Casual speech processes: L1 knowledge and L2 speech perception

Annelie Tuinman[1] and Anne Cutler[1,2]

[1]Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands
[2]MARCS Auditory Laboratories, University of Western Sydney, Australia
Annelie.Tuinman@mpi.nl, Anne.Cutler@mpi.nl

ABSTRACT

Every language manifests casual speech processes, and hence every second language too. This study examined how listeners deal with second-language casual speech processes, as a function of the processes in their native language. We compared a match case, where a second-language process (/t/-reduction) is also operative in native speech, with a mismatch case, where a second-language process (/r/-insertion) is absent from native speech. In each case native and non-native listeners judged stimuli in which a given phoneme (in sentence context) varied along a continuum from absent to present. Second-language listeners in general mimicked native performance in the match case, but deviated significantly from native performance in the mismatch case. Together these results make it clear that the mapping from first to second language is as important in the interpretation of casual speech processes as in other dimensions of speech perception. Unfamiliar casual speech processes are difficult to adapt to in a second language. Casual speech processes that are already familiar from native speech, however, are easy to adapt to; indeed, our results even suggest that it is possible for subtle difference in their occurrence patterns across the two languages to be detected, and to be accommodated to in second-language listening.

Keywords: L1, L2, casual speech, reduction, insertion.

1. INTRODUCTION

The perception of speech in a second language (L2) cannot be understood except in the light of the perceiver's first-language (L1) knowledge; so much has been known to speech and language researchers for at least the better part of a century (see, e.g., Polivanov, 1931: "The phonological representations of our native language are so tightly coupled to our perception that even when we hear words or sentences from a language with quite different phonology, we tend to analyse these words in terms of the phonemic representations of the native language"). Experimental reports now provide abundant documentation of L1 effects on speech perception in L2, and models of the L1-to-L2 influence (e.g., Best & Tyler, 2007; Flege, 1995) offer detailed accounts of its genesis.

Like models of speech perception in general, however, models of L2 speech perception are understandably based on a somewhat idealised situation. The mapping of a phoneme or sequence of phonemes to stored representations can be predicted very well by the perceptual models, but the modelled situation will only arise if the input actually presents an acoustic form corresponding to each proposed segment. As listeners and speech researchers know only too well, however, real speech abounds with casual speech processes such as assimilation, reduction, deletion and intrusion, all of which lead to phonetic forms which deviate drastically from the canonical pronunciation of the words intended by the speaker.

In recent years, psycholinguistics has turned increasingly to investigation of real speech, and how listeners deal with the non-canonical forms it presents. A grossly over-simplified summary of the accrued results to date is that 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, but that 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 berries*; Gow, 2002), in

neutralisation (e.g., to distinguish the final /p/ of Dutch *slip* from the devoiced final sound of *slib*; Warner, Jongman, Sereno & Kemps, 2004), and in liaison (e.g., to distinguish the word-initial /p/ in French *trop partisan* from the liaison realisation 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 realisation (e.g., Alphen & McQueen, 2006; Sumner & Samuel, 2005). Despite all this success at dealing with real-speech forms, however, 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 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).

All of this research has, of course, been carried out with L1 listeners. But if even these experienced listeners are misled, what is going to happen when L2 listeners hear the same sort of input? Hear it they will, because all languages manifest casual speech processes, and 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. In the current investigation we make a start on charting the perceptual effects of casual speech processes in L2 listening.

Interestingly for our purposes, there are some casual speech processes which are found in many languages, and some which are found in only few. Thus assimilation is widespread across languages, with assimilation of place being extremely common (indeed obligatory in some languages, such as Japanese), and assimilation of voice also quite frequent; but patterns such as the Hungarian assimilation of /lr/ sequences to /rr/ are much rarer (Mitterer, Csépe, & Blomert, 2006). Listeners are better at dealing with the processes that are more common in their L1; e.g., for English speakers, assimilation of place is easier to deal with in word recognition tasks than assimilation of voice, while for French speakers the reverse is the case (Darcy, Peperkamp, & Dupoux, 2007). It is therefore interesting to ask whether this advantage can be transferred to L2 input, i.e., whether listeners will also find it easy to deal with an L2 process if they already have experience with the same process in their L1.

One process that is found in many languages is /t/-reduction (Guy, 1980). In the Germanic languages English, German and Dutch, the process patterns very similarly. For instance, /t/ is highly 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/. We compare this common process with a far less common 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. For the /t/-reduction case, we used Dutch speech presented to L1 speakers and to speakers of German with proficient L2 Dutch; since both languages have this reduction process, we refer to this as a case in which the L1 and L2 match on this feature. For the /r/-insertion case, we used British English speech presented to L1 listeners and to speakers of Dutch with proficient L2 English; since English has this process but Dutch does not, we refer to this case as one in which L1 and L2 mismatch.

The experiments which we carried out were broadly similar, in that in each case we constructed a phonetic continuum of stimuli, in which the phoneme under investigation (/t/, /r/) varied from effectively absent through partially realised to indubitably present. In each case, the listeners' task amounted to judging whether the phoneme was present or absent. In each case we compared the L2 listeners' judgements with the judgements given by L1 listeners presented with the same continuum of stimuli. In the match case, native speakers of German with proficient L2 Dutch and native speakers of Dutch decided in two perception experiments whether or not target words ended in /t/; the target words were verbs in the first experiment and nouns and adjectives in the second. In the mismatch case, native speakers of Dutch with proficient L2 English and native speakers of English decided whether target words were occurrences of *ice* or of *rice*.

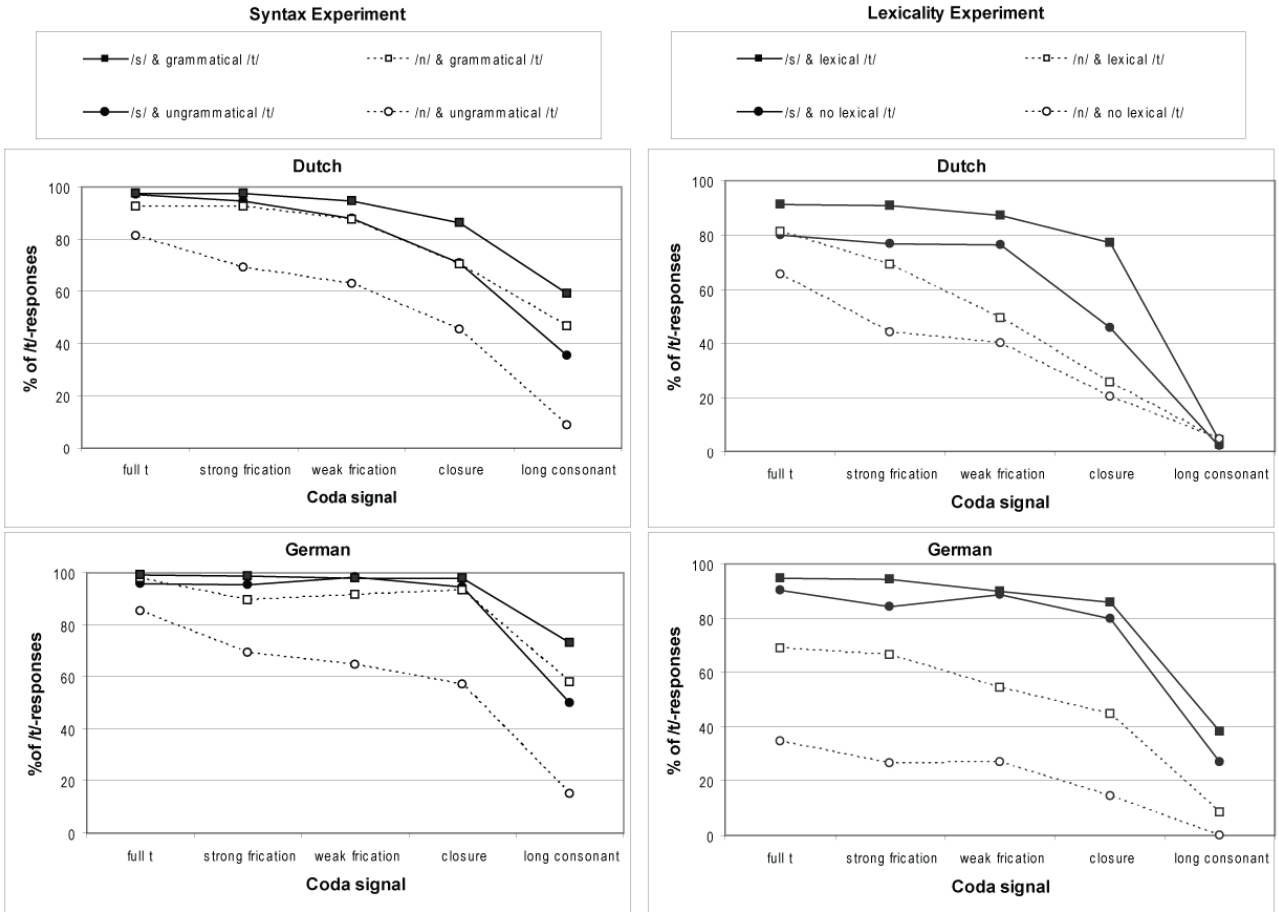
2. PERCEPTION OF REDUCED /t/

2.1. Method

Twenty-one native speakers of Dutch and 16 native speakers of German participated in the first experiment and 16 native speakers of Dutch and 16 native speakers of German in the second experiment. The German participants had a high level of proficiency in Dutch as L2.

Five realizations of /t/, from full production to complete deletion, were presented in two acoustic contexts, after /n/ (where /t/-reduction is unlikely) and after /s/ (where /t/-reduction occurs frequently). The selection of the /s/ and /n/ context and the different forms of reductions were based on patterns observed in a corpus study (Mitterer & Ernestus, 2006). In each sentence, listeners judged whether the target word ended in /t/ or not. In the Syntax Experiment, 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. In the Lexicality Experiment, target words were nouns and adjectives and lexical information produced the same result: interpreting /t/ made the target word a correct word (*charmant* 'charming') or not (*kanon*[t] 'gun').

Figure 1: Percentages of /t/-responses for Dutch listeners and German listeners



2.2. Results

The results for the Syntax Experiment with verbs and the Lexicality Experiment with nouns and adjectives were analyzed separately, in each case with a linear-mixed effects model. For the Syntax Experiment, participant was entered as a random factor, and Native Language, Coda Signal (from full production to complete deletion), Preceding Context (/n/ vs. /s/) and Grammar (/t/ or no /t/ predicted) as fixed factors. For

the Lexicality Experiment, participant was entered as a random factor, and Native Language, Coda Signal, Preceding Context and Lexicality (/t/ or no /t/ predicted) as fixed factors. For the Native Language variable, the German speaker group was mapped on the intercept. Figure 1 shows the mean /t/-response percentages.

Analysis started with a full model and in stepwise fashion insignificant interactions were pruned. For the Syntax Experiment, there was an overall significant tendency for more /t/-responses by German participants. This tendency, however, 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 Grammar on all five levels of Coda Signal. This analysis showed that for the full /t/ and strong frication coda signal there were no overall difference between Dutch and German listeners, but for the full /t/ Germans had a larger effect of Grammar and a smaller effect of Preceding Context than Dutch listeners. For the weak frication coda signal the Germans gave more /t/-responses than the Dutch and the effect was enlarged if the Grammar predicts the presence of a /t/. Again, for the closure coda signal there was an overall effect of Native Language which was moderated by Grammar and Preceding Context, so that Germans gave more /t/-responses in all cases except when the Preceding Context was /n/ and the Grammar predicted no /t/. For the long consonant coda signal—when there is no hint of /t/, but actually an extra long /n/ or /s/—German overall gave more /t/-responses than the Dutch.

The Lexicality Experiment showed 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 L2 listeners. However, the reported effects were moderated by various interactions. Therefore, we examined the effects of Native Language, Preceding Context and Lexicality on all five levels of Coda Signal. The strong Coda Signals showed a consistent pattern with main effects for Preceding Context and Lexicality and a significant interaction between Native Language and Preceding Context—L2 listeners had a larger context effect than L1. For the weak Coda Signals the effect of Lexicality was larger for L2 listeners, but depended on the Preceding Context.

In summary, the German L2 listeners' responses are, overall, sensitive to the same factors as the Dutch L1 responses; however, in some cases the L2 responses include significantly more /t/ reports than the L1 baseline. A comparison with the production facts suggests an explanation for this pattern. In German, reduction of morphologically functional /t/ (such as a verb inflection) is inhibited, whereas in Dutch, reduction is equally likely for morphological and for non-morphological /t/. Separate production experiments with non-Dutch-speaking Germans and Dutch native speakers confirmed this pattern (Tuinman, 2006). In the light of this comparison, the tendency of the L2 listeners to produce more /t/ responses, especially when Grammar predicted a /t/, suggests that they were sensitive to the L1-L2 difference, and tried to compensate for it.

3. PERCEPTION OF INTRUSIVE /r/

3.1. Method

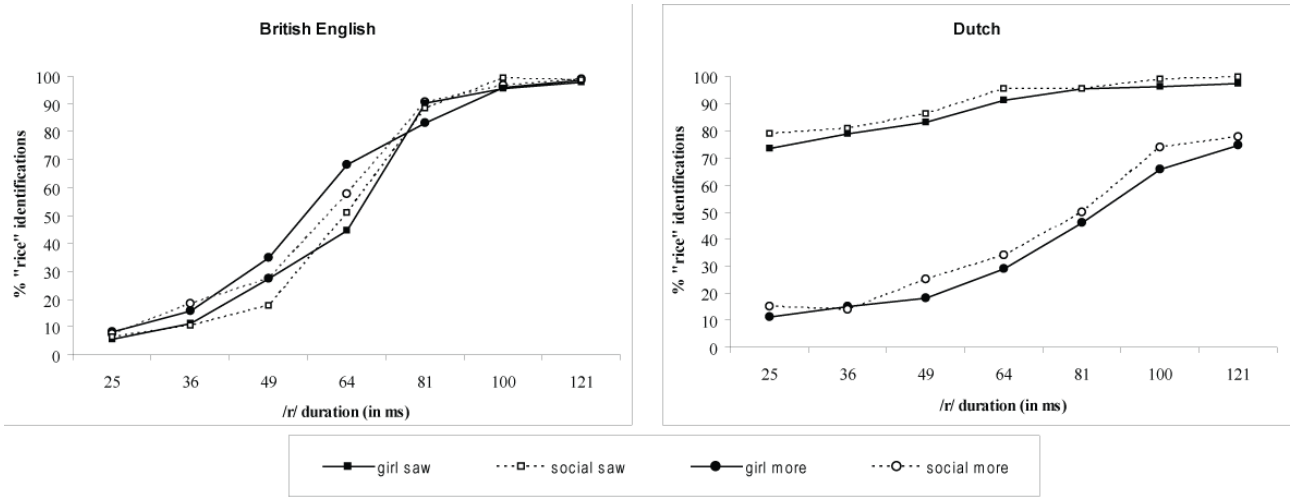
Eighteen native speakers of Dutch and 18 native speakers of British English took part in the experiment. The Dutch participants had a high level of proficiency in English as a second language. On average, they had received 7 years of English instruction in primary and secondary education.

Seven realizations of /r/, from short (25 ms) to long (121 ms), were presented in four sentence contexts with a contextual bias and 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* and more biased towards *ice*. 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, the perception of an /r/ in the speech signal can be attributed to *more*, while in the case of *saw ice*, perception of an /r/ sound cannot be mapped to the spelling; the bias should therefore manifest itself in terms of more reports of *rice* after *saw* than after *more*.

3.2. Results

Figure 2 shows the percentage of rice judgments as a function of the three independent variables, separately for the L2 and L1 listeners. The principal result is clear at a glance: the response patterns are different from each other. British English L1 listeners based their responses mainly on the durational variation rather than on the other variables, while the L2 listeners barely used the durational information, but show a clear separation between the responses in sentences with saw (squares) versus more (circles).

Figure 2: Percentages of "rice" identifications for British English and Dutch listeners



The results were analyzed with a linear-mixed effects model, with Participant 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 Native Language, the L1 group was mapped on the intercept. Analysis started with a full model and in stepwise fashion, insignificant interactions were pruned. The L1 group had a significant effect of /r/ Duration, no effect of Context, and a small negative effect of Orthography. That is, L1 listeners in fact gave somewhat more *rice* responses when the preceding word was *more* than when the preceding word was *saw*. The L2 listeners showed a significant effect of Orthography, and gave more *rice* responses when the preceding word was *saw* than when the preceding word was *more*. The L2 listeners were also less categorical in their responses than the L1 listeners. Additionally, the L2 group was influenced by Context, and made less use of /r/ Duration when there was an Orthographic Bias.

4. CONCLUSION

The combined results of our two sets of studies motivate the inescapable conclusion that casual speech processes are subject to the same tight coupling with L1 listening experience as every other aspect of L2 speech perception. The L2 listeners' performance in the match case (/t/-reduction experiment) related quite differently to the native performance baseline than the performance of the L2 listeners in the mismatch case (/r/-intrusion experiment). In the match case, the German listeners were broadly sensitive to the same range of factors in the Dutch input as the Dutch L1 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. In the mismatch case, however, the Dutch listeners showed quite a different pattern of sensitivity than the English L1 listeners: while the L1 listeners based their responses overall on the acoustic characteristics of the stimuli, were quite insensitive to the sentence meaning and certainly did not incline to report /r/ when there was none in the orthographic representation, the L2 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.

We conclude, therefore, that the casual speech patterns of the L1 influence interpretation of casual speech processes in an L2. A process familiar from the L1 can be processed easily; an unfamiliar L2 process with no corresponding L1 experience is very hard to adapt to.

However, this is not the end of the story, because in some respects the L2 listeners did deviate from L1 performance in the match case as well. Especially they were more likely than the L1 listeners to report a /t/ when the sentence grammar predicted it, i.e. when it was a morphological inflection. This is precisely the kind of /t/ that is less likely to be reduced in their L1. 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 L2 that matches an L1 process, but they are even sensitive to subtle differences in its distribution of occurrence across the languages, and can accommodate their perceptual responses accordingly.

5. ACKNOWLEDGEMENTS

This research was supported by a doctoral grant from the Max Planck Society and by a Fulbright travelling fellowship, both to the first author. We are grateful to Holger Mitterer for advice on all aspects of the project.

6. REFERENCES

Alphen, P. M. v., 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., Tyler, M. D. 2007. Nonnative and second-language speech perception: Commonalities and complementarities. In: Munro, M., Bohn, O.-S. (eds.), *Second language speech learning Amsterdam*: John Benjamins.

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

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

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

Flege, J. E. 1995. Second language speech learning: Theory, findings, and problems. In: Strange, W. (ed.), *Speech perception and linguistic experience: Issues in cross-language research*. Baltimore: York Press. 233-277.

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. 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: Labov, W. (ed.), *Locating language in time and space*. New York: Academic Press. 1-36

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.

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.

Polivanov, E. 1931. La perception des sons d'une langue étrangère. *Travaux du Cercle Linguistique de Prague*, 4, 79-96.

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.

Tuinman, A. 2006. Overcompensation for /t/-reduction in Dutch by German-Dutch bilinguals, Abstracts of *Laboratory Phonology 10*. Paris, France.

Warner, N., Jongman, A., Sereno, J., Kemps, R. J. 2004. Incomplete neutralization and other sub-phonemic durational differences in production and perception of Dutch. *Journal of Phonetics*, 32, 251-276.

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