in the vowel preceding the /p/ (Gow, 2002). And French listeners are able to distinguish the word-initial /p/ in French *trop partisan* from the liaison realization of a word-final /p/ in *trop artisan* based on consonant duration (Spinelli, McQueen, & Cutler, 2003). Another source of information that helps listeners to compensate for reductions is the phonological context. Most reductions are conditioned by phonological context and listeners make use of this information in compensation for assimilation (Gaskell & Marslen-Wilson, 1996; Gow, 2003; Mitterer & Blomert, 2003) and compensation for /t/-reduction (Mitterer & Ernestus, 2006; Mitterer & McQueen, 2009).

However, all of this research has been carried out with native listeners. Given that even these experienced listeners are often burdened by reductions, what is going to happen when non-native listeners hear the same sort of input? Hear it they will, because all languages manifest casual speech processes. Moreover, L2 listeners cannot permanently confine themselves to speech situations in which the input is as close to canonical perfection as it is in the classroom or on language tapes. Therefore, in the present study, we investigate what implications casual speech processes can have for L2 listening.

Obviously, not all non-native casual speech processes may be equally difficult for a non-native listener, just as not all non-native phoneme contrasts are equally difficult for L2 listeners. With regard to phoneme contrasts, for instance, the Perceptual Assimilation Model (PAM, see Best, 1994, 1995) assumes different types of contrast relations, which are more or less difficult for the learner. The most difficult is the same-category contrast, in which two non-native contrasts are assimilated to the same native category. A contrast is easier to acquire if both non-native sounds are assimilated to two different native phonemes. An example of these two types of relation is the English /r/-/l/ contrast for Dutch and Japanese learners. The English /r/-/l/ contrast is not problematic for Dutch listeners as Dutch has two phoneme categories very similar to English /r/ and /l/. As Japanese has only one phoneme in the perceptual space of the English /r/ and /l/, PAM predicts that it is among the most difficult contrasts for Japanese listeners, and as mentioned before, it indeed is.

What holds for segmental contrasts in a non-native language, may also hold for casual speech processes. If this is the case, casual speech processes that are unique to the L2 should be particularly hard for non-native listeners. In line with this assumption, Tuinman, Mitterer and Cutler (under revision) show that Dutch listeners have trouble with the process of /r/-intrusion in their L2 English, in which an /r/ is inserted between two vowels at a word boundary. A famous example for this process is provided by the phrase "I saw a film today" in the Beatles' song *A day in the life*, which is pronounced as "I soar a film today". As this /r/-insertion does not occur in Dutch, it should hence be difficult for Dutch learners of English, and, indeed, it is. However, it is unclear if the reverse is also true. Does a L2 casual speech process cause less of a problem if it does occur in the L1? Hence, we investigated whether listeners will find it easy to deal with an L2 process if they already have experience with the same process in their L1.

As a case study we test whether German advanced learners of Dutch are able to compensate for /t/-reduction in Dutch. There are two reasons to choose the process of /t/-reduction. The first is that the process of /t/-reduction is found in many languages (Guy, 1980) and patterns very similarly in the Germanic languages English, German and Dutch. Second, /t/-reduction in Dutch has been intensively studied (e.g., Janse, Nooteboom, & Quené, 2007; Mitterer & Ernestus, 2006; Mitterer & McQueen, 2009). These studies show that Dutch listeners are highly sensitive to the patterning of /t/-reduction in their native language. Based on these results, we focus on three aspects that have been shown to influence compensation for /t/-reduction: phonetic detail, preceding phonological context and higher-level knowledge. First of all, Dutch allows various gradations of /t/-reduction, and listeners are highly sensitive to the subtle differences in phonetic detail that result from this. Secondly, preceding context is important as /t/-reduction is more likely after /s/ than after /n/ in Dutch and listeners take this into account in perception. Finally, listeners also make use of higher-level knowledge, and are more likely to restore a reduced /t/ if this "leads to a word". That is, they are more likely to report a /t/ at the end of "fros..." (*frost* being a word) than at the end of "blis..." (*blist* being an English nonword).

However, before we can use /t/-reduction as a case study we need to know how similar it patterns in Dutch and German. We are not aware of a quantitative comparison between Dutch and German /t/-reduction patterns. Therefore, we conducted a production study with very similar set-ups for German and Dutch participants to investigate how similar or dissimilar the reduction patterns really are. Critically, we will focus on whether /t/-reduction in German is more likely after /s/ than after /n/, which is the pattern attested for Dutch (Mitterer & Ernestus, 2006). Additionally, we investigated the pattern of /t/-reduction in nouns—in which the /t/ was part of the word's stem—and /t/-reduction in verbs, in which the /t/ is the morphological marker for the third-person singular. There is evidence to suggest that morphological variables influence reduction processes (Guy & Boyd, 1990). The morphological influence may be different for German than for Dutch, because German is morphologically richer with many different verb conjugations, a three- rather than a two-gender system, and more grammatical cases for nouns.

Experiment 1

Method

Participants

Ten Dutch and 10 German speakers took part in this experiment. The Dutch participants were students at the University of Nijmegen, the Netherlands and members of the Max Planck Institute's subject pool. Some of the German participants were also taken from this population; others were employees at the Max Planck Institute with basic knowledge of Dutch. None reported any hearing loss. All were volunteers and received a small fee for participation.

Materials, design and procedure

The experiment consisted of two production tasks and was constructed in such a way that the Dutch and German participants performed exactly the same tasks, with as similar target items as possible.

The first part of the experiment examined /t/-reduction in verbs with a sentence generation task. The stem of the critical verb in all sentences ended on either /n/ or /s/ and the verb was always presented in its full form (e.g., *rennen* 'run'), and the participants had to produce a sentence with the third-person singular present (e.g., *rent* 'runs'; note that the third-person singular inflection in German is the same as in Dutch, i.e., /t/).

The second part of the experiment tested /t/-reduction after /n/ and after /s/ in proper names, using a blending task. Participants saw two non-existent place names (e.g., *Toestwoud* and *Liekbeek* for Dutch), and made a new place name with the first part of the first place name and the second part of the second place name (in this case *Toestbeek*) and produced this place name in a sentence frame. Again the acoustic context was manipulated in that the first part of the place name ended in either /nt/ or /st/.

Mitterer and Ernestus (2006) found that both preceding and following context influence the likelihood of /t/-reduction. In perception there were only effects of preceding context in their study. As we were interested in the relation between perception and production, we focused on the preceding context only.

The experiment was run on a standard PC running with the NESU package. Participants were tested one at a time in a sound-proof booth. They sat at a comfortable reading distance from the computer screen and had a microphone with a recording device and a two-button response box in front of them. Dutch participants received instructions in Dutch and German participants in German; these informed participants that for the first part of the experiment they would see words, in random order, on the computer screen. The Dutch participants, for example, would see *bij/Maarten/de bushalte/wonen* and the Germans would see *bei/Martin/der Bushaltestelle/wohnen*. They were instructed to produce a sentence with the third-person singular present using the words. In this example the correct response for the Dutch was *Maarten woont bij de bushalte* and the correct German response *Martin wohnt bei der Bushaltestelle* ('Maarten/Martin lives near the bus stop'). Both Dutch and German participants received 40 different stimulus sentences. The presentation of the sentences was random and different for every participant.

This part of the experiment, the sentence generation task, started with four practice trials. Each trial (experimental and practice trials) began with a blank screen. Then, the words were presented in the middle of the screen. After 1100 ms the message 'Press the right button to continue' was displayed on the screen in either Dutch or German, so that participants could continue with the next sentence as soon as they were ready.

For the second part of the experiment, on each trial, participants would see the name of a store (e.g., *groenteboer/Gemüsehändler*, 'greengrocer'), a product (e.g., *appels/Äpfel*, 'apples') and two non-existing placenames (e.g., *Klestfoort/Klestfurt* and *Roenveen/Ruhnfehn*). Again, they were instructed to produce a sentence with the words on the screen; in particular, they had to make a new place name with the first part of the first place name and the second part of the second place name (in this case *Klestveen/Klestfehn*). The correct response in this example was *Bij de groenteboer in Klestveen koop ik appels* for the Dutch participants and the correct German response was *Beim Gemüsehändler in Klestfehn kaufe ich Äpfel* ('At the greengrocer in *Klestveen/Klestfehn* I buy apples'). Again, both Dutch and German participants received 40 different sentence frames and the presentation of the sentences was random and different for every participant.

This part of the experiment, the blending task, also started with four practice trials. Each trial (experimental and practice trials) began with a blank screen. Then, the words were presented on the screen: the name of the shop in the upper left corner, the product in the upper right corner and the two place names just below the center. After 1500 ms the message 'Press the right button to continue' was displayed on the screen in either Dutch or German, so that participants could continue with the next sentence.

Results

The 1600 sentences (10 speakers of each language x 80 tokens) were analyzed for /t/-reduction in the critical words. On the basis of visual inspection of the sound files using PRAAT (Boersma & Weenink, 2005), the productions of /t/ were classified and it was judged whether the /t/ was present or not.

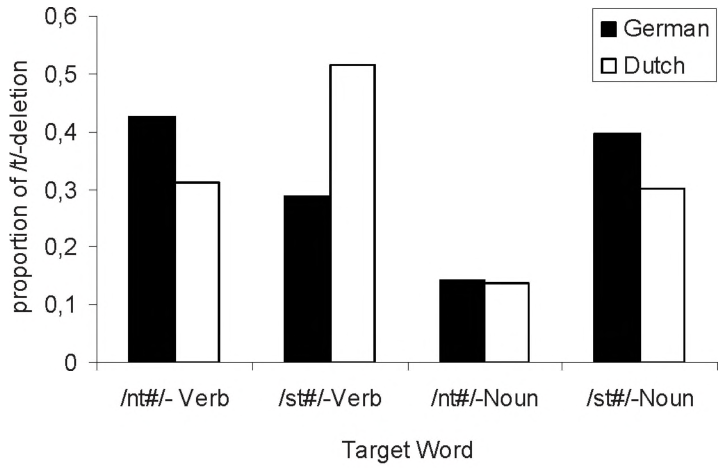

Figure 1: *Experiment 1: Proportion of /t/-deletion by German and Dutch native speakers.*

Figure 1 shows the results of the transcription. Dutch participants tend to delete /t/ more often if the /t/ was preceded by an /s/ rather than an /n/. This pattern was consistently observed for both nouns and verbs. German participants, however, show a different pattern for verbs, with /t/-deletion in fact being more likely in verbs with an /nt/-coda than in verbs with a /st/-coda.

Table 1: *Experiment 1: Regression weights for the final models*

Type of word	Effect	Regression Weight
Proper name	(Intercept)	2.9554***
	NativeLanguage = Dutch	0.5352
	Preceding Context	-1.6906***
Verb	(Intercept)	0.9280
	NativeLanguage = Dutch	0.8143
	Preceding Context	0.5038
	NativeLanguage : Preceding Context	-2.3404*

Note: *** = $p < 0.001$; ** = $p < 0.01$; * = $p < 0.05$

For the statistical analysis, the results for the proper names and verbs were analyzed separately with a linear mixed-effects model with a binomial linking function to account for the categorical nature of the dependent variable (cf. Dixon, 2008). Participant and item were entered as random factors and Native Language and Preceding Context as fixed factors. The fixed factors were contrast coded, with the preceding context /n/ and the Native Language German coded as -0.5. A positive regression weight hence would indicate that more /t/s were produced in the /s/-context, and by the Dutch participants. In contrast coding for binary variables, one level is coded as -0.5 and the other as 0.5, so that regression weights for simple effects in the regression models show the overall effect of a given variable for the complete data set (cf. Barr, 2008).

Analysis started with a full model and in stepwise fashion insignificant interactions were pruned. Table 1 shows the regression weights for the final models for the proper names and verbs. The analyses show that for proper names there were no significant differences between Dutch and German speakers. The Preceding Context did have a significant effect in that /t/ was more often reduced after /s/ than after /n/ (the regression weight is negative).

For verbs, there were no main effects of Native Language or Preceding Context, but a significant interaction. To gain insight in what drives this interaction we analyzed the data separately for German and Dutch speakers. This showed that only the Dutch speakers have a significant effect of Preceding Context and reduced /t/ more often after /s/ than after /n/ (b = -1.89, $p < 0.05$), while no such effect was observed for German speakers ($p > 0.1$).

Discussion

The results show that Dutch and German are very similar with regard to /t/-reduction in proper nouns. In both languages, /t/-reduction was more likely after /s/ than after /n/. However, in line with our suspicion, reduction of a morphological /t/ was different in German than in Dutch. In Dutch, verbs patterned just as nouns, with more deletion after /s/ than after /n/. In German, however, /t/-reduction was independent of the preceding context. A possible explanation for this pattern is that the morphological /t/ has no special status in Dutch, a morphologically less rich language than German. Hence, a morphological /t/ does not behave differently than a /t/ in a noun. In the morphologically richer language German, however, the morphological status seems to block the influence of the preceding phonological context.

What do these results mean for the perception of /t/-reduction in Dutch by German learners? If the similarity of the L1 and the L2 pattern governs the perception of casual speech processes in L2, German listeners may be able to perform very similarly to native Dutch listeners for cases in which the /t/ is part of the stem of a content word. However, /t/-reduction in verbs may be more of a challenge for German learners of Dutch.

To investigate this, we presented German and Dutch listeners with varying amounts of acoustic evidence for word-final /t/ in verbs and nouns and adjectives. Five realizations of /t/, from full production to complete deletion, are presented in two acoustic contexts, after /n/ (where /t/-reduction is unlikely) and after /s/ (where /t/-reduction occurs frequently). These 5 levels of the /t/-∅ continuum are based on findings from a corpus study on word-final /t/ in Dutch (Mitterer & Ernestus, 2006). In each sentence, listeners judged whether the target word ended in /t/ or not.

If non-native listeners have difficulty in exploiting the phonological cues for /t/, they might base their judgments of whether the target word ended in a /t/ or not on higher-level information such as lexical status and syntax in verbs. Therefore, we also added as a factor whether the /t/ is "prescribed" by syntax or lexical status. In

Experiment 2, the /t/ had no morphological role, but was part of the stem of a noun or an adjective, and lexical information prescribed the presence of a /t/ (*charmant* 'charming', *charman* being a Dutch nonword) or not (*kanon*, 'gun', *kanont* being a Dutch nonword). In Experiment 3, target words were verbs (e.g., *ren* 'run', *kus* 'kiss') which makes it possible for listeners to use grammar to predict whether or not the ending should be /t/. The Dutch present tense third-person singular inflection is /t/ (e.g., *zij rent*, 'she runs') while the first-person inflection is null (e.g., *ik ren*, 'I run'). Hence, listeners should be biased to expect a /t/ at the end of a verb that is preceded by the third-person singular pronoun *zij* and biased to expect no /t/ at the end of a verb that is preceded by the first-person singular pronoun *ik*.

Experiment 2

Method

Participants

Sixteen native speakers of German participated in the experiment. The participants were students at the University of Nijmegen, the Netherlands and members of the Max Planck Institute's subject pool. None reported any hearing loss. All were volunteers and received a small fee for participation. The German participants had a high level of proficiency in Dutch as L2 as they followed an intensive Dutch language course and passed an exam as a requirement to enter a Dutch university. To compare the data to native performance, control data of Dutch native speakers were taken from Mitterer and Ernestus (2006).

Materials

The materials for the experiment were those of Mitterer and Ernestus (2006), Experiment 2. The stimulus sentences were created by synthesis via a Klatt synthesizer (Klatt, 1980) and had the following format: *Jan sprak "kanon" moeilijk uit* ('John pronounced "canon" with difficulty'). Two different male names served as subjects and five adverbs were used. The target stimuli were the nonwords *orkes* and *charman*, as well as the Dutch words *moeras* and *kanon*, followed by one of five coda signals (see also Figure 2): full /t/, strong frication /t/, weak frication /t/, closure only, and a long consonant. These five signals constitute a /t/-∅ continuum in which the first step contains the strongest phonetic information for the presence of /t/ and the fifth step (the long consonant /n/ or /s/) no information for the presence of /t/. Because consonants tend to be longer in simple codas than in complex codas, a long consonant can be considered as evidence for the absence of a following /t/ (Lehiste, 1970).

The first signal was similar to a full /t/ with a 25 ms closure and a 45 ms transient-frication sequence. To prevent an unnatural flat line in the signal, the closure was synthesized with 20 dB amplitude of frication (AF) and a 30 dB amplitude bypass (AB) of the parallel branch of the synthesizer. At the release of the initial burst, AB was set to zero, and AF increased to 40 dB, and decreased again to 15 dB at the end of the 65 ms signal. For this and all the other target signals, amplitude changes were loglinear in dB to achieve a linear amplitude envelope. The transient-frication signal was dominated by the fifth and sixth formats (55dB) starting at 5700Hz (bandwidth of 500 Hz) and 7500 Hz (700 Hz), respectively. These formants fell to 5200 and 6850 Hz at 65 ms. The second, third and fourth formant stayed constant throughout the transient-frication sequence at 1434 Hz (200), 2212 Hz (500) and 3840 Hz (600), respectively, and their amplitude increased from 20 to 25 dB in order to mimic the increasing low-amplitude frication in the model [t].

The second coda signal was a 65 ms frication noise, as often found in /st/ codas. The amplitude of frication started and ended at 15 dB with a 40dB maximum at 30 ms. The settings for the second, third and fourth formant were the same as the settings in the full [t]. The fifth and sixth formants started at 5130 Hz (bandwidth: 500 Hz) and 6750 Hz (700 Hz) and fell to 4620 Hz and 6080 Hz at the end of the signal (*A*s = 55 dB). The weak-frication signal was derived from this signal by reducing the overall amplitude of fricative noise by a factor of 5 (= 14 dB). The closure-only signal was synthesized with the same settings as for the closure of the full /t/.

The final coda signal was a coda that was not followed by any signal implying a /t/. Instead, it was created by elongating the consonant /n/ by 45 ms and /s/ by 65. These durations were based on the measurements of the coda duration in the recording made of the Dutch native speaker reading the test materials. These measurements also showed that consonants in consonant clusters are shorter than in simple codas.

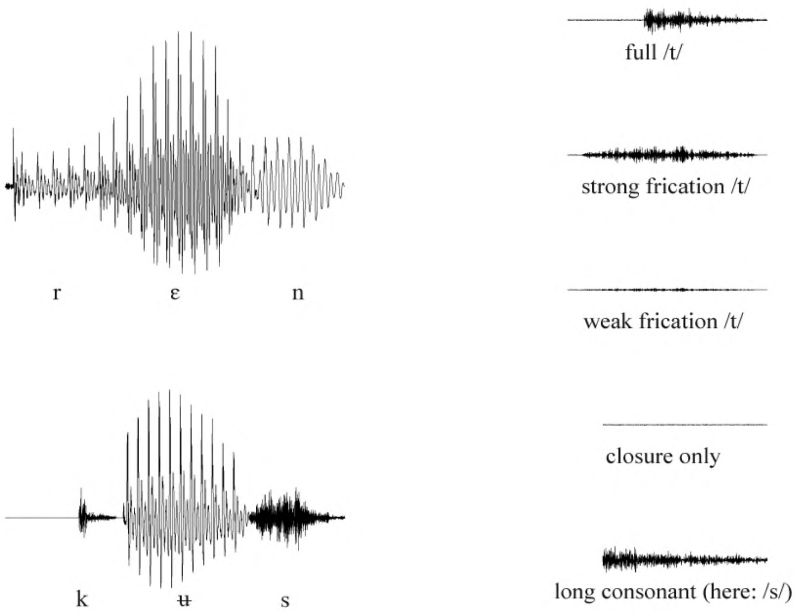

Figure 2: *Right: spectrograms of the different target signals presented in Experiment 2 and 3. Left: spectrogram of the target words* ren *and* kus *as used in Experiment 3.*

It is important to note that the targets *orkes* and *charman* are Dutch nonwords, but *orkest*, 'orchestra' and *charmant*, 'charming' are existing Dutch words, whereas the targets *moeras* 'swamp' and *kanon* 'canon' are Dutch words, but become nonwords when a /t/ is added (i.e., *moerast* and *kanont* are not Dutch words).

Design and procedure

There were three independent variables: (1) the coda signal (from full production to complete deletion); (2) the context preceding the coda signal (/n/ vs. /s/); (3) the lexicality manipulation: target words were existing or non-existing words if they ended in /t/. The between-subjects variable was native language and the dependent variable was the percentage of /t/-responses in each cell of the design.

The experiment was run on a standard PC running with the NESU package. Participants were tested one at a time in a sound-attenuated booth. They wore Sennheiser headphones, sat at a comfortable reading distance from the computer screen and had a two-button response box in front of them. Instructions were given in Dutch (also to the non-native participants); these informed participants that on each trial they would hear a Dutch sentence and see two words on the computer screen, one that ended in a /t/ and one that did not. They were asked to press the right button if the sentence they heard contained a target word that ended in a /t/, and to press the left button if the target word did not end in a /t/. The 300 different stimulus sentences (6 target words x 5 coda signals x 5 adverbs x 2 subjects) were presented only once to each participant. The presentation of the sentences was random and different for every participant.

The experiment started with four practice trials. Each trial (experimental and practice trials) began with 150 ms of blank screen. Then, the response alternative without a /t/ in the coda (e.g., "orkes") was presented in the upper left corner of the screen whereas the other (e.g., "orkest") was presented in the upper right corner. After another 450 ms the sentence was played. From the onset of the target word, participants had 2.5 seconds to press one of the buttons. After responding, the chosen alternative was moved further in respectively the upper right or upper left corner while the other alternative was removed from the screen, so that the participants could see that their answer had been registered by the computer. If a participant did not react within 2.5 seconds, a stopwatch was shown on the screen to remind participants to respond more rapidly. The three feedback signals—no /t/-response, /t/-response and no response—stayed on the computer screen for 1 second before the next trial started. Participants were free to take a break after each 50^{th} trial and could continue as soon as they were ready.

Results

The results of Experiment 2 are shown in Figure 3. A comparison of the panels for native Dutch listeners and non-native German listeners shows that the overall pattern is quite similar. Participants give, independent of native language, more /t/-responses if there is more acoustic-phonetic evidence for a /t/ (an effect of the /t/-∅ continuum), if assuming a /t/ results in an existing word (an effect of Lexicality), and if the preceding context is /s/ rather than /n/.

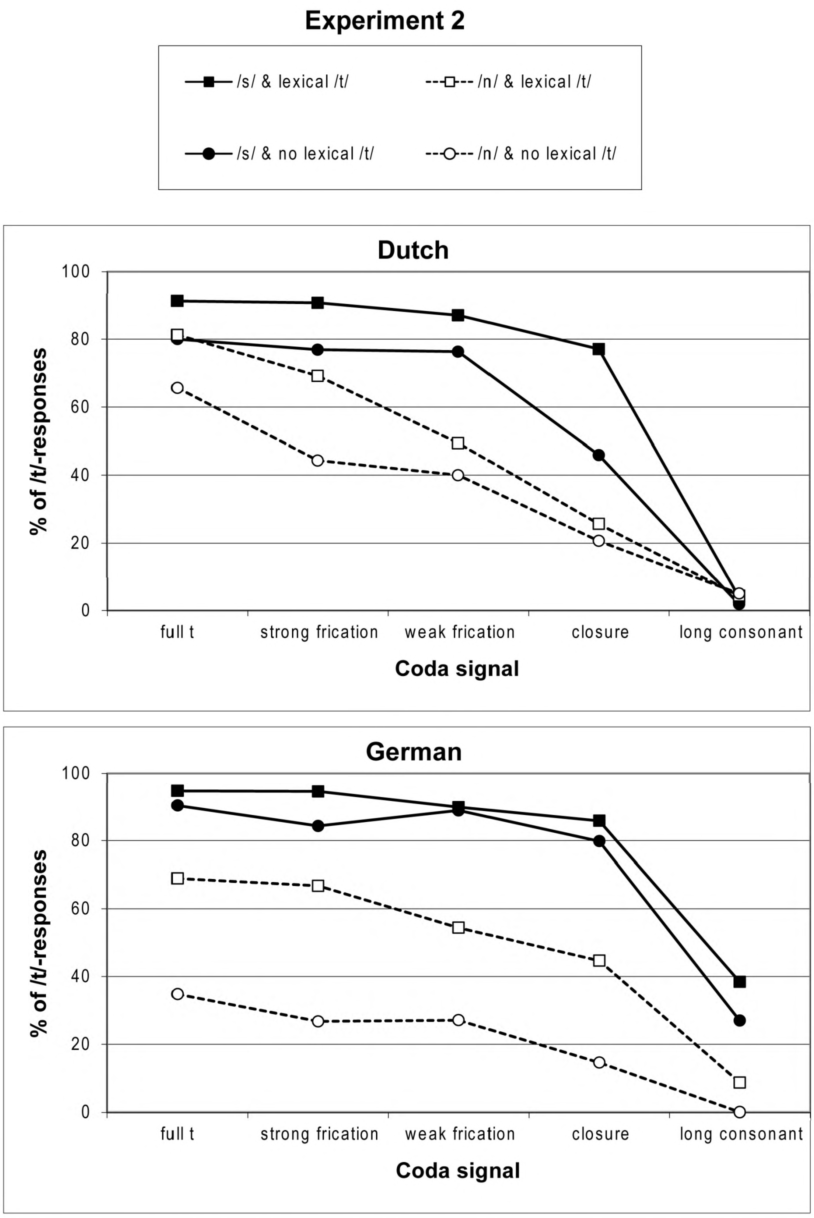

Figure 3: *Percentage of /t/-responses in Experiment 2 for each Coda Signal. The lines indicate the two different Preceding Contexts /n/ and /s/ and the two different Lexicality conditions.*

Table 2: *Experiment 2: Regression weights for the final models on each level of the continuum.*

Effect	Coda signal				
	Full /t/	Strong frication	Weak frication	Closure	Long consonant
(Intercept)	1.91***	1.29***	0.04	0.94	-3.42***
PrecedingContext	1.87***	2.19***	2.35***	2.52***	1.14***
Lexicality	1.08***	1.45***	0.89***	0.67***	0.84***
NativeLanguage	-0.40	-0.27	0.63*	0.17***	1.75**
PrecedingContext: Lexicality	-	-	0.11	-0.31	-
PrecedingContext: NativeLanguage	1.84***	0.74*	0.73**	0.67*	3.00***
Lexicality: NativeLanguage	-	-	0.14	0.19	-
PrecedingContext: Lexicality: NativeLanguage	-	-	-1.81**	-1.40*	-

Note: *** = $p < 0.001$; ** = $p < 0.01$; * = $p < 0.05$

Statistical analysis of Experiment 2 started with a full linear mixed-effects model and in stepwise fashion insignificant interactions were pruned. Participant was entered as a random factor, and Native Language, Coda Signal (from full production to complete deletion), Preceding Context (/n/ vs. /s/) and Lexicality (/t/ or no /t/ predicted) as fixed factors. The binary variables Native Language, Preceding Context were contrast coded. The value of -0.5 was assigned to the Preceding Context /n/, the no-lexical /t/-condition and the native Dutch listener group. Overall, there were main effects of Preceding Context and Lexicality; more /t/-responses were given after /s/ than after /n/, and more /t/-responses were given if an existing word resulted. Overall, the effect of Preceding Context was larger for the German listeners. However, the reported effects were moderated by various interactions. Therefore, we also examined the effects of Native Language, Preceding Context and Lexicality on all five levels of Coda Signal. Table 2 shows the regression weights for the final models on each step of the continuum. Note that simple effects remained in the model even if insignificant if they were involved in higher-order interactions. Regression weights missing in Table 2 were not included in the final model.

The strong Coda Signals (see the first two columns of Table 2) showed a consistent pattern with main effects for Preceding Context and Lexicality and a significant interaction between Native Language and Preceding Context—German listeners had a larger context effect than the Dutch listeners. The same pattern is attested for the Long Consonant Condition (the last column of Table 2). However, a different pattern is observed in the middle of the /t/-∅ continuum. Here, we also see a three-way interaction between Preceding Context, Lexicality and Native Language. To understand these three-way interactions, additional models were run on each step for each level of Preceding Context separately, with the predictors Lexicality and Native Language. An inspection of Figure 3 suggests a source for this interaction: The lexical effect in the /n/-context is visibly larger for German participants than for Dutch participants. In the /s/-context, the opposite pattern is observed; the lexical effect is larger for Dutch than for German participants. Statistically, three of these four comparisons revealed Lexicality by Native Language interactions ($p < 0.05$). Only for the weak-frication signal in the /s/-context the overall lexical effect (b(Lexicality) = 0.54, $p < 0.05$) was not moderated by Native Language (b(Lexicality by Native Language) = -0.62, $p > 0.2$).

To sum up the results of Experiment 2, we observed that both listener groups pattern to some extent similarly; it is not the case that some variables influenced only one group of listeners. However, there is a tendency for German listeners to give more /t/-responses when there is little evidence for a /t/ (i.e., for the three coda signals at the "∅" end of the /t/-∅ continuum). Moreover, there are some interactions that indicate that the Native Language influenced the degree of the effects. The effect of Preceding Context is consistently larger for German than for Dutch participants. However, there is little evidence to suggest that German participants rely overall more strongly on lexical cues to the presence of /t/. Even though there are interactions involving Lexicality and Native Language, these are not consistently in the same direction with a larger effect of Lexicality for the German listeners. As laid out in detail above, in some conditions the native Dutch participants show a stronger lexical effect, in others the non-native German participants show a stronger lexical effect.

Discussion

The results of Experiment 2 indicate that German listeners are able to make use of the phonological cues to the presence of /t/. Even though they tend to perform less categorically and, as a consequence, give more /t/-responses when there is little evidence for a /t/, they do not show a pattern of less reliance on phonological cues and more reliance on lexical cues. In fact, they even are more strongly influenced by phonological context than the Dutch listeners.

This indicates that the German listeners are generally well able to deal with /t/-reduction in their second language Dutch. This ability may be grounded in the experience German listeners have with /t/-deletion in their native language. As we argued in the introduction, previous research on segmental contrasts in a second language has indicated that a segmental contrast in a second language is easiest to learn if the contrast is similar to a native contrast. That is, Dutch listeners of English have less problems with the /r/-/l/ contrast than Japanese learners, because the Dutch language has a /r/-/l/ contrast as well, while Japanese does not. Similarly, German learners may be able to rely on their L1 knowledge when dealing with /t/-reduction in Dutch, because /t/-deletion in proper names patterns similar in both Dutch and German.

However, there is another account for the current finding. The relatively native-like performance of German learners may simply be due to them having learned the Dutch reduction pattern rather than transfer of knowledge from their L1. Fortunately, these two accounts can be distinguished by looking at the pattern for verbs. If German learners are indeed able to learn the pattern of /t/-reduction in a different language, they should also show near-native like performance in verbs. If, however, differences between L1 and L2 influence the mastery of a casual speech process in a second language, German learners should be less proficient in dealing with /t/-reduction in verbs, where their L1 knowledge does not match the pattern of the L2 target.

Experiment 3

Method

Participants

Sixteen native speakers of German, none of whom had participated in the previous experiments, took part in Experiment 3. Participants were from the same participant pool as those of Experiment 2. To compare these data with the native pattern, control data of 21 native speakers of Dutch were taken from Tuinman, Mitterer and Cutler (submitted).

Materials

For this experiment, three Dutch verbs with a stem ending on /n/ and three verbs with a stem ending on /s/ were chosen with a matched lemma log frequency using CELEX (Baayen, Piepenbrock, & Gulikers, 1995). These verbs were placed in sentences and a male native speaker of Dutch (the same speaker as in Mitterer and Ernestus, 2006) recorded the sentences several times. Table 3 shows the combinations of the target verbs and preceding and following words. According to the reading list, the verbs in the sentences were pronounced with or without a final /t/. These utterances were used as templates for synthesis via a Klatt synthesizer (Klatt, 1980). Stimuli were created in a similar fashion to Mitterer and Ernestus, which were also the basis for the stimuli in Experiment 2. Formants and bandwidths were measured at the beginning, in the middle and at the end of each segment and at major formant transitions point within segments. When the formant could not be estimated reliably by Linear Predictive Coding (using Praat, Boersma & Weenink, 2005), values recommended by Klatt (1980) were chosen. Additionally, nasalization for a nasal was carried over into the vowel before it was reduced completely, and anticipated for postvocalic nasals to match the natural utterances. Parameters for synthesis were generated by interpolating linearly between measurements points. Occasionally, values were altered in order to prevent clicks and other transients in the synthesized signal, which can occur if control parameters change too quickly. Amplitude values were iterated to imitate the amplitude envelope of the natural utterances.

Table 3: *Sentence frame and target words for the stimuli in Experiment 3*

Connection word	Subject	Target word	Adverb
		blaas /bla:s/ (1.62)	nauwelijks /ˈnɑuʋələks/
	zij /zɛɪ/	kreun /krø:n/ (1.46)	langzaam /ˈlɑŋza:m/
		bloos /blo:s/ (1.21) ..t	moeizaam /ˈmujza:m/
Maar /ma:r/		zoen /zun/ (1.26) ...Ø	soms /ˈsɔms/
	Ik /ɪk/	ren /rɛn/ (1.95)	vaak /ˈva:k/
		kus /kʉs/ (1.75)	

Note: English translations are *Maar* 'but', *ik* 'I', *zij* 'she', *blaas* 'blow', *kreun* 'moan', *bloos* 'blush', *zoen* 'kiss', *ren* 'run', *kus* 'kiss', *nauwelijks* 'hardly', *langzaam* 'slowly' *moeizaam* 'with difficulty', *soms* 'sometimes', *vaak* 'often'. The numbers between brackets are the lemma log frequencies for the words taken from CELEX (Baayen, et al., 1995).

The verb stems were followed by one of five different synthesized signals for the coda, just as in Experiment 2 (see also Figure 2). The long consonant target signals were adapted to the following adverb to make the transition as natural as possible between the target verb and the subsequent adverb. Thus, for the long consonant /n/ the amplitude remained high if another nasal followed (e.g., if "nauwelijks" or "moeizaam" was the following adverb), but it fell before other phonemes. For the long consonant /s/, the amplitude did not fall before another /s/ (e.g., if "soms" was the following adverb), but decreased before other phonemes. A 25 ms closure was inserted after "Maar ik" and a 50 ms closure before verbs starting with a /k/ (e.g., "kus") to make the synthesized materials more like the natural utterances.

Thus, there were five coda signals for the target verbs: full /t/, strong frication /t/, weak frication /t/, closure only, and a long consonant. All these different signals were put together with the 6 target verbs resulting in 30 stimuli for which the participants decided whether they heard a coda with or without a /t/. The 30 stimuli were placed in sentences starting with either "Maar ik …" or "Maar zij…" and five adverbs following the target words. This resulted in 300 different sentences (see also Table 3).

Design and procedure

There were three independent variables for Experiment 3: (1) the coda signal (from full production to complete deletion); (2) the context preceding the coda signal (/n/ vs. /s/); (3) the syntactic manipulation: target words would be grammatically correct or incorrect if ending in /t/. The between-subjects variable was native language and the dependent variable was the percentage of /t/-responses in each cell of the design.

The procedure of the experiment was as described for Experiment 2 with the exception that the stimuli and target words differed.

Results

The results of Experiment 3 are presented in Figure 4. As can be seen in Figure 4, there are similar types of main effects in both panels. There are clear effects in both panels based on the /t/-∅ continuum (more /t/-responses with more /t/-like signal), syntax (more /t/-responses if the syntax prescribes a /t/), and preceding phonological context (more /t/-responses after /s/ than after /n/). However, there are also clear differences. The German participants hardly ever give a no-/t/-response if either the syntax or the preceding phonological context creates a bias towards /t/. The identification functions stay near ceiling for the first four levels of the /t/-∅ continuum in all conditions except the one in which neither the syntax nor the phonological context bias towards a /t/-percept: The condition in which the possible /t/ would be ungrammatical and preceded by an /n/.

To analyse the results statistically, we again ran linear mixed-effects models. Participant was entered as a random factor, and Native Language, Coda Signal (from full production to complete deletion), Preceding Context (/n/ vs. /s/) and Syntax (/t/ or no /t/ predicted) as fixed factors. The binary variables Native Language, Preceding Context were contrast coded. The value of -0.5 was assigned to the Preceding Context /n/, the ungrammatical /t/-condition and the native Dutch listener Group. A positive regression weight hence would indicate that more /t/-responses are given in the /s/-context condition, the grammatical /t/ condition, and by the non-native German listeners.

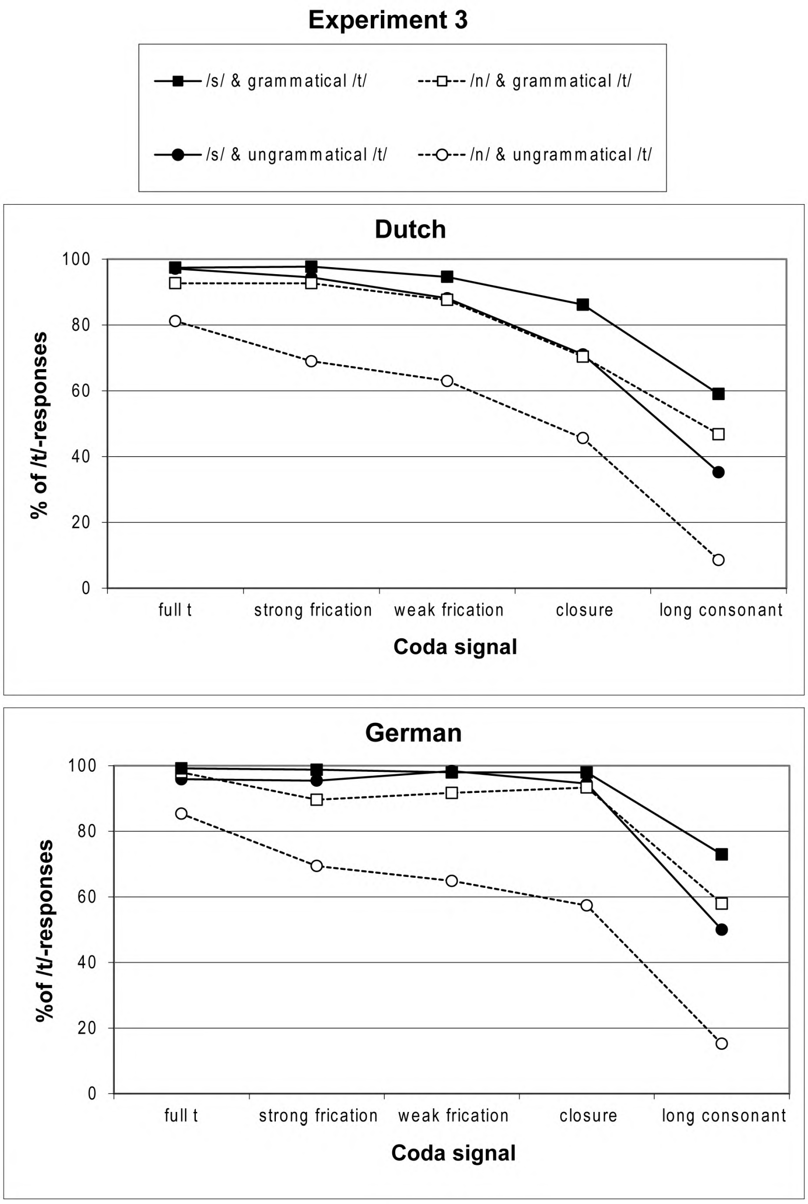

Figure 4: *Percentage of /t/-responses in Experiment 3 for each Coda Signal. The lines indicate the two different Preceding Contexts /n/ and /s/ and the two different Syntax conditions.*

Table 4: *Experiment 3: Regression weights for the final models on each level of the continuum.*

	Coda Signal				
Effect	Full /t/	Strong frication	Weak frication	Closure	Long consonant
(Intercept)	3.89***	2.95***	2.59***	1.79***	1.64***
PrecedingContext	1.55***	2.23***	1.98***	1.57***	1.34***
Syntax	1.37***	1.66***	1.24***	1.45***	1.33***
NativeLanguage	0.33	0.05	0.92*	1.65***	1.27***
PrecedingContext: Syntax	-1.09*	-	-1.05**	-0.84*	-0.65**
PrecedingContext: NativeLanguage	-0.98*	-	-	0.85*	-
Syntax : NativeLanguage	1.02*	-	1.38***	0.67	-
PrecedingContext: Syntax: NativeLanguage	-	-	-	-1.39*	-

Note: *** = $p < 0.001$; ** = $p < 0.01$; * = $p < 0.05$

Analysis of Experiment 3 also started with a full model and, in a stepwise fashion, insignificant interactions were pruned. There was an overall significant tendency for more /t/-responses by German participants. This tendency, however, was not of the same size in all conditions, that is, it was moderated by various interactions with the other experimental variables. To understand the nature of the interactions, we examined the effects of Native Language, Preceding Context and Syntax on all five levels of Coda Signal separately. For each level, the analysis started with a full model with all possible interactions of the three independent variables Native Language, Preceding Context and Syntax. Subsequently, insignificant interactions were pruned. Table 4 shows the result of this process, indicating which effects were included in the final models. Note that simple effects remained in the model even if insignificant if they were involved in higher-order interactions. Again, regression weights missing in Table 4 were not included in the final model.

In what follows, we discuss the final models for each step. For the full /t/ and the strong frication Coda signal, the analyses showed no overall difference between Dutch and German listeners but overall effects of Syntax and Preceding Context (see the main effects in the first two columns of Table 4). But for the full /t/, German participants had a larger effect of Syntax and a smaller effect of Preceding Context than Dutch listeners (see the two-way interactions including Native Language, which is positive for Syntax x Native Language but negative for Preceding Context x Native Language). Additionally, we find an additional attenuation of the effect of Preceding Context if the Syntax already facilitates a /t/-response. This interaction is also observed on all other levels except the strong-frication Coda signal. These interactions involving Preceding Context are probably due to ceiling effects: If the syntactic frame already prescribes a /t/-response, preceding Context cannot exert a strong influence anymore.

For the weak frication coda signal, the German participants gave more /t/-responses than the Dutch and the effect was enlarged if the Syntax predicts the presence of a /t/ (a significant Native Language by Syntax interaction with a positive regression weight). For the Closure coda signal there was an overall effect of Native Language which was moderated by Syntax and Preceding Context together in a three-way interaction. This three-way interaction is due to the fact that the German participants were strongly influenced by the Syntax variable in the /n/-context (b(Syntax x NativeLanguage) = 1.41, $p < 0.001$), but not in the /s/-context ($p > 0.2$). If we focus on the /s/-context, there only remained significant main effects of Native language (with more /t/-responses by German listeners, b(Native Language) = 2.03, $p < 0.001$) and Syntax (with more /t/-responses if the syntax prescribes a /t/, b(Syntax) = 1.00, $p < 0.001$). Finally, for the long consonant Coda signal—when there is no hint of /t/, but actually a long /n/ or /s/—German participants overall gave more /t/-responses than the Dutch. As Table 4 also indicates, there were also significant main effects of Preceding Context and Syntax at these levels of the /t/-∅ continuum.

To sum up, the data analysis shows that both listener groups were affected by the independent variables Syntax and Preceding Context in a similar fashion. However, the size of the effects differed between the two groups on some levels of the /t/-∅ continuum. The most consistent pattern is that the German participants were more likely to give /t/-responses, and this main effect tended to be enlarged if the Syntax prescribed a /t/ (see the interactions of Native Language and Syntax).

Discussion

The results of Experiment 3 show clear differences between native and non-native listeners. The non-native German listeners give more /t/-responses than the Dutch participants for every cell of the design. Moreover, they are less influenced by acoustic detail and more by syntactic constraints than the native-Dutch listeners. Figure 4 shows that, if grammar prescribes it, German listeners become insensitive to phonetic detail information. There is hardly any influence of type of Coda signal on the amount of /t/-responses for the first four steps of the /t/-∅ continuum. In contrast, the effect of the Syntax variable tends to be larger in the non-native German than in the Dutch listeners.

How can we account for these differences between German and Dutch listeners? An obvious candidate explanation is transfer from the L1. Experiment 1 had shown that the two groups of speakers differed in how often they reduced /t/ at the end of verbs. However, the difference in the production experiment was opposite to the direction one would expect given the results of the perception experiment. If L1 transfer would be at the basis of the perception results, German speakers should have been more likely to reduce /t/ at the end of verbs than Dutch speakers. In that case, transfer would lead them to restore /t/ more often at the end of verbs. But the production pattern was different. We observed that Dutch speakers were more likely to reduce /t/ at the end of verbs, and, in fact, only in verbs ending on /st/. This suggests that the German listeners have learned that reductions are more likely in Dutch than in German, but fail to appreciate the exact pattern of the difference. As a consequence, they overcompensate for /t/-reduction in Dutch verbs and tend to give more /t/-responses than native listeners.

Importantly, the fact that non-native listeners behave quite differently from native listeners in Experiment 3 allows us to answer the question that motivated this experiment. Is the relatively good performance of German learners in Experiment 2 due to learning of the non-native pattern or due to transfer from the L1? If the first would be the case, they should perform relatively well independent of differences between the production patterns in their L1 and their L2. However, when there are subtle differences in how /t/-reduction patterns in the two languages, German learners do not show a relatively native-like performance. This seems to indicate that the relatively good performance in Experiment 2 is due to transfer from the L1.

General Discussion

In three experiments, we investigated how L2 learners deal with a casual speech process with which they are already familiar with from their L1. The case under study was /t/-reduction in Dutch, and its perception by German learners of Dutch. Experiment 1 was necessary to establish to what extent /t/-reduction in Dutch and German patterns similarly, and thus, whether German learners could rely on their L1 knowledge in dealing with /t/-reduction in Dutch. If /t/ is a morphological marker, /t/-reduction differs between the languages, with /t/-reduction being especially likely in Dutch if the /t/ is preceded by an /s/. In German, in contrast, /t/-reduction was in this case not conditioned by the preceding phonological context. If /t/ was part of the word's stem—and hence not a morphological marker—both languages showed similar patterns with more /t/-reduction after /s/ than after /n/.

Experiments 2 and 3 then investigated to what extent these similarities and differences are reflected in perception. Experiment 2 found that German learners behaved quite similarly to native listeners for non-morphological /t/. German listeners made use of phonological cues to the presence of /t/ and did not rely more strongly on lexical knowledge than Dutch listeners. Experiment 3 investigated the perception of /t/-reduction for morphological /t/ in verbs. In this case, there were clearer differences between native Dutch listeners and German learners of Dutch. The German learners relied more strongly on grammatical cues and paid less attention to phonetic cues. These results may be taken to suggest that the German participants are less sensitive to the phonetic and/or phonological variables and rely more on higher-level information. However, the results of Experiment 2 show that this is not the complete picture. Here, non-native German participants show a similar identification function over the /t/-∅ continuum as the Dutch participants; they are also more strongly influenced by the preceding phonological context than Dutch listeners. Moreover, the German participants did not rely more strongly on lexical cues than the Dutch listeners.

This indicates that the German listeners are generally well able to deal with /t/-reduction in their second language Dutch. This ability may be grounded in the experience German listeners have with /t/-deletion in their native language. Previous research on segmental contrasts in a second language has indicated that a segmental contrast in a second language is easiest to learn if the contrast is similar to a native contrast. Similarly, the fact that the German listeners perform well in their compensation for /t/-reduction in Dutch might be due to the experience they already have with this casual speech process from their native language.

However, there is another possibility to account for the relatively good performance of the German listeners. Maybe /t/-reduction is very easy to deal with for a German—or in fact any other—listener independent of language experience. The process of /t/-reduction occurs in several different languages and may be constrained by universal perceptual biases. In line with this proposal, Mitterer, Yoneyama, and Ernestus (2008) found that even Japanese listeners, without any experience with /t/-deletion, show some basic context effects, so that the reduction of /t/ is more difficult to notice—and therefore less problematic for word recognition—after /s/ than after /n/.

With this line of reasoning, we join a lively debate about the role of native-language learning in compensation for casual speech processes. Gow and Im (2004), for instance, argue that language experience is not necessary for listeners to compensate for assimilation. Native and non-speakers of Hungarian and Korean monitored for segments in assimilated and non-assimilated control contexts related to language-specific assimilation processes in those languages. Native and non-speakers of Hungarian showed similar context effects related to Hungarian voicing assimilation, whereas native and non-speakers of Korean both showed no context effect related to Korean labial place assimilation. Gow and Im's findings support the view that progressive assimilation context effects are produced by mechanisms that do not require familiarity with a specific assimilation process.

Looking at regressive context effects, Mitterer, Csépe, Honbolygo and Blomert (2006) showed that Dutch listeners can compensate for Hungarian liquid assimilation, despite the fact that no similar process exists in Dutch. In this study, native Hungarian and non-native Dutch listeners behaved identically. However, the data in Mitterer, Csépe, Blomert (2006) show a more nuanced picture. Next to language independent effects, which made the consequences of the assimilation difficult to perceive, language experience seemed to help listeners to deal with the potentially ambiguous stimuli arising from assimilation. A similar conclusion is also reached by Darcy, Peperkamp and Dupoux (2007). They also found that some compensation for assimilation is possible without language experience. However, they showed that native listeners show a stronger compensation effect than non-native listeners, and more advanced L2 learners show larger compensation effects than beginning learners.

The current data at the very least indicate that compensation for /t/-reduction is not completely independent of language experience, given that we observed differences between native Dutch listeners and German L2 learners. Similarly, the results of Mitterer et al. (2008) had already shown clear differences between native Dutch and naive non-native Japanese listeners. In this context, it is interesting to note that Mitterer and Ernestus (2006) already observed that lexical effects seem to be more pronounced in compensation for /t/-reduction than in compensation for assimilation. These data all suggest that compensation for /t/-reduction is quite reliant on higher-level processing.

As a consequence, non-native listeners show worse performance, when they cannot rely on L1 knowledge. German learners can rely on L1 knowledge for compensation for /t/-reduction in nouns and adjectives, and they seem to do so quite successfully. However, they cannot rely on L1 knowledge when dealing with the reduction of a morphological /t/ in verbs. The production data revealed clear differences in the reduction patterns between Dutch and German: Dutch speakers reduce /t/ more often when it is a verb inflection and is preceded by /s/ than German speakers do. This seems to have consequences for perception. Our data suggest that, in this case, German learners do not rely directly on their L1 knowledge. They seem to have learned the difference but overgeneralize this knowledge and as a result overcompensate for /t/-reduction in Dutch verbs and fail to exploit the acoustic cues to the full extent; even when the acoustic signal has almost no cues for the presence of a /t/, the German listeners mainly give /t/-responses for Dutch verbs with a stem ending in /s/.

Even though the non-native listeners in our study may not perform completely native-like in their compensation for /t/-reduction, it is clear that it is easier to deal with an L2 casual speech process if it patterns similarly in the L1. Our German listeners showed a native-like performance in their compensation for /t/-reduction in nouns and adjectives, which can be explained by the fact that the /t/-reduction patterns for these types of words is similar in Dutch and German. For verbs, the reduction patterns differ between Dutch and German, and consequently it becomes more difficult for German listeners to compensate for /t/-reduction in Dutch in a native-like manner. In other words, with less similarities between the casual speech process in the

L1 and the L2, it becomes harder to effectively compensate for it. Or, vice versa, a casual speech process in the L2 is relatively easy if it occurs in a similar fashion in the L1.

This line of reasoning is buttressed if we look at a case in which a casual speech process does not exist at all in the native language. It seems that L2 listeners then have much more difficulties in coping with it. As mentioned in the introduction, Tuinman et al. (under revision) investigated the process of /r/-intrusion in British English—a process that does not occur in Dutch—and found that Dutch listeners showed a completely different pattern than native British English listeners. This study also employed a 2AFC task, and listeners had to decide whether a phrase such as "saw(r)ice" meant "saw rice" or "saw ice". Tuinman et al. varied the duration of the [r], and British English listeners showed a great sensitivity to this subphonemic detail. They showed a clear sigmoid identification function over the continuum. The identification function of the Dutch learners, however, was essentially flat. The Dutch performed somewhat better if there was an orthographic cue to the presence of /r/, in a phrase such as "more ice" (Note that in British English, "saw" and "more" rhyme, that is, "more" is canonically produced without a final /r/). In this case, the decision was between a "singelton /r/" ("more ice"), and a geminate /r/ ("more rice")—a distinction Dutch listeners are familiar with from their native language ("meer reis" vs. "meer ijs")—and Dutch listeners showed a more categorical identification function, that was, however, still not as steep as the identification function by British English listeners.

The data points discussed seem to converge on the conclusion that connected speech processes pose similar problems in learning a second language as new phonemes do. Processes that are unique to the L2, such as /r/-insertion in British English for Dutch learners, lead to major perceptual problems. Processes that subtly differ between the L1 and the L2 lead to moderate problems. Examples are /t/-deletion in Dutch verbs for German listeners, and the singleton-geminate distinction for Dutch learners of English. Finally, processes that are quite similar in the L1 and L2, such as /t/-deletion in nouns and adjectives in Dutch and German, are easy to master for L2 learners.

As with speech perception in noise, the adverse condition created by casual speech processes is less problematic for native listeners than non-native listeners. However, when non-native listeners can transfer their L1 knowledge of a casual speech process to their L2, such a process is relatively easy to deal with. When listeners cannot completely rely on their L1 knowledge to compensate for a casual speech process such as /t/-reduction this can be problematic in L2 speech perception.

References

Baayen, H., Piepenbrock, R., & Gulikers, L. (1995). The CELEX lexical database (release 2) [CD-ROM]. Philadelphia, PA: Linguistic Data Consortium, University of Pennsylvania [Distributor].

Barr, D. J. (2008). Analyzing 'visual world' eyetracking data using multilevel logistic regression. *Journal of Memory and Language, 59*, 457-474.

Best, C. T. (1994). The emergence of native-language phonological influences in infants: a perceptual assimilation model. In H. C. Nusbaum (Ed.), *The development of speech perception: The transition from speech sounds to spoken words*. Cambridge, MA: MIT.

Best, C. T. (1995). A direct realist view of cross-language speech perception. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 171-204). Baltimore: York Press.

Boersma, P., & Weenink, D. (2005). Praat: doing phonetics by computer (Version 4.3.07) [Computer program]. Retrieved April 2005, from http://www.praat.org/.

Bohn, O.-S., & Munro, M. J. (Eds.). (2007). *Language experience in second language speech learning: In honor of James Emil Flege*. Amsterdam: John Benjamins.

Bradlow, A. R., Pisoni, D. B., Akahane-Yamada, R., & Tohkura, Y. i. (1997). Training Japanese listeners to identify English /r/ and /l/: IV. Some effects of perceptual learning on speech production. *Journal of the Acoustical Society of America, 101*(4), 2299-2310.

Broersma, M., & Cutler, A. (2008). Phantom word activation in L2. *System, 36*, 22-34.

Broersma, M., & Cutler, A. (2011). Competition dynamics of second-language listening. *Quarterly Journal of Experimental Psychology, 64*, 74-95.

Cutler, A., & Otake, T. (2004). Pseudo-homophony in non-native listening. *Journal of the Acoustical Society of America, 115*, 2392.

Cutler, A., Weber, A., & Otake, T. (2006). Asymmetric mapping from phonetic to lexical representations in second-language listening. *Journal of Phonetics, 34*, 269-284.

Darcy, I., Peperkamp, S., & Dupoux, E. (2007). Bilinguals play by the rules: Perceptual compensation for assimilation in late L2-learners. In J. Cole & J. Hualde (Eds.), *Papers in Laboratory Phonology 9* (pp. 411-442). Berlin: Mouton de Gruyter.

Davis, M., Marslen-Wilson, W. D., & Gaskell, M. G. (2002). Leading up the lexical garden path: Segmentation and ambiguity in spoken word recognition. *Journal of Experimental Psychology: Human Perception and Performance, 28*, 218-244.

Dilley, L., & Pitt, M. A. (2007). A study of regressive place assimilation in spontaneous speech and its implications for spoken word recognition. *Journal of the Acoustical Society of America, 122*, 2340-2353.

Dixon, P. (2008). Models of accuracy in repeated-measures designs. *Journal of Memory and Language, 59*, 447-456.

Ernestus, M., Baayen, H., & Schreuder, R. (2002). The recognition of reduced word forms. *Brain and Language, 81*, 162-173.

Gaskell, M. G., & Marslen-Wilson, W. D. (1996). Phonological variation and inference in lexical access. *Journal of Experimental Psychology: Human Perception and Performance, 22*, 144-158.

Gow, D. W. (2002). Does English coronal place assimilation create lexical ambiguity? *Journal of Experimental Psychology: Human Perception and Performance, 28*, 163-179.

Gow, D. W. (2003). Feature parsing: Feature cue mapping in spoken word recognition. *Perception and Psychophysics 65*, 575-590.

Gow, D. W., & Im, A. M. (2004). A cross-linguistic examination of assimilation context effects. *Journal of Memory and Language, 51*, 279-296.

Guy, G. R. (1980). Variation in the group and the individual. In W. Labov (Ed.), *Locating language in time and space* (pp. 1-36). New York: Academic Press.

Guy, G. R., & Boyd, S. (1990). The development of a morphological class. *Language Variation and Change, 2*, 1-18.

Ingram, J. C. L., & Park, S.-G. (1998). Language, context, and speaker effects in the identification and discrimination of English /r/ and /l/ by Japanese and Korean listeners. *Journal of the Acoustical Society of America, 103*(2), 1161-1174.

Janse, E., Nooteboom, S. G., & Quené, H. (2007). Coping with gradient forms of /t/-deletion and lexical ambiguity in spoken word recognition. *Language and Cognitive Processes, 22*, 161-200.

Klatt, D. H. (1980). Software for a cascade/parallel formant synthesizer. *Journal of the Acoustical Society of America, 67*, 971-995.

Lecumberri, M. L. G., Cooke, M., & Cutler, A. (2010). Non-native speech perception in adverse conditions: a review. *Speech Communication, 52*, 864-886.

Lehiste, I. (1970). *Suprasegmentals*. Cambridge, MA: MIT Press.

Mitterer, H., & Blomert, L. (2003). Coping with phonological assimilation in speech perception: Evidence for early compensation. *Perception and Psychophysics, 65*, 956-969.

Mitterer, H., Csépe, V., & Blomert, L. (2006). The role of perceptual integration in the perception of assimilation word forms. *Quarterly Journal of Experimental Psychology, 59*, 1395-1424.

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.

Mitterer, H., & Ernestus, M. (2006). Listeners recover /t/s that speakers reduce: Evidence from /t/-lenition in Dutch. *Journal of Phonetics, 34*, 73-103.

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.

Mitterer, H., Yoneyama, K., & Ernestus, M. (2008). How we hear what is hardly there: Mechanisms underlying compensation for /t/-reduction in speech comprehension. *Journal of Memory and Language, 59*, 133-152.

Oostdijk, N. (2000). The Spoken Dutch Corpus project. *ELRA Newsletter, 5*, 4-8.

Pallier, C., Colomé, A., & Sebastián-Gallés, N. (2001). The influence of native-language phonology on lexical acces: Exemplar-based versus abstract lexical entries. *Phonological Science, 12*, 445-449.

Pitt, M. A., Dilley, L., Johnson, K., Kiesling, S., Raymond, W., Hume, E., et al. (2007). Buckeye Corpus of Conversational Speech (2nd release) [www.buckeyecorpus.osu.edu]. Columbus, OH: Department of Psychology, Ohio State University (Distributor).

Pluymaekers, M., Ernestus, M., & Baayen, H. (2005). Lexical frequency and acoustic reduction in spoken Dutch. *Journal of the Acoustical Society of America, 118*, 2561-2569.

Spinelli, E., McQueen, J. M., & Cutler, A. (2003). Processing resyllabified words in French. *Journal of Memory and Language, 48*, 233-254.

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

Sumner, M., & Samuel, A. G. (2005). Perception and representation of regular variation: The case of final /t/. *Journal of Memory and Language, 52*, 322-338.

Sumner, M., & Samuel, A. G. (2009). The effect of experience on the perception and representation of dialect variants. *Journal of Memory and Language, 60*, 487-501.

Torreira, F., Adda-Decker, M., & Ernestus, M. (2010). The Nijmegen corpus of casual French. *Speech Communication, 52*, 201-212.

Tucker, B. V., & Warner, N. (2007). *Inhibition of processing due to reduction of the American English flap.* Paper presented at the 16th International Congress of Phonetic Sciences, Saarbrücken, Germany.

Tuinman, A., Mitterer, H., & Cutler, A. (under revision). Cross-language and cross-dialect differences in perception of intrusive /r/ in English. *Journal of the Acoustical Society of America*. *Chapter 3 of this dissertation.*

Tuinman, A., Mitterer, H., & Cutler, A. (submitted). Syntactically-based compensation for phonological reduction without top-down feedback. *Chapter 5 of this dissertation.*

Underbakke, M., Polka, L., Gottfried, T. L., & Strange, W. (1988). Trading relations in the perception of /r/-/l/ by Japanese learners of English. *Journal of the Acoustical Society of America, 84*(1), 90-100.

Warner, N., Fountain, A., & Tucker, B. V. (2009). Cues to perception of reduced flaps. *Journal of the Acoustical Society of America, 125*, 3317-3327.

Cross-language and cross-dialect differences in perception of intrusive /r/ in English

Chapter 3

Tuinman, A., Mitterer, H., & Cutler, A. (under revision). Cross-language and cross-dialect differences in perception of intrusive /r/ in English. *Journal of the Acoustical Society of America*.

Abstract

In sequences such as *law and order*, speakers of British English often insert /r/ between *law* and *and*. Intrusive /r/ could create speech perception difficulties, especially for listeners unfamiliar with this phonological process. In a 2AFC experiment, British and American English listeners and Dutch listeners proficient in English heard British English sentences in which the duration of /r/ was manipulated across a word boundary (*saw (r)ice* or *saw more (r)ice*), and orthographic and semantic biases favoring the r-initial interpretation were present or absent. British English 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. American and Dutch listeners relied less on durational cues than the British listeners, and were affected by orthography, reporting /r/ more often after *saw* than after *more*. Additionally, Dutch listeners were susceptible to semantic bias in the sentences. Thus although intrusive /r/ causes perceptual problems for listeners with another language or another dialect, a language difference induces greater difficulties. Neither Dutch nor most American English dialects have intrusive /r/, but American listeners may be more sensitive to this phenomenon in the British version of their own language.

Introduction

Margaret Thatcher's legendary nickname *Laura Norder* not only reflected her political preferences, but also the fact that, like most of her compatriots, she pronounces *law and order* with an intrusive /r/ between *law* and *and*. Most British English dialects are non-rhotic, i.e., they have the phonotactic constraint whereby /r/ can occur in word onsets but not at the end of words; thus the /r/ in *real* /ril/ is pronounced whereas the /r/ in a citation-form utterance of *hear* /hɪə/ is not, even though the spelling of *hear* ends with the letter *r*. However, word-final /r/ occurs under two conditions. First, it appears when the following word begins with a vowel so that the /r/ can be realized in syllable onset position (e.g., it can surface in *hear it* /hɪərɪt/). This phenomenon is known as 'linking /r/'. Second, as in the *law and order* case, there is 'intrusive /r/', whereby /r/ may be inserted after a non-high vowel (e.g., [ə, a, ɔ] and diphthongs ending in [ə]) and before a vowel-initial word. In contrast to linking /r/, intrusive /r/ is not represented in spelling (Cruttenden & Gimson, 1994, pp. 262-264; Giegerich, 1992, pp. 66, 282-283).

The *law and order* example shows that an intrusive /r/ can be perceived as an onset /r/, and that native listeners realize that although *law and* was meant, the phonemic sequence is also equivalent to *Laura N*. Connected speech processes such as intrusive /r/ have received widespread attention in speech perception research, and native (L1) listeners prove to be highly able to derive correct interpretations of connected speech (Connine, Ranbom, & Patterson, 2008; Mitterer, Csépe, & Blomert, 2006; Mitterer & Ernestus, 2006). Less attention, however, has been paid to whether connected speech processes cause difficulties for second language (L2) listeners (and indeed listeners with a different dialect, too). Are such listeners led into misperceptions by connected speech processes that form no part of their native variety? Specifically, will listeners who do not know that *law and order* can attract an intrusive /r/ be most likely to perceive this phrase as *Laura Norder*?

Research on perceptual difficulties in L2 listening has focused principally on inventory differences between languages (e.g., Best, McRoberts, & Goodell, 2001; Guion, Flege, Akahane-Yamada, & Pruitt, 2000; Polka, 1995). The /r/-/l/ contrast in English, for example, causes well-known difficulties for Japanese listeners, whose L1 has only a single phonetic category matching well to neither /r/ nor /l/ (e.g., Bradlow, Pisoni, Akahane-Yamada, & Tohkura, 1997; Ingram & Park, 1998; Underbakke, Polka, Gottfried, & Strange, 1988). In an eye-tracking study of Cutler, Weber and Otake (2006), Japanese listeners instructed to click on a picture of a rocket experienced interference when a picture of a locker was present; they tended to look at the locker before eventually focusing on the rocket. In natural listening as in this experiment, L2 listeners have difficulties perceiving phonemes that they are unfamiliar with from their L1, and their ultimate success in comprehending what the speaker meant is often achieved by the use of disambiguating information that follows.

Listening to a different dialect than your own can also prove to be a challenge. Sumner and Samuel (2009), for instance, found that speakers of a General American dialect can have problems in recognizing words produced with a New York accent. Floccia, Goslin, Girard and Konopczynski (2006) also found an initial processing cost for a different dialect in word recognition tasks, and Otake and Cutler (1999) found that cross-dialect perception of Japanese words exhibited lower sensitivity to information in the signal (*d'*) and a higher degree of bias (β) towards lexical knowledge. Nevertheless, cross-dialect difficulties in phoneme perception are not always large. Cutler, Smits and Cooper (2005) asked Australian, American and Dutch listeners to identify American English vowels in meaningless CV and VC syllables, and found that overall, the Australian and American listeners performed equally well, although the Australians were systematically affected by the tendency to greater vowel tenseness in their native dialect (reporting, for example, /ɔ/ as /ɑ/ much more than vice versa). The Dutch listeners' overall performance, however, was significantly worse, suggesting that language differences have greater consequences than dialect differences for the perception of phonemes.

In our experiment we investigate whether this also holds for the interpretation of speech that involves phonological processes. We are not aware of any studies of the effect of dialect differences in connected speech processes on the interpretation of speech sounds or words, although Scott and Cutler (1984) showed that British English listeners were unable to use patterns of palatalization and intervocalic flapping in American English as cues to syntactic structure. There have been cross-language studies at the word level, however; thus Darcy, Peperkamp and Dupoux (2007) tested the perception of consonant-to-consonant voicing assimilation—which occurs in

French, but not in English—by native French listeners and English learners of French. The English were either beginning or advanced learners. Advanced learners were able to compensate for the assimilation just as well as the native listeners. The beginning learners also showed some compensation, though this may be due to general auditory processes which contribute to compensation for assimilation (Mitterer, Csépe, & Blomert, 2006).

Our study addresses the /r/-insertion found in non-rhotic British English and how it is perceived, by native and by non-native listeners. The insertion process induces acoustic evidence for /r/ that cannot be attributed to an underlying or orthographic representation. Non-native listeners who are unfamiliar with this phonological process may find the intrusive /r/ difficult to interpret. Besides listeners with British English as their native dialect, we test listeners whose language is different, and listeners whose dialect is different: Dutch listeners proficient in English, and American English listeners. We chose these listeners groups because, in contrast to British English, both the Dutch language and most American English dialects are rhotic; i.e., both Dutch and American English speakers do produce word-final /r/s.

Dutch dialects exhibit conspicuous variation in the way /r/ is produced (van Bezooijen, 2005; van de Velde, 1996), but no dialect is non-rhotic in the way British English is. There is thus no role for linking /r/ in Dutch; in the three cases *meer* ('more'), *meer tijd* ('more time') and *meer appels* ('more apples') the final /r/ in *meer* is always pronounced. Further, in no variety of Dutch is /r/ ever inserted in contexts resembling *law and*, e.g., *na appels* 'after apples' (Collins & Mees, 1999, pp. 178-181; Gussenhoven & Broeders, 1997, pp. 154-155). All our participants, like all Dutch students, had high proficiency in English. Exposure to both British and American varieties of English is widely available in the media in the Netherlands; British television and radio can be received in all households, and where British or American movies or series are shown on Dutch television channels, they are subtitled, never dubbed. The target English pronunciation taught in Dutch schools and universities is actually British English (see e.g., Collins & Mees, 1999, p. vii; Gussenhoven & Broeders, 1997, pp. 16-17).

Likewise, in most dialects of American English, the /r/ in *more* is always pronounced in the English translation equivalents of the above Dutch utterances: *more*, *more time, more apples.* A few American dialects are non-rhotic, i.e., referred to as showing "r-dropping" (a term that well demonstrates the normative status of rhoticity in the dialect area!). However, even the Boston accent, perhaps the most widely referred to among such cases, is actually known to be undergoing a change towards rhoticity (Irwin & Nagy, 2007; Trudgill, 1986, pp. 76-77; Wells, 1982, p. 520). Nevertheless, we did not test speakers of such a dialect; our listeners were speakers of rhotic varieties, and were tested in Philadelphia, where a rhotic variety is the norm. To confirm that rhotic varieties of American English are unlikely to exhibit intrusive /r/ even in casual speech contexts, we also conducted a complete search of the Buckeye Corpus of Conversational Speech (Pitt, et al., 2007). We isolated from this corpus all 4698 instances in which /r/-insertion would be possible between a word ending in a non-high vowel followed by a vowel-initial word (the licit context for the intrusion in British English). In none of these, there is a clear insertion. In three cases, an /r/ is transcribed at the word boundary, but this stems from the words themselves rather than from an insertion (e.g., "camera either" produced as "camreither"). Like our Dutch participants, however, we expect that our American participants will be familiar with British English speech via exposure in the media.

Thus our two non-native listener groups are both expected to have experience with the perception of British English, but differ significantly from British English listeners in their expectations about /r/ occurrence based on their native variety. In British English versions of the examples we have cited, of course, *more* would have no /r/ at all, *more apples* would evince a (linking) /r/, whereas in *saw apples* there could be an intrusive /r/.

At issue in our study is the extent to which the listeners will acknowledge the presence of a phonetic segment /r/. We thus first conduct acoustic analyses of naturally produced intended onset versus intrusive /r/ segments, to establish whether these two types of /r/ differ acoustically, and, if so, in what manner. On the basis of earlier reports we expected that the duration would be longer, and the intensity decrement in dB larger, for onset /r/ tokens than for intrusive /r/ tokens (Cruttenden & Gimson, 1994, pp. 187, 262-264). However, segment pronunciations, and especially casual speech phenomena, may change with time, and it is important to have current measurements; moreover, intrusive /r/ is an optional process and it is important to know whether the speaker whose productions stand model for our study actually produces significant acoustic differences. On the basis of the acoustic measurements we then construct materials varying in amount of acoustic evidence for an /r/ segment, and present these materials to our three listener groups. The context in which this evidence is presented is at a word boundary, and we further vary whether or not there is orthographic support for a final /r/ (i.e., we contrast the word *more* versus the word *saw* before the boundary), and whether or not there is semantic support for a syllable-onset /r/ interpretation (a context favoring *rice* after the boundary, versus a context offering no such support).

With this design we can ascertain first how native listeners decide whether an /r/ sound is an intended phoneme versus an intrusive sound: whether they base their judgements on acoustic evidence alone, or are inclined to take orthographic and semantic evidence into consideration as well. By comparing both non-native listener groups with the native group, we can then determine whether either non-native group can attain the standard of performance we find with the native listeners, and whether the cues used by the non-native listeners in making their decisions are the same as those used by the native listeners. Note that there is evidence in previous literature that non-native listeners may rely to a greater extent than native listeners on both orthographic information (Escudero, Hayes-Harb, & Mitterer, 2008) and semantic information (Bradlow & Alexander, 2007). Finally, we can determine whether, as with phoneme perception, it makes a difference whether the mismatch in the native and non-native systems is at the level of language or of dialect.

Acoustic analyses

Materials

We constructed 27 pairs of English sentences contrasting an intended onset /r/ and an intrusive /r/. An example sentence is "*My brother likes extra rice/ice 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. Trivially, the r-initial member of the pair in the sentence will trigger pronunciation of an /r/. More importantly, the vowel-initial member of the pair together with the preceding low vowel (in this case, the last vowel of *extra*) creates a context in which an intrusive /r/ can be produced. These sentences are listed in the Appendix.

A list was constructed in which all sentences occurred, in random order and, in order to obscure the purpose of the study, interspersed among 208 filler sentences. The entire list was recorded by a female native speaker of British English from the

London area who produces intrusive /r/s in casual speech. Each crucial sentence with /r/ was recorded at least twice.

Measurements and discussion

We measured both the duration of each intrusive or onset /r/, and the decrease in intensity from the vowel preceding the /r/ to the lowest point in the /r/. All 126 /r/ tokens were measured by the first author and a subset of these sentences was also measured by the second author as a reliability check. The correlation between the two measurements was high (duration: $r = .72$; intensity difference: $r = .95$). The item-by-item measurement data can also be found in the Appendix.

Both predicted differences appeared in this speaker's productions. Overall, onset /r/s were longer [$F(1,124) = 8.74$, $p < 0.01$]; the measured onset /r/s were on average 89 ms, while the intrusive /r/s averaged 69 ms. Onset /r/s also displayed a larger intensity decrement from the preceding vowel to the lowest point [$F(1,124) = 10.79$, $p < 0.01$]; the mean intensity decrement for onset /r/s was 7.9 dB, and for intrusive /r/s 2.2 dB.

To ascertain the perceptual usefulness of these patterns we calculated a power estimate, the Cohen's d difference score (Mean Difference divided by Standard Deviation; Cohen, 1992). Values of Cohen's d above 0.8 are held to indicate a large (and hence perceptually relevant) effect size. Cohen's d for the durational difference was 1.6, and for the intensity decrement 1.9. Thus these measured differences between onset and intrusive /r/s can be perceptually of use to listeners. Whether listeners can indeed distinguish between the two types of /r/ on the basis of acoustic evidence was then tested in a 2AFC experiment.

Perception experiments

Method

Participants

The three perception experiments involved respectively 18 native speakers of Dutch, 18 native speakers of British English, and 14 native speakers of American English. The Dutch participants had a high level of proficiency in English as a second language; on average, they had received seven years of English instruction in primary and secondary education. The British and American participants did not have any knowledge of Dutch. The Dutch participants were recruited from the Max Planck Institute participant pool, the British participants from the participant pool of the Laboratory of Experimental Psychology of the University of Sussex, and the American participants from the participant pool of the Institute for Research in Cognitive Science of the University of Pennsylvania. None reported any hearing impairment. All were volunteers and were paid a small fee for their participation.

Materials

Four experimental sentences (see Table 1) were constructed, crossing a contextual bias with an orthographic bias for the perception of /r/. In each sentence, listeners judged the critical word *(r)ice* and had to decide whether they heard *ice* or *rice*. Sentences with the context *the social worker* and *given to the poor* were intended to have a semantic bias towards *rice* rather than *ice*, while sentences with *the little girl* and *given to her brother* were assumed to be less biased towards *rice*. The orthographic bias was established by the words *saw* and *more* preceding the target word *(r)ice*. As the phrase *more ice* includes an /r/ in the spelling, a perceived /r/ in the speech signal can be attributed to *more*, while in the case of *saw ice*, an /r/ sound cannot be ascribed to spelling. The bias should therefore manifest itself in terms of more reports of *rice* after *saw* than after *more*.

Table 1: *Sentence materials for perception experiments*

	Orthography	
Context	***ice*** bias	***rice*** bias
less *rice* bias	The little girl **saw more** ice/rice was given to her brother	The little girl **saw** ice/rice was given to her brother
more *rice* bias	The social worker **saw more** ice/rice was given to the poor	The social worker **saw** ice/rice was given to the poor

Although our acoustic measurements showed that both the duration of /r/ and the intensity drop into /r/ differed across intrusive versus onset cases, intensity and duration are highly correlated in the duration range in which phonemic distinctions are realized (Cohen, Ward, & Enns, 1999, pp. 234-235). Increasing the amount of intensity decrement thus has the same effect as increasing duration where there is an intensity decrement. For this reason, we varied our materials along a single parameter, namely duration. Note that our duration measurements led to a lower Cohen's d than the intensity measurements, and thus potentially offer a more sensitive test of perceptual utilization.

To maintain full control of the acoustic manipulation we constructed our sentences using MBROLA (Dutoit, Pagel, Pierret, Bataille, & Vreken, 1996), a speech synthesizer based on diphone concatenation, with the same female native speaker of British English who produced the materials for the acoustic measurements serving as a model. This speaker first recorded the four sentences in Table 1 several times. The most fluent recordings without hesitations were chosen as the model sentences. The durations of all segments and the course of the pitch contour of each sentence were measured. These measurements were then used to create synthetic tokens of the same sentences, with MBROLA taking as input the list of phonemes plus prosodic information (phoneme durations, pitch contour) to produce speech samples. Four native British listeners then judged these output tokens and further improved them by adjusting the phoneme durations and pitch contour to render them even more natural. Final versions were then created in which all constant portions were identical in duration across the four sentences. Thus every occurrence of *saw* had exactly the same duration in each sentence, and so did every occurrence of *more*. Similarly, the duration of the phonemes /ai/ and /s/ was constant in all versions of *ice* and *rice*. The only non-constant parameter was the duration of /r/; this varied from 25 ms to 121 ms in seven exponential steps (see Figure 1). This range of durations was, as noted above, based on our measurements of the same speaker's onset and intrusive /r/s.

Design and procedure

There were three independent variables: (1) the duration of /r/ at the critical word boundary; (2) the orthographic manipulation: sentences with *more (r)ice* versus with *saw (r)ice*; (3) the semantic manipulation: sentences with *the little girl* versus *the* *social worker* as subject. The dependent variable was the percentage of *rice* responses in each cell of the design.

In all three countries, the experiments were conducted on a standard PC running the NESU software. Participants were tested one at a time in a quiet room. They wore Sennheiser headphones, sat at a comfortable reading distance from the computer screen and had a two-button response box in front of them. Instructions were given in English (also to the non-native participants); these informed participants that on each trial they would hear an English sentence and see the words "ice" and "rice" on the computer screen. They were asked to press the right button labeled "rice" if the sentence they heard contained the word "rice", and to press the left button labeled "ice" if the sentence contained the word "ice". The 280 stimulus sentences (seven /r/-durations x four sentences x 10 repetitions) were presented in a random order which was different for every participant.

The experiments started with four practice trials. Each trial (practice or experimental) began with 150 ms of blank screen, after which the words "ice" and "rice" appeared in the upper left and upper right corners of the screen. After a further 450 ms the sentence was presented over the headphones. From the onset of /r/ in the sentence, participants had four seconds to press one of the buttons. Their response caused the other word to be removed from the screen, so that it could be seen that the answer had been registered by the computer. If participants did not respond within this time limit, a stopwatch appeared on the screen to remind them to react more rapidly. After each button press, the next trial started after a one-second interval. Participants could take a break after every 50 trials and continued when they were ready. Additionally, reaction times were recorded from onset of the target word *(r)ice*, although the percentage of *rice* judgments was the principal dependent variable.

Results and discussion

Figure 1 shows the percentage of *rice* judgments as a function of the three independent variables, separately for each experimental population. It is clear at a glance that the three groups' response patterns all differed from each other. British English listeners produced four closely similar functions for the four sentences, with a smooth categorical curve suggesting that they based their responses mainly on the durational variation rather than on the other variables. In contrast, for both non-native

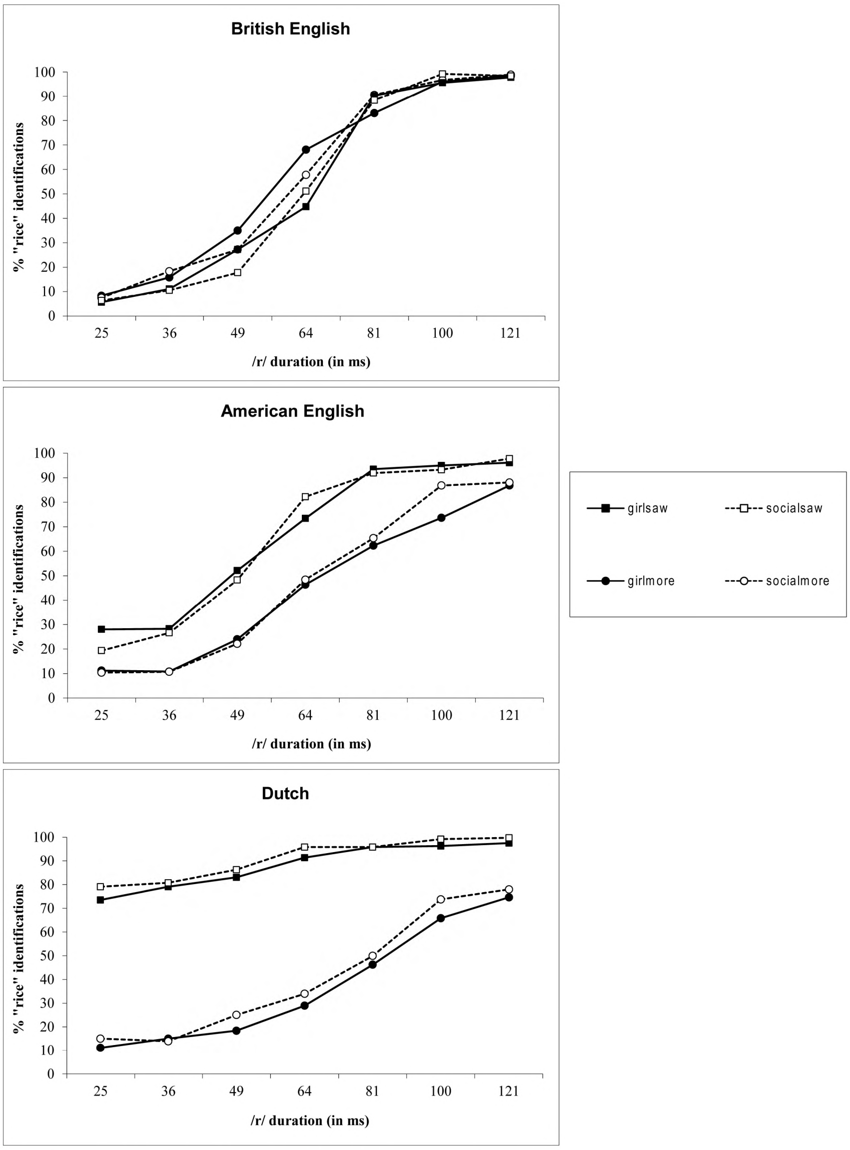

Figure 1: *Percentages of "rice" identifications for British English, American English and Dutch listeners.*

groups there was a clear separation between the response functions for sentences with *saw* (squares) versus *more* (circles), suggesting that each of these groups was affected by the presence of an orthographic r in the pre-boundary word. Further, the two non-native groups' functions are also not identical; they differ in relative categorical shape, in degree of separation between the *more* and the *saw* sentences, and in whether or not there is a difference between the *little girl* and the *social worker* sentences.

The results were analyzed with a linear mixed-effects model (Baayen, 2008). Participant was entered as a random factor and Native Language, Orthography, Context, and Duration as fixed factors. Duration was entered as a numerical variable, centered around zero, and Orthography and Context were contrast-coded (-/r/ bias = -0.5, +/r/ bias = 0.5). For the Native Language variable, the British English group was mapped on the intercept (see Baayen, 2008, pp. 254-255).

Analysis started with a full model and, in stepwise fashion, insignificant interactions were pruned. Table 2 shows the regression weights for the final model. Note that the terms not qualified with a Native Language term are those for the British English group (mapped on the intercept). The analysis shows that the effects readily visible in Figure 1 are statistically significant (marked with asterisks in the Table). The British English Group shows a clear effect of /r/ Duration, no effect of Context, and a small negative effect of Orthography. That is, British English listeners in fact gave somewhat more *rice* responses when the preceding word was *more* than when the preceding word was *saw*.

Table 2: *Regression weights for the final model*

Effect	Regression Weight
(Intercept)	0.51367***
NativeLanguage = Dutch	0.48597**
NativeLanguage =American English	-0.03413
/r/ Duration	1.30925***
Orthography	-0.32732***
Context	-0.04242
NativeLanguage = Dutch:/r/ Duration	-0.73329***
NativeLanguage = American English:/r/ Duration	-0.44165***
NativeLanguage = Dutch: Orthography	3.60064***
NativeLanguage = American English: Orthography	1.66664***
/r/ Duration: Orthography	0.15115*
/r/ Duration: Context	0.06244*
NativeLanguage = Dutch: Context	0.30691**
NativeLanguage = American English: Context	0.10611
NativeLanguage = Dutch:/r/ Duration: Orthography	-0.26804**
NativeLanguage = American English:/r/ Duration: Orthography	-0.05934

Note: *** = $p < 0.001$; ** = $p < 0.01$; * = $p < 0.05$

Both the American English group and the Dutch group showed a large positive effect of Orthography, i.e., gave, as predicted, more *rice* responses when the preceding word was *saw* than when the preceding word was *more* (see the Native Language by Orthography terms). Both non-native groups also were less categorical in their responses (see the negative terms for Native Language by /r/ Duration) than the British English group. Additionally, only the Dutch group was influenced by Context (made more *rice* responses in the context of the *social worker* sentences), and only the Dutch group made less use of /r/ Duration when there was an Orthographic Bias (three-way interaction of Native Language: Dutch by Duration by Orthography; flatter Dutch function for the *saw* sentences than for the *more* sentences).

The negative findings for the British English listeners (lack of an Orthographic effect in the expected direction and lack of a Context effect) could in principle result from faster responding compared to the other listener groups. However, the mean reaction times of the British English and American English listeners did not differ (688ms vs. 690ms), although the responses of Dutch listeners, unsurprisingly, were much slower (mean 952ms). Sub-analyses of the British English data revealed no difference between participants with slow versus fast mean response times; neither of these subsets showed either an effect of Context or an effect of Orthography in the expected direction.

General discussion

Intervocalic /r/ in British English sentences is interpreted differently by native versus non-native listeners. Both listeners with another language, and listeners with English as a native language but from another dialect area, produced response patterns in our *ice/rice* judgment task which differed significantly from those of the native baseline group. This suggests that both dialect and language differences lead to problems in the perception of the phonological process of /r/-insertion; in natural conversations, speakers of British English may therefore cause perceptual problems (in particular misapprehension of word boundary placement) for their non-native interlocutors.

The speaker whose utterances we analysed made consistent and substantial acoustic distinctions between the tokens of intended /r/, in word onset position, and the tokens of /r/ she intruded between vowels at a word boundary. This suggests that in practice the difference between the two types of /r/ should be easily accessible for listeners. In this respect, intrusive /r/ patterns similarly to other linking phenomena that have been tested in both speech production and perception. Consider the phenomenon of liaison in French, whereby segments that are not pronounced in citation-form utterances surface in running speech when the following word begins with a vowel; the liaison segments are also significantly shorter than the same segments when in an intended (word onset) pronunciation, and listeners can attend to the difference and use it where necessary to resolve ambiguity (e.g., the difference between *trop artisan* and *trop partisan*; Spinelli, Cutler, & McQueen, 2002; Spinelli, McQueen, & Cutler, 2003). Whether there is a separate difference in British English between linking /r/ and intrusive /r/ is a question for a more extended later study; but the difference between intrusive and onset /r/ is quite clear.

These clearly perceptible differences found in our acoustic measurements formed the basis of our perceptual study, and the expected result appeared with the British English native listeners: their responses were highly categorical and not affected by any other factors than the acoustic differences. Short /r/s were reported by these listeners as intrusive /r/s, and long /r/s as onset /r/s. No use was made of context. A small difference did surface on our orthography comparison; however, this effect was actually not in the direction indicative of an orthographic sensitivity, because these listeners gave more *rice* responses after *more* than after *saw*.

We suggest that this somewhat paradoxical finding is in fact to be explained in terms of the expected location of a prosodic boundary. In natural speech occurrences of sentences such as those used in our study, the beginning of the sentence complement, unmarked by a conjunction, is likely instead to be marked by a prosodic break, to prevent listeners from interpreting *ice/rice* as the object of the verb *saw*. This prosodic boundary would give rise to pre-boundary lengthening and initial strengthening of the onset phoneme following the boundary (Cho, McQueen, & Cox, 2007). After *saw*, but not after *more*, the critical word *(r)ice* in our materials occurred at the beginning of the complement. We hypothesize that the British English listeners might have interpreted a long /r/ after *saw* as a pre-boundary-lengthened intrusive /r/ rather than as an (insufficiently lengthened) onset /r/. Since the prosodic effects are not specific to British English, note that this explanation relies on the assumption that the native listening group's durational expectations are more refined than those of the non-native listeners.

Although the native listeners clearly showed a sensitive appreciation of the acoustic realisation of /r/, to the exclusion of our other manipulations, this was not at all the case for the two non-native listener groups. Both differed from the British listeners in a qualitatively similar way, but these differences were quantitatively larger in the Dutch group. First, both non-native groups produced shallower identification functions than the British group over the /r/ duration continuum; that is, although they too attended to the acoustic manipulation, their use of it was less sensitive than that of the native listeners. Second, both American and Dutch listeners were affected by the written form of the words they heard, and gave fewer onset /r/ responses when the /r/ sound could have been attributed to the letter *r* in *more*, which, as we saw, the native listeners did not do at all.

However, there were also qualitative differences between American and Dutch listeners. Dutch listeners used semantically biased context information, in that they gave more /r/ responses for sentences starting with *the social worker* than for sentences starting with *the little girl*. American listeners, in contrast, did not rely on such semantic factors. Additionally, Dutch listeners used durational information even less when the orthography suggested that the /r/ must have been an onset /r/, whereas the American listeners were influenced by orthography, but not to the extent that it weakened their use of durational information.

With respect to cross-language differences, our study suggests that unfamiliar phonological processes can cause difficulty for L2 listeners. As we noted in the introduction, Darcy et al. (2007) showed that advanced L2 learners are able to compensate for a type of assimilation that does not occur in their native language. Our results, however, showed that advanced learners of English did not attain native listener levels in perceiving intrusive /r/. Although, as already noted, British English is the target pronunciation taught in Dutch schools and universities (see e.g., Collins & Mees, 1999, p. vii; Gussenhoven & Broeders, 1997, pp. 16-17), and British English is widely available in the media in the Netherlands, the process of /r/-insertion is, along with all other casual speech phenomena, not explicitly taught, and the process is completely absent from Dutch. The reason why the advanced learners in Darcy et al.'s study were able to compensate for a type of assimilation that does not occur in their L1, and our advanced learners were not, might be that the process of assimilation is more natural than the phonological process of /r/-insertion as it occurs in many different languages, whereas /r/-insertion does not. Even the English beginning learners in Darcy et al.'s study show some compensation for voicing assimilation in French. Moreover, Mitterer, Csépe, Honbolygo and Blomert (2006) found that Dutch listeners compensate for a Hungarian assimilation rule; this shows that compensation for assimilation can even occur in the absence of experience with a particular assimilation rule. Though /r/-insertion in British English is widespread (Foulkes & Docherty, 1999, pp. 51, 76, 111, 133, 147, 174), some speakers variably suppress intrusive /r/ for socially motivated reasons (Broadbent, 1991; Gussenhoven & Broeders, 1997, p. 155). If the insertion of /r/ can be and is avoided in speaking, it is likely to be more difficult for non-native listeners to interpret an intrusive /r/ correctly when they hear one.

With respect to cross-dialect differences, our study suggests that phonological processes may cause some listening difficulty, but this is not of the same order as the difficulty experienced by the L2 listeners. In this our results are in line with those of Cutler et al. (2005), who found that when the speech input mismatches the native dialect slightly, the difficulty in perception is much less than that which arises when speech input completely mismatches the native language in terms of available phonemic categories. In our experiment, American listeners did not attain native listener levels in their perception of intrusive /r/. Nevertheless, they were able to use the durational cues in much the same way as the native listeners did, so that their responses certainly showed categorization, although the rhoticity in their native dialect made them susceptible to the influence of orthography.

Based on this experiment only, it cannot be concluded that the American listeners have significant difficulties dealing with intrusive /r/. Similar conclusions have arisen from other studies of cross-dialect perception. It has been repeatedly shown that different varieties of the same language encourage attention to different cues to the same contrast (e.g., Kirby, 2010; Miller & Grosjean, 1997); but the flexibility that speakers of the same language show in speech perception allows rapid readjustment given adequate exposure. Thus within regional dialects of British English, Evans and Iverson (2004) showed that listeners could alter their goodness rating for vowels according to the dialect (northern vs. southern) of the carrier phrase in which the tokens occurred; and within regional dialects of American English, Clopper and Bradlow (2006) found that perception of predictable sentences in noise was not adversely affected by a speaker/listener mismatch in regional dialect. Otake and Cutler (1999) found that suprasegmental structure differences across accent did not adversely affect accuracy of interpretation in minimal-pair decisions, although, as already mentioned, signal-detection measures showed that listeners with a mismatching dialect were both less sensitive to acoustic cues to the distinction in question, and more susceptible to bias. Further efficient learning was demonstrated by Floccia et al. (2006), who observed that adjustment to regional accent occurred during their experiments, and by Sumner and Samuel (2009), who found that General American speakers' difficulties in recognizing words produced with a New York accent could easily be removed by semantic priming. In Cutler et al. 's (2005) study, the identification of American-English vowels by listeners with another (but less widely spoken) dialect did not differ significantly from the performance of American listeners. Again, it is an empirical issue whether listeners from the more widely spoken variety would also perform well with the less widely spoken variety, or would require additional exposure.

With the phonological process of /r/-insertion, however, our results clearly show that listeners of another language perform differently from native listeners; in deciding whether they are hearing an onset /r/ or not, they are disposed to draw on semantic context and on orthographic information where available, and crucially, they rely less on the durational cues which are the true signals of the distinction in that they are what native listeners use. Listeners of a different dialect resemble the L2 listeners when it comes to reliance on orthographic information. But in all other respects—the use of semantic context and, most importantly, the central reliance on durational cues—the cross-dialect listeners are more native-like. Clearly, native listeners keep *Laura Norder* in her place; we now know that listeners with another dialect also know where *Laura* should go, and only non-native listeners are not so sure what to do with her.

References

Baayen, H. (2008). *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge: Cambridge University Press.

Best, C. T., McRoberts, G. W., & Goodell, E. (2001). Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system. *Journal of the Acoustical Society of America, 109*(2), 775-794.

Bradlow, A. R., & Alexander, J. A. (2007). Semantic and phonetic enhancements for speech-in-noise recognition by native and non-native listeners. *Journal of the Acoustical Society of America, 121*, 2339-2349.

Bradlow, A. R., Pisoni, D. B., Akahane-Yamada, R., & Tohkura, Y. i. (1997). Training Japanese listeners to identify English /r/ and /l/: IV. Some effects of perceptual learning on speech production. *Journal of the Acoustical Society of America, 101*(4), 2299-2310.

Broadbent, J. (1991). Linking and intrusive r in English. *UCL Working Papers in Linguistics, 3*, 281-302.

Cho, T., McQueen, J. M., & Cox, E. A. (2007). Prosodically driven phonetic detail in speech processing: The case of domain-initial strengthening in English. *Journal of Phonetics, 35*, 210-243.

Clopper, C. G., & Bradlow, A. R. (2006). Effects of dialect variation on speech intelligibility in noise. *Journal of the Acoustical Society of America, 119*(5), 3424.

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*, 155-159.

Collins, B., & Mees, I. M. (1999). *The phonetics of English and Dutch* (4th ed.). Leiden: Brill.

Connine, C. M., Ranbom, L. J., & Patterson, D. J. (2008). Processing variant forms in spoken word recognition: The role of variant frequency. *Perception and Psychophysics, 70*(3), 403-411.

Cruttenden, A., & Gimson, A. C. (1994). *Gimson's pronunciation of English* (5th ed.). London: Arnold.

Cutler, A., Smits, R., & Cooper, N. (2005). Vowel perception: Effects of non-native language vs. non-native dialect. *Speech Communication, 47*, 32-42.

Cutler, A., Weber, A., & Otake, T. (2006). Asymmetric mapping from phonetic to lexical representations in second-language listening. *Journal of Phonetics, 34*, 269-284.

Darcy, I., Peperkamp, S., & Dupoux, E. (2007). Bilinguals play by the rules: Perceptual compensation for assimilation in late L2-learners. In J. Cole & J. Hualde (Eds.), *Papers in Laboratory Phonology 9* (pp. 411-442). Berlin: Mouton de Gruyter.

Dutoit, T., Pagel, V., Pierret, N., Bataille, F., & Vreken, O. v. d. (1996). *The MBROLA Project: Towards a Set of High-Quality Speech Synthesizers Free of Use for Non-Commercial Purposes.* Paper presented at the 4th International Conference on Spoken Language Processing (ICSLP 1996), Philadelphia.

Escudero, P., Hayes-Harb, R., & Mitterer, H. (2008). Novel second -language words and asymmetric lexical access. *Journal of Phonetics, 36*, 345-360.

Evans, B. G., & Iverson, P. (2004). Vowel normalization for accent: An investigation of best exemplar locations in northern and southern British English sentences. *Journal of the Acoustical Society of America, 115*, 352-361.

Floccia, C., Goslin, J., Girard, F., & Konopczynski, G. (2006). Does a regional accent perturb speech processing? *Journal of Experimental Psychology: Human Perception and Performance, 32*, 1276-1293.

Foulkes, P., & Docherty, G. (Eds.). (1999). *Urban Voices: Accent studies in the British Isles*. London: Arnold.

Giegerich, H. J. (1992). *English phonology: An introduction*. Cambridge, UK: Cambridge University Press.

Guion, S. G., Flege, J. E., Akahane-Yamada, R., & Pruitt, J. C. (2000). An investigation of current models of second language speech perception: The case of Japanese adults' perception of English consonants. *Journal of the Acoustical Society of America, 107*(5), 2711-2724.

Gussenhoven, C., & Broeders, A. (1997). *English pronunciation for student teachers*. Groningen: Wolters-Noordhoff.

Ingram, J. C. L., & Park, S.-G. (1998). Language, context, and speaker effects in the identification and discrimination of English /r/ and /l/ by Japanese and Korean listeners. *Journal of the Acoustical Society of America, 103*(2), 1161-1174.

Irwin, P., & Nagy, N. (2007). Bostonians /r/ speaking: A quantitative look at (R) in Boston. *Penn Working Papers in Linguistics, 13*, 135-148.

Kirby, J. (2010). Dialect experience in Vietnamese tone perception. *Journal of the Acoustical Society of America, 127*, 3749-3757.

Miller, J. L., & Grosjean, F. (1997). Dialect effects in vowel perception: The role of temporal information in French. *Language and Speech, 40*, 277-288.

Mitterer, H., Csépe, V., & Blomert, L. (2006). The role of perceptual integration in the perception of assimilation word forms. *Quarterly Journal of Experimental Psychology, 59*, 1395-1424.

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.

Mitterer, H., & Ernestus, M. (2006). Listeners recover /t/s that speakers reduce: Evidence from /t/-lenition in Dutch. *Journal of Phonetics, 34*, 73-103.

Otake, T., & Cutler, A. (1999). Perception of suprasegmental structure in a non-native dialect. *Journal of Phonetics, 27*, 229-253.

Pitt, M. A., Dilley, L., Johnson, K., Kiesling, S., Raymond, W., Hume, E., et al. (2007). Buckeye Corpus of Conversational Speech (2nd release) [www.buckeyecorpus.osu.edu]. Columbus, OH: Department of Psychology, Ohio State University (Distributor).

Polka, L. (1995). Linguistic influences in adult perception of non-native vowel contrasts. *Journal of the Acoustical Society of America, 97*(2), 1286-1296.

Scott, D. R., & Cutler, A. (1984). Segmental phonology and the perception of syntactic structure. *Journal of Verbal Learning and Verbal Behavior, 23*, 450-466.

Spinelli, E., Cutler, A., & McQueen, J. M. (2002). Resolution of liaison for lexical access in French. *Revue Française de Linguistique Appliquée, 7*, 83-96.

Spinelli, E., McQueen, J. M., & Cutler, A. (2003). Processing resyllabified words in French. *Journal of Memory and Language, 48*, 233-254.

Sumner, M., & Samuel, A. G. (2009). The effect of experience on the perception and representation of dialect variants. *Journal of Memory and Language, 60*, 487-501.

Trudgill, P. (1986). *Dialects in contact*. Oxford: Basil Blackwell.

Underbakke, M., Polka, L., Gottfried, T. L., & Strange, W. (1988). Trading relations in the perception of /r/-/l/ by Japanese learners of English. *Journal of the Acoustical Society of America, 84*(1), 90-100.

van Bezooijen, R. (2005). Approximant /r/ in Dutch: Routes and feelings. *Speech Communication, 47*, 15-31.

van de Velde, H. (1996). Variatie en verandering in het gesproken Standaard-Nederlands (1935-1993) [Variation and change in spoken standard Dutch (1935-1993)]: Dissertation Nijmegen University.

Wells, J. C. (1982). *Accents of English*. Cambridge: Cambridge University Press.

Resolving ambiguity in familiar and unfamiliar casual speech

Chapter 4

Tuinman, A., Mitterer, H., & Cutler, A. (under revision). Resolving ambiguity in familiar and unfamiliar casual speech. *Journal of Memory and Language*.

Abstract

In British English, the utterance "Canada aided" can sound like "Canada raided" if the speaker links the two vowels at the word boundary with an intrusive /r/. There are, however, subtle phonetic differences between "Canada raided" with onset /r/ versus intrusive /r/. Three cross-modal priming experiments tested whether these differences suffice to resolve the ambiguity for three listener groups: Native British listeners, American English listeners, and advanced Dutch learners of English. British and American listeners exploited the phonetic cues very effectively, but Dutch listeners did not. An additional eye-tracking study showed that the latter group tended to interpret any /r/ as an onset /r/. Thus the potential ambiguity arising from intrusive /r/ can be resolved by efficient extraction of information from the speech signal. Native experience with the language enables this, but listeners with a different native language, lacking the intrusive-/r/ process, are likely to misinterpret the intended utterance.

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 *eyesore*, since all these words have an underlying /r/ sound which, in any version of English, can be pronounced before a vowel. But given that none of those interpretations makes sense in combination with *a film today*, most listeners should quickly realize that Lennon must have meant *I saw*. This type of re-computation is one with which British English listeners should be highly familiar.

British English dialects in general, and certainly the standard forms such as Received Pronunciation, are non-rhotic, i.e., they have the phonological constraint that /r/ can occur word-initially but not word-finally. However, word-final /r/ can appear under two circumstances (Cruttenden & Gimson, 1994; Giegerich, 1992). First, a 'linking /r/' surfaces when a word ending with an underlying /r/ precedes a word beginning with a vowel (e.g., *soar up into the sky*). Second, there is 'intrusive /r/', whereby /r/ may be inserted after a non-high vowel ([ə, a, ɔ] and diphthongs ending in [ə]) and before a vowel-initial word, as in *I saw a film today*. A linking /r/ is thus represented in the spelling, but 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.

Phonological processes such as this have over recent years received widespread attention in research on the perception of connected speech, and a summary of the accumulated findings is that native listeners are certainly able to derive correct interpretations of speech in spite of the alterations which the processes cause (e.g., Connine, Ranbom, & Patterson, 2008; Mitterer, Csépe, & Blomert, 2006; Mitterer & Ernestus, 2006). Little attention, however, has been paid to the issue of how listeners cope with connected speech processes that they are not familiar with. Second language (L2) listeners, for example, may have no comparable process in their own native language (L1). Likewise, dialects of the same language notoriously differ in phonological constraints and processes, so that listeners with one dialect may be quite unfamiliar with a given process in another dialect. How difficult, then, is it to hear and interpret *I saw a* correctly if one is unfamiliar with the process of intrusive /r/?

Connected speech processes may, in fact, quite significantly complicate speech perception for L2 listeners. L2 listeners already face difficulties due to mismatches in the segmental inventory of the L1 and L2, a smaller vocabulary, and lesser familiarity with language structure at all levels. Spoken-word recognition involves, for any listener, multiple activation of word candidates, with the simultaneously activated word forms competing for recognition (see, e.g., McQueen, 2007, for a review). Listening to an L2 actually involves more such activation and competition between word forms than listening to the L1, and often words are activated that were not intended by the speaker. Broersma and Cutler (2008, 2011), for example, showed that L2 listening can involve the phantom activation of words which are not actually in the input. In English auditory lexical decision studies, native speakers of Dutch accepted spoken nonwords such as *lemp*, *dask*, *groof* and *flide* as real English words. Unsurprisingly, native English listeners did not do this. As Dutch has only one vowel where English has the two which contrast in words such as *had* versus *head*, and further has no contrast between voiced and voiceless sounds in word-final position, the Dutch listeners presumably mistook these nonwords for the words *lamp*, *desk*, *groove* and *flight* respectively. Similarly, Cutler, Weber, and Otake (2006) showed that Japanese listeners instructed in an eye-tracking study to click on a picture of a *rocket* experienced interference when a picture of a *locker* was present; they tended to look at the *locker* before eventually focusing on the *rocket*. These experiments show that L2 listeners have difficulties with the perception of words containing phonemic contrasts they are unfamiliar with from their L1.

Listening to a different dialect than that of the L1 is not necessarily much easier than listening to another language. Floccia, Goslin, Girard and Konopczynski (2006), for example, found that regional accent normalization is typified by an initial temporary processing cost. They reasoned that this is because the short-term adjustment mechanism develops only after a certain amount of accented speech has been processed. Otake and Cutler (1999) found that cross-dialect perception of Japanese words involved less attention to the information in the speech signal and more reliance on knowledge of lexical patterns. Sumner and Samuel (2009) found that speakers of a General American dialect can have problems in recognizing words produced with a New York accent, although experience with a dialect influences the ability of listeners to process dialect variants. Scott and Cutler (1984) similarly found evidence that comprehension difficulties decrease as familiarity with a speaker of a different dialect of the native language increases, in their case for British English listeners processing American English medial /t/ in words like *liter*, which, produced with flapping, is virtually homophonous with *leader*. Scott and Cutler tested a group of British English listeners who had lived in England throughout their entire lives, and another group who had been living in the United States for several years; the latter group had less difficulty processing flapped /t/s than the former.

Both listeners with a different language and listeners with a different dialect are thus at risk of misinterpreting speech. Running speech requires listeners to deal with ambiguities due to phonological processes. For instance, the phrase *once paid* can sound very like *one spade*, and *two lips* can sound like *tulips*. Native listeners efficiently exploit phonetic detail to disambiguate such phrases (Gow, 2002; Kemps, 2004; Mitterer & McQueen, 2009; Quené, 1992; Shatzman & McQueen, 2006). Using cross-modal priming, Gow and Gordon (1995) examined recognition of ambiguous sequences that could be interpreted as two words or as one longer word (e.g., *two lips*/*tulips*). They found priming to the second word (e.g., *lips*) when it was preceded by the two-word sequence (*two lips*), but not when it was preceded by the one-word sequence (*tulips*). It appeared that word-initial phonemes (e.g., the /l/ in *two lips*) were longer than corresponding non-initial phonemes (e.g., the /l/ in *tulips*). Gow and Gordon concluded that listeners used the difference in duration as cues for word segmentation and lexical access. In an eye-tracking study of Shatzman and McQueen (2006), Dutch participants listened to Dutch sentences in which a stop-initial target word (e.g., *pot* 'jar') was preceded by an /s/, causing ambiguity as to whether the word *pot* or *spot* 'spotlight' was the speaker's intention. The participants took longer to fixate on the target picture *pot* when the sentence was manipulated such that the target and preceding /s/ were spliced from a recording of *spot* than when the same sequence was spliced from another recording of the sentence with *pot*. Acoustic measurements showed that the duration of /s/ differed in the two source sentences, and this correlated with the perceptual effect; a longer /s/ was interpreted as a word-initial /s/.

Spinelli, McQueen, and Cutler (2003) investigated the phonological process of liaison in French, a linking process similar to the one described above for British English; in French, many word-final underlying consonants are pronounced only when a following word begins with a vowel. Again, consonant duration proved to be important in disambiguation, for instance in distinguishing between *dernier oignon* ('last onion') and *dernier rognon* ('last kidney'). Spinelli et al. found the onset /r/ in *dernier rognon* to be longer than the liaison /r/ in *dernier oignon*, and in a perception study, visually presented target words (e.g., ROGNON) were recognized most rapidly when a preceding auditory prime had been pronounced with the intention of expressing the same word (e.g., *dernier rognon*).

The British English process of intrusive /r/ is similar to French liaison, in that it too can create ambiguous sequences. Thus *extra ice* with /r/ inserted between *extra* and *ice* can strongly resemble *extra rice*. In contrast to French liaison, though, the English intrusive /r/ is optional and can be avoided by the speaker without sounding completely unnatural—even though the use of intrusive /r/ is widespread, it is often avoided in formal speech (Collins & Mees, 1999; Giegerich, 1992). In French, failing to implement liaison is perceived as wrong. Note that the fact that intrusive /r/ in English is optional might make its correct perception more difficult than perception of the obligatory liaison in French. Especially non-native listeners might find it difficult not to be able to rely on complete regularity of where and when the intrusive /r/ might appear.

However, Darcy, Peperkamp and Dupoux (2007) provided evidence that the interpretation of speech containing phonological processes is not always problematic for L2 listeners. They examined how native French listeners and English learners of French (beginning or advanced) deal with voicing assimilation in French (such when the second phoneme of *absurde* 'absurd' becomes /p/, to match the voicelessness of the following /s/). Crucially, such consonant-to-consonant voicing assimilation does not occur in English, although place assimilation does occur, so that the L2 listeners were not unfamiliar with assimilation in general, but did not know this particular process from their L1. In word detection, the advanced learners were able to compensate for the assimilation just as well as the native listeners. The beginning learners also showed some compensation for assimilation, although this is possibly due to general auditory processes which contribute to compensation for assimilation (Mitterer, Csépe, & Blomert, 2006).

More strongly even, sometimes language experience is not at all necessary for listeners to compensate for a casual speech process. Gow and Im (2004) tested perception, by native and non-speakers of Hungarian and Korean, of language-specific assimilation phenomena in these two languages. Their participants monitored for segments in assimilated and non-assimilated control contexts. Gow and Im found that 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. These findings suggest that experience with a language-specific assimilation process is not necessary to compensate for assimilation. Mitterer et al. (2006) likewise showed that language experience has only a minor role to play in perceptual compensation for assimilation. In an identification and a discrimination task, they presented native speakers of Hungarian and Dutch non-speakers of Hungarian with Hungarian words and nonwords in which a viable and an unviable liquid assimilation was applied. The results of their study suggested that viably changed forms were difficult to distinguish from canonical forms irrespective of whether listeners had had experience with the assimilation rule applied in the utterances. All of these findings allow the possibility that the kind of connected speech processes we are concerned with may actually be equally easy to process by all listeners, whether the speech is in their L1, or a different dialect of their L1, or in an L2.

In a phonetic study of intrusive /r/ (Tuinman, Mitterer, & Cutler, under revision), we found that 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 that these differences are strong enough to be of use to listeners. We then investigated whether native and non-native listeners can exploit the phonetic differences in phoneme identification, comparing Dutch listeners proficient in English and American English listeners with British English listeners. Like most American dialects, the Dutch language is rhotic, and does not display the British pattern of intrusive /r/. Thus just like American English speakers, Dutch speakers produce /r/s word-finally, and in Dutch /r/ is never inserted in intervocalic contexts equivalent to *saw a* (Collins & Mees, 1999; Gussenhoven & Broeders, 1997).

In a 2AFC categorization experiment, these three listener groups heard 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 indeed 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. Both the American and Dutch groups relied less on durational cues than the British listeners, however, and both were affected by orthography, reporting /r/ more often after *saw* than after *more*. In addition, Dutch listeners were susceptible to semantic bias in the sentences. Thus intrusive /r/ does indeed cause phoneme perception problems for listeners with another language or another dialect (with the difficulties being greater in the case of a language difference). 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 dialect or a non-native language make much less use of the available phonetic cues.

So far, then, there is evidence that listeners can have word recognition difficulties in non-native language and in non-native dialect, but there is also evidence that some connected speech processes are very easy for all listeners to deal with. In the case of the process of /r/-intrusion, however, our phonetic investigations have shown that only British English listeners deal with it by relying exclusively on the acoustic-phonetic characteristics. If such a way of categorizing /r/ is the only way to avoid word recognition problems, then we may predict that British English intrusive /r/ will induce lexical ambiguity for both Dutch and American English listeners.

In the present study, we test this by extending our investigations of /r/-insertion to word recognition in connected speech. We use a cross-modal priming task with natural speech material, again presented to British, American 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: British English listeners

Method

Participants

Seventy-two native speakers of British English participated in the experiment. The participants were recruited from the participant pool of the Laboratory of Experimental Psychology of the University of Sussex. None reported any hearing loss. All were volunteers and received a small fee or course credits for participation.

Stimuli

Twenty-seven pairs of English sentences were constructed. An example sentence is "*My 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. Trivially, the r-initial member of the pair in the sentence will trigger pronunciation of an /r/. More importantly, 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 in the pronunciation. All sentences are listed in the Appendix.

The sentences were recorded by a female native speaker of British English, from London, who was unaware of the purpose of the study and normally produces intrusive /r/s in casual speech. All sentences were recorded at least twice without disfluencies. Measurements showed that the /r/ was always longer and had a larger intensity decrease in the sentences with onset /r/ than in similar sentences with intrusive /r/ (cf. Cruttenden & Gimson, 1994; Tuinman, et al., under revision). In each experimental sentence, an /r/-like sound could be interpreted as either an intrusive /r/ or an onset /r/. Thus, the experimental sentences are potentially ambiguous for listeners who do not exploit the phonetic difference between intrusive /r/ and onset /r/. The prime materials for the experiment were created by truncating the sentences directly after the (potential) prime word (e.g., *My brother likes extra ice* -).

Besides the experimental sentences, 27 control sentences were constructed with target words matched for English frequency to those in the experimental sentences (e.g., *My brother likes extra pages*...), as well as 108 filler prime sentences with 54 unrelated-word and 54 nonword targets. Further, we constructed 27 "semi-experimental" prime sentences with nonword targets starting with either /r/ or a vowel. Nine of these sentences contained a potential linking /r/ (e.g., *I think that you*r *explanation*...) and a partially overlapping visual nonword target (*explint* or *rexplint*). Another nine sentences contained geminate /r/ (e.g., *And then my neighbou*r r*efused*...) and their nonword targets were partially overlapping (e.g., *effint* or *reffint*). These 18 sentences were included to prevent an association of phonological relatedness between prime and target with "yes" responses in the visual lexical decision task. Nine further sentences without /r/ near the place of truncation (e.g., *I heard on the news that taxes*...) and with unrelated nonword targets (e.g., *ilems* or *rilems*) were also included.

Finally, 18 question trials were constructed. In these trials, participants would hear a complete sentence followed by a yes/no question. These trials served as a check on whether 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 and Procedure

The target words for visual presentation were the r-initial words (e.g., *rice*) in version 1 of the experiment, and the vowel-initial words (e.g., *ice*) in version 2. This separation was chosen to avoid calling attention to the potential ambiguity. For the semi-experimental sentences, the nonword targets with /r/ were used in version 1 and the targets with a vowel in version 2.

Thus the experimental design has two main factors; Prime type (within-subjects: r-prime, vowel-prime, control) and Target type (between-subjects: r-initial, vowel-initial target words).

Each participant was presented with each experimental cross-modally primed visual target only once, with nine targets in each of the three prime conditions. In version 1 of the experiment, where the targets were the r-initial words (e.g., *rice*), r-primes matched the target and vowel primes mismatched. In version 2, where the target words were vowel-initial (e.g., *ice*), the vowel primes matched and the r-primes mismatched. In the control condition in both versions, targets were preceded by a phonologically and semantically unrelated prime.

Participants were tested one at a time in a sound-attenuated booth. The participants received English instructions on a computer screen, informing them that on each trial they would hear a portion of an English sentence, directly after which an English word or nonword would appear on the screen. They were instructed to press a green response button labeled "yes" with their dominant hand if they thought the visually presented item was an English word, and a red response button labeled "no" with their other hand if they thought the visually presented item was not an English word. Participants were asked to try to respond as rapidly as possible without making many errors. The experiment started with seven practice trials and one practice question trial.

Each participant was presented with all filler word and filler nonword trials, for a total of 81 word and 81 nonword decisions. Additionally, each participant received the 18 yes/no questions that they could only answer if they had paid attention to the previous auditory sentence. The 27 experimental trials thus formed 15% of all trials (excluding practice). Item and question trials were randomized for each participant. After every 50 trials the participants could take a short break.

Results

Lexical decision reaction times (RTs) were measured from onset of the visual presentation of the target words. Responses slower than 1500 ms were excluded; together with errors this resulted in 12.1% missing data. Figure 1 (top panel) shows the priming effects in Experiment 1.

For the statistical analysis, the RTs were log-transformed to reduce the skew of the RT distribution. The log RTs were then subjected to a linear mixed-effects model analysis with Target type (r-initial, vowel-initial targets), Prime type (r-prime, vowel-prime, control), the interaction between Target type and Prime type, and Trial number as fixed factors, and Participants and Target word as random factors. The interaction between Target type and Prime type proved to be significant [$F(2, 1580) = 20.72$, $p < .001$], so the two conditions of Target type were analyzed separately in analyses with the same random factors and only trial and Prime type as fixed factors.

In linear mixed-effects modelling, factors such as Prime type (with in this case three levels) are analyzed with one level mapped on the intercept and a regression weight calculated for all other levels, indicating the difference from the intercept condition. We mapped the control condition on the intercept to get regression coefficients which showed the difference of the matching and mismatching prime conditions from the control condition. For the r-initial targets, there was a significant main effect of Prime type [$F(2, 796) = 14.33$, $p < .001$]. RTs were 67 ms shorter in the r-prime (match) condition than in the control condition [$t(796) = -4.94$, $p < .001$]. The 10 ms difference between the vowel-prime (mismatch) and control conditions was not significant [$t(796) = -.70$, $p > .1$].

For the vowel-initial targets, there also was a significant main effect of Prime type [$F(2, 783) = 11.7$, $p < .001$]. RTs were 47 ms shorter in the vowel-prime condition (match) than in the control condition [$t(783) = -3.48$, $p < .01$]. The 29 ms difference between the r-prime (mismatch) and control conditions was not significant [$t(783) = 1.20$, $p < 1$].

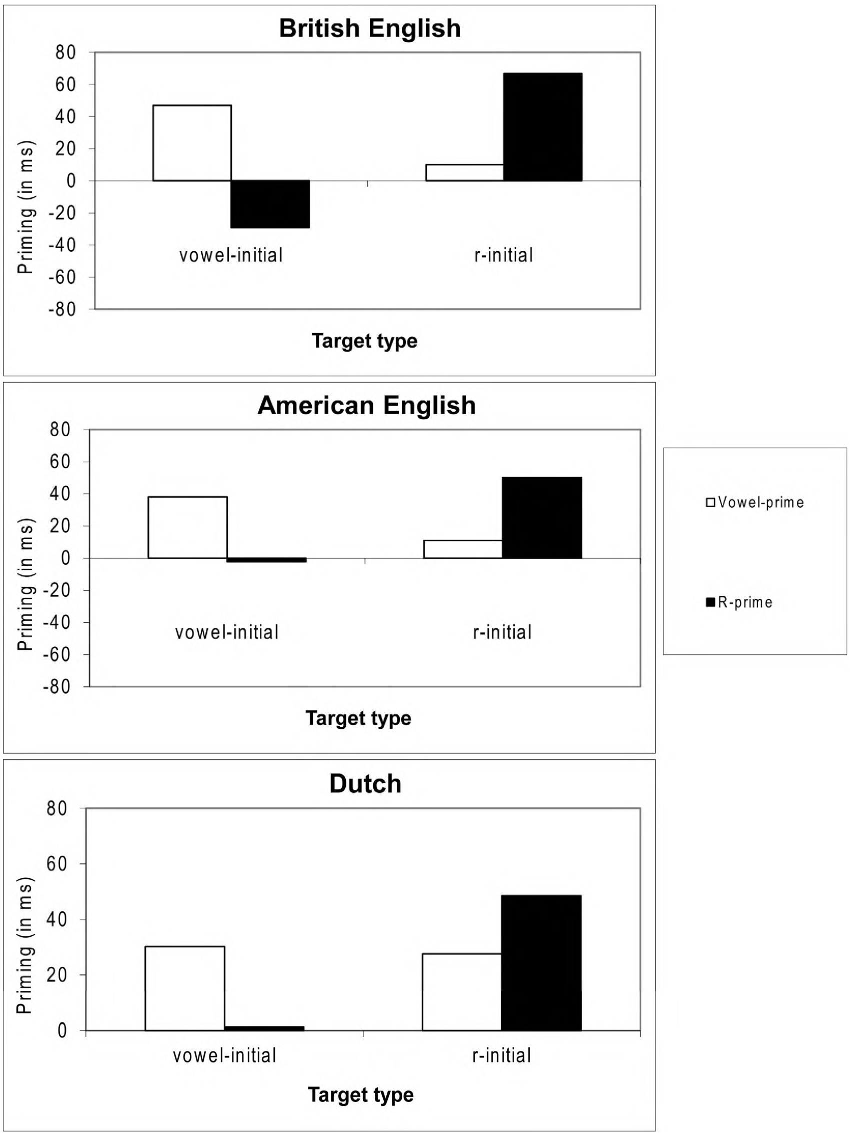

Figure 1: *Priming effects (mean RT in control-prime condition minus mean RT in vowel-prime and r-prime conditions), separately for vowel-initial and r-initial targets, for British English listeners (Experiment 1, top panel) American English listeners (Experiment 2, middle panel) and Dutch listeners (Experiment 3, bottom panel).*

Discussion

The native listeners perform as expected: hearing a vowel-initial word preceded by an intrusive /r/ primes vowel-initial target words, but not r-initial target words. Conversely, hearing an r-initial word primes r-initial words, but not vowel-initial words. These results replicate and extend findings by Spinelli et al. (2003) who also found that native listeners are able to use subtle phonetic cues to resolve apparent ambiguities caused by a phonological process, in that case liaison in French. In fact the effect that we find is much stronger than that found by Spinelli et al., in that the native listeners in our experiment only show a significant priming effect for the intended phrase and no hint of a priming effect when there is a subphonemic mismatch between the prime and the target (see Figure 1). This is as predicted from the results of our phonetic categorization study, which showed that British English listeners based their interpretation of /r/ as onset or intrusive exclusively on the acoustic-phonetic evidence.

In that categorization study, American English listeners were able to make use of durational cues to a certain extent, but did not show completely native-like performance in their categorization. In our next cross-modal priming experiment, we examine whether American English listeners can succeed in distinguishing between intrusive and onset /r/ for the purposes of word recognition.

Experiment 2: American English listeners

Method

Participants

Seventy-two native speakers of American English, recruited from the participant pool of the Institute for Research in Cognitive Science of the University of Pennsylvania, took part in Experiment 2. None reported any hearing impairment. All were volunteers and were paid a small fee for their participation.

Stimuli, design and procedure

The stimuli, design and procedure in Experiment 2 were identical to those in Experiment 1.

Results

Again, lexical decision reaction times (RTs) were measured from onset of the visual presentation of the target words, and responses slower than 1500 ms again excluded (for a total of 13.5% missing data). Figure 1 (middle panel) shows the priming effects in Experiment 2.

The log RTs were analyzed as in Experiment 1. Again the interaction of Target type and Prime type was significant [$F(2, 1552) = 6.97$, $p < .01$] and thus the two conditions of Target type were again analyzed separately. For r-initial targets there was a significant main effect of Prime type [$F(2, 793) = 5.99$, $p < .01$]. RTs were 50 ms shorter after r-primes (match) than after control primes [$t(794) = -3.29$, $p < .01$]. The 11 ms difference between the vowel-prime (mismatch) and control conditions was not significant [$t(794) = -.71$, $p > .1$]. For vowel-initial targets too, the main effect of Prime type was significant [$F(2, 757) = 4.39$, $p < .05$]. RTs were 38 ms shorter after vowel-primes (match) than control primes [$t(759) = -2.50$, $p < .05$], but the 2 ms difference between the r-prime (mismatch) and control conditions was not significant [$t(756) = .11$, $p > .1$].

We also conducted a cross-experiment comparison between the British and American English participants. This analysis revealed no effects of language group: neither the main effect of language group nor its interaction with any of the within-subjects variables was significant.

Discussion

In the cross-modal priming results, the American English listeners pattern like the British English listeners: they show a priming effect for vowel-initial targets (e.g., *ice*) preceded by matching sentences with an intrusive /r/, but show no priming when the same target words are preceded by mismatching sentences with an onset /r/. Conversely, they show a priming effect for r-initial target words (e.g., *rice*) after a matching sentence with an onset /r/ but no priming effect when they heard a mismatching sentence with an intrusive /r/.

Thus the relatively good but not perfect performance of the American English listeners in our previous categorization experiment may underestimate the ability of these non-native dialect listeners to distinguish intrusive from onset /r/s. We return to this issue in the general discussion.

The non-native language listeners in the categorization study were further from the native baseline. Nevertheless, the word recognition effects of intrusive /r/ appear not to be simply predictable from phonetic categorization results. In our third cross-modal priming study, we therefore test Dutch listeners to English in the same task, to establish the effects of the intrusive /r/ process on their recognition of words in British English.

Experiment 3: Dutch listeners

Method

Participants

Eighty-four[1] native speakers of Dutch participated in Experiment 3. The Dutch participants had a high level of proficiency in English as a second language. Note that the target variety of English taught in Dutch schools is British (Gussenhoven & Broeders, 1997; Collins & Mees, 1999). The participants were recruited from the Max Planck Institute participant pool; none reported any hearing loss, and all were volunteers and received a small fee for participation.

Stimuli, design and procedure

These were again as in Experiments 1 and 2.

Results

Again, lexical decision reaction times (RTs) were measured from onset of the visual presentation of the target words, and responses slower than 1500 ms excluded (for in total 24.4% missing data). Figure 1 (bottom panel) displays the priming effects in Experiment 3.

Log RTs were again analyzed in the same way as in the two preceding experiments, and again an interaction between Target type and Prime type [$F(2, 1575) = 3.40$, $p < .05$] led to the two conditions of Target type being analyzed separately. For the r-initial targets, there was a significant main effect of Prime type [$F(2, 797) =$

[1] More Dutch listeners were tested to produce a similar number of data points as for the British English and American English listeners.

4.66, $p < .05$]. RTs were 48 ms shorter in the r-prime condition (match) than in the control condition [$t(796) = -3.05$, $p < .01$]. The RT comparison between the vowel-prime (mismatch) and control conditions showed an effect in the same direction, but this difference of 28 ms did not achieve significance [$t(796) = -1.71$, $p = .088$]. For the vowel-initial targets, there was no significant main effect of Prime type [$F(2, 776) = 2.63$, $p = .073$].

Another cross-experiment comparison compared the Dutch to the British participants. This analysis revealed clear effects of language group. The Dutch participants were not only overall slower [$F(1, 3102) = 117.3$, $p < 0.001$], but the significantly different patterns of priming that Figure 1 shows resulted in a three-way interaction of Target Type, Prime Type, and Language Group [$F(2, 3524) = 15.7$, $p < 0.001$].

Discussion

The fact that the prime type (onset /r/ vs. intrusive /r/) influenced the patterns of priming, as is evident in the Prime type by Target type interaction, shows that the Dutch listeners are not completely insensitive to the pronunciation differences. However, they are less sensitive than the British and American English listeners. For r-initial targets such as *raided*, the matching prime (*Canada raided*) produces the expected significant priming, but the mismatching intrusive-/r/ prime (*Canada aided*) also produces a facilitation of RTs that was not observed in the other data sets. Moreover, the Dutch listeners show no significant priming effect for vowel-initial targets (e.g., *aided*) after hearing a matching vowel-initial word preceded by intrusive /r/ (*Canada aided*), although both English-native groups did. Apparently, the non-native language listeners consider the intrusive /r/ an equivalent match to the r-initial targets and the vowel-initial targets, but a really good match to neither (see Figure 1).

These results are in line with our findings from phonetic categorization: Dutch listeners make almost no use of durational information to categorize /r/, but by preference base their decision on orthographic information—giving more *rice* responses after *saw* than after *more*. Lack of familiarity with the phenomenon of intrusive /r/ and with the cues that can differentiate it from onset /r/ is presumably the basis of this response pattern.

However, although the priming results show that in recognizing words, just as in categorizing phonemes, the Dutch listeners fail to use the evidence concerning intrusive /r/ in a native-like manner, we still do not have a complete picture of their processing. For that the RT patterns were insufficiently conclusive. It is possible, though, 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, & Sedivy, 1995). This we therefore did in our next study.

Most visual world studies measured looks to pictures (Tanenhaus & Spivey-Knowlton, 1996), and the target words used in our priming experiments do not lend themselves easily to pictorial representation; but recently, a number of studies have successfully made use of this paradigm with printed-word materials. McQueen and Viebahn (2007), for instance, examined the influence of mismatching phonetic evidence on spoken-word recognition in a variant of a study by Allopenna, Magnuson and Tanenhaus (1998), but with printed words instead of pictures. Allopenna et al. found that participants looked more at pictures of offset-mismatch competitors (e.g., *beetle*, given the target *beaker*) than of onset-mismatch competitors (e.g., *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).

Given that eye tracking with printed word targets is able to show subtle effects in speech perception, we could use it to gain more information on how the Dutch listeners perceive intrusive /r/. In fact, this eye-tracking method is more powerful than cross-modal priming, as both target types (r-initial and vowel-initial words) are tested on a single trial and listeners are asked to choose between them.

Experiment 4: Eye tracking with Dutch listeners

Method

Participants

Twenty-four native speakers of Dutch meeting the description given in Experiment 3 took part. None had participated in the previous experiment.

Stimuli

Participants watched a screen with four printed words while hearing an auditory sentence. The critical experimental materials were 54 auditory sentences, all of which contained a member of a minimal pair such as *ice/rice*—the same 27 pairs as were used in the previous experiments. For each word pair, there was thus a pronunciation with an onset /r/ (e.g., *rice*) and a pronunciation with an intrusive /r/, where the speaker intended the vowel-initial word (e.g., *ice*). These were in fact the same recordings as used in Experiments 1-3, with the only difference being that the sentences had not been truncated. Some sentences were disambiguated after the r-initial and vowel-initial words by the following context, while others remained ambiguous. Additionally, 54 filler sentences (again from Experiments 1-3) were used, with filler words inserted in the slot of the r-initial or vowel-initial word (e.g., *pages*, *space*).

For the visual stimuli, 27 quadruplets of words were presented four times each. The words were centered respectively on 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.

Apparatus and procedure

Participants were tested individually in a quiet room. They were seated in front of the computer screen at a comfortable reading distance. The head-mounted eye tracker was mounted and calibrated. Eye movements were monitored using a SMI EyeLink eye-tracking system, sampling at 250Hz. The experiment was controlled by a Compaq 486 computer. Spoken sentences were presented over headphones using NESU software. 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 instructed to click on one of the four words even when they were not sure which word they heard, and were told that they did not have to wait till the end of the sentence to click. After the mouse click the next trial was initiated. 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 54 experimental and 54 filler items. In each list, 27 experimental sentences had an r-initial target word, 27 experimental sentences had a vowel-initial target word and in 54 filler trials these target words were used as distractors. Over the 108 trials per participant, the target appeared 27 times on each of the four screen positions. With 27 items per condition, it was not possible to exactly counterbalance target position within condition for each participant. However, over participants, target positions were counterbalanced within each condition. Four additional trials served as practice trials.

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. Looks were measured as follows: For both target and competitor, a fixation was counted as "on" the word if it fell within 150 pixels of the center of the word in the horizontal direction and within 100 pixels in the vertical direction. The fixation proportion for a given time window was the proportion of time points in that window for which the fixation was on a given printed word. Fixation proportions were then logistically transformed (Dixon, 2008), and the difference between proportions for the r-initial word and the vowel-initial word determined.

Results

Figure 2 shows the fixation proportions over time and the distribution of clicks for intrusive /r/ trials (vowel target: left panels) versus onset /r/ trials (r target: right panels). It can be seen that neither looks nor clicks were differentiated by the target; instead, participants always had a strong preference for r-initial words, even when hearing an intrusive /r/.

Analysis of the proportion of correct clicks revealed this clearly. When the auditory sentence contained an intended onset /r/, participants clicked correctly on the r-initial word in 96.5% of cases. When the auditory sentence contained an intrusive /r/, i.e., the vowel-initial word was intended, participants still clicked erroneously on the r-initial word in 70.9% of the cases. Analyzed with a linear mixed-effects model with subject and item as random factors and type of /r/ as a fixed factor, this difference in correct response proportion was highly significant [$b_{\text{intrusive /r/}} = -5.05$, $p < 0.001$]. We additionally investigated whether participants improved across the experiment, by using trial number, normalized to a range from zero to one, as a covariate. While this was not the case overall [$b_{\text{trial}} = -0.43$, $p > 0.1$], participants showed a change in bias and made over the course of the experiment more errors on onset /r/ stimuli [$b_{\text{trial x onset /r/}} = -2.02$, $p < 0.5$] and fewer errors on intrusive /r/ stimuli [$b_{\text{trial x ntrusive /r/}} = 1.97$, $p < 0.5$]. Furthermore, we analyzed whether participants made fewer erroneous clicks on disambiguated sentences than on ambiguous sentences. This appeared to be the case for both sentences with an intrusive /r/ (23.9% vs. 33.3% correct) and an onset /r/ (94.6% vs. 98.0% correct).

The preference for an interpretation of the intrusive /r/ as an onset /r/ is also clearly visible in the eye-tracking results in Figure 2. Even if the stimulus contains only an intrusive /r/, participants look more at the r-initial word; in the left panel, the majority of looks are made to the competitor, not to the target, while in the right panel, the majority of looks are to the target. 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-effects model with subject and item as random factors and type of /r/ as a fixed factor. If, in such model, the intercept is significantly different from zero, this indicates a preference for one visual word over the other. Positive values reflect more

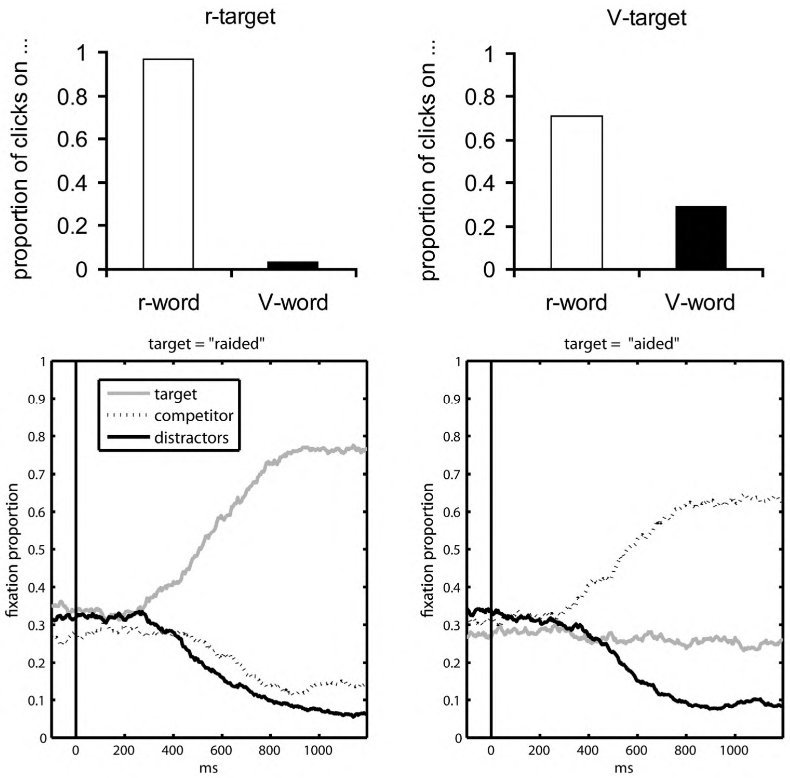

Figure 2: *Results from the eye-tracking experiment with Dutch listeners. Upper panels: Proportion of clicks on r-initial versus vowel-initial response alternatives; Lower panels: Fixation proportions over time to target, competitor and averaged distractors, in each case for trials in which the sentence contained intrusive /r/ (vowel target: left panels) versus trials in which the sentence contained onset /r/ (r target: right panels). Note that the proportion of clicks to distractor response alternatives was zero on all experimental trials.*

looks to r-initial words, negative values more looks to vowel-initial words. We tested this for time windows with a width of 100 ms, from 100ms to 700ms. Figure 3 shows the beta weights of this analysis. In the 100-200 ms window, neither intercept nor type of /r/ had a significant influence. This is in line with the assumption that it takes about 200 ms for eye movements to reflect an influence of the speech material (e.g., Matin, Shao, & Boff, 1993). For all other time windows there is a significant preference for r-initial words, while the effect of type of /r/ in the stimulus is only significant from the time window 500-600 ms onwards.

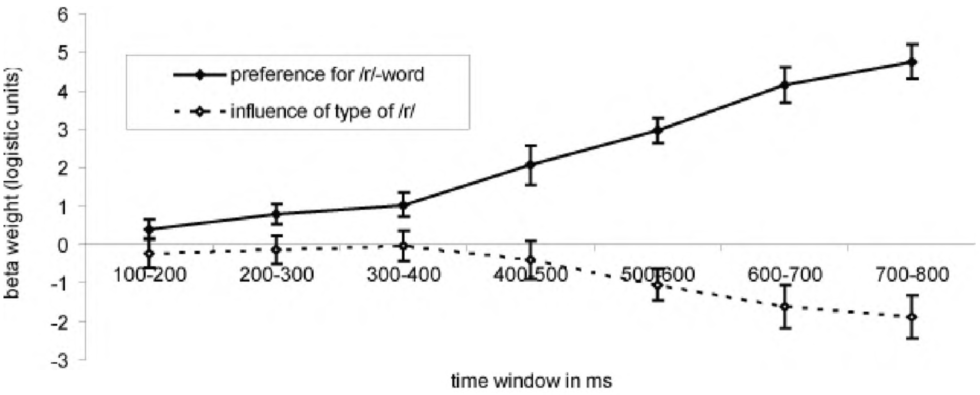

Figure 3: *Beta weights of the analysis for each of seven time windows. Overall preference for /r/ word indicates to what extent participants look at the r-initial word rather than the vowel-initial word, independent of the stimulus. The beta weights for "type of /r/" indicate to what extent the overall preference for /r/ is moderated by the strength of the /r/ (onset vs. intrusive) in the acoustic input.*

Note that the effect of "type of /r/" reflects to what degree the participants are looking at the intended word more than at the competitor. It is important to note here that while some of our sentences were "minimal pairs" (e.g., *Canada (r)aided* ...), some sentences diverged after the critical word (*Sheila elates everybody with her presence* vs. *Sheila relates well to other people*). The effect of type of /r/ may hence be attributed either to the acoustic-phonetic differences between an intrusive versus onset /r/, or to the semantic disambiguation of the sentences. To test the potential role of semantic cues, we assessed whether the pattern differed between ambiguous and unambiguous sentences. The effect of type of /r/ was not moderated by sentence ambiguity in any time window (max(|t|) = 1.2). The preference for the intended word in the late time windows must therefore be due to the acoustic-phonetic differences between onset /r/ and intrusive /r/.

Discussion

The eye-tracking data indicate that Dutch listeners are likely to misinterpret an intrusive /r/ as an onset /r/. On the majority of trials with an intrusive /r/, their looks go to an r-initial word. The main rationale for this eye-tracking experiment was further investigation of why, in cross-modal priming, primes with intrusive /r/

facilitated Dutch listeners' responses neither to vowel-initial nor to r-initial targets. To understand this, we needed to know how such primes are actually perceived by these listeners. Experiment 4 has provided an answer: primes with intrusive /r/ are overwhelmingly taken to be words beginning with /r/.

The eye-tracking data suggest that at the first hint of a rhotic in the signal, Dutch listeners directed their gaze toward the printed word with initial /r/. That is, we see a preference for r-initial words in the time window 200-300 ms after word onset, the first time window in which a signal-driven effect can be expected given the latency of eye movements. Although the phonetic difference between intrusive and onset /r/ also influences eye movements, this occurs only 300 ms later.

One possible explanation for this lag might be that the difference between an intrusive /r/ and an onset /r/ is one of duration, so that the listeners need to wait for the /r/ to end to make a decision. However, onset and intrusive /r/ are also differentiated by the amount of intensity decrement during the /r/, i.e., how much the amplitude decreases (for instance between the schwa of *extra* and [ɑɪ] of *ice* in *extra (r)ice*). This differentiating information is immediately available in the speech signal, but Dutch listeners fail to make immediate use of it.

General discussion

The resolution of ambiguity in casual speech depends crucially on experience with the process which has produced it. Unfamiliar casual speech processes can cause listeners to be seriously misled, and interpret an utterance quite wrongly. A British speaker describing how Canada aided her neighbours can cause a proficient Dutch listener to English to draw the startling conclusion that Canada raided her neighbours.

In our four experiments, we investigated the extent to which native and non-native listeners can resolve ambiguity caused by the connected speech process of intrusive /r/ in British English. There are pronunciation differences that potentially allow listeners to distinguish whether the /r/ in *Canada (r)aided* is an onset /r/ (i.e., *raided* is the intended verb) or an intrusive /r/ (i.e., *aided* is the intended verb). In the first three experiments, the impact of this difference on perception was tested with a cross-modal priming task in which we could test priming of the visual targets *raided* and *aided* by an intended verb *aided*, preceded by an intrusive /r/, versus by an intended verb *raided*. To assess further whether resolving the ambiguity depends on experience with intrusive /r/, we tested British English listeners, American English listeners, and Dutch learners of English.

Our results show that native British English listeners effectively exploit the subtle phonetic differences between an intrusive /r/ and an onset /r/, and are thus able to resolve potentially ambiguous phrases such as *Canada (r)aided*. American English listeners, with less experience with intrusive /r/, but with the same language, were also able to distinguish intrusive /r/ from onset /r/, and their processing of the potentially ambiguous sequences was indistinguishable from that of the British participants. The Dutch learners of English, however, although they too showed some sensitivity to the phonetic cues, appeared unable to effectively distinguish between intrusive and onset /r/s. An onset /r/ only primed r-initial words, but an intrusive /r/ showed, in this priming task, a tendency to be equivalently relevant to both vowel-initial and r-initial target words.

To understand the source of this pattern, we conducted an additional eye-tracking study. In this experiment, Dutch learners of English were presented with the same ambiguous sentences as in the cross-modal priming experiment in a visual world paradigm with printed words. That is, they saw four printed words on the screen and were instructed to click on the word that they heard in the sentence. For example, they heard "*Canada(r)aided....*" and the printed words on the screen were *aided*, *raided* and two phonologically unrelated distractors. Regardless of whether the sentence they heard contained an intrusive /r/ or an onset /r/, the Dutch listeners had a strong preference for r-initial target words. Over time, they did begin to show influence of the type of /r/ in the sentence, but even then this influence remained much smaller than their preference for r-initial words.

Our eye-tracking results partly replicate and partly extend previous eye-tracking studies with non-native listeners. Several studies (Cutler, et al., 2006; Escudero, Hayes-Harb, & Mitterer, 2008; Weber & Cutler, 2004) have investigated whether the eye movements of non-native listeners are influenced by distinctions that are known to be difficult for them to process (such as English /æ/ vs. /ɛ/ for Dutch listeners, or English /r/ vs. /l/ for Japanese listeners). In general, these studies revealed little influence of phonetic form of the input on the L2 listeners' fixations. Thus Japanese listeners looked at the picture of a locker independently of whether the spoken instruction contained the word *locker* or *rocket* (until, of course, the word-final consonant disambiguated the input (Cutler, et al., 2006), and Dutch listeners first looked at a picture of a pencil irrespective of whether *pencil* or *panda* was named. Moreover, in both cases there also was an asymmetry, so that when instructed to click on the locker or the pencil, the respective listener groups were unlikely to look at the rocket, or the panda. That is, the sounds /r/ and /l/ were both taken to be /l/ by Japanese listeners, and the sounds /æ/ and /ɛ/ were both taken to be /ɛ/ by Dutch listeners (see also Escudero et al., 2008).

In our data too, an asymmetry appeared. Dutch listeners hardly distinguished between phrases with an intended vowel-initial word and an intended r-initial word, but they did not consider the two lexical candidates as equivalent. Overwhelmingly, the word they clicked on was the r-initial candidate. They were apparently biased to interpret any hint of rhoticity as evidence for an r-initial word, so that they hardly ever considered the vowel-initial words. Whatever the intended input, if it began with an /r/ sound, it was taken to be an r-initial word.

This is not to gainsay the small but clear effect that we observed of signal-driven looks to the intended word. The beta weights plotted in Figure 3 capture the difference which can also readily be seen in the two eye-tracking graphs in Figure 2; in the left panel, the vowel-initial word (here, the target) continues to receive more looks than the distractors right to the last measured time window, whereas in the right panel, looks to the vowel-initial word (here, the competitor) rapidly drop in parallel with the looks to the distractor. Thus although the pattern of clicks would suggest that Dutch listeners unambiguously choose the r-initial response alternative, their looks betray that they give consideration to the phonetic characteristics of the signal as well. This seems to suggest that Dutch listeners can at least to some degree appreciate the acoustic-phonetic correlates of the unfamiliar connected speech process of intrusive /r/ in British English. Although the presence of a rhotic sound leads for them to activation of an r-initial word, the level of this word's activation is perhaps less when the rhotic sound is less acoustically strong, and, in consequence, the alternative interpretation of the signal (the vowel-initial word) is considered.

Our results shed light on several theoretical issues, which have not previously been brought together. First, they extend our understanding of how inter-word phonetic processes affect listening to speech. Second, they offer a new source of evidence on listening to a non-native language, and the pitfalls this can involve. And third, they further illuminate the process of listening to non-native dialect, and how dialect mismatch differs from language mismatch.

With respect to the effect of the inter-word phonetic processes in natural speech, our experiments have provided even stronger evidence for native listeners' use of phonetic detail than the study of liaison by Spinelli et al. (2003). Spinelli et al. consistently found priming for unintended words over the course of four cross-modal priming experiments with native French listeners. The reaction times to vowel-initial words (OIGNON) when the consonant-initial word (ROGNON) was intended, and the reaction times to the consonant-initial words when the vowel-initial word was intended, were both always faster than the baseline. The speaker's intention—and the associated phonetic details—only modulated the priming effects, with more priming for the intended than the unintended word. Our current results show more dramatic effect of the speaker's intention and the resulting differences in phonetic detail. In the data of the British English and American English listeners, there was virtually no indication of any priming of the unintended word. The phonetic details were hence strong enough to fully inhibit activation of the unintended word, a pattern not observed by Spinelli et al. A strong /r/ clearly activated an r-initial word for these listeners, and an acoustically weaker /r/ was accurately interpreted as a word-boundary process consistent with the presence of a vowel-initial word.

Second, with respect to the processing of non-native language, our results suggest that casual speech processes can induce significant difficulties. Even advanced learners of a language are not necessarily able to cope with connected speech processes they are unfamiliar with from their native language. Recall that, as described in the introduction, Darcy et al. (2007) came to a different conclusion on the basis of their studies of the perception of voicing assimilation. Advanced learners, they found, were able to compensate for an unfamiliar assimilation just as well as native listeners, and beginning learners also showed some compensation. However, recall also that Mitterer et al. (2006) found compensation for Hungarian liquid assimilation in listeners with little to no experience with a language (Dutch listeners, who performed similarly to native Hungarian listeners). Mitterer et al. argue that compensation for assimilation can often draw on general auditory processes, in the light of which Darcy et al.'s findings of compensation for voice assimilation would not necessarily rest on effects of language experience. Another difference between the process of assimilation and the process of r/-intrusion, of course, is that assimilation alters a segment while /r/-intrusion adds a segment to the speech stream; these two effects may give rise to different processing implications. There is abundant evidence that changing a segment leads to processing costs for native listeners. Andruski, Blumstein and Burton (1994) found that when an initial stop was shortened to only one-third of its normal duration (so that for instance a prime word like *king* would sound more like *ging*), responses to the related words in a priming task were less facilitated. Tucker and Warner (2007) found that American English words with an intervocalic /d/ were also less effective cross-modal primes when the consonant was reduced, that is, more approximant- or vowel-like, than when it was a clearly consonantal flap [ɾ]. Adding a segment, in contrast, may even lead to a processing benefit for native listeners. Insertion of an optional epenthetic schwa in word-final consonant clusters in Dutch, whereby the Dutch word *film* 'film' /fɪlm/ becomes bisyllabic /fɪləm/, produces a form that listeners recognize more rapidly, in both lexical decision and word spotting, than the canonical form without vowel epenthesis (van Donselaar, Kuijpers, & Cutler, 1999). It is clear that the addition of a segment in the case of /r/-intrusion, however, did not benefit either native or non-native listeners; the native listeners, could cope with it effectively, but the listeners with a different native language were significantly misled.

Third, our experiments shed light on the comparison between a different dialect and a different language. Other studies have also shown that difficulties arising from cross-dialect perception are in many cases not significant. Floccia et al. (2006), for example, found an initial processing cost for a regional accent in word recognition tasks, but listeners rapidly adjusted to it. Sumner and Samuel (2009) found that American listeners can have initial problems with recognition of words in another American accent, but these problems could easily be removed by semantic priming. Cutler, Smits and Cooper (2005) found that noise-masked American English vowels in simple syllables could be identified as well by Australian speakers of English as by speakers of American English, but, notably, proficient second-language users of English with Dutch as native language performed significantly worse. As further attested in the present study, language differences have greater consequences than dialect differences for speech perception.

What requires explanation in this context, however, is the asymmetry between the American listeners' excellent use of the acoustic-phonetic information in the present study, where their word recognition performance paralleled that of the native British listeners, and the finding from our earlier study of phonetic categorization (Tuinman, et al., under revision) that American listeners made less use of acoustic-phonetic (durational) cues than the British participants in that study. We propose that task characteristics may have played a role in the underestimation of the cross-dialect performance in that task. The 2AFC categorization task required participants to focus on whether they heard *ice* or *rice* (because they had to make a decision on which word they heard). Also, only one of the cues that indicate the difference between intrusive and onset /r/, namely duration, was manipulated in that study, and furthermore the speech input was synthesized rather than being natural speech. In the present case, the range of cues that natural speech provides to differentiate onset and intrusive /r/ were all available, and in this case, the American listeners' performance was indistinguishable from that of the native British listeners. An obvious possibility, then, is that while native listeners are so flexible that they can rely solely on acoustic-phonetic cues even if only one such cue is available, cross-dialect listeners are less flexible but still quite able to cope effectively when the full (and potentially redundant) range of cues characteristic of natural speech is on offer.

In this light our results confirm those found in the other studies of non-native dialect and non-native language that we summarized: listening across dialects may be partially impaired, but the impairments are relatively minor and adjustment is rapid. Listening across languages is another story entirely: it is fragile and easily disrupted by unfamiliarity. Unfamiliar casual speech processes simply add one more source of disruption to the non-native listener's lot.

Prior literature has established that learning is needed to cope with different sets of connected speech processes (cf. Mitterer & McQueen, 2009), and also that connected speech processes differ in the intrinsic difficulty that they pose to listeners. Gow and Im (2004), for instance, in their comparisons of types of assimilation, found that Hungarian voice assimilation—a typologically frequent form of assimilation—was easier for listeners to process than Korean labial-to-velar assimilation—a typologically less common process. The critical test in these experiments was whether listeners could anticipate the segment that triggered the assimilation, and this proved to be possible for the Hungarian case but not for the Korean case. Where does /r/-intrusion then fit into this processing continuum?

The process has been claimed to have an articulatory source—Broadbent (1991), for example, argues that it can be interpreted as a type of glide formation, the process which gives rise to intervocalic /j/ and /w/, as between the first and second vowels of *kiosk* or *hoeing*. Unlike these two cases, however, the articulatory target for /r/ does not follow naturally from the gestural transition between a mid or low vowel to another vowel. Also note that the intrusive /r/ is not dependent on the place or height of the following vowel, which also indicates that it does not naturally arise from a gestural transition. The process has further been argued to have a historical source (Gimson, 1980). 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 accents (Giegerich, 1992). Consider again *soar* and *saw*. In rhotic accents, these words are clearly distinguished. In British English, in contrast, the two are homophonous in isolation and in many contexts, that is the canonical pronunciation for both words is /sɔː/. In connected speech in prevocalic position (*soar up* versus *saw up*), though, they are not homophonous; the /r/ in *soar* is pronounced but *saw* undergoes no change. Giegerich proposes that this process of /r/-insertion in words like *soar* was overgeneralized to other words (like *saw*) with mid or high vowels, making intrusive /r/ a historically motivated process.

Although both linking and intrusive /r/ are common in nonrhotic accents of English, the latter may be stigmatized and can provoke a negative reaction from English speakers themselves (Collins & Mees, 1999; Foulkes & Docherty, 1999). A final indication that intrusive /r/ is not strongly articulatory motivated stems from the fact that other languages have developed other strategies to realize a transition between a vowel-final and a vowel-initial word, such as the glottal stops, sometimes reduced to creaky voice, found in German (Kohler, 1994). Glottal stops are indeed employed by some British English speakers to avoid intrusive /r/ (Brown, 1988). In general this intrusion process thus contrasts with assimilations, which are not consciously and specifically avoided, as they may be absent if the speaker chooses a formal speaking style. These considerations suggest that /r/-insertion is a less natural process than, for instance, place and voicing assimilation. Because the process is very specific with respect to language and context, it is thus likely to be more difficult for non-native listeners to correctly interpret intrusive /r/s than to interpret more common and phonetically grounded connected speech processes.

Our experiments confirm this probability. The process of /r/-intrusion misled our Dutch listeners in Experiments 3 and 4. These non-native listeners contrasted with the native listeners, who efficiently extracted the subtle acoustic-phonetic cues to the intrusive versus onset status of /r/ tokens, thereby resolving a potential ambiguity, in just the signal-driven way that was shown to be possible for other connected speech processes that have previously been studied (e.g., assimilation, /t/-reduction, liaison). Not only the truly native listeners exploited this phonetic detail to the fullest, though; so did the listeners with the same language whose native dialect does not have /r/-intrusion. Only non-native listeners fail at this task and perceive any evidence for an /r/ as an intended /r/. For the intrusive /r/ it suggests that non-native listeners hear John Lennon sing *I soar a film today* and are puzzled for a moment, whereas native language and native dialect listeners know how to deal with Lennon's intrusive /r/.

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.

Andruski, J. E., Blumstein, S. E., & Burton, M. (1994). The effect of subphonetic differences on lexical access *Cognition, 52*, 163-187.

Broadbent, J. (1991). Linking and intrusive r in English. *UCL Working Papers in Linguistics, 3*, 281-302.

Broersma, M., & Cutler, A. (2008). Phantom word activation in L2. *System, 36*, 22-34.

Broersma, M., & Cutler, A. (2011). Competition dynamics of second-language listening. *Quarterly Journal of Experimental Psychology, 64*, 74-95.

Brown, A. (1988). Linking, intrusive, and rhotic /r/ in pronunciation models. *Journal of the International Phonetic Association, 18*(2), 144-151.

Collins, B., & Mees, I. M. (1999). *The phonetics of English and Dutch* (4th ed.). Leiden: Brill.

Connine, C. M., Ranbom, L. J., & Patterson, D. J. (2008). Processing variant forms in spoken word recognition: The role of variant frequency. *Perception and Psychophysics, 70*(3), 403-411.

Cruttenden, A., & Gimson, A. C. (1994). *Gimson's pronunciation of English* (5th ed.). London: Arnold.

Cutler, A., Smits, R., & Cooper, N. (2005). Vowel perception: Effects of non-native language vs. non-native dialect. *Speech Communication, 47*, 32-42.

Cutler, A., Weber, A., & Otake, T. (2006). Asymmetric mapping from phonetic to lexical representations in second-language listening. *Journal of Phonetics, 34*, 269-284.

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. Hualde (Eds.), *Papers in Laboratory Phonology 9* (pp. 411-442). Berlin: Mouton de Gruyter.

Dixon, P. (2008). Models of accuracy in repeated-measures designs. *Journal of Memory and Language, 59*, 447-456.

Escudero, P., Hayes-Harb, R., & Mitterer, H. (2008). Novel second -language words and asymmetric lexical access. *Journal of Phonetics, 36*, 345-360.

Floccia, C., Goslin, J., Girard, F., & Konopczynski, G. (2006). Does a regional accent perturb speech processing? *Journal of Experimental Psychology: Human Perception and Performance, 32*, 1276-1293.

Foulkes, P., & Docherty, G. (Eds.). (1999). *Urban Voices: Accent studies in the British Isles*. London: Arnold.

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.

Gow, D. W. (2002). Does English coronal place assimilation create lexical ambiguity? *Journal of Experimental Psychology: Human Perception and Performance, 28*, 163-179.

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.

Gow, D. W., & Im, A. M. (2004). A cross-linguistic examination of assimilation context effects. *Journal of Memory and Language, 51*, 279-296.

Gussenhoven, C., & Broeders, A. (1997). *English pronunciation for student teachers*. Groningen: Wolters-Noordhoff.

Kemps, R. J. (2004). *Morphology in auditory lexical processing: Sensitivity to fine phonetic detail and insensitivity to suffix reduction.* Doctoral dissertation, Radboud University Nijmegen (MPI Series in Psycholinguistics, Vol. 28). Wageningen: Ponsen and Looijen.

Kohler, K. J. (1994). Glottal stops and glottalization in German data and theory of connected speech processes. *Phonetica, 51*, 38-51.

Matin, E., Shao, K. C., & Boff, K. R. (1993). Saccadic overhead: Information processing time with and without saccades. *Perception and 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., & Viebahn, M. C. (2007). Tracking recognition of spoken words by tracking looks to printed words. *Quarterly Journal of Experimental Psychology, 60*, 661-671.

Mitterer, H., Csépe, V., & Blomert, L. (2006). The role of perceptual integration in the perception of assimilation word forms. *Quarterly Journal of Experimental Psychology, 59*, 1395-1424.

Mitterer, H., & Ernestus, M. (2006). Listeners recover /t/s that speakers reduce: Evidence from /t/-lenition in Dutch. *Journal of Phonetics, 34*, 73-103.

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.

Otake, T., & Cutler, A. (1999). Perception of suprasegmental structure in a non-native dialect. *Journal of Phonetics, 27*, 229-253.

Quené, H. (1992). Durational cues for word segmentation in Dutch. *Journal of Phonetics, 20*, 331-350.

Scott, D. R., & Cutler, A. (1984). Segmental phonology and the perception of syntactic structure. *Journal of Verbal Learning and Verbal Behavior, 23*, 450-466.

Shatzman, K. B., & McQueen, J. M. (2006). Segment duration as a cue to word boundaries in spoken-word recognition. *Perception and Psychophysics, 68*, 1-16.

Spinelli, E., Cutler, A., & McQueen, J. M. (2002). Resolution of liaison for lexical access in French. *Revue Française de Linguistique Appliquée, 7*, 83-96.

Spinelli, E., McQueen, J. M., & Cutler, A. (2003). Processing resyllabified words in French. *Journal of Memory and Language, 48*, 233-254.

Sumner, M., & Samuel, A. G. (2009). The effect of experience on the perception and representation of dialect variants. *Journal of Memory and Language, 60*, 487-501.

Tanenhaus, M. K., & Spivey-Knowlton, M. J. (1996). Eye-tracking. *Language and Cognitive Processes, 11*, 583-588.

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

Tucker, B. V., & Warner, N. (2007). *Inhibition of processing due to reduction of the American English flap.* Paper presented at the 16th International Congress of Phonetic Sciences, Saarbrücken, Germany.

Tuinman, A., Mitterer, H., & Cutler, A. (under revision). Cross-language and cross-dialect differences in perception of intrusive /r/ in English. *Journal of the Acoustical Society of America. Chapter 3 of this dissertation.*

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.

Syntactically-based compensation for phonological reduction without top-down feedback

Chapter 5

Tuinman, A., Mitterer, H., & Cutler, A. (submitted). Syntactically-based compensation for phonological reduction without top-down feedback.

Abstract

Syntactic context effects on the identification of word-final reduced /t/s in Dutch verbs were measured in two 2 AFC experiments in which listeners decided whether or not the verb in a sentence such as *Maar zij ren(t) soms* ('But she run(s) sometimes') ended in /t/. In Dutch, the presence or absence of /t/ after the verb stem is the difference between third- and first-person singular present. In both experiments, listeners perceived more /t/s in sentences in which this is syntactically correct. In Experiment 1, the biasing syntactic context preceded the target verb; in Experiment 2, it followed the target verb. The syntactic bias was larger for fast than for slow responses in Experiment 1, but no such difference was found in Experiment 2. This suggests that sentential context can influence phoneme decision, but does not influence prelexical representations, given that the influence of context can also decrease again.

Introduction

What we expect to see or hear influences what we think that we perceive. Try this example: A 13 C and 12 13 14. For most people the first sequence is the start of the alphabet and the second sequence three consecutive numbers. What might go unnoticed if the examples are presented alone, however, is that the second item is actually the same in both sequences. But '13' is perceived as a B when it is encircled by other letters and as 13 when it is surrounded by numbers. The same item may be perceived differently as a result of expectations raised by the context.

Ganong (1980) developed a paradigm to show how lexical expectancies influence the perception of phonemes, with a paradigm in which listeners had to identify syllable-initial ambiguous stimuli on a voicing continuum (e.g., a continuum between [d] and [t]). In one continuum the voiced endpoint /d/ formed a word (e.g., *dice-tice*) and in the other the voiceless endpoint /t/ formed a word (e.g., *dype-type*). Ganong showed that listeners tend to label the perceptually ambiguous stimuli such that the utterance they heard formed an existing word (as [d] in *dice-tice*, but as [t] in *dype-type*). Others have replicated this effect in word-medial position (e.g., [d] and [g] in *cradle-cragle* and *badel-bagel*, Connine, 1990) and word-final position (e.g., [s] and [ʃ] in *kiss-kish* and *fiss-fish*; McQueen, 1991).

Most models of spoken-word recognition, including TRACE (McClelland & Elman, 1986), the distributed cohort model (Gaskell & Marslen-Wilson, 1997), the modular Race model (Cutler & Norris, 1979), and Shortlist (Norris, 1994), account for these lexical context effects by assuming that the word recognition system has two basic levels. The first is the prelexical level, where an abstract representation of the speech input is made. Secondly, there is the lexical level at which all words that match the prelexical (phonemic) representation of the input to a certain extent are activated and compete for recognition. However, the models have different assumptions about the nature and possible directions of the information flow, and thus different explanations for lexical context effects. One explanation for lexical effects, incorporated in interactive models such as TRACE (McClelland & Elman, 1986), is that information can flow from the lexicon back to the prelexical level and directly influence perception. The ambiguous signal [?ype], for example, would activate the lexical representation of *type* on the lexical level, which feeds back support to the representation of the phoneme [t] at the prelexical level. As a result, more [t] responses are expected in the [?ype] context than in the [?ice] context (in which feedback would produce a bias toward [d]).

Autonomous models of speech perception (Cutler & Norris, 1979; Forster, 1976) in contrast, argue that context effects are not due to feedback in speech processing, but represent integration of independent information sources. In these models, information can only flow from the bottom up. Norris, McQueen, and Cutler (2000) argue that lexical knowledge can influence perceptual decision making without influencing prelexical processing. In their *Merge* model (McQueen, Norris, & Cutler, 1999; Norris, et al., 2000), information flows from the prelexical level to the lexicon without feedback. Phonemic decisions are made at dedicated decision units that continuously receive input from both the prelexical and lexical levels of processing. At the decision stage, information from the two levels is merged and they both influence the final decision. The Merge model assumes that a signal like [?ype] activates the decision units for both [t] and [d] through connections from the prelexical level to the phoneme decision units. At the same time, the signal also activates the word representation for *type*, which feeds support to the phoneme decision unit [t], creating a lexical bias toward [t]. Accordingly, more [t] responses are expected in a [?ype] context than in a [?ice] context. A model like Merge can thus account for lexical context effects in speech recognition without feedback of lexical information.

The debate about which of these accounts is correct has not been limited to lexical context effects as they arise in the Ganong paradigm (1980) but also encompassed sentential context effects. Several studies have shown that listeners are more likely to identify the word that makes most sense in the sentence they have heard compared with words that are less sensible (Borsky, Shapiro, & Tuller, 2000; Borsky, Tuller, & Shapiro, 1998; Connine, 1987; Connine, Blasko, & Hall, 1991; Gaskell & Marslen-Wilson, 2001). In Borsky et al. 's (1998) study, for example, listeners heard stimuli from a *goat-coat* continuum embedded in sentences that biased interpretation to either "goat" (*The busy farmer hurried to milk the* [?]oat *in the drafty barn*) or "coat" (*The careful laundress had to dry-clean the* [?]oat *in the cluttered attic*). At the offset of the target word listeners saw a visual probe (*goat* or *coat*) and were instructed to decide whether the visual word matched what they had just heard. Borsky et al. found a sentential context effect in that listeners gave more *goat* identifications in sentences with a *goat* interpretation bias and more *coat* responses for sentences with a *coat* bias.

As in the lexical case, the question arises whether sentential context effects are due to feedback or to a merging of information from different sources. Both the feedback account (Borsky et al., 1998) and the account based on a decision bias (Samuel, 1981; Connine, 1987; Norris et al., 2000) have had their followers.

Van Alphen and McQueen (2001) attempted to distinguish between these two possible explanations of sentential context effects. They used reaction time (RT) range analyses in which they divided the categorization responses obtained in their experiments into fast, medium, and slow RT ranges. According to van Alphen and McQueen, the reasoning behind this kind of analysis is that if sentential context effects are indeed caused by feedback processes, they should first take time to build up as higher level processes may need time to complete, and then be quite stable over time. Indeed, modeling studies using the TRACE model of speech perception (McClelland & Elman, 1986) showed such a pattern for phoneme decisions in word-initial (McClelland & Elman, 1986) and word-final position (McClelland, 1987). Hence, these context effects are not expected to die away rapidly under the assumption of interactive processing. Once information from the sentential level of processing flows to lower lexical and prelexical levels and this information modifies phonological representations, these modifications are expected to continue to influence responses made after that time, because top-down information overrides bottom-up information with no chance of recovery. However, if sentential context effects are the result of a decision bias it is less clear what kind of results can be expected. Van Alphen and McQueen propose that a decision bias can influence responses at one moment in time, but not necessarily later in time. Because the decision bias does not change the bottom-up encoding of speech, the information from the original speech signal remains available to influence later processing.

Van Alphen and McQueen (2001)'s study examined sentential context effects on the identification of the Dutch function words *te* (to) and *de* (the). In each experiment they created three different sentences in which either *te* or *de* or both *te* and *de* were logical interpretations of an ambiguous sound on the [tə]-[də] continuum. In their first experiment, for example, they selected an unambiguous plural noun, *schoenen* (shoes), that can be preceded by the definite article *de* but not by the infinite marker *te*; an unambiguous verb infinitive, *schieten* (to shoot), that can be preceded by *te* but not by *de*; and another word, *schaatsen*, that can either be a plural noun (skates) or an infinitive verb (to skate) and can thus be preceded by both *de* and *te*. These three words were put in the same sentence context (e.g., *Ik probeer* [?ə] *schoenen* "I try to/the shoes", *Ik probeer* [?ə] *schieten* "I try to/the shoot", and *Ik probeer* [?ə] *schaatsen* "I try to/the skate[s]"). Listeners heard several repetitions of these sentences and were instructed to decide as fast as possible whether they had heard *te* or *de*. Note that the disambiguating information in these sentences follows the ambiguous sound. In the other experiments, the relation of the place of disambiguation and the ambiguous sound was varied; in one experiment it followed even later, in the second word after [?ə], and in another experiment disambiguation occurred on the word before [?ə]. For all experiments, van Alphen and McQueen divided the responses in three categories: fast, medium, and slow RTs. This way, they could test whether the sentential context effect is stable over time—which would be in line with a feedback account—or not. When the disambiguating information appeared directly after the target sound, the sentential context affected phonetic identification in both the fast and medium RT range, but the effect was weaker and not significant in the slow RT range. As the effect of the sentential context dies away, van Alphen and McQueen concluded that this finding is inconsistent with a feedback account of speech processing, because a sentential bias effect caused by feedback is expected to rise or remain stable over time. When the crucial information for decoding the ambiguous sound was presented even later in the sentence, no significant sentential bias effect was found in any RT range (although the effect was almost significant in the slow RT range). In the last experiment, when the disambiguating information appeared before the target sound, the sentential bias effect was significant in the fast and medium RT ranges, but not in the slow RT range. From this series of experiments, van Alphen and McQueen concluded that sentential context influences the identification of function words, but only within a limited time frame. Furthermore, their results challenge explanations of sentential context effects based on feedback, because if feedback influences the lexical or prelexical processes, then sentential context effects should not have become weaker over time.

Studies of artificially created ambiguous phonemes have greatly extended our knowledge on how information flows through the speech recognition system and whether this information strictly flows bottom-up or can also flow back from higher levels to lower ones. Nevertheless, it remains a fact that listeners are hardly ever confronted with such ambiguous phonemes. Therefore, it is important to know whether the results also hold when listeners are confronted with ambiguous signals that are more natural. Natural ambiguities are abundant in spontaneous speech due to segment deletions and reductions. In such circumstances, the speech signal may be more impoverished which in turn may necessitate more top-down processing to restore the intended meaning.

The present study therefore investigates how ambiguities that arise due to reduction in spontaneous speech are resolved and whether in such cases contextual information influences phoneme recognition through feedback or not. As an example of reduction in spontaneous speech, we used word-final /t/-reduction (e.g., *post box* realized as [pəʊsbɒks]) which occurs frequently in English (Guy, 1980; Sumner & Samuel, 2005), German (Kohler, 1990) and Dutch (Janse, Nooteboom, & Quené, 2007; Mitterer & Ernestus, 2006). Furthermore, Mitterer and Ernestus found that the process of /t/-reduction is gradient in the sense that it is not always complete. The gradient nature of /t/-reduction makes it possible to create ambiguous signals for which syntactic context effects are expected to be strongest. This makes /t/-reduction an interesting test case for the influence of sentence context on the identification of phonemes. With the process of /t/-reduction we can test the hypothesis that true top-down effects may occur in the restoration of reduction in spontaneous speech.

If a listener has to decide whether a Dutch speaker intended (but reduced) a word-final /t/, a syntactic context effect may arise in the perception of inflected verbs. In Dutch, the presence or absence of a /t/ at the end of an inflected verb is the difference between the third- and first-person singular present form. Syntax requires a /t/ to be present in the sentence "zij ren_" ('she runs')—at the position of the underscore—but not in the sentence "ik ren_" ('I run').

This was exploited in the present study, in which listeners decided whether a verb contained a word-final /t/ or not in sentences with either a first-person singular subject, inducing a no-/t/ bias, or a third-person singular subject, inducing a /t/-present bias. Following van Alphen and McQueen (2001), we investigated the time course of the syntactic context effect in different RT ranges. We also varied the point in the sentence at which the syntactic bias arose in relation to the critical word. In the first experiment, the sentence subject, which generates a bias, appeared before the critical word for which participants had to decide whether it ended on a /t/ or not. In the second experiment the syntactic bias occurs after the critical word. To ensure that the participants made a perceptual judgment, we also introduced two independent variables which, in the studies of Mitterer and Ernestus (2006), influenced the perception of reduced /t/: preceding phonological context and acoustic-phonetic evidence for /t/.

Experiment 1

Method

Participants

Twenty-one members of the Max Planck Institute's subject pool participated in this experiment. The participants were native speakers of Dutch and none reported any hearing impairment. All were volunteers and were paid a small fee for their participation.

Materials

Three Dutch verbs with a stem ending on /n/ and three verbs with a stem ending in /s/ were matched for lemma log frequency (Baayen, Piepenbrock, & Gulikers, 1995) and used as target verbs. These verbs were placed in sentences and a male native speaker of Dutch (the same speaker as in Mitterer and Ernestus, 2006) recorded the sentences several times. Table 1 shows the combinations of the target verbs and preceding and following words. The verbs in the sentences were pronounced with or without a final /t/ according to the reading list. Subsequently, these utterances were used as templates for synthesis via a Klatt synthesizer (Klatt, 1980). Stimuli were created in a similar fashion to Mitterer and Ernestus. Formants and bandwidths were measured at the beginning, in the middle and at the end of each segment and at major formant transition points within segments. When the formant could not be estimated reliably by Linear Predictive Coding (using Praat, Boersma & Weenink, 2005), values recommended by Klatt (1980) were chosen. In addition, nasalization for a nasal was carried over into the vowel before it was reduced completely, and anticipated for postvocalic nasals to match the natural utterances. Parameters for synthesis were generated by interpolating linearly between measurement points. Occasionally values were altered in order to prevent clicks and other transients, which can occur if control parameters change too quickly. Amplitude values were iterated to imitate the amplitude envelope of the natural utterances.

Table 1: *Sentence frame for the stimuli in Experiment 1*

Connection word	Subject	Target word	Adverb
		blaas /blaːs/ (1.62)	nauwelijks /ˈnɑuʋələks/
	zij /zɛɪ/	kreun /krøːn/ (1.46)	langzaam /ˈlɑŋzaːm/
		bloos /bloːs/ (1.21) ..t	moeizaam /ˈmujzaːm/
Maar /maːr/		zoen /zun/ (1.26) ...Ø	soms /ˈsɔms/
	Ik /ɪk/	ren /rɛn/ (1.95)	vaak /ˈvaːk/
		kus /kʉs/ (1.75)	

Note: English translations are *Maar* 'but', *ik* 'I', *zij* 'she', *blaas* 'blow', *kreun* 'moan', *bloos* 'blush', *zoen* 'kiss', *ren* 'run', *kus* 'kiss', *nauwelijks* 'hardly', *langzaam* 'slowly' *moeizaam* 'with difficulty', *soms* 'sometimes', *vaak* 'often'. The numbers between brackets are the lemma log frequencies for the words taken from CELEX (Baayen, et al., 1995).

The verb stems were followed by one of five different synthesized signals for the coda (see Figure 1) (cf. Mitterer & Ernestus, 2006). The first signal was similar to a full /t/ with a 25 ms closure and a 45 ms transient-frication sequence. To prevent an unnatural flat line in the signal, the closure was synthesized with 20 dB amplitude of frication (AF) and a 30 dB amplitude bypass (AB) of the parallel branch of the synthesizer. At the release of the initial burst, AB was set to zero, and AF increased to 40 dB, and decreased again to 15 dB at the end of the 65 ms signal. For this and all the other target signals, amplitude changes were loglinear in dB to achieve a linear amplitude envelope. The transient-frication signal was dominated by the fifth and sixth formats (55dB) starting at 5700Hz (bandwidth of 500 Hz) and 7500 Hz (700 Hz), respectively. These formants fell to 5200 and 6850 Hz at 65 ms. The second, third and fourth formant stayed constant throughout the transient-frication sequence at 1434 Hz (200), 2212 Hz (500) and 3840 Hz (600), respectively, and their amplitude increased from 20 to 25 dB in order to mimic the increasing low-amplitude frication in the model [t].

The second coda signal was a 65 ms frication noise, as often found in /st/ codas. The amplitude of frication started and ended at 15 dB with a 40dB maximum at 30 ms. The settings for the second, third and fourth formant were the same as the settings in the full [t]. The fifth and sixth formants started at 5130 Hz (bandwidth: 500 Hz) and 6750 Hz (700 Hz) and fell to 4620 Hz and 6080 Hz at the end of the signal (As = 55 dB). The weak-frication signal was derived from this signal by reducing the overall amplitude of fricative noise by a factor of 5 (= 14 dB). The closure-only signal was synthesized with the same settings as for the closure of the full /t/.

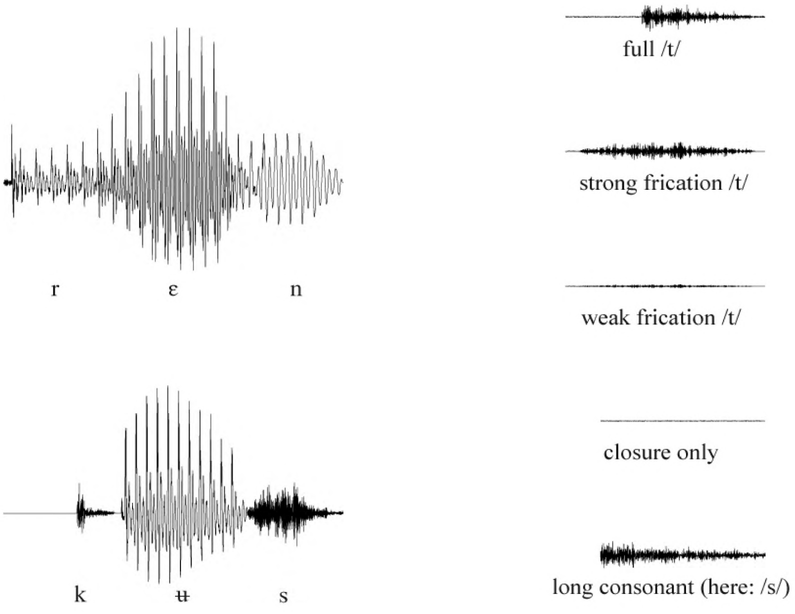

Figure 1: *The different target signals presented in Experiment 1 and 2.*

The final coda signal was a coda that was not followed by any signal implying a /t/. Instead, it was created by elongating the consonant /n/ by 45 ms and /s/ by 65. These durations were based on the measurements of the coda duration in the recording made of the Dutch native speaker reading the test materials. These measurements also showed that consonants in consonant clusters are shorter than in simple codas (Lehiste, 1970).

The long consonant target signals were adapted to the following adverb to make the transition as natural as possible between the target verb and the subsequent adverb. Thus, for the long consonant /n/ the amplitude remained high if another nasal followed (e.g., if "nauwelijks" or "moeizaam" was the following adverb), but it fell before other phonemes. For the long consonant /s/, the amplitude did not fall before another /s/ (e.g., if "soms" was the following adverb), but decreased before other phonemes. A 25 ms closure was inserted after "Maar ik" and a 50 ms closure before verbs starting with a /k/ ("kus" and "kreun") to make the synthesized materials more like the natural utterances.

Overall, there were five coda signals for the target verbs: full /t/, strong frication /t/, weak frication /t/, closure only, and a long consonant. All these different signals were combined with the six target verbs to give 30 stimuli for which the participants had to decide whether they heard a coda with or without a /t/. The duration of the complete target verbs varied from 241 ms ("zoen…") to 430 ms ("blaas…"). This durational difference is partly explained by vowel length. In consequence, the speaking rate varied from 2 to 4 syllables per second.

The 30 target verbs were placed in ten carrier sentences that were created by combining the two possible initial parts of the sentences ("Maar ik …" and "Maar zij…") with the five possible final parts (the five adverbs following the target verbs). This resulted in 300 different sentences (see also Table 1).

Procedure

The experiment was conducted on a standard PC running the NESU experimental control software. Participants were tested one at a time in a sound-attenuated booth, which was normally lit. They wore Sennheiser closed headphones, sat at a comfortable reading distance from the computer screen and had a two-button response box in front of them. Participants received instructions in Dutch asking them to press the right button if the sentence they heard contained the verb form presented in the upper right corner on the computer screen, and to press the left button if the sentence contained the verb form in the upper left corner. Participants were further instructed to ignore the fact that some sentences would be ungrammatical with or without the interpretation of the /t/ at the end of the verb and to just respond whether they had in fact heard a /t/ or not. The 300 different stimulus sentences were presented only once to each participant, in a random order which was different for every participant.

After the instructions, four practice trials followed to familiarize participants with the procedure. The practice trials were sentences made up of the natural sentences used as templates for re-synthesis.

The practice and experimental trials had the same structure. Each trial began with 150 ms of blank screen. The response alternative without a /t/ in the coda (e.g., "ren") then appeared in the upper left corner of the screen and the other (e.g., "rent") appeared in the upper right corner. After another 450 ms the sentence was played. From the onset of the target verb, participants had 2.5 seconds to press one of the buttons. Their response caused the other alternative to be removed from the screen, so that the participants could see that their answer had been registered by the computer. If a participant did not react within 2.5 seconds, a stopwatch was shown on the screen to remind participants to respond more rapidly. The three feedback signals—no /t/-response, /t/-response and no response—stayed on the computer screen for one second before the next trial started. Participants were free to take a break after every 50 trials and could continue when they were ready.

Design

There were three independent variables. The first independent variable was the nature of the Coda Signal at the end of the target verb (full /t/, strong /t/ frication, weak /t/ frication, closure only, and long consonant). The second variable was the Preceding Context of the target signal (/n/ or /s/). The third independent variable was the Syntax of the sentence (is the sentence syntactically correct with or without a /t/ at the end of the verb). The dependent variable was the percentage of /t/-responses in each cell of the design.

Results

Analysis of number of /t/-responses

RTs faster than 100ms (after offset of the target verb) were discarded from this analysis. Prior to the analyses, the RTs were corrected to reflect the response time from offset of the target verb. All cases in which no response was given were also removed from the analyses. In total, 1.4 % of the trials were removed.

The mean percentages of /t/-responses are shown in Figure 2. A repeated measures analysis on the percentages of /t/-responses was performed with Syntax, Coda Signal, and Preceding Context as predictors. As is evident from Figure 2, listeners are more likely to report the presence of a /t/ if this interpretation leads to a syntactically correct sentence [Syntax: $F(1,20)= 55.4$, $p < 0.001$; ik: 65.3%, zij: 82.5%] and if the Coda Signal carries more acoustic information for the presence of a /t/ [$F(4,80)= 133.0$, $p < 0.001$]. Figure 2 also clearly indicates that Preceding Context /s/ elicited more /t/-responses than a preceding /n/ [$F(1,20)= 40.7$, $p < 0.001$; /s/: 82.1%, /n/: 65.8 %]. There were significant interactions of Syntax and Preceding Context [$F(1,20)= 30.4$, $p < 0.001$], and Syntax and Coda Signal [$F(4,80)= 9.0$, $p < 0.001$]. Other interactions failed to reach significance [Preceding Context by Coda Signal: $F < 1$; Syntax by Preceding Context by Coda Signal: $F(4,80)= 1.64$, $p > 0.1$].

To investigate the nature of the interaction between Syntax and Preceding Context, we examined the effect of Syntax on both levels of Preceding Context, namely /n/ or /s/ by *t*-testing. Both *t*-tests showed a significant effect of Syntax for the Preceding Context /n/ and for the Preceding Context /s/ [$p_{max} < 0.001$]. Thus, the interaction must reflect the difference in size of the effect of Syntax for each Context. To examine this, we computed the differences between the percentages of /t/-responses with Syntax 'ik' or 'zij' for the two levels of Preceding Context (/n/ and /s/). The Syntax effect was significantly larger for Preceding Context /n/ (24.6%) than for the /s/ context (9.8%).

The *t*-tests of the effect of Syntax were significant for all levels of the /t/-Ø series [$p_{max} < 0.01$], so that the interaction of Syntax by Coda Signal must again reflect a difference in effect size. The values for the differences between the percentages of /t/-responses with Syntax 'ik' or 'zij' for all five levels of the Coda Signal were analysed with a one-factorial ANOVA [$F(1,20) = 9.0$, $p < 0.001$]. Post-hoc tests demonstrated that the differences increased significantly from the full /t/ to the long consonant /n/ or /s/. The strong frication /t/, weak frication /t/ and closure did not differ significantly from each other, but did differ significantly from the full /t/ and the long consonant signal.

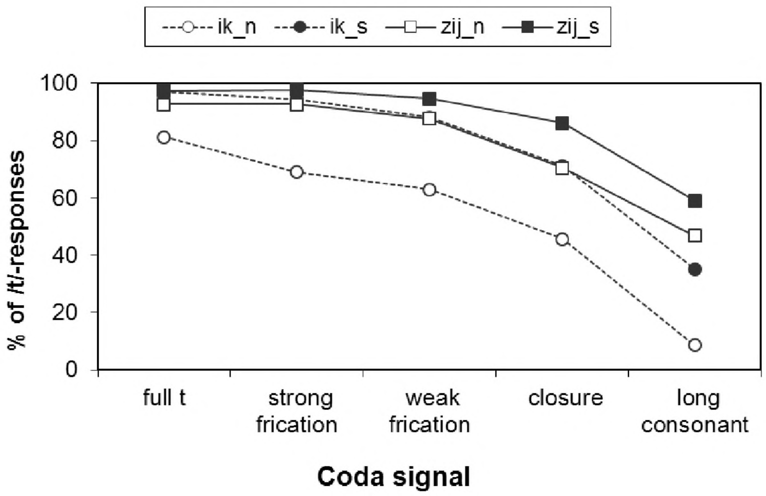

Figure 2: *Percentages of /t/-responses in Experiment 1 for each Coda Signal. The lines indicate the two different Preceding Contexts /n/ and /s/ and the two different Syntax conditions, first person singular 'ik' and third person singular 'zij'.*

Figure 2 shows that even the long consonant /s/ as a preceding context elicited a considerable number of /t/-responses in the first-person singular Syntax condition. This can be explained by the fact that the following fricative contexts "vaak" and "soms" received many /t/-responses, presumably because the slope at the end of the consonant /s/ before a following fricative can easily be interpreted as a frication /t/. (Note that the mean percentage of /t/-responses for the long consonant Coda Signal was 37.4% with following fricative contexts "vaak" and "soms" present in the analysis and 22.5% without).

RT range analyses

To examine whether the influence of Syntax on the perception of the Coda Signal is stable over time, or increases or reduces, a reaction time split was applied. Because reaction times differed considerably for the different conditions (varying from 526 ms for *zij* Syntax, /s/ Context and strong frication coda to 760 ms for *ik*, /n/, strong frication) we did not use a fixed cut-off point. With a fixed cut-off point "difficult" cells with a long reaction time would hardly contribute any data points to the fast-reaction-time range. Therefore, reaction times were ranked on speed for each participant individually and for all cells of the three-factor design (Syntax, Preceding Context and Coda Signal). Subsequently, these ranked RTs were divided into two RT ranges: the 40% fastest responses were considered Fast and the 40% slowest responses were considered Slow. Reaction times in the "middle" (40-60%) were omitted from this analysis. A repeated measures analysis on the logistically transformed percentages of /t/-responses was performed with Syntax, Coda Signal, Preceding Context, and RT range as predictors. This analysis showed that responses in the fast RT range were significantly more often a /t/-present response than slow responses [RT range: $F(1,20)= 16.8$, $p = .001$; fast: 76.1%; slow: 71.2%]. There were also significant interactions of Syntax by RT range [$F(1,20)= 9.3$, $p < .01$], and of Syntax by Preceding Context by Coda Signal by RT range [$F(4,80)= 4.4$, $p < .01$]. All other interactions with RT range were insignificant ($p > .05$).

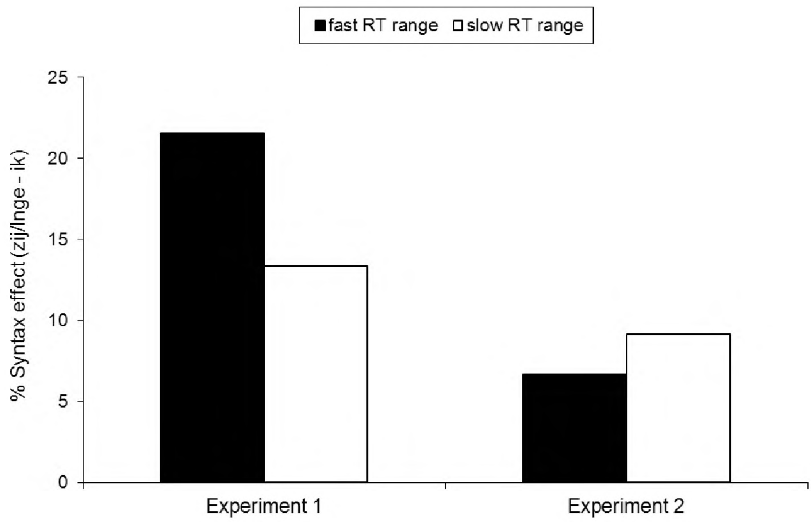

Figure 3: *Syntax effect in fast and slow reaction time (RT) ranges in Experiment 1 and Experiment 2.*

The nature of the interaction between Syntax and RT range was determined by examining the effect of Syntax on both levels of RT range. Both *t*-tests showed a significant effect of Syntax in the fast and slow RT ranges [$p_{max} < 0.001$]. Thus, the interaction must be due to the difference in size of the effect of Syntax for the two different RT ranges. We calculated the size of the syntax-effect by subtracting the percentages of /t/-responses with Syntax *ik* from those with Syntax *zij* for the two levels of RT range (Fast and Slow, see Figure 3). This revealed the Syntax effect to be significantly larger for the fast response times (21.2%) than for the slow response times (13.4%) [$t(20) = 3.0$, $p < 0.01$]. In other words, the sentential context effect, syntax in this case, does not remain stable over time, but decreases.

The Syntax by RT range interaction was further qualified by a four-way interaction of Syntax by RT range by Preceding Context by Coda Signal. Table 2 shows the source of this interaction. The effect of Syntax is larger in the fast RT range than in the slow RT range for all combinations of Preceding Context and Coda Signal but the Coda Signal "closure only" with the Preceding Context /s/, where there is a trend in the opposite direction.

Table 2: *Effect of Syntax on the mean percentage of /t/-responses for different Preceding Contexts /n/ and /s/, Coda Signals in the fast and slow RT ranges in Experiment 1.*

		fast RT range	slow RT range	
Preceding Context	Coda Signal	Syntax effect (zij-ik)	Syntax effect (zij-ik)	Syntax*RT range (F-values and significance)
/n/	full t	13.5	7.9	1.1
	strong frication	29.5	18.2	4.5*
	weak frication	29.4	27.0	0.1
	closure	35.1	13.5	5.6 *
	long consonant	43.9	29.4	4.7*
/s/	full t	1.7	0	0.2
	strong frication	4.3	0.8	0.6
	weak frication	7.3	7.0	0.1
	closure	5.9	20.2	3.8
	long consonant	41.5	10.3	9.1 *

Note: * = $p < 0.05$

Discussion

The probability of /t/-responses was influenced by the syntactic context of the sentence. There were more /t/-responses in the sentences that were /t/-biased by the preceding third-person singular pronoun *zij* than in the sentences that were biased towards the absence of a /t/ by the preceding first-person singular pronoun *ik*. Furthermore, this experiment replicated the two findings of Mitterer and Ernestus (2006): listeners took the preceding context into account in their perception and gave more /t/-responses after /s/ than after /n/, and they took the subphonemic detail into consideration in that coda signals with clear acoustic cues for the presence of a /t/ received more /t/-responses and codas with clear acoustic cues for the absence of a /t/ received fewer /t/-responses.

Of crucial importance is the evidence of the reaction time split testing the influence of sentential context on the identification of /t/s in fast and slow responses. Explanations of sentential context effects based on feedback suppose that lexical information flows back down and permanently modifies the phonological encoding processes. This implies that it is not possible for effects of this information feedback to become weaker over time. But this is exactly what the current result shows: the effect of sentential context becomes weaker in slower responses.

Although the results are at odds with the predictions of a feedback account, they do not—by themselves—provide support for autonomous models. As stated in the introduction, no clear prediction could be derived for the time course of the syntactic context effect. Van Alphen & McQueen (2001), who found a similar pattern of results, argued that a possible reason for the decrease in the influence of contextual information is that this information can only be used for a limited time. If syntactic processing has finished before the identification decision is reached, a syntactic bias may no longer be available to influence phonetic decisions. The mean reaction time for the slow responses was 898 ms (*SD* = 417 ms) after the offset of the coda signal, and the crucial syntactic information (the pronoun *ik* or *zij*) was already presented before the target verb. Therefore, it is very likely that syntactic processing was already completed before a phonetic decision was reached in case of a slow reaction time, so that the syntactic information would have influenced the identification of the coda signal to a lesser extent. If this account is correct, we can now derive a prediction for the time course of the syntactic context effect: It should depend on the relation of the timing of syntactic processing relative to phonetic decision-making. Hence, the time course of the syntactic context effect should change if this relation is changed. One way to test this experimentally is to change the word order of the sentences so that the syntactic information is presented *after* the target sound. This was tested in Experiment 2.

Experiment 2

In this experiment, we changed the word order so that the syntactic bias occurred after the critical target word in the sentence frame. Because of the limitations of Dutch grammar, the crucial syntactic information (e.g., the subject of the sentence in the form of a pronoun or common name) cannot appear very late in the sentence. In Dutch, the subject and the finite verb are usually directly adjacent (unless a subordinate clause intervenes), but they can trade places. For instance, in sentences starting with an adverb, the order of the subject and verb is inverted: *Soms ren(t) ik naar huis* ('Sometimes run(s) I home'). The pronoun *ik* cannot be placed further from the verb *ren*. To keep the phonological context following the reduced /t/ identical, the common name *Inge* was used instead of the pronoun *zij* ('she'), so that the parallel sentence with a /t/ bias was *Soms ren(t) Inge naar huis* ('Sometimes run(s) Inge home').

With this manipulation, it possible to examine whether the influence of syntactic information is smaller when it is presented later and can thus only be processed later. We can also again examine whether syntactic information can only influence phonetic decisions in a limited time window, and whether the syntactic context effect rises, decreases or remains stable over time.

Method

Participants

Twenty-one members of the Max Planck Institute's subject pool, none of whom had participated in Experiment 1, took part in this experiment. The participants were native speakers of Dutch and none reported any (history of) hearing problems. All were volunteers and were paid a small fee for their participation.

Materials

The same target verbs were used as in Experiment 1. Again, these verbs were placed in sentences with the crucial difference that the subject of the sentence occurred after the target verb. The same male native speaker of Dutch as in Experiment 1 recorded the sentences, schematized in Table 3, several times. The verbs in the sentences were pronounced with or without a final /t/ according to the reading list given to the speaker. As in Experiment 1, the utterances were used as templates for synthesis via a Klatt (1980) synthesizer. For the words that also occurred in Experiment 1, the parameter settings from Experiment 1 were used as a starting point for the settings in Experiment 2 and adjusted wherever needed to model the natural utterances. The main difference for the target verbs and adverbs across the two experiments was their duration and the pitch contour. For example, the adverbs were now utterance-initial instead of utterance-final. Therefore, the adverbs were shorter than in Experiment 1, in which they were subject to phrase-final lengthening. Otherwise, stimuli were created in the same way as in Experiment 1 and the same five coda signals were used.

Table 3: *Sentence frame for the stimuli in Experiment 2*

Adverb	Target verb	Subject	Place/object
nauwelijks /ˈnɑuʋələks/	blaas /blaːs/ (1.62)	ik /ɪk/	een noot /ən̩not/
langzaam /ˈlɑŋzaːm/	kreun /krøːn/ (1.46)		op maandag /ɔpmaːndɑɣ/
moeizaam /ˈmujzaːm/	bloos /bloːs/ (1.21) ..t		op school /ɔp sxol/
soms /ˈsɔms/	zoen /zun/ (1.26) ...Ø	Inge /ɪŋ̣ə/	mijn moeder /mɛɪn mudər/
vaak /ˈvaːk/	ren /rɛn/ (1.95)		naar huis /naːr hœys/
	kus /kʏs/ (1.75)		haar tante /haːr tɑntə/

Note. English translations are *nauwelijks* 'hardly', *langzaam* 'slowly' *moeizaam* 'with difficulty', *soms* 'sometimes', *vaak* 'often', *blaas* 'blow', *kreun* 'moan', *bloos* 'blush', *zoen* 'kiss', *ren* 'run', *kus* 'kiss', *ik* 'I', *Inge* 'Inge' (common Dutch first name), *een noot*' 'a note', *op maandag* 'on Monday', *op school* 'at school', *mijn moeder* 'my mother', *naar huis* 'home', *haar tante* 'her aunt'. The numbers in brackets are the lemma log frequencies for the words taken from CELEX (Baayen, et al., 1995).

Procedure and design

These were as for Experiment 1.

Results

Analysis of number of /t/-responses

As in Experiment 1, reaction times faster than 100ms (after offset of the target verb), and all cases in which no response was given were discarded, leading to the removal of 1.6 % of the trials.

The mean percentages of /t/-responses are shown in Figure 4. A repeated measures analysis on the percentages of /t/-responses was again performed with Syntax, Coda Signal, and Preceding Context as predictors. As Figure 4 clearly indicates, Preceding Context /s/ elicited more /t/-responses than Preceding Context /n/ [$F(1,20)= 72.8$, $p < 0.001$; /s/: 71.8%, /n/: 35.8 %]. Listeners were also more likely to report the presence of a /t/ if this interpretation led to a syntactically correct sentence [Syntax: $F(1,20)= 33.8$, $p < 0.001$; ik: 49.9%, Inge: 57.7%] and if the Coda Signal carried more acoustic information for the presence of a /t/ [$F(4,80)= 96.7$, $p < 0.001$].

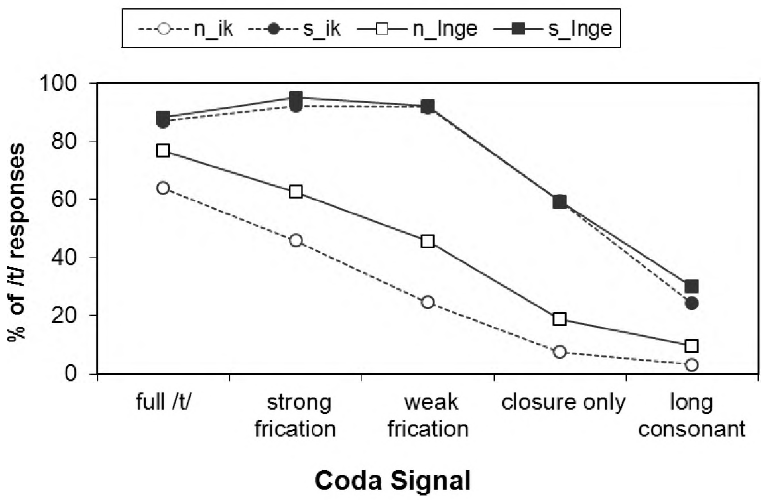

Figure 4: *Percentages of /t/-responses in Experiment 2 for each Coda Signal. The lines indicate the two different Preceding Contexts /n/ and /s/ and the two different Syntax conditions, first person singular 'ik' and third person singular common name 'Inge'.*

Additionally, there were significant interactions of Syntax and Preceding Context [$F(1,20)= 28.57$, $p < 0.001$] and Preceding Context and Coda Signal [$F(4,80)= 28.8$, $p < 0.001$] . The interaction between Syntax and Coda Signal and the three-way interaction failed to reach significance [$F<1$ and $F(4,80) = 2.2$, $p = 0.75$]. To examine the interaction of Syntax and Preceding Context, we tested the effect of Syntax on both levels of the factor Preceding context. This showed that the effect of Syntax was highly significant for the /n/-context ($t(20) = 6.9$, $p < 0.001$) but not for the /s/-context ($t(20) = 1.4$, $p > 0.1$). The interaction between Preceding Context and Coda Signal is due to the effect of Context being larger on the more ambiguous steps in the middle of the /t/-∅ continuum than at the endpoints.

RT range analyses

Again, we also examined whether the influence of syntactic information on the identification of the ambiguous target signal remains stable over time. The same reaction time split as in Experiment 1 was applied to divide slow and fast responses, and a repeated measures analysis on the percentages of /t/-responses was performed with Syntax, Coda Signal, Preceding Context, and RT range as predictors. This analysis showed neither a main effect of RT range nor any significant interactions with Syntax and RT range [Fs < 1]. While there were significant interactions of Preceding Context and RT range [$F(1,20)= 12.4$, $p < 0.01$], Coda Signal and RT range [$F(4,80)= 6.7$, $p < 0.001$], and of Preceding Context by Coda Signal by RT range [$F(4,80)= 2.9$, $p < 0.05$], the effect of Syntax was stable over time.

Discussion

Once again, syntax, preceding context and subphonemic detail influenced the probability of /t/-responses. Importantly, syntax influenced the /t/-responses even though the crucial syntactic information appeared after the ambiguous signal that had to be identified. Thus, more /t/-responses were given for sentences that became /t/-biased because of the following common name *Inge* than for sentences that continued with a first person singular pronoun *ik*. Listeners also took the preceding context into account, giving more /t/-responses after /s/ than after /n/, and they took subphonemic detail into account, giving more /t/-responses for coda signals with clear acoustic cues for the presence of a /t/ than for codas with clear acoustic cues for the absence of a /t/.

Our principal interest was again whether the sentential context effect on the identification of /t/s differed between fast and slow responses. No difference in the influence of Syntax was found between fast and slow responses in this experiment. This contrasts with Experiment 1 in which such an effect of RT range was found.

When the two experiments were analysed together in a repeated measurements analysis with Experiment as a between subjects factor, an interaction between Syntax and Experiment [$F(1,40) = 11.8$, $p < 0.01$] was found; the Syntax effect was larger in Experiment 1 (17.3%) than in Experiment 2 (7.9%). There was also an interaction between Syntax, RT range and Experiment [$F(1,40) = 7.7$, $p < 0.01$] showing that the influence of Syntax for the fast and slow response RT ranges differed across the two experiments. This shows that the time course of the syntactic context effect differed—as predicted by the autonomous account—between the two experiments.

General discussion

Syntactic information influences the identification of reduced /t/s, but the size of this influence varies over time and depends on when the syntactic information is presented: before or after the ambiguous sound. This finding is not in line with a feedback account of speech processing which predicts the influence of syntactic information to remain stable over time. In Experiment 1, when the disambiguating syntactic information in the sentence appeared in the word preceding the ambiguous verb-final /t/, listeners used that information in their identification decisions but did so to a greater extent in their fast responses than in their slow responses. There were more /t/-responses when the preceding word was a third-person singular pronoun (*zij* 'she') than when the preceding word was the first-person singular pronoun (*ik* 'I'). This effect was significant in both the fast and slow RT ranges, but larger in the fast RT range than in the slow RT range. However, when the disambiguating information appeared directly after the ambiguous sound, as in Experiment 2, listeners did use this information, but to a lesser extent than in Experiment 1.

Together, the results of the two present experiments suggest that the influence of contextual information on the identification of verb-final reduced /t/s is not stable over time. Moreover, the time-course of the syntactic bias effect differs depending on when the syntactic information is presented in relation to the ambiguous target signal. A feedback account cannot explain this kind of result. If the syntactic bias effect is the consequence of information feedback from higher levels to the prelexical level, then it should increase until it reaches an asymptote, and it should not have different time-courses in different experiments. In both experiments the syntactic information seems to have been available, as listeners used it to make their (biased) decisions. However, in Experiment 1, where syntactic information became available early, the influence of sentential information became weaker over time, whereas in Experiment 2, where syntactic information became available after the target signal, the influence of sentential context did not change over time. A feedback account cannot explain the current results as they seem to depend on when the information becomes available.

It needs to be stressed that our stance against on-line feedback does not mean that we argue against the use of higher-level knowledge or that we want to marginalise its role in on-line sentence comprehension. Another line of research clearly shows that sentential context effects are real and occur in on-line processing. Consider the following example from a study by van Berkum, Brown, Zwitserlood, Kooijman, and Hagoort (2005): *The burglar had no trouble locating the secret family safe. Of course it was situated behind a* . . . In two event-related brain potential (ERP) experiments and one self-paced reading study van Berkum et al. (2005) showed that listeners and readers indeed make predictions and anticipate upcoming words in discourse (in this case *painting*). In their first ERP experiment, participants listened to Dutch stories that supported the prediction of a certain noun (like the sentence above). To investigate whether listeners indeed anticipated the expected noun (e.g., *painting*) by the time they had heard the indefinite article, stories in the test condition were continued with a gender-marked adjective, the inflectional suffix of which did not correspond with the expected noun's syntactic gender. Compared with consistently inflected adjectives, this elicited a reliable positive deflection in the ERP waveforms which appeared directly at the inflection. When participants heard the same sentences without the wider context that supported a particular prediction, the ERP effect disappeared. In addition, when prediction-inconsistent adjectives were again presented in a self-paced reading study, readers slowed down before the noun. These results suggest that listeners use world-knowledge in sentence comprehension on-line. We maintain, however, that this use does not imply top-down feedback.

Finally, comparing our results with natural ambiguous phonemes with those of van Alphen and McQueen (2001) with artificial ambiguous phonemes, we found that the overall size of the context effect is larger in our experiments than in the experiments of van Alphen and McQueen. This means that the syntactic biases are given greater leverage to influence the final percept when restoring reduction in spontaneous speech. This greater leverage, however, still arises in an integration of phonological processing and syntactic processing and not in a top-down influence of syntactic processing on phonological processing.

Despite the difference in effect size, our results converge with those of van Alphen and McQueen (2001) with respect to the question of top-down feedback. Our first experiment (when the contextual information preceded the ambiguous target sound) showed a context effect that became weaker over time, just as in van Alphen and McQueen's Experiment 1 and 4, which were similar in construction to our study. When the contextual information followed the target sound (as in our Experiment 2 and van Alphen and McQueen's Experiment 2), we found a context effect which became larger, though not significantly, over time, but this effect was smaller than for Experiment 1. Van Alphen and McQueen found no significant context effect in any RT range in their second experiment, although the context effect seemed to increase over time. Overall, these patterns are quite consistent with each other. This shows that the challenges of reduction in spontaneous speech do not invoke more use of top-down processing.

In sum, it can be concluded that feedback accounts of context effects cannot explain the current results. The results suggest that sentential context can influence phoneme decision, but it does not influence or change the prelexical representations as the influence of context can also decrease again. This is consistent with a model like Merge (Norris, et al., 2000). In the Merge model, phoneme decisions can be biased by syntactic context, but this bias can be undone since no prelexical representations have been changed. Furthermore, Merge predicts that the strength of this decision bias depends on when the contextual information is available and can be merged with other pieces of information. This is exactly what we found when the two experiments were compared. The decision bias is present in both experiments, but larger when the biasing information was available earlier.

References

Alphen, P. van., & McQueen, J. M. (2001). The time-limited influence of sentential context on function word identification. *Journal of Experimental Psychology: Human Perception and Performance, 27*(5), 1057-1071.

Baayen, H., Piepenbrock, R., & Gulikers, L. (1995). The CELEX lexical database (release 2) [CD-ROM]. Philadelphia, PA: Linguistic Data Consortium, University of Pennsylvania [Distributor].

Boersma, P., & Weenink, D. (2005). Praat: doing phonetics by computer (Version 4.3.07) [Computer program]. Retrieved April 2005, from http://www.praat.org/.

Borsky, S., Shapiro, L. P., & Tuller, B. (2000). The temporal unfolding of local acoustic information and sentence context. *Journal of Psycholinguistic Research, 29*, 155-168.

Borsky, S., Tuller, B., & Shapiro, L. P. (1998). "How to milk a coat": The effects of semantic and acoustic information on phoneme categorization. *Journal of the Acoustical Society of America, 103*, 2670-2676.

Connine, C. M. (1987). Constraints on interactive processes in auditory word recognition: The role of sentence context. *Journal of Memory and Language, 26*, 527-538.

Connine, C. M. (1990). Effects of sentence context and lexical knowledge in speech processing. In G. T. M. Altmann (Ed.), *Cognitive models of speech processing: Psycholinguistic and computational perspectives* (pp. 281-294). Cambridge, MA: MIT Press.

Connine, C. M., Blasko, D. G., & Hall, M. (1991). Effects of subsequent sentence context in auditory word recognition: Temporal and linguistic constraints. *Journal of Memory and Language, 30*, 234-250.

Cutler, A., & Norris, D. (1979). Monitoring sentence comprehension. In W. E. Cooper & E. C. T. Walker (Eds.), *Sentence processing: Psycholinguistic studies presented to Merrill Garrett*. Hillsdale, NJ: Erlbaum.

Forster, K. I. (1976). Accessing the mental lexicon. In R. J. Wales & E. Walker (Eds.), *New approaches to language mechanisms* (pp. 257-287). Amsterdam: North-Holland.

Ganong, W. F. (1980). Phonetic categorization in auditory word perception. *Journal of Experimental Psychology: Human Perception and Performance, 6*, 110-125.

Gaskell, M. G., & Marslen-Wilson, W. D. (1997). Integrating form and meaning: A distributed model of speech perception. *Language and Cognitive Processes, 12*, 613-656.

Gaskell, M. G., & Marslen-Wilson, W. D. (2001). Lexical ambiguity resolution and spoken word recognition: Bridging the gap. *Journal of Memory and Language, 44*, 325-349.

Guy, G. R. (1980). Variation in the group and the individual. In W. Labov (Ed.), *Locating language in time and space* (pp. 1-36). New York: Academic Press.

Janse, E., Nooteboom, S. G., & Quené, H. (2007). Coping with gradient forms of /t/-deletion and lexical ambiguity in spoken word recognition. *Language and Cognitive Processes, 22*, 161-200.

Klatt, D. H. (1980). Software for a cascade/parallel formant synthesizer. *Journal of the Acoustical Society of America, 67*, 971-995.

Kohler, K. J. (1990). Segmental reduction in connected speech in German: Phonological facts and phonetic explanations. In W. J. Hardcastle & A. Marchal (Eds.), *Speech production and speech modelling* (pp. 69-92). Dordrecht, the Netherlands: Kluwer Academic Publishers.

McClelland, J. L. (1987). The case for interactionism in language processing. *Attention and Performance, 12*, 3-36.

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

McQueen, J. M. (1991). The influence of the lexicon on phonetic categorization: Stimulus quality in word-final ambiguity. *Journal of Experimental Psychology: Human Perception and Performance, 17*, 433-443.

McQueen, J. M., Norris, D., & Cutler, A. (1999). Lexical influence in phonetic decision-making: Evidence from subcategorical mismatches. *Journal of Experimental Psychology: Human Perception and Performance, 25*, 1363-1389.

Mitterer, H., & Ernestus, M. (2006). Listeners recover /t/s that speakers reduce: Evidence from /t/-lenition in Dutch. *Journal of Phonetics, 34*, 73-103.

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

Norris, D., McQueen, J. M., & Cutler, A. (2000). Merging information in speech recognition: Feedback is never necessary. *Behavioral and Brain Sciences, 23*, 299-325.

Sumner, M., & Samuel, A. G. (2005). Perception and representation of regular variation: The case of final /t/. *Journal of Memory and Language, 52*, 322-338.

van Berkum, J. J. A., Brown, C., Zwitserlood, P., Kooijman, V., & Hagoort, P. (2005). Anticipating upcoming words in discourse: Evidence from ERPs and reading times. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*, 443-467.

Summary and conclusions

Chapter 6

Summary

This dissertation investigated the perception of casual speech in a second language. Its main topic is the role of familiarity in the perception of casual, connected speech: does it matter whether second language listeners are familiar with a casual speech process from their native language or not? First, German listeners' perception of /t/-reduction in Dutch was examined. As the casual speech process of /t/-reduction occurs in both Dutch and German it is expected that Germans with Dutch as a second language do not have major difficulties in coping with /t/-reduction in Dutch. Second, this dissertation investigated how second language listeners perceive casual speech processes in their second language with which they are not familiar from their native language. The insertion of /r/ between two vowels, as in the phrase *idea is*, occurs frequently in British English, but does not occur in Dutch. Therefore, this study assessed Dutch and English listeners' perception of intrusive /r/ in British English speech.

Given the suspicion that second language learners may have problems with compensation for casual speech processes, it is reasonable to assume that they will make more use of higher-level knowledge to decode the signal. Higher-level knowledge such as lexical and sentence context information can be used by listeners—also in the first language—to resolve ambiguities in the speech input created by casual speech processes. Second language listeners might (need to) use higher-level knowledge more than native listeners to compensate for casual speech processes, because they are not be able to make use of the bottom-up signal to the same extent as native listeners. Therefore, this dissertation also investigated what the influence was of higher-level knowledge on the perception of both /t/-reduction and intrusive /r/. Finally, this dissertation investigated how higher-level knowledge is actually integrated in spoken-word recognition when ambiguity arises due to a casual speech process as this is still a matter of debate.

(Un)familiarity of a casual speech process

The production and perception of /t/-reduction—the familiar casual speech process—was studied in a production experiment and two 2-alternative forced-choice (2AFC) tasks reported in Chapter 2. First, a production experiment investigated how similar or dissimilar the reduction patterns really are in Dutch and German. In Experiment 2.1, native speakers of German and native speakers of Dutch performed in two production tasks. This experiment compared incidence of /t/-reduction in Dutch and German in verbs and proper nouns after the preceding context /n/ or /s/. A previous corpus study on Dutch casual speech had shown that /t/-reduction occurs frequently after /s/ and is not likely after /n/. The results showed that Dutch and German are very similar with regard to /t/-reduction in proper nouns. In both languages, /t/-reduction was more likely after /s/ than after /n/. However, reduction of a morphological /t/ in verbs was different in German than in Dutch. In German, /t/-reduction was independent of the preceding context when the /t/ was a verbal inflection, whereas in Dutch, verbs patterned just as nouns, with more deletion after /s/ than after /n/.

Subsequently, Experiment 2.2 and 2.3 investigated the perception of /t/-reduction in Dutch by native speakers of German proficient in Dutch as a second language and native speakers of Dutch. In both experiments, listeners decided whether or not target words ended in /t/. In Experiment 2.2, the target words were nouns and adjectives. Lexical information produced a prediction: interpreting /t/ made the target item a correct Dutch word or not: *charmant* ('charming') and *kanon* ('gun') are real words, but *charman* and *kanont* are not. In Experiment 2.3, the target words were verbs (e.g., *ren* 'run', *kus* 'kiss'). This made it possible to use grammar (preceding *ik* 'I', *zij* 'she') to predict whether or not the ending should be /t/; the Dutch present tense third person singular inflection is /t/ while the first person inflection is null, so that after *ik* the grammatical forms are *ren* and *kus*, while after *zij* the grammatical forms are *rent* and *kust*. The relative sensitivity of native and second language listeners to the acoustic evidence for /t/, and their sensitivity to the preceding phoneme, in contexts where these types of higher-level information require versus rule out a /t/, was compared.

The German non-native listeners' responses were, overall, sensitive to the same factors as the Dutch native listeners' responses. They used the "bottom-up" information—phonetic form of the coda and phonological context—to a similar degree as the Dutch listeners. However, the German second language learners had a tendency to report significantly <u>more</u> /t/ than the native listener baseline when the /t/ was a verbal inflection. Additionally, the German learners were more influenced by higher-level constraints only when /t/ was a verbal inflection. A comparison with the production Experiment 2.1 suggests an explanation for this pattern. Reduction of morphologically functional /t/ (such as a verbal inflection) shows a different pattern in German than in Dutch. In the light of this comparison, the tendency of the second language listeners to produce more /t/-responses, especially when the syntax predicted a /t/, suggests that they were sensitive to the differences between German and Dutch, and tried to compensate for it. Interestingly, this also seems to indicate that small differences in the first and second language patterns of casual speech processes seem to be sufficient to hinder second language listening.

The perception of intrusive /r/ in British English—a completely unfamiliar casual speech process for Dutch listeners—was studied in a 2AFC task, a cross-modal priming and an eye tracking experiment reported in Chapter 3 and 4. In Chapter 3, acoustic analyses first established that there are indeed differences between an onset /r/ as in *extra rice* and an intrusive /r/ as in *extra[r]ice*: onset /r/s were longer and displayed a larger intensity decrement from the preceding vowel to the lowest point. Then, three 2AFC experiments reported whether native British English, American English and Dutch listeners proficient in English can indeed distinguish between the two types of /r/ on the basis of acoustic evidence. Neither Dutch nor most American English dialects have intrusive /r/, but American listeners were expected to be more sensitive to this phenomenon in the British version of their own language, as previous studies indicate that dialect differences are generally easy to overcome. In the present experiments participants heard British English sentences in which the duration of /r/ was manipulated across a word boundary (*saw (r)ice* or *saw more (r)ice*), and orthographic and semantic biases favoring the r-initial interpretation were present or absent. British English 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. American and Dutch listeners relied less on durational cues than the British listeners, and were affected by orthography, reporting /r/ more often after *saw* than after *more*. Additionally, Dutch listeners were susceptible to semantic bias in the sentences. Thus although intrusive /r/ causes perceptual problems for listeners with another language or another dialect, a language difference induces greater difficulties.

Chapter 4 investigated whether non-native dialect and non-native language listeners could exploit the subtle phonetic differences between an onset /r/ and intrusive /r/ in lexical processing. As shown in Chapter 3, there are pronunciation differences that potentially allow listeners to distinguish whether the /r/ in *Canada (r)aided* is an onset /r/ (i.e., *raided* is the intended verb) or an intrusive /r/ (i.e., *aided* is the intended verb). The impact of this difference on perception was tested with cross-modal priming experiments in which the priming was tested of the visual targets *raided* and *aided* by an intended verb *aided*, preceded by an intrusive /r/, versus by an intended verb *raided*. To assess whether resolving the ambiguity created the casual speech process of intrusive /r/ depends on familiarity with this process, British English listeners, American English listeners, and Dutch learners of English were tested.

The results showed that native British English listeners effectively exploit the subtle phonetic differences between an intrusive /r/ and an onset /r/, and are thus able to resolve potentially ambiguous phrases such as *Canada (r)aided*. American English listeners, with less experience with intrusive /r/, but with the same language, were also able to distinguish intrusive /r/ from onset /r/, and their processing of the potentially ambiguous sequences was indistinguishable from that of the British participants. The Dutch learners of English, however, although they too showed some sensitivity to the phonetic cues, appeared unable to effectively distinguish between intrusive and onset /r/s. An onset /r/ only primed r-initial words (e.g., *raided*), but an intrusive /r/ primed both vowel-initial (e.g., *aided*) and r-initial target words (e.g., *raided*), to a similar degree.

To learn more about this pattern, an additional eye-tracking study was conducted. In this experiment, Dutch learners of English were presented with the same ambiguous sentences as in the cross-modal priming experiment in a visual world paradigm with printed words. That is, they saw four printed words on the screen and were instructed to click on the word that they heard in the sentence. For example, they heard "*Canada(r)aided....*" and the printed words on the screen were *aided*, *raided* and two phonologically unrelated distractors. Regardless of whether the sentence they heard contained an intrusive /r/ or an onset /r/, the Dutch listeners had a strong preference for r-initial target words. With more processing time, they began to show influence of the type of /r/ in the sentence, but even then this influence remained much smaller than their preference for r-initial words.

Higher-level information in spoken-word recognition

As noted above, casual speech processes might induce second language listeners to make more use than native listeners of higher-level knowledge such as lexical and sentence context information, thereby compensating for an inability to effectively use the bottom-up signal to resolve these processes, in the way native listeners do. It is still a matter of debate how higher-level knowledge is integrated in the process of spoken-word recognition. Two basic types of models have been proposed in the literature. The first type of model assumes that information can flow from higher levels back to the prelexical level and directly influence perception. According to the other type of model, information flows from the prelexical level to the lexicon without feedback; in other words, information can only flow from the bottom up. Research focusing on this issue has usually employed artificially generated ambiguous speech input, but it might be the case that when listeners have to compensate for ambiguous speech due to a casual speech process they do this differently than when they have to compensate for arbitrary ambiguous signals. This dissertation therefore investigated how the integration of information works in the case of a casual speech process. In two 2AFC experiments the casual speech process of /t/-reduction was used to investigate syntactic context effects on the identification of word-final reduced /t/s in Dutch verbs. Native listeners of Dutch decided whether or not the verb in a sentence such as *Maar zij ren(t) soms* ('But she run(s) sometimes') ended in /t/. In Dutch, the presence or absence of /t/ after the verb stem is the difference between third- and first-person singular present. In both experiments, listeners perceived more /t/s in sentences in which this is syntactically correct. In Experiment 5.1, the biasing syntactic context preceded the target verb; in Experiment 5.2, it followed the target verb. In both experiments a syntactic bias effect was found. It was examined whether this bias was larger or smaller in fast vs. slow reactions. Modeling studies using top-down models predicted that the influence of the higher-level knowledge should be larger in slow than in fast responses. However, this bias was larger for fast than for slow responses in Experiment 5.1, and no difference between fast and slow responses was found in Experiment 5.2. The results hence contradict the predictions of top-down models.

This suggests that sentential context can influence phoneme decision, but does not influence prelexical representations, given that the influence of context can also decrease again. In other words, information feedback is not necessary to compensate for the ambiguities created by /t/-reduction.

Conclusions

This dissertation investigated the perception of casual speech in a second language. It provides new insights into the processing of non-native speech. Further, the results from this dissertation raise some interesting new questions.

First, the findings provide new insights into the way non-native listeners cope with casual speech processes in their second language—processes that are either familiar or unfamiliar. The combined results of the studies with /t/-reduction and intrusive /r/ motivate the inescapable conclusion that casual speech processes are subject to the same tight coupling with native language experience as every other aspect of second language speech perception. The second language listeners' performance in the experiments with the familiar casual speech process (/t/-reduction in Chapter 2) related quite differently to the native performance baseline than the performance of the second language listeners in the experiments with the unfamiliar casual speech process (/r/-intrusion in Chapter 3 and 4). The listeners' familiarity and experience with the casual speech processes of their native language crucially determined the way they dealt with the effects of casual speech processes in their second language.

In the case of the familiar casual speech process, the non-native German listeners were broadly sensitive to the same range of factors in the Dutch input as the Dutch native listeners: they were in general more likely to report the presence of a /t/ when it followed /s/ rather than /n/, when it formed a grammatical string, and when it made a real word. Their similar sensitivity to the acoustic realisation of /t/, and to the probabilities as a function of preceding phonetic context, reflect the parallelism between Dutch and German in the patterning of /t/-deletion. But here it can also be seen that familiarity even matters on a "microscopic" level, as the German listeners displayed less native-like performance for /t/-reduction in Dutch verbs, putatively because it patterns differently in German. In the case of the unfamiliar casual speech process, however, the non-native Dutch listeners showed quite a different pattern of sensitivity than the British English native listeners: while the native listeners based their responses overall on the acoustic characteristics of the stimuli, were quite insensitive to the sentence meaning and certainly were not inclined to report /r/ if there was none in the orthographic representation, the second language listeners mimicked none of these patterns. They made relatively little use of the acoustic information, they were significantly influenced by the sentence meaning, and they were far more likely to report /r/ when the orthography contained none. The research described in Chapter 4 showed that in tasks requiring more lexical processing than phonetic processing, Dutch listeners also could not process intrusive /r/ in a native-like manner.

It can be concluded, therefore, that the casual speech patterns of the native language influence interpretation of casual speech processes in a second language. A process familiar from the native language can be processed easily; an unfamiliar process in the second language with no corresponding process in the native language is very hard to adapt to. This has significant implications for the understanding of speech, since native listeners are very good at adjusting their spoken-word recognition to compensate for the effects of casual speech processes. The present results predict that German listeners to Dutch should be able to judge whether they are hearing Dutch *kas* ('greenhouse') or *kast* ('cupboard') with similar efficiency to native Dutch listeners. The same should be true for any second language listener confronted with a casual speech process that they are already familiar with from their native language. Dutch listeners to English, however, are unable to tell an intrusive /r/ from an intended /r/ (as shown in Chapter 3), and this leads them into misrecognition of words (as shown in Chapter 4). Again, this pattern will hold likewise for other cases in which a casual speech process in a second language is unfamiliar to the second language listener.

However, the second language listeners also deviated from the native listeners' performance in the case of the familiar casual speech process in one respect: they were more likely than the native listeners to report /t/ when it was a morphological inflection. This is precisely the kind of /t/ that is less likely to be reduced in their native language. This pattern therefore raises the intriguing possibility, certainly worthy of future investigation, that listeners not only can deal easily with a casual speech process of the second language that matches a native process, but they are even sensitive to subtle differences in its distribution of occurrence across the languages, and accommodate their perceptual responses accordingly.

What does this research mean for second language learners? The dissertation brings them both good and bad news. On the positive side, casual speech processes they are already familiar with from their native language do not create major problems for speech perception in their second language. On the negative side, if a casual speech process does not occur in their native language, non-native listeners are likely to have trouble with the ambiguous speech in their second language created by this casual speech process. Thus, the current study adds another item to the long list of complications that makes speech recognition in a second language such a challenge.

Samenvatting

Zoals veel mensen weten uit eigen ervaring, kan het leren van een vreemde taal behoorlijk lastig zijn. Het leren van een nieuwe taal betekent niet alleen het je eigen maken van nieuwe woorden en uitdrukkingen, maar ook een andere grammatica en misschien zelfs een ander schrift of alfabet. Nederlanders worden vaak geprezen om de hoeveelheid vreemde talen die ze spreken, maar wat daarbij al gauw vergeten wordt is dat ze voornamelijk talen van buurlanden beheersen, die bovendien erg op het Nederlands lijken: Duits en Engels. Voor Nederlanders is het relatief gemakkelijk om Duits en Engels te leren omdat het alfabet, het vocabulair en de structuur van de talen erg op elkaar lijken. Spaans en Italiaans zijn voor Nederlanders al moeilijker te leren omdat de verschillen in vocabulaire en grammatica groter zijn. Japans en Arabisch zijn een nog grotere uitdaging, omdat Nederlanders dan ook nog een anders schrift moeten leren. Kortom, het leren van een andere taal is het makkelijkst als deze veel lijkt op je moedertaal; de nieuwe taal is dan al wat vertrouwd.

Behalve in vocabulaire of grammatica kan een tweede taal ook verschillen in de zogenoemde 'spontanespraakprocessen' die ze heeft. Dit zijn processen die voorkomen in spontane spraak die er voor zorgen dat woorden anders klinken dan wanneer ze zeer nauwkeurig worden uitgesproken. Bij spontane spraak—die het tegendeel is van de formele spraak van bijvoorbeeld een nieuwslezer—kan het bijvoorbeeld gebeuren dat klanken zich aanpassen aan een naburige klank. Het woord *tuinbank* wordt dan bijvoorbeeld uitgeproken als *tuimbank*; de *n* heeft zich aangepast aan de *b* en klinkt dan als een *m*. Dit noemt men 'assimilatie'. Daarnaast worden in spontane spraak bepaalde klanken soms niet of nauwelijks uitgesproken: zo kan *postbode* klinken als *posbode*. Dit wordt 'reductie' genoemd. Ook kan het voorkomen dat in spontane spraak juist extra klanken in een woord worden ingevoegd. Een voorbeeld is het woord *film* dat soms met een extra 'u'-klank wordt uitgesproken als *fillum*. Dit fenomeen heet 'insertie'.

In de afgelopen jaren heeft de psycholinguïstiek veel onderzoek gedaan naar de manier waarop moedertaalluisteraars omgaan met spontane spraak en de klankveranderingen die daar in voorkomen. Kort samengevat blijkt uit het onderzoek tot nu toe dat luisteraars erg goed zijn in het gebruik maken van de fonetische details in de spraak en het herkennen van de juiste woorden, zelfs als die woorden veranderd zijn ten opzichte van hun 'woordenboekuitspraak'. De kleine verschillen tussen de bedoelde klanken en de klanken die het resultaat zijn van een spontane spraakproces worden herkend door luisteraars, bijvoorbeeld bij assimilatie. Engelse luisteraars onderscheiden onder andere de /p/ uit het Engelse woord *ripe* in *ripe berries* van de naar /p/ geassimileerde /t/ in *right berries* (wat dan dus klinkt als *ripe berries*). Maar moedertaalluisteraars worden ook wel eens misleid. Engelse luisteraars reageerden bijvoorbeeld op de klank /p/ die was toegevoegd in het nonchalant uitgesproken *something* ('*sompthing*'). Als zelfs moedertaalluisters soms in de war worden gebracht door spontanespraakprocessen, hoe zal het tweedetaalluisteraars dan vergaan? Het onderzoek dat in dit proefschrift wordt gepresenteerd, ging in op deze vraag.

Eerst is gekeken naar hoe moedertaalluisteraars en bekwame tweedetaalluisteraars omgaan met /t/-reductie op het einde van een Nederlands woord. Het spontanespraakproces van /t/-reductie komt voor in verschillende talen, waaronder het Engels, Duits en Nederlands. De letter /t/ wordt bijvoorbeeld vaak niet of nauwelijks uitgesproken ná een /s/ en vóór een /b/ of /m/, zodat in de uitspraak van het Engelse *postman*, het Duitse *Postbeamter* en het Nederlandse *postbode* vaak amper de klank /t/ is waar te nemen. In een productie-experiment werd eerst onderzocht hoe vergelijkbaar de reductiepatronen eigenlijk zijn voor /t/-reductie in het Nederlands en Duits. Moedertaalsprekers van het Nederlands en Duits namen deel aan twee productietaken. Er werd gekeken naar het voorkomen van /t/-reductie in werkwoorden en eigennamen na een voorafgaande /n/ of /s/. Uit eerder onderzoek van Nederlandse spontane spraak bleek dat /t/-reductie vaak voorkomt na een /s/ en niet waarschijnlijk is na een /n/. De resultaten van het productie-experiment tonen aan dat het Nederlands en het Duits erg vergelijkbaar zijn met betrekking tot /t/-reductie in eigennamen. In beide talen is /t/-reductie waarschijnlijker na een /s/ dan na een /n/. Het voorkomen van reductie van een /t/ die een werkwoordsuitgang is, zoals de /t/ in *woont*/*wohnt* blijkt te verschillen in het Nederlands en het Duits. In het Duits was /t/-reductie onafhankelijk van de voorafgaande klank. In het Nederlands daarentegen had /t/-reductie hetzelfde patroon als voor de eigennamen, met meer reductie na een /s/ dan na een /n/.

Vervolgens werd onderzocht hoe /t/-reductie in het Nederlands wordt waargenomen door Nederlandse luisteraars en door Duitse luisteraars die ook goed Nederlands spreken. In twee perceptie-experimenten hoorden luisteraars Nederlandse zinnen en moesten ze bepalen of een zogeheten 'targetwoord' dat ze hoorden eindigde in de klank /t/ of niet. De laatste klank in dit targetwoord verschilde steeds in de mate waarin het klonk als een duidelijke /t/; er waren 5 stappen van een overduidelijke /t/-klank tot een volledig gereduceerde /t/. Met andere woorden, de fonetische vorm van de laatste klank in het targetwoord verschilde steeds. In het eerste experiment waren de targetwoorden zelfstandige en bijvoeglijke naamwoorden. Lexicale informatie (kennis van Nederlandse woorden) zorgde voor een voorspelling: het interpreteren van een /t/ maakte het targetwoord een correct Nederlands woord of niet: *charmant* en *kanon* zijn Nederlandse woorden, maar *charman* en *kanont* niet. In het tweede experiment waren de targetwoorden werkwoorden, zoals *ren* en *kus*. Dit maakte het mogelijk om aan de hand van grammatica te voorspellen of het werkwoord moest eindigen op een /t/ of niet. Als het werkwoord werd voorafgegaan door *ik* dan was *ren* grammaticaal correct en *rent* niet, en in het geval dat het werkwoord werd voorafgegaan door *zij* was *rent* juist correct en *ren* niet. Het doel van de experimenten was om te kijken wat de invloed was van akoestische evidentie voor een /t/, de voorafgaande klank (/n/ of /s/) en de lexicale of grammaticale informatie op het waarnemen van de /t/-klank en of dit verschillend was voor de Nederlandse moedertaalluisteraars en de Duitse tweedetaalluisteraars.

De Duitse tweedetaalluisteraars waren over het algemeen gevoelig voor dezelfde factoren als de Nederlandse moedertaalluisteraars. Ze maakten in dezelfde mate gebruik van zogenaamde "bottom-up" informatie—de fonetische vorm (van overduidelijk een /t/ tot een compleet gereduceerde /t/) en de voorafgaande klank (/n/ of /s/)—als de Nederlandse luisteraars. Echter, de Duitse luisteraars hadden wel de neiging om vaker /t/ als antwoord te geven als de /t/ een werkwoordsuitgang was dan de Nederlanders. Bovendien ondervonden de Duitse proefpersonen alleen meer invloed van grammaticale informatie en niet van lexicale informatie. De uitkomsten van het productie-experiment geven een verklaring voor dit patroon. Reductie van een /t/ met een grammaticale functie zoals bij een werkwoordsuitgang liet een verschillend patroon zien voor het Duits en het Nederlands. De twee perceptie-experimenten suggereren dat de Duitse tweedetaalluisteraars gevoelig waren voor de verschillen in /t/-reductie tussen het Duits en het Nederlands en hiervoor overcompenseerden door meer /t/-antwoorden te geven, vooral wanneer de grammatica dat voorschreef.

Vervolgens is er gekeken naar een ander spontanespraakproces, namelijk de zogenoemde 'intrusive /r/'. In het Brits Engels komt het voor dat een extra /r/-klank wordt ingevoegd tussen woorden, zodat *extra ice* klinkt als *extra rice*. Het Nederlands kent evenwel geen intrusive /r/. Er werd onderzocht hoe Nederlanders die ook goed Engels spreken deze intrusive /r/ waarnemen: horen zij een 'echte' /r/, of hebben ze door dat deze /r/ eigenlijk genegeerd kan worden?

Eerst werd door middel van akoestische analyses vastgesteld dat er inderdaad verschillen zijn tussen een echte /r/ zoals in *extra rice* of een intrusive /r/ zoals in *extra[r]ice*. Dit bleek inderdaad het geval te zijn: een echte /r/-klank duurt bijvoorbeeld langer. Daarna is in drie experimenten gekeken of moedertaalsprekers van het Brits, Amerikaans en het Nederlands ook inderdaad de twee soorten /r/ kunnen onderscheiden op basis van akoestische informatie. Het Nederlands en de meeste Amerikaanse dialecten hebben geen intrusive /r/, maar het was de verwachting dat de Amerikanen gevoeliger waren voor dit fenomeen in de Britse versie van hun moedertaal aangezien eerdere studies hebben uitgewezen dat verschillen tussen dialecten over het algemeen weinig problemen opleveren. In de experimenten kregen de proefpersonen Brits Engelse zinnen te horen waarin de duur van de /r/ was gemanipuleerd over een woordgrens (*saw (r)ice* of *saw more (r)ice*). Spelling en semantische informatie zorgden daarbij voor een voorkeur voor een /r/ interpretatie of niet. De Brits Engelse proefpersonen antwoordden categorisch; ze antwoordden *ice* na een korte /r/ en *rice* na een lange /r/ en ze werden niet beïnvloed door de semantische en spellingsmanipulaties. De Amerikaanse en Nederlandse luisteraars maakten minder gebruik van de duurinformatie dan de Britten, maar werden wel beïnvloed door spelling en gaven dus meer *rice* antwoorden na *saw* dan na *more*. Daarbij waren de Nederlanders ook gevoelig voor de semantische manipulatie in de zinnen. Dus hoewel de intrusive /r/ zowel interpretatieproblemen oplevert voor luisteraars met een ander dialect als voor luisteraars met een andere moedertaal, heeft die laatste groep er grotere moeite mee.

Ook is er gekeken of Britse, Amerikaanse en Nederlandse luisteraars gebruik konden maken van de subtiele verschillen tussen een 'echte' /r/ en een intrusive /r/ bij het verwerken van woorden. In de experimenten kregen de proefpersoen zinnen te horen met een intrusive /r/, bijvoorbeeld *Canada (r)aided…* of met een echte /r/ zoals in *Canada raided*. Vervolgens zagen ze op een computerscherm het woord *aided* of *raided*. De verwachting was dat als een proefperson *Canada (r)aided* had gehoord, en de intrusive /r/ juist had verwerkt, hij of zij sneller reageert op het woord *aided* dan op een controlewoord, en minder snel of juist langzamer op het woord *raided*. Voor

Cananda raided geldt precies het omgekeerde. De resultaten van de experimenten lieten zien dat de Britse proefpersonen het potentieel dubbelzinnige *Canada (r)aided* correct interpreteren en dus blijkbaar gebruik kunnen maken van de subtiele verschillen tussen de twee soorten /r/. Ook de Amerikaanse luisteraars konden de intrusive /r/ onderscheiden van de 'echte' /r/. Echter, de Nederlanders bleken geen goed gebruik te maken van de subtiele verschillen. Als ze *Canada raided* hoorden reageerden ze sneller op *raided*, zoals was te verwachten, maar als ze *Canada (r)aided* hoorden (dus met een intrusive /r/), dan reageerden in dezelfde mate op zowel *aided* als *raided*.

Om dit patroon bij de Nederlandse proefpersonen beter te begrijpen werd nog een oogbewegingsexperiment gedaan. In dit experiment kregen de Nederlanders dezelfde dubbelzinnige Brits Engelse zinnen te horen als in het eerdere experiment. Op het computerscherm zagen ze vier woorden. Ze kregen de opdracht om op het woord te klikken dat ze hoorden in de zin. Ze hoorden bijvoorbeeld "*Canada(r)aided*" (dus met een intrusive /r/) en zagen *aided*, *raided* en twee ongerelateerde woorden op het scherm. De Nederlandse luisteraars hadden een sterke voorkeur voor de woorden die begonnen met een /r/, ongeacht of ze een zin met een intrusive /r/ of een 'echte' /r/ hadden gehoord.

Wat betekent dit onderzoek nu voor tweedetaalluisteraars? Dit proefschrift bevat zowel goed als slecht nieuws. Het goede nieuws is dat spontane spraakprocessen waarmee luisteraars al bekend zijn vanuit hun moedertaal voor weinig problemen zorgen bij het begrijpen van de tweede taal. Het slechte nieuws is dat tweedetaalluisteraars moeite hebben met dubbelzinnige spraak in de tweede taal als dit wordt veroorzaakt door een spontanespraakproces dat niet voorkomt in de moedertaal. Het huidige onderzoek voegt nog een punt toe aan de al lange lijst van moeilijkheden die spraakperceptie in een tweede taal zo'n opgave maakt.

Appendix

Sentence	mean /r/ duration (ms)	mean intensity decrement (dB)
A few days ago, I saw aces when I looked at my cards.	56,5	1,3
A few days ago, I saw races from the British Superbike Championship.	69,8	11,1
I heard that Canada aided the area of Lesotho in Africa.	63,3	2,0
I heard that Canada raided the area of Lesotho in Africa.	84,9	7,1
I really thought that I saw air burning.	71,8	2,8
I really thought that I saw rare animals.	93,6	6,2
Did you know that the terracotta ear of the statue was broken?	64,7	1,1
Did you know that the terracotta rear of the statue was broken?	80,8	8,7
My brother likes extra ice when he has dinner.	64,7	2,8
My brother likes extra rice when he has dinner.	83,3	8,0
I think that Rebecca owes a lot to London.	64,8	2,8
I think that Rebecca rows a lot to London.	86,8	8,9
In north Malaysia itches are number one on the list of annoyances.	65,8	4,4
In north Malaysia riches are frowned upon.	100,5	7,5
And then Emma ejected the cassette.	60,6	1,3
And then Emma rejected the cassette.	69,3	5,2
I read that people from China etch whenever they feel like doing so.	64,6	1,3
I read that people from China retch whenever they feel like doing so.	88,1	7,0
My youngest sister saw odes to Rome made by many different people.	68,6	0,2
My youngest sister saw roads to Rome made by many different people.	96,5	9,1
I saw on Discovery Channel that people in Panama ate cows' eyes in former times.	91,3	-2,2
I saw on Discovery Channel that people in Panama rate David as the nicest city.	107,7	1,3
My favourite grandma aged because her dog suddenly died.	87,4	2,4
My favourite grandma raged because her dog suddenly died	112,8	4,2
Do you think it was sepia ash on her body?	72,6	0,3
Do you think it was sepia rash on her body?	88,1	6,1

We had not expected it, but suddenly Natasha emitted a cry of pain.	58,8	2,9
We had not expected it, but suddenly Natasha remitted the money to us.	71,4	3,6
Usually, Sheila elates everybody with her presence.	69,4	7,8
Usually, Sheila relates well to other people.	79,6	12,4
The Broadway musical "Clarissa" evolves into a deeply moving metaphor for the struggles of an entire generation.	66,8	1,1
The Broadway musical "Clarissa" revolves around a young American girl in London.	100,4	8,1
I read that the president of Algeria evokes the idea of using nuclear weapons.	66,2	3,2
I read that the president of Algeria revokes his decision to cut taxes.	87,3	7,8
I heard that the Australia alley is a nice street to live in.	74,5	2,0
I heard that the Australia rally lasts two weeks.	94,1	10,8
For Buddhists in India enunciation is of great importance.	73,5	1,4
For Buddhists in India renunciation is part of their daily lives.	96,7	7,8
We asked the children to draw apt presents for Mother's Day.	66,2	5,2
We asked the children to draw wrapped presents for Mother's Day.	80,4	11,9
Obviously, the bourgeois eye expert was bragging about his salary to his Harvard friends.	88,0	0,0
Obviously, the bourgeois rye expert was bragging about his salary to his Harvard friends.	96,9	8,5
Davidson and Brooks claim to be thorough ale-manufacturers.	77,7	4,9
Davidson and Brooks claim to be thorough rail-manufacturers.	98,4	12,9
The president of Russia eagerly awaits his caviar.	62,4	4,1
The president of Russia regally decorated his new office.	100,7	10,6
In this area anglers are still ice fishing.	78,6	1,5
In this area wranglers are experienced horseback tour guides.	86,2	5,4
In postwar America aches and pains are usually ignored by people without health insurance.	68,2	0,5
In postwar America rakes and planes are sold in hardware shops.	91,3	6,3
You should use extra old cheese in this recipe.	62,3	3,2
You should use extra rolled oats in this recipe.	83,9	8,2
I think that the extra ink cartridges were too expensive.	53,2	1,5
I think that the extra rink did not result in more ice skaters visiting it.	80,3	4,8

Curriculum vitae

Levensloop

Annelie Tuinman werd in 1981 in Groningen geboren. Ze doorliep de OBS De Zuidwester in Zuidlaren en het Maartens College te Haren. In 1998 begon ze haar studie Engelse Taal en Cultuur aan de Rijksuniversiteit Groningen, nadat ze was uitgeloot voor Tandheelkunde. In 1999 begon ze aan haar propedeuse Tandheelkunde om uiteindelijk toch haar studie Engels te vervolgen. In 2002 studeerde ze in het kader van een uitwisselingsprogramma enkele maanden aan de University of Salford in Manchester (GB). In 2003 rondde ze haar studie Engels af en begon een Master Toegepaste Taalwetenschappen, ook aan de Rijksuniversiteit Groningen. Haar onderzoeksstage deed ze aan het Nijmegen Institute for Cognition and Information (nu Donders Institute for Brain, Cognition and Behaviour). In 2004 rondde ze haar Master *cum laude* af en werd haar scriptie genomineerd voor de Anéla Scriptieprijs. In 2005 begon ze op het Max Planck Instituut voor Psycholinguïstiek te Nijmegen aan haar promotieonderzoek. In 2007 ontving ze een Fulbright Award om een half jaar onderzoek te doen aan de University of Pennsylvania in Philadelphia (VS). Momenteel is ze als statistisch onderzoeker werkzaam bij het Centraal Bureau voor de Statistiek te Den Haag.

Publicaties

Tuinman, A. (2004). Visuele woordherkenning in tweetaligen [Visual word recognition in bilinguals]. *Toegepaste Taalwetenschap in Artikelen*, 71, 93-101.

Tuinman, A. & Dijkstra, T. (2005). Herkenning van homografen door tweetaligen: De invloed van taakeisen en stimuluslijstcompositie [Recognition of homographs by bilinguals: The influence of task demands and stimulus list composition]. *Toegepaste Taalwetenschap in Artikelen*, 73, 31-40.

Tuinman, A., Mitterer, H., & Cutler, A. (2005). *L1 knowledge and perception of reduced forms in continuous L2 speech.* Presentation at the 10th Winter Conference of the Dutch Society of Psychonomics (NVP), Egmond aan Zee.

Tuinman, A., Mitterer, H. & Cutler, A. (2005). *Perceptie van gereduceerde woorden: Een uitdaging voor de tweetalige luisteraar?[Perception of reduced words: a challenge for the second language listener?]* Presentation for the Dutch Society for Phonetics, Utrecht.

Tuinman, A. (2006). *Overcompensation for /t/-reduction in Dutch by German-Dutch bilinguals*. Poster presented at Laboratory Phonology 10, Paris, France.

Tuinman, A., Mitterer, H., & Cutler, A. (2006). *Variable influence of grammatical context on identification of reduced /t/s.* Poster presented at the 12th Annual Conference on Architectures and Mechanisms for Language Processing (AMLaP), Nijmegen.

Tuinman, A., Mitterer, H. & Cutler, A. (2007). Speakers differentiate English intrusive and onset /r/, but L2 listeners do not. *Proceedings of the Sixteenth International Congress of Phonetic Sciences*, Saarbrücken, pp. 1905-1908.

Tuinman, A. & Cutler, A. (2010). Casual speech processes: L1 knowledge and L2 speech perception. *Proceedings of the Sixth International Symposium on the Acquisition of Second Language Speech: New Sounds 2010*, Poznan, Poland.

Cutler, A., Mitterer, H., Brouwer, S. & Tuinman, A. (2010). Phonological competition in casual speech. *Proceedings of the 5th Workshop on Disfluency in Spontaneous Speech and 2nd International Symposium on Linguistic Patterns in Spontaneous Speech*, Tokyo, Japan, pp. 43-46.

Tuinman, A. & Cutler, A. (2011). L1 knowledge and the perception of casual speech processes in L2. In K. Dziubalska-Koaczyk, M. Wrembel & M. Kul (Eds.), *Achievements and Perspectives in the Acquisition of Second Language Speech: New Sounds 2010*. Bern: Peter Lang; 289-301.

Tuinman, A., & Mitterer, H. (under revision). Transfer of L1 knowledge helps in doubly adverse conditions: Perception of casual speech in L2. *Language and Cognitive Processes. Chapter 2 of this dissertation.*

Tuinman, A., Mitterer, H. & Cutler, A. (under revision). Cross-language and cross-dialectal differences in perception of intrusive /r/ in English. *Journal of the Acoustical Society of America. Chapter 3 of this dissertation.*

Tuinman, A., Mitterer, H. & Cutler, A. Resolving ambiguity in familiar and unfamiliar casual speech. (under revision). *Journal of Memory and Language. Chapter 4 of this dissertation.*

Tuinman, A., Mitterer, H. & Cutler, A. (submitted). Syntactically-based compensation for phonological reduction without top-down feedback. *Chapter 5 of this dissertation.*

MPI Series in Psycholinguistics

1. The electrophysiology of speaking: Investigations on the time course of semantic, syntactic, and phonological processing. *Miranda van Turennout*
2. The role of the syllable in speech production: Evidence from lexical statistics, metalinguistics, masked priming, and electromagnetic midsagittal articulography. *Niels O. Schiller*
3. Lexical access in the production of ellipsis and pronouns. *Bernadette M. Schmitt*
4. The open-/closed-class distinction in spoken-word recognition. *Alette Haveman*
5. The acquisition of phonetic categories in young infants: A self-organising artificial neural network approach. *Kay Behnke*
6. Gesture and speech production. *Jan-Peter de Ruiter*
7. Comparative intonational phonology: English and German. *Esther Grabe*
8. Finiteness in adult and child German. *Ingeborg Lasser*
9. Language input for word discovery. *Joost van de Weijer*
10. Inherent complement verbs revisited: Towards an understanding of argument structure in Ewe. *James Essegbey*
11. Producing past and plural inflections. *Dirk Janssen*
12. Valence and transitivity in Saliba: An Oceanic language of Papua New Guinea. *Anna Margetts*
13. From speech to words. *Arie van der Lugt*
14. Simple and complex verbs in Jaminjung: A study of event categorisation in an Australian language. *Eva Schultze-Berndt*
15. Interpreting indefinites: An experimental study of children's language comprehension. *Irene Krämer*
16. Language-specific listening: The case of phonetic sequences. *Andrea Weber*
17. Moving eyes and naming objects. *Femke van der Meulen*
18. Analogy in morphology: The selection of linking elements in Dutch compounds. *Andrea Krott*
19. Morphology in speech comprehension. *Kerstin Mauth*
20. Morphological families in the mental lexicon. *Nivja H. de Jong*

21. Fixed expressions and the production of idioms. *Simone A. Sprenger*
22. The grammatical coding of postural semantics in Goemai (a West Chadic language of Nigeria). *Birgit Hellwig*
23. Paradigmatic structures in morphological processing: Computational and cross-linguistic experimental studies. *Fermín Moscoso del Prado Martín*
24. Contextual influences on spoken-word processing: An electrophysiological approach. *Daniëlle van den Brink*
25. Perceptual relevance of prevoicing in Dutch. *Petra M. van Alphen*
26. Syllables in speech production: Effects of syllable preparation and syllable frequency. *Joana Cholin*
27. Producing complex spoken numerals for time and space. *Marjolein Meeuwissen*
28. Morphology in auditory lexical processing: Sensitivity to fine phonetic detail and insensitivity to suffix reduction. *Rachèl J. J. K. Kemps*
29. At the same time...: The expression of simultaneity in learner varieties. *Barbara Schmiedtová*
30. A grammar of Jalonke argument structure. *Friederike Lüpke*
31. Agrammatic comprehension: An electrophysiological approach. *Marlies Wassenaar*
32. The structure and use of shape-based noun classes in Miraña (North West Amazon). *Frank Seifart*
33. Prosodically-conditioned detail in the recognition of spoken words. *Anne Pier Salverda*
34. Phonetic and lexical processing in a second language. *Mirjam Broersma*
35. Retrieving semantic and syntactic word properties. *Oliver Müller*
36. Lexically-guided perceptual learning in speech processing. *Frank Eisner*
37. Sensitivity to detailed acoustic information in word recognition. *Keren B. Shatzman*
38. The relationship between spoken word production and comprehension. *Rebecca Özdemir*
39. Disfluency: Interrupting speech and gesture. *Mandana Seyfeddinipur*
40. The acquisition of phonological structure: Distinguishing contrastive from non-contrastive variation. *Christiane Dietrich*
41. Cognitive cladistics and the relativity of spatial cognition. *Daniel B.M. Haun*

42. The acquisition of auditory categories. *Martijn Goudbeek*
43. Affix reduction in spoken Dutch. *Mark Pluymaekers*
44. Continuous-speech segmentation at the beginning of language acquisition: Electrophysiological evidence. *Valesca Kooijman*
45. Space and iconicity in German Sign Language (DGS). *Pamela Perniss*
46. On the production of morphologically complex words with special attention to effects of frequency. *Heidrun Bien*
47. Crosslinguistic influence in first and second languages: Convergence in speech and gesture. *Amanda Brown*
48. The acquisition of verb compounding in Mandarin Chinese. *Jidong Chen*
49. Phoneme inventories and patterns of speech sound perception. *Anita Wagner*
50. Lexical processing of morphologically complex words: An information-theoretical perspective. *Victor Kuperman*
51. A grammar of Savosavo, a Papuan language of the Solomon Islands. *Claudia Wegener*
52. Prosodic structure in speech production and perception. *Claudia Kuzla*
53. The acquisition of finiteness by Turkish learners of German and Turkish learners of French: Investigating knowledge of forms and functions in production and comprehension. *Sarah Schimke*
54. Studies on intonation and information structure in child and adult German. *Laura de Ruiter*
55. Processing the fine temporal structure of spoken words. *Eva Reinisch*
56. Semantics and (ir)regular inflection in morphological processing. *Wieke Tabak*
57. Processing strongly reduced forms in casual speech. *Susanne Brouwer*
58. Ambiguous pronoun resolution in L1 and L2 German and Dutch. *Miriam Ellert*
59. Lexical interactions in non-native speech comprehension: Evidence from electro-encephalography, eye-tracking, and functional magnetic resonance imaging. *Ian FitzPatrick*
60. Processing casual speech in native and non-native language. *Annelie Tuinman*