The direct and indirect effects of the phonotactic constraints in the listener's native language on the comprehension of reduced and unreduced word pronunciation variants in a foreign language

Mirjam Ernestus[a,b,*], Huib Kouwenhoven[a], Margot van Mulken[a]

[*] Corresponding author; telephone number: +31-24 3521511; email address: m.ernestus@let.ru.nl
[a] Centre for Language Studies, Radboud University, Erasmusplein 1, 6525HT, Nijmegen, the Netherlands
[b] Max Planck Institute for Psycholinguistics, Wundtlaan 1, 6525XD, Nijmegen, the Netherlands

Abstract

This study investigates how the comprehension of casual speech in foreign languages is affected by the phonotactic constraints in the listener's native language. Non-native listeners of English with different native languages heard short English phrases produced by native speakers of English or Spanish and they indicated whether these phrases included *can* or *can't*. Native Mandarin listeners especially tended to interpret *can't* as *can*. We interpret this result as a direct effect of the ban on word-final /nt/ in Mandarin. Both the native Mandarin and the native Spanish listeners did not take full advantage of the subsegmental information in the speech signal cueing reduced *can't*. This finding is probably an indirect effect of the phonotactic constraints in their native languages: these listeners have difficulties interpreting the subsegmental cues because these cues do not occur or have different functions in their native languages. Dutch resembles English in the phonotactic constraints relevant to the comprehension of *can't*, and native Dutch listeners showed similar patterns in their comprehension of native and non-native English to native English listeners. This result supports our conclusion that the major patterns in the comprehension results are driven by the phonotactic constraints in the listeners' native languages.

Keywords
Foreign language speech comprehension; Phonotactic constraints; Subsegmental information; Speech reduction; Non-native listeners

The direct and indirect effects of the phonotactic constraints in the listener's native language on the comprehension of reduced and unreduced word pronunciation variants in a foreign language

1.0 Introduction

Words are often produced with weaker articulatory gestures and with fewer segments in informal than in formal conversations (see Ernestus & Warner, 2011, for an overview of the phenomenon and many examples in several languages). Johnson (2004), for instance, on the basis of part of the Buckeye Speech Corpus (Pitt et al., 2007), showed that in conversational American English over 20% of the words lack at least one segment. Native listeners have been shown to understand reduced word pronunciation variants well. Non-native listeners, in contrast, seem to experience problems recognizing these variants, even if they have high proficiency in the foreign language (e.g., Nouveau, 2012; Wong et al., 2015; ten Bosch et al., 2016). This article contributes to answering the question of why non-native listeners have difficulties understanding reduced word pronunciation variants, focusing on the role of the phonotactic constraints in the listener's native language.

1.1 Previous studies on the comprehension of casual speech

A number of studies have shown that native listeners rely on several types of cues for the recognition of reduced word pronunciation variants. They rely on the meaning of the word's context (e.g., van de Ven et al., 2011), on the probability of the word given its co-text (the preceding and following words, e.g., van de Ven et al., 2012), on the syntactic structure of the sentence (Tuinman et al., 2014; Viebahn et al., 2015), and on speech rate (e.g., Dilley & Pitt, 2010).

Native listeners may also rely on the sound patterns in their language. They easily reconstruct missing segments whose absence leads to phoneme sequences that are phonotactically illegal (e.g., Spinelli & Gros-Balthazard, 2007). Furthermore, native listeners more frequently reconstruct a missing segment in that segmental context in which it is most frequently reduced. For instance, native Dutch listeners reconstruct word-final /t/ more often after /s/ than after /n/, mirroring the fact that /t/ is more often acoustically weak after /s/ than after /n/ (e.g., Mitterer et al., 2008). They also reconstruct these segments when they are completely absent in the speech signal and there are no acoustic traces (e.g., Janse et al., 2007).

Finally, native listeners may use all types of subsegmental properties of the acoustic signal. These cues may be traces of the reduced segment itself, temporally located between the preceding and the following segment, for instance weak frication of a reduced /t/ (e.g., Mitterer & Ernestus, 2006). In addition, these cues may be captured in the articulation of neighboring segments. A well-known finding is that native listeners of English can distinguish *support* from *sport*, even if *support* is produced without schwa, on the basis of the duration of the aspiration on the following stop consonant (Manuel, 1991). Another example comes from German, in which the /s/ in word-final /st/ tends to be longer if the /t/ is missing in the speech signal (Zimmerer et al., 2011; Zimmerer & Reetz, 2014) and accordingly native listeners tend to reconstruct a missing /t/ after a word-final [s] more often if the [s] is longer (Zimmerer & Reetz, 2014). In the remainder of this paper, these traces of reduced segments will be called acoustic traces.

These studies addressed the question how native listeners understand reduced word pronunciation variants. Research on non-native speech processing, in contrast, has nearly exclusively focused on the question whether reduced word variants are well understood by non-native listeners (e.g., Nouveau, 2012). Only a few studies have also investigated why non-native listeners, even those with high proficiency levels in the non-native language, have great difficulties understanding these variants. Van de Ven et al. (2010) showed that reduced word pronunciation variants do not prime semantically related words in non-native listeners at the interstimulus intervals at which they prime in native listeners. Furthermore, Wong et al. (2015) showed that non-native listeners recognize reduced pronunciation variants better if they have larger receptive vocabularies in the foreign language.

1.2 Direct and indirect effects of phonotactic constraints on the comprehension of foreign casual speech

The present study investigates the direct and indirect effects of the phonotactic constraints in the listener's *native* language on the comprehension of casual speech in a foreign language. Languages differ in the phonotactic constraints on the possible sequences of segments. For instance, whereas English words can end in /nt/, Spanish and Mandarin words cannot (e.g., Coe, 2001; Chang, 2001). We know that non-native speakers may show a direct effect of the phonotactic constraints of their native languages in speech production, by adding, changing, or deleting segments in the words of their foreign language. Thus, speakers of English with Spanish or Mandarin as native language tend to produce English words ending in /nt/ without the final /t/ (e.g., Coe, 2001; Chang, 2001).

We formulated two competing hypotheses for the *direct* effects of phonotactic constraints in the native language on word comprehension in a foreign language. We know that when non-native speakers of a language change one segment for another during production because of the phonotactic constraints in their native languages, they may also apply these substitutions during speech comprehension (e.g., Moreton, 2002). The question as to whether comprehension also patterns with production when the phonotactic constraints of the native language make the non-native speaker omit a segment remains open. That is, it is unknown whether non-native speakers who tend to omit a segment (e.g., the /t/ of word-final /nt/) in foreign languages as a result of the phonotactic constraints of their native languages also tend to ignore this segment when listening to these foreign languages, and thus have a bias towards the interpretation of words as not containing that segment (e.g., as ending in /n/ rather than in /nt/). The competing hypothesis for the direct effect is that in a similar way to native listeners of a language (e.g., Mitterer et al., 2008), non-native listeners may tend to reconstruct segments during perception in those segmental positions where they tend to omit those segments themselves in production. They then have learnt that the segment may often be acoustically absent in some segmental contexts and compensate for its absence during perception. Following this hypothesis, the mental representations of the words do not differ substantially between native and non-native listeners of a language with respect to the presence versus absence of segments. This hypothesis predicts that non-native listeners may also reconstruct segments in words where these segments are not intended (e.g., they may interpret words ending in [n] as ending in /nt/).

The phonotactic constraints in the listeners' native languages may also have a more indirect effect on their ability to interpret reduced word pronunciation variants. These constraints may indirectly affect their sensitivity to some (subtle) characteristics of the speech signal. When non-native listeners are not familiar with a given segment sequence from their native language, they are also not familiar with the (subtle) characteristics of the speech signal that may cue the segment sequence after it has been reduced. These characteristics may include acoustic traces of the reduced segments themselves and the exact acoustic properties of neighboring segments.

The acoustic cues to a reduced segment that is phonotactically illegal in the listener's native language may nevertheless occur in that same language, albeit in other positions in the word. For instance, native Spanish listeners may be familiar with acoustic traces of /t/ in syllable onset position from their native language, but not with these same acoustic cues to /t/ in word-final /nt/. In addition, acoustic cues may occur in the listener's native language but with different functions. For instance, in the listener's native language, a lengthened vowel may only cue the presence of word stress or word final lengthening, rather than also the absence of a following consonant. Listeners then have to acquire the new functions of the subsegmental information in the speech signal in order to become efficient listeners of the foreign language.

One study has made a start at investigating non-native listeners' sensitivity to acoustic cues to reduced word pronunciation variants. Mitterer and Tuinman (2012) showed that German learners of Dutch rely more on the (subtle) cues for (reduced) /t/ when the segment is part of the stem of a content word, for which the reduction patterns in Dutch and in German are similar, than when /t/ is a marker of verbal inflection, which is more often reduced in Dutch than in German. German learners of Dutch report the presence of the verbal suffix /t/ more often than native Dutch listeners (and thus show less sensitivity to the exact information in the acoustic signal) if the absence of the /t/ would result in an ungrammatical sentence. These findings suggest that non-native listeners mostly rely on subsegmental cues if these cues occur where they also occur in the listeners' native languages. However, the data could also be interpreted as pointing to differences between native listeners of German and Dutch in the weighting of grammatical and acoustic cues.

The question of whether the phonotactic constraints of listeners' native languages affect their sensitivity to subsegmental information in a foreign language is relevant for theories of perceptual learning. Research has shown that both native and non-native listeners may adjust their interpretation of the acoustic signal in order to correctly classify a sound as one phoneme or another. For instance, Japanese learners of English may learn to rely on (subtle) properties of the acoustic signal to distinguish between /l/ and /r/ (e.g., Pisoni et al., 1994). The question then arises whether non-native listeners can also learn to interpret subsegmental cues to reduced phoneme sequences that do not occur in their native languages.

In addition, this question is relevant for models of word comprehension. Many current models of word comprehension do not reflect how subsegmental information affects word comprehension (e.g., TRACE: McClelland & Elman, 1986) or do not reflect whether and how listener groups may differ in their sensitivity to subsegmental cues. For instance, exemplar theory (e.g., Goldinger, 1998; Johnson, 2004) postulates that many tokens of a word produced or perceived by the language user are mentally stored with all their acoustic detail, and these mental representations include the subsegmental cues. Exemplar theory thus explains the role of subsegmental cues in speech comprehension. Importantly, it predicts that all listeners take all relevant subsegmental cues into account, which may be contrary to fact.

We thus hypothesize both direct and indirect effects of the listener's native language on the comprehension of casual speech in a foreign language. We also hypothesize that these effects are especially noticeable if the listener has a low or intermediate proficiency in the foreign language. The effects may disappear if the listener becomes more proficient and is more experienced in processing casual speech in the foreign language.

1. 3 The present study

In this study, we investigate how non-native listeners of English comprehend reduced and unreduced word pronunciation variants in English, focusing on /t/ reduction. Reduction of /t/ is a frequent and well-studied phenomenon, especially in English (e.g., Labov, 1972; Guy, 1991; Pitt, 2009). Several studies have shown that native listeners easily reconstruct reduced /t/. For instance, Sumner and Samuel (2005) showed that in native English listeners, a word activates semantically related words, independently of how its word-final /t/ is articulated (i.e., as a fully articulated canonical /t/; a coarticulated, glottalized stop; or a glottal stop). Similarly, Pitt (2009) showed that native English listeners recognize words with an /nt/ cluster (e.g., *counter*) as easily when the cluster is pronounced in full (i.e., as [nt]) or as a single nasal flap, provided that the word is often pronounced with the nasal flap. All these studies focus on native listeners. We also studied how listeners with different native languages interpret reduced word-final /t/ in English.

More specifically, we researched how listeners of different language backgrounds and with intermediate to advanced proficiencies in English interpret tokens of *can't* with and without clearly audible /t/ (i.e., unreduced and reduced tokens of *can't*) and compared their comprehension scores for reduced and unreduced *can't* with those for unreduced *can*. We studied whether differences in comprehension scores for unreduced and reduced *can't* and for *can* among the listeners could be ascribed to the presence versus absence of a ban on word-final /nt/ in their native languages.

In several respects, this study is very different from previous studies on (non-)native listeners' comprehension of reduced word pronunciation variants. Most importantly, absence of /t/ in *can't* leads to another real English word (*can*), which can occur in the same syntactic, morphological, and phonological contexts as *can't*. As a consequence, when listeners are presented with reduced tokens of *can't*, they cannot disambiguate the meanings of the tokens on the basis of, for instance, lexical or grammatical cues, as they could, for instance, in Mitterer and Tuinman (2012).

Furthermore, all our stimuli, consisting of a personal pronoun followed by *can* or *can't* and an infinitive, were spliced from natural conversations. To our knowledge, all previous studies on the comprehension of reduced word pronunciation variants with non-native listeners are based on resynthesized speech or on read aloud speech in which the speaker (unnaturally) incorporated reduced word pronunciation variants. Although these studies provide relevant information about what non-native listeners can perceive, they do not necessarily show how these listeners process naturally occurring reduction. We believe that the results of our study are more ecologically valid.

The stimuli in our experiment were produced by native speakers of American English or by native speakers of Spanish with proficiencies in English at the A2 – B1 / B2 level according to the Common European Framework of Reference for Languages (Council of Europe, 2011). Native speakers of American English are known to produce tokens of *can* and (reduced) *can't* with (slightly) different vowel qualities, segment durations and pitch, among other subsegmental cues. The word *can't* often carries contrastive focus, in which case it maintains its full vowel, and its segments are relatively long. Even when *can't* is pronounced without contrastive focus, it usually keeps most of its vowel quality and its segments are relatively long. In contrast, in unaccented *can* (without focus), the vowel is often reduced in quality (to schwa) and in duration, and may even be completely absent. Similarly, its consonants may be reduced. The quality and durations of the velar stop, the vowel, and the nasal may therefore provide cues as to whether a native speaker of American English intended *can* or *can't*. In addition, any acoustic traces of the /t/ itself in reduced *can't* may form valuable cues.

Many cues to the identity of a *can* / *can't* token that can be found in American English are likely to be less prominent in *can* / *can't* tokens produced by native speakers of Spanish, even if these speakers have advanced proficiencies in English. Since Spanish has neither schwa nor strong vowel reduction (for a discussion see e.g., Cobb & Simonet, 2015), *can* and *can't* tokens pronounced by native speakers of Spanish are likely not to differ much in the spectral properties or the duration of the vowel. Furthermore, if these non-native speakers delete /t/ in order to bring the pronunciation of *can't* in line with Spanish phonotactics, they are less likely to leave acoustic traces of /t/ than native speakers of American English, who probably reduce /t/ for articulatory reasons.

A direct effect of a phonotactic constraint on word final /nt/ in listeners' native languages is expected to arise in their comprehension of both the stimuli produced by native speakers of American English and by native speakers of Spanish. If acoustic traces of /t/ in reduced *can't* are more prominent in the speech produced by native speakers of English than by native speakers of Spanish, an indirect effect of the constraint may especially arise when the non-native listeners hear the reduced *can't* stimuli produced by native speakers of English. Nonnative listeners' diminished sensitivity to acoustic cues to reduced /t/, due to the ban on word-final /nt/ in their native languages, then hinders them to rely on the cues that are especially present in native English.

We compared the performance of four listener groups. The first group was formed by native English listeners. We expected these listeners to perform excellently on the unreduced *can* and *can't* tokens. We also expected them to correctly interpret the reduced tokens of *can't* without clearly audible /t/ produced by native speakers of American English, using the subsegmental information in the signal. In contrast, we thought that they may have difficulty comprehending reduced tokens of *can't* produced by native speakers of Spanish if these tokens contain fewer subsegmental cues to the reduced /t/.

In addition, we tested three groups of non-native listeners of English: native listeners of Spanish, Mandarin, and Dutch. Henceforth, we will refer to these participants as native Spanish, Mandarin, or Dutch listeners, respectively, instead of as learners, because many of them are no longer actively trying to improve their proficiency levels in English. The native Spanish listeners had the same backgrounds as the native speakers of Spanish of the stimuli. Their average proficiency level in English was slightly lower than the average proficiency level of the native Dutch listeners. The native Mandarin listeners had the lowest average proficiency level in English. We investigated the role of the phonotactic constraints in the listeners' native languages, taking differences in proficiency in English into account.

Spanish and Mandarin share some important phonotactic constraints: as mentioned above, neither Spanish nor Mandarin allow word-final /nt/ (e.g., Bent et al., 2007; Chang, 2001). Native listeners of both languages may therefore show a bias towards either *can* or *can't*. Moreover, they may show indirect effects of the phonotactic constraints of their native languages. Since both Spanish and Mandarin lack word-final /nt/, schwa, and strong vowel reduction (e.g., Gut, 2003; Cobb & Simonet, 2015), native listeners of these languages are not familiar from their native languages with the most important cues to reduced word-final /nt/ of English *can't*. As a consequence, they may rely less on these cues than native English listeners and they may tend to interpret reduced tokens of *can't* similarly to tokens of *can*, both in native American English and in English produced by native speaker of Spanish.

Dutch is much more similar to English in that it has many words ending in /nt/ and has phonological/phonetic processes that are (nearly) identical to English /t/ reduction and vowel reduction (e.g., van Bergem, 1993; Schuppler et al., 2011). If the phonotactic constraints in the listener's native language play an important role in the perception of *can't* and *can*, we expect native Dutch listeners to perform similarly to native English listeners. That is, we expect their performance to differ significantly from the performance by the native Spanish and Mandarin listeners, by showing no bias for either *can* or *can't,* and by showing sensitivity to subtle acoustic cues in English reduced *can't*.

The experiment not only contained an auditory comprehension part and a proficiency assessment part, but also a frequency rating part. Listeners may correctly identify a reduced *can't* token as *can't,* rather than as *can*, by taking into account the probabilities of *can* and *can't* in the co-text. Importantly, non-native English listeners may have different expectations than native English listeners, due to their different cultural backgrounds (e.g., fewer people may be able to cycle in Spain than in the United States and native Spanish listeners may therefore differ from native English listeners in their estimation of the probability of *can* versus *can't* in a sentence like *he can cycle*) and to their limited exposure to American English. We wished to take these differences into account in the analysis of our comprehension data and therefore established with the rating experiment the listeners' expectations of *can* versus *can't* in the phrases presented in the comprehension experiment.

2.0 Comprehension experiment

2.1 Method

2.1.1 Participants

A total of 127 participants took part in the experiment, divided in four listener groups. Thirty-six native English listeners (24 females, mean age of 19.78 years, *SD* = 1.80) and 21 native Mandarin listeners (14 females, mean age of 20.05 years, *SD* = 1.94) from the participant pool of the Department of Linguistics, University of Alberta[1] received course credits for their participation. Forty native listeners of European Spanish (18 females, mean age of 21.93 years, *SD* = 2.27) were recruited at the Escuela Técnica Superior de Ingenieros de Telecomunicación of the Universidad Politécnica de Madrid. Finally, thirty native Dutch listeners (20 females, mean age of 20.50 years, SD = 1.65) were recruited from the participant pool of the Max Planck Institute for Psycholinguistics in Nijmegen, the Netherlands. Except for the native Mandarin listeners, all listener groups were thus tested in countries where their native languages are spoken. The native listeners of Spanish and Dutch received a small financial reward for their participation.

We assessed all participants' proficiencies in English with the LexTALE task (*Lexical Test for Advanced Learners of English*, Lemhöfer & Broersma, 2012). Although LexTALE is a visual lexical decision task focusing on vocabulary knowledge, it has been shown to correlate substantially with a general proficiency measure (Lemhöfer & Broersma, 2012), and therefore provides some insights into the participants' proficiency levels. A participant's score in the test is the average of the percentage of correct responses to the real words and to the nonwords, corrected for the unequal proportion of real words and nonwords in the test (i.e., ((number of correct words / 40 * 100) + (number of correct nonwords / 20 * 100)) / 2). A linear regression model, reported in Table 1, revealed that the LexTALE scores differed between all listener groups, with the native English listeners having the highest and the native Mandarin listeners the lowest LexTALE scores. According to Lemhöfer and Boersma (2012), the scores of the

[1] We acknowledge that our native English listeners are not from the same dialect group as the speakers in the Buckeye corpus (Canadian and North Midlands dialect groups, respectively; Labov et al., 2005), from which we drew our stimuli. To our knowledge, however, the differences between the dialect groups cannot be expected to affect the Canadian listeners' ability to effortlessly perceive the *can-can't* contrast in our stimuli.

Dutch and Spanish participants correspond with an Upper Intermediate level (B2) on to the European Framework, and the Chinese participants have on average a score that corresponds with a lower level (B1 or lower).

In addition to an analysis of the full dataset, we also analyzed a subset of the participants who were most comparable in English proficiency, to allow us to investigate whether differences in English proficiency explained differences in performance in the *can* / *can't* comprehension task over and above the differences in their native language. This subset included the 23 native Spanish listeners with the highest LexTALE scores and the 20 native Dutch listeners with the lowest LexTALE scores, together with a random selection of 20 native English listeners and all 21 native Mandarin listeners. The mean LexTALE score in the subset was significantly higher for the native Spanish listeners than for the native Dutch listeners (see the rightmost column of Table 1). Importantly, the statistical analyses on the full *can* / *can't* comprehension dataset (as presented below) and on this subset yield similar results, which means that the effects reported below also hold for native listeners of Dutch and Spanish with similar general proficiency levels in English.

Table 1: ***Results of two linear regression models predicting LexTALE scores as a function of listener group (full dataset and subset). The intercept represents native Dutch listeners.***

	Full dataset			Subset		
Predictor	*β*	*t(122)*	*p*	*β*	*t(79)*	*p*
Intercept	75.83	835.12	< .001	71.25	606.13	< .001
Listener group (Spanish)	-8.05	-67.03	< .001	1.14	7.10	< .001
Listener group (Mandarin)	-19.29	-136.28	< .001	-14.70	-89.51	< .001
Listener group (Native English)	14.65	119.17	< .001	17.13	103.01	< .001

2.1.2 Materials

The stimuli in the comprehension experiment all contained tokens of *can* or *can't* from either the Buckeye corpus (Pitt et al., 2007) or the Nijmegen Corpus of Spanish English (NCSE; Kouwenhoven et al., to appear). The Buckeye corpus contains conversational American English from native, mostly monolingual speakers. The entire Buckeye Corpus has been phonetically annotated in two steps: an automatic speech recognizer generated phonetic transcriptions, which were then hand corrected by human transcribers (see

Pitt et al., 2005). We considered /t/ in *can't* to be present if it was transcribed as a canonical /t/, as a glottal stop, as a flap, or as a /d/ or /p/, which may arise due to co-articulation. We considered /t/ to be absent, and thus the *can't* token to be reduced, if the complete /nt/-cluster was realized as a nasal (flap).

The NCSE contains conversational speech in English by native speakers of Spanish who lived in Madrid at the time of the recording and were thus exposed to Castilian Spanish on a daily basis. Their proficiency in English ranged from A2 to B1 / B2 according to the Common European Framework of Reference for Languages (Council of Europe, 2011). The NCSE is not phonetically annotated and we automatically created a phonetic transcription as described in Appendix 1. This transcription does not distinguish between different variants of /t/ and we considered *can't* to be reduced if it was transcribed without /t/.

Each stimulus in the comprehension experiment consisted of three words (i.e., was a trigram): a token of *can*, unreduced *can't*, or reduced *can't*, preceded by a pronoun, and followed by an infinitive (e.g., *I can't imagine* or *I can think*). As stimuli for our experiment, we selected 93 trigrams with reduced *can't* and 147 trigrams with unreduced *can't*, such that the infinitive occurred at least once in combination with a full and once with a reduced token of *can't*, and this infinitive was pronounced at least once by a native speaker of American English and at least once by a native speaker of Spanish. We also included as stimuli in our experiment 218 trigrams with *can* followed by infinitives that also occur in the *can't* stimuli. The stimuli thus represented 29 different infinitives, which occurred between three and 53 times. The *can* and *can't* tokens were preceded by six different pronouns, which occurred between four and 185 times. The stimuli were produced by 29 different native speakers of Spanish (193 tokens) and by 35 different native speakers of American English (265 tokens).Table 2 provides an overview of the stimuli.

Table 2: ***Number of stimuli per Stimulus Type and Type of Speech.***

Stimulus type	Native speakers of American English (Buckeye corpus)	Native speakers of Spanish (NCSE)	Total
Can	123	95	218
Unreduced *can't*	99	48	147
Reduced *can't*	43	50	93
Total	265	193	458

We verified in a pretest whether the reduced *can't* tokens had been correctly orthographically transcribed as *can't*. Eight native speakers of English were presented with the orthographic transcriptions of the context (i.e., 25 preceding and 25 following words for the Buckeye tokens; eight preceding and eight following chunks for the NCSE tokens, with each chunk containing on average 4.2 words) of all 93 reduced tokens of *can't* and of 50 randomly selected tokens of *can* in randomized order. The participants were asked to indicate whether they thought *can* or *can't* fitted the given context best. We found that for 79 of the 93 reduced tokens of *can't* at least six participants agreed on *can't*, which we accepted as sufficient. For the remaining 14 tokens, we created sound files of about 30 seconds long, from about 22 seconds before to about eight seconds after the token of *can't*. A phonetically trained, native listener of American English evaluated these sound files and confirmed that *can't* had been uttered in each case.

We resampled the Buckeye stimuli from 16 000 Hz to 44 100 Hz so that they matched the sampling frequency of the NCSE stimuli. Then, we normalized all stimuli in amplitude.

We produced six pseudo-randomized lists of stimuli produced by the native speakers of American English, and six lists of stimuli produced by the native speakers of Spanish, ensuring that no more than two stimuli of the same type (i.e., with *can*, unreduced *can't*, or reduced *can't*) followed each other. We exhaustively combined each 'American' list with each 'Spanish' list, which resulted in 36 experimental lists. Each experimental sublist with speech from native speakers of one language (either American English or Spanish) was divided into two blocks. We varied the order in which the four blocks of an experimental list were presented, such that in some lists the blocks with stimuli produced by native speakers of American English alternated with the blocks with stimuli produced by native speakers of

Spanish, while in other lists these blocks followed each other. Each experimental block was preceded by the same six familiarization trials in the same order. These familiarization trials were trigrams containing clear tokens of *can* or *can't* that could not be used as stimuli because they did not meet all the inclusion criteria. Each participant was presented with one list.

Figure 1 shows the mean word and phoneme durations of the *can* and *can't* tokens in the stimuli. These durations indicate that the reduced *can't* tokens produced by the native speakers of American English are less ambiguous than those produced by the native speakers of Spanish. In American English, the velar stop and the vowel of unreduced and reduced *can't* are, on average, longer than the corresponding segments in *can*, probably because *can't* tends to carry some type of accent, whereas *can* does not. In contrast, in the stimuli produced by the native speakers of Spanish each segment of reduced *can't* is more similar in duration to the corresponding segment in *can* than to the one in unreduced *can't.* The segments in reduced *can't* thus seem to be more similar to those in unreduced *can't* in the stimuli produced by native speakers of American English and to those in *can* in the stimuli produced by native speakers of Spanish.

This pattern is no coincidence. We analyzed all 309 unreduced and 170 reduced *can't* tokens and all 1094 *can* tokens from the Buckeye Corpus and the NCSE that were not utterance final and that appeared without background noise or overlapping speech. Of these tokens, 823 were produced by 40 different native speakers of American English (97 reduced *can't*, 199 unreduced *can't*, and 527 *can*). The remaining 750 tokens were produced by 34 different native speakers of Spanish (73 reduced *can't*, 110 unreduced *can't*, and 567 *can*). These tokens show the same durational patterns. We analyzed the duration of the velar consonant, the duration of the vowel, and the duration of the nasal consonant using linear mixed effects models that incorporated the fixed independent variables Type of Speech (Buckey Corpus versus NCSE) and Word type (unreduced *can't*, reduced *can't*, and *can*), and the random effects Speaker and Following Word. The model for the duration of the vowel showed an interaction between the two fixed predictors ($\chi 2(2) = 30.75$, $p < .001$; this type II Wald chi-square test was produced by the Anova function from the Car package for R, Fox & Weisberg, 2011, which we ran over the final linear mixed effects models). Further analysis of the durations of the vowels in the tokens produced by the native speakers of American English revealed that the vowel is significantly longer in reduced *can't* than in *can* (β = 47.89 ms, *t* = -13.07, p < .001), while there is no statistically significant difference between reduced and unreduced *can't* (β = -2.55, *t* = -0.64, p > 0.1). The tokens from the NCSE show the opposite pattern: the vowel is significantly longer in unreduced *can't* than in reduced *can't* (β = 21.50, *t* = 3.35, p < .001), while there is no statistically significant difference in vowel duration between reduced *can't* and *can* (β = 2.06, *t* = 0.38, p > 0.1).

The difference in vowel duration between unreduced and reduced *can't* on the one hand and *can* on the other in the stimuli produced by native speakers of American English strongly correlates with differences in vowel quality. The majority (over 70%) of the *can* tokens selected for our stimuli were transcribed in the Buckeye Corpus with the symbols “ih” or “en” to represent the vowel, whereas the vast majority (90%) of unreduced *can't* tokens in our stimuli were transcribed with “ae” and “aen”. Importantly, the vowel of the reduced *can't* tokens in our stimuli received similar transcriptions as the vowel in the unreduced *can't* tokens: most of them (79%) were also transcribed with “ae” and “aen”. The durational differences in the stimuli thus correlate with differences in vowel reduction.

We also tried to obtain more information about the exact pronunciation of the vowel in the stimuli produced by the native speakers of Spanish. We estimated the first formant (F1) and the second formant (F2) of the vowel in each token of *can*, unreduced *can't*, and reduced *can't* on the basis of a window of 25 ms around the center of the vowel. We created several pairs of F1 – F2 plots, applying different types of normalization (Flynn & Foulkes, 2011). Each pair consisted of one plot for the male speakers and one for the female speakers. These plots suggest that the type of stimulus (reduced *can't*, unreduced *can't*, *can*) hardly affected the quality of the vowel. These formant estimations have to be interpreted with care, however, because the formants were not normalized on the basis of the speakers' other vowels and, more importantly, some vowels were very short and did not contain stable central parts of 25 ms. The results, however, strengthen our impression that the native speakers of Spanish did not vary the quality of the vowel as a function of whether the intended word was *can* or *can't*. They typically pronounced *can* and *can't* with the same vowel, that is, with a vowel between /æ/ and /ɑ/.

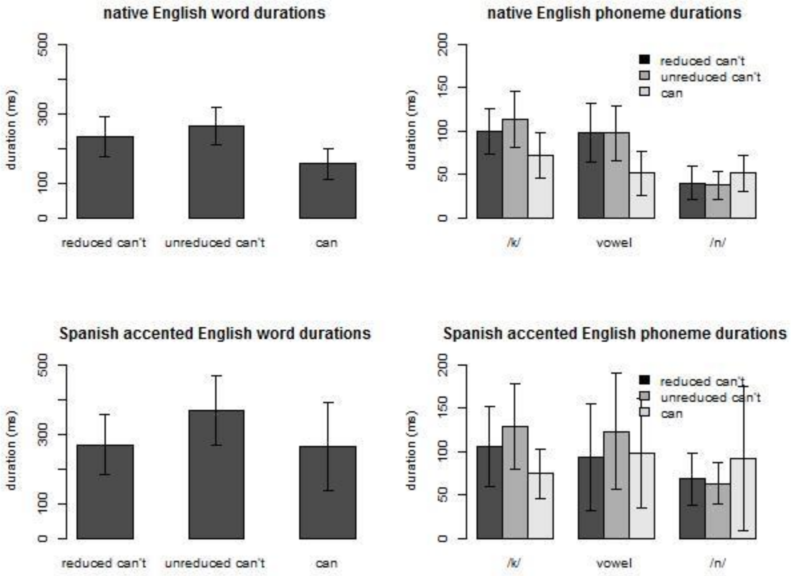

Figure 1: ***Mean durations (in ms) of reduced and unreduced*** **can't** ***and of*** **can*****, and of their phonemes other than /t/, in the stimuli produced by the native speakers of American English (top) and in the stimuli produced by the native speakers of Spanish (bottom, referred to as "Spanish accented English"). The error bars represent one standard deviation.***

In order to obtain information about the participants' expectations of the likelihoods of *can* versus *can't* in the stimuli, we invited them to read each trigram on screen together with its positive or negative counterpart (e.g., *I can go* and *I can't go*) and to rate the probability of the positive versus the negative counterpart. The ninety-six unique trigrams in the experiment were presented in one of seven randomized lists.

2.1.3 Procedure

Participants were tested individually in a sound-attenuated booth. As mentioned above, the experiment consisted of three parts: an auditory comprehension, a frequency rating, and a proficiency assessment.

For the auditory comprehension part, participants received instructions on the screen that they were going to hear short audio fragments, but not that they would hear native and non-native English.

After each fragment they were asked to indicate whether the second word was *can* or *can't*, and to do so as quickly and as accurately as possible. Participants gave their answers by means of button presses on a button box (*can*-responses with the dominant hand). They listened to the auditory stimuli through headphones. A trial contained one stimulus: participants saw a fixation cross for 400 ms in the middle of the screen, followed by a 200 ms pause before the stimulus was played. After the participant's button press, or 3650 ms after stimulus onset if the participant did not press a button, another 200 ms pause followed before the start of the next trial. Participants took a short break at the end of each block. This auditory comprehension part lasted about 20 minutes.

In the second part of the experiment, participants estimated the relative frequency of occurrence of *can* versus *can't* for each trigram in the auditory comprehension part. For each trigram, they saw a seven point scale with the trigram with *can* (e.g., *I can remember*) on the left end of the screen and the trigram with *can't* (e.g., *I can't remember*) on the right end. The instructions read: "Please indicate which of the two occurs more frequently in English". Participants used the 1-7 keys at the top of a keyboard in order to indicate how frequently they thought that the positive trigram occurs in English relative to the negative trigram, and vice versa. If, for example, a participant typed a '6', this number implied that the participant estimated that *can* seldom, and *can't* very frequently occurs in the given context in English. There was no time limit and the next trial appeared on the screen when the participant pressed the button. This part consisted of two blocks and participants took a short break between the two. The frequency rating task lasted about 20 minutes.

The third part of the experiment consisted of the LexTALE task (Lemhöfer & Broersma, 2012), a visual lexical decision task. It consists of three familiarization items, 40 real English words and 20 non-words that are orthographically legal and pronounceable in English. Participants gave their answers by means of button presses on a button box (*yes*-responses with the dominant hand). There was no time limit and the next trial appeared on the screen when the participant pressed the button. The LexTALE task took approximately 5 minutes.

2.2 Results: frequency ratings

We first investigated the homogeneity of each listener group with respect to the probability ratings of *can* versus *can't* collected in the second part of the experiment. We determined the interrater agreement for each listener group with Kendall W (see Table 3). The results show little difference between the groups: all groups show low but statistically significant interrater agreements, meaning the non-native groups did not show substantially more variation than the native group.

In order to establish whether the ratings provided by the non-native listeners deviated substantially from those provided by the native English listeners, we determined the correlations of the trigram ratings averaged over participants between each of the groups of non-native listeners on the one hand and the group of native English listeners on the other. The native Spanish and Dutch listeners showed almost identical significant correlations of 0.56 with the native English listeners (the native Spanish listeners: $t(94) = 6.4797$, p < 0.0001; the native Dutch listeners: $t(94) = 6.4761$, $p < 0.0001$). The native Mandarin listeners showed no significant correlation ($p > 0.1$), possibly because of lack of power (this group only consisted of 21 participants versus, for instance, 40 participants in the group of native Spanish listeners).

Table 3: ***Interrater agreement for the four listener groups.***

Listener Group	Wt	$\chi^2(95)$	*p*
English	0.10	328	< 0.0001
Spanish	0.09	342	< 0.0001
Dutch	0.12	331	< 0.0001
Mandarin	0.08	157	< 0.0001

2.3 Accuracy in the comprehension task

2.3.1 Description of the statistical analyses

We compared listener groups' accuracies in the comprehension test by means of logistic mixed effects models with the binomial link function. We tested for fixed effects of three predictors of interest and the interactions between the three: Listener Group (native listeners of English, Spanish, Mandarin, or Dutch), Stimulus Type (reduced *can't*, unreduced *can't*, or *can*), and Type of Speech (stimuli produced by native speakers of American English or by native speakers of Spanish). We also included the Relative Frequency

Rating of each trigram as indicated by the relevant participant. For the negative trigram, this Relative Frequency Rating equaled the number typed in by the participant, while for the positive trigram, we calculated this Relative Frequency Rating as eight minus the number typed in by the participant. If, for example, a participant typed a ‘6’, the positive trigram (e.g., *I can remember*) received a score of ‘2’ and the negative trigram (e.g., *I can’t remember*) received a score of ‘6’, which means that the participant estimated that *can* seldom occurs in the given context in English, while *can’t* occurs very frequently. Furthermore, we tested for three random factors: Participant, Speaker of the Stimulus, and Stimulus.

We tested for more fixed control predictors (e.g., the Participant's LexTALE score, Trial Number, Stimulus Duration) as well as for random slopes, but in the final models that we report below, these control predictors are not included for the following reasons: first, and most importantly, none of the additions impacted the effects of the four main predictors to such an extent that we would have come to different conclusions. In other words, the effects of the main predictors were sufficiently robust to remain statistically significant also in the presence of other fixed predictors and in the presence of random slopes. Since the addition of more fixed predictors and of random slopes had no impact on the effects of the main predictors, the models including these additional predictors were unnecessarily complex. Secondly, we wanted to avoid the risk of over-fitting the models to our specific dataset, which would reduce the generalizability of our findings. Lastly, the R statistical package (R Core Team, 2014) provided warning messages for some models including additional predictors, stating that it failed to produce a reliable model.

We also built models with Vowel Duration instead of Type of Speech as fixed predictor. These models could show whether the listener groups differed in how much they relied on vowel duration in their comprehension of the stimuli. Unfortunately, none of the models converged; so we do not report the results of these models.

Fixed effects and interactions were only included in a model if they were statistically significant ($p < .05$). Random factors were only included if they significantly improved the model, which was tested by means of likelihood ratio tests. We established the random structure based on a simple model including only Listener Group as fixed factor before we added more fixed factors.

2.3.2 Participants' accuracy for all stimulus types

Figure 2 shows the mean accuracies of the four Listener Groups on the three Stimulus Types for each Type of Speech. It shows that the native English listeners had little difficulty classifying the stimuli produced by native speakers of American English, including those with reduced tokens of *can't*. The native Dutch listeners produced similar results, although their error rates were a little higher overall. Both groups had more difficulty correctly identifying the reduced *can't* tokens produced by native speakers of Spanish than those produced by native speakers of American English. The native listeners of Spanish and Mandarin showed much higher error rates, especially for reduced tokens of *can't* produced by native speakers of American English. The native listeners of Mandarin also showed a bias towards *can*.

We first performed statistical analyses on the full dataset. Table 4 presents our final model in an analysis of deviance table, produced by the Anova function from the Car package (Fox & Weisberg, 2011) for R.

We found a simple effect of Relative Frequency Rating, showing that the four participant groups based their decisions to the same extent on their estimations of the likelihoods of *can* and *can't* in the trigrams presented. Further analyses showed that the contribution of the Relative Frequency Ratings to the participants' responses was low (as indicated, for instance, by the low coefficients of Relative Frequency Rating in Tables 5 and 7, presented below).

In addition, we found simple effects of and a three-way interaction between Listener Group, Stimulus Type, and Type of Speech. To further explore these interactions, we performed additional analyses on subsets of our data split by Stimulus Type.

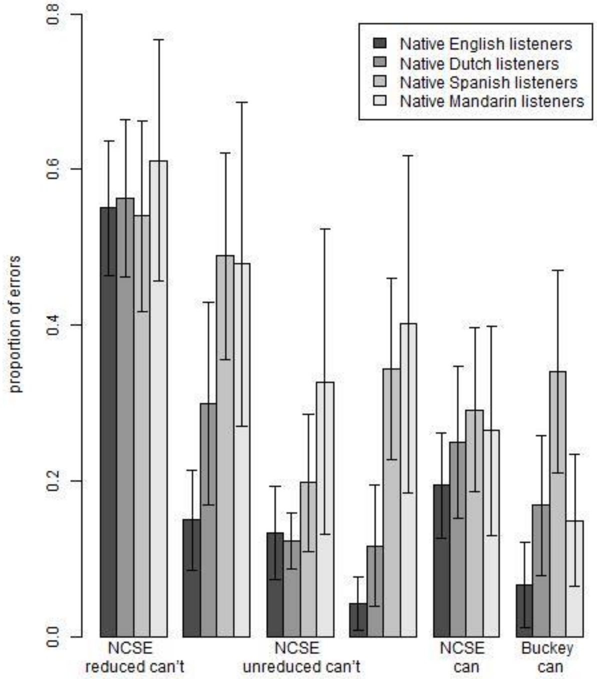

Figure 2: ***Proportions of incorrect responses to the*** **can** ***and*** **can't** ***stimuli, split by Type of Speech (English produced by native speakers of Spanish and of American English), Stimulus Type, and Listener Group. The error bars represent one standard deviation.***

Table 4: ***Analysis of deviance table (Type II Wald chi-square tests) for the fixed effects in our final overall model predicting the accuracies of participants' responses.***

Fixed effects	χ^2	*Df*	*p*
Listener Group	228.19	3	< .001
Stimulus Type	194.89	2	< .001
Type of Speech	5.17	1	< .05
Relative Frequency Rating	66.23	1	< .001
Listener Group x Stimulus Type	550.99	6	< .001
Listener Group x Type of Speech	769.68	3	< .001
Stimulus Type x Type of Speech	34.32	2	< .001
Listener Group x Stimulus Type x Type of Speech	48.46	6	< .001

2.3.3 Participants' accuracy on reduced can't *stimuli*

Table 5 shows the final model for the accuracies on the reduced *can't* stimuli. The interaction between Listener Group and Type of Speech shows that the Type of Speech affected the listener groups differently. In order to investigate this interaction pattern in more detail, we split the reduced *can't* data by Type of Speech.

Table 5: ***Statistical model for the accuracy of participants' responses to reduced*** **can't** ***stimuli. The intercept represents native English listeners, listening to stimuli produced by native speakers of American English. The term 'Spanish English' refers to the English stimuli produced by native speakers of Spanish.***

Fixed effects	*B*	*Z*	*p*
Intercept	1.68	9.68	< .001
Listener Group (Dutch)	-0.93	-6.17	< .001
Listener Group (Spanish)	-1.86	-13.32	< .001
Listener Group (Mandarin)	-1.77	-10.75	< .001
Type of Speech (Spanish English)	-2.14	-11.21	< .001
Relative Frequency Rating	0.05	3.77	< .001
Listener Group (Dutch) x Type of Speech (Spanish English)	0.88	7.06	< .001
Listener Group (Spanish) x Type of Speech (Spanish English)	1.90	16.40	< .001
Listener Group (Mandarin) x Type of Speech (Spanish English)	1.49	10.89	< .001

Analyses of only the accuracies for the reduced *can't* stimuli produced by native speakers of American English showed that the native English listeners outperformed all three non-native listener groups ($\beta_{Spanish}$ = -1.89, z = -11.83, p < .001; β_{Dutch} = -0.93, z = -5.38, p < .001; $\beta_{Mandarin}$ = -1.76, z = -9.30, p < .001). Fitting the same model with the native Dutch listeners on the intercept revealed that these listeners performed more accurately than both the native Spanish listeners (β = 0.96, z = 5.99, p < .001) and native Mandarin listeners (β = 0.83, z = 4.40, p < .001). The same model with the native Mandarin listeners on the intercept showed that these listeners did not differ from the native Spanish listeners.

Our analyses of the accuracies for the reduced *can't* stimuli produced by native speakers of Spanish showed that there was no difference in performance between the native English, Spanish and Dutch listeners. The native English listeners (β = 0.33, z = 2.07, p < .05) and the native Spanish listeners (β = 0.37, z = 2.43, p < .05) outperformed the native Mandarin listeners.

2.3.4 Participants' accuracy on unreduced can't *stimuli*

Table 6 shows the final model for the accuracies on the unreduced *can't* stimuli. Again, we found an interaction between Listener Group and Type of Speech, and therefore split the data by Type of Speech.

Table 6: ***Statistical model for the accuracy of participants' responses to unreduced* can't *stimuli. The intercept represents native English listeners, listening to stimuli produced by native speakers of American English. The term 'Spanish English' refers to the English stimuli produced by native speakers of Spanish.***

Fixed effects	*B*	*Z*	*p*
Intercept	3.36	16.33	< .001
Listener group (Dutch)	-1.16	-5.83	< .001
Listener group (Spanish)	-2.74	-14.91	< .001
Listener group (Mandarin)	-3.00	-14.02	< .001
Type of Speech (Spanish English)	-1.01	-3.68	< .001
Listener group (Dutch) x Type of Speech (Spanish English)	1.29	7.95	< .001
Listener group (Spanish) x Type of Speech (Spanish English)	2.06	14.49	< .001
Listener group (Mandarin) x Type of Speech (Spanish English)	1.43	9.25	< .001

Separate analyses of the accuracies obtained for the stimuli produced by native speakers of American English show that the native English listeners performed more accurately than all three non-native listener groups ($\beta_{Spanish}$ = -2.77, *z* = -14.17, *p* < .001; β_{Dutch} = -1.17, *z* = -5.50, *p* < .001; $\beta_{Mandarin}$ = -3.03, *z* = -13.31, *p* < .001). An additional analysis with the native Dutch listeners on the intercept revealed that these listeners performed more accurately than the native listeners of Spanish (*β* = 1.60, *z* = 8.33, *p* < .001) and of Mandarin (*β* = 1.86, *z* = 8.73, *p* < .001). The same analysis with the native Spanish listeners on the intercept revealed that these listeners performed as accurately as the native Mandarin listeners.

Our analyses of the accuracies for the unreduced *can't* stimuli produced by native speakers of Spanish revealed that the native English listeners and the native Dutch listeners performed equally accurately, and more accurately than the native Spanish listeners (*β* = -0.68, *z* = -3.42, *p* < .001) and the native Mandarin listeners (*β* = -1.62, *z* = -6.96, *p* < .001). Fitting the same model with the native Spanish listeners on the intercept showed that these listeners outperformed the native Mandarin listeners (*β* = 0.93, *z* = 4.20, *p* < .001).

2.3.5 Participants' accuracy on can *stimuli*

Table 7 shows the final model for the response accuracies for the *can* stimuli. Again, we found an interaction between Listener Group and Type of Speech, and therefore split the data by Type of Speech.

Table 7: ***Statistical model for the accuracy of participants' responses to*** **can** ***stimuli. The intercept represents native English listeners, listening to stimuli produced by native speakers of American English. The term 'Spanish English' refers to the English stimuli produced by native speakers of Spanish.***

Fixed effects	*B*	*Z*	*P*
Intercept	2.72	19.93	< .001
Listener Group (Dutch)	-1.08	-6.70	< .001
Listener Group (Spanish)	-2.09	-14.09	< .001
Listener Group (Mandarin)	-0.91	-5.12	< .001
Type of Speech (Spanish English)	-1.22	-9.68	< .001
Relative Frequency Rating	0.03	3.04	< .01
Listener Group (Dutch) x Type of Speech (Spanish English)	0.74	7.39	< .001
Listener Group (Spanish) x Type of Speech (Spanish English)	1.52	16.67	< .001
Listener Group (Mandarin) x Type of Speech (Spanish English)	0.49	4.48	< .001

We first analyzed the accuracies to the *can* stimuli produced by native speakers of American English, which showed that, again, the native English listeners outperformed all non-native listener groups ($\beta_{Spanish}$ = -2.16, *z* = -12.65, *p* < .001; β_{Dutch} = -1.13, *z* = -6.11, *p* < .001; $\beta_{Mandarin}$ = -0.95, *z* = -4.66, *p* < .001). The same analysis with the native Dutch listeners on the intercept revealed that these listeners performed as accurately as the native Mandarin listeners, but were more accurate than the native Spanish listeners (β = -1.03, *z* = -6.01, *p* < .001).

We then analyzed the accuracies to the *can* stimuli produced by native speakers of Spanish. These analyses showed that the native English listeners performed better than all non-native listener groups (β_{Dutch} = -0.36, *z* = -2.41, *p* < .05; $\beta_{Spanish}$ = -0.59, *z* = -4.28, *p* < .001; $\beta_{Mandarin}$ = -0.42, *z* = -2.59, *p* < .01). Fitting the model again with the native Dutch listeners on the intercept showed that the accuracies of the three non-native listener groups did not differ from each other.

3.0 General discussion and conclusion

Previous studies have shown that non-native listeners of a language have difficulty understanding reduced word pronunciation variants in that language (e.g., Nouveau, 2012; Wong et al. 2015; ten Bosch et al., 2016). The aim of the present study was to contribute to answering the question of why this is the case, by focusing on the direct and indirect effects of the phonotactic constraints in the listener’s native language. Direct effects would be non-native listeners’ bias to hear or not to hear segments that are phonotactically illegal in their native languages, while indirect effects would be insensitivity to subsegmental cues signaling these segments. We compared how native English listeners and different groups of non-native listeners with intermediate to advanced proficiencies in English identify tokens of unreduced *can* and *can’t* and of reduced *can’t* (without clear /t/) in stretches of casual conversations (consisting of a pronoun, the target item, and an infinitive) in English.

The stretches were produced by native speakers of American English or by native speakers of Spanish who lived in the area of Madrid and had proficiency levels in English at the A2 – B1 / B2 level according to the Common European Framework of Reference for Languages (Council of Europe, 2011). The speech from the native and non-native speakers differ from each other in whether the vowel of reduced *can't* is more similar in duration to the vowel of unreduced *can't* (as in the stimuli produced by native speakers of American English) or to the vowel of *can* (as in the stimuli produced by native speakers of Spanish). Furthermore, in the stimuli produced by native speakers of American English, reduced *can’t* was typically produced with a full vowel, like unreduced *can*’*t*, whereas *can* often contained a reduced vowel. We could not find a similar difference in vowel quality for the tokens of reduced and unreduced *can’t* versus *can* produced by native speakers of Spanish. Future detailed phonetic analyses may reveal more subsegmental differences between *can’t* and *can* produced by the two speaker groups. The general pattern, however, seems clear: the reduced *can’t* tokens produced by native speakers of Spanish contained fewer subsegmental cues to their identity than the reduced *can’t* tokens produced by native speakers of American English.

Native English listeners and three groups of non-native listeners (native listeners of Spanish, Mandarin, and Dutch) identified the tokens of reduced and unreduced *can’t* and of *can*.
If the phonotactic constraints of the listeners’ native languages play an important role, we expect the native Spanish listeners and Mandarin to produce similar comprehension patterns because both Spanish

and Mandarin have very few words (if any) ending in /nt/, and there is no strong vowel reduction (which may function as a cue to *can* versus *can't* in English). In contrast, both English and Dutch have many words ending in /nt/ and both have strong vowel reduction.

The native English listeners had little difficulty classifying the tokens produced by native speakers of American English, including reduced tokens of *can't*. This result shows that native English listeners are able to rely on subsegmental cues to distinguish between reduced *can't* and c*an*, probably including the duration and the quality of the vowel and acoustic traces of the reduced /t/ itself. This result is in line with findings by Pitt (2009) and by Sumner and Samuel (2005), among others, who also showed that native English listeners can easily comprehend words with reduced /t/s. Moreover, this finding is in line with earlier evidence that subsegmental properties of the acoustic signal facilitate the comprehension of reduced word pronunciation variants by native listeners (e.g., Manuel 1991; Mitterer & Ernestus, 2006; Zimmerer & Reetz, 2014).

In general, the native English listeners were also well able to understand the unreduced *can* and *can't* tokens produced by native speakers of Spanish, although they performed less well on these stimuli than on the stimuli produced by native speakers of American English. Our native English listeners thus had little difficulty understanding English produced by native speakers of Spanish. This result is as expected: the stimuli produced by the native speakers of Spanish were presented in blocks, and previous research has shown that native listeners can quickly adapt to speech produced with a foreign accent if it does not substantially deviate from native speech (e.g., Clarke & Garrett, 2004; Witteman et al., 2013). Our data suggest that this adaptation also occurs when the presented speech consists of short phrases spliced from casual conversations.

In contrast, the native English listeners often misinterpreted reduced *can't* tokens produced by native speakers of Spanish. Since these tokens contained less clear acoustic cues to *can't* than the reduced *can't* tokens produced by native speakers of American English, this finding supports the hypothesis that listeners of English rely on these subsegmental cues. Further research could investigate the absence of which type of subsegmental cue (e.g., vowel duration, vowel quality, or acoustic traces of the reduced /t/) is most cumbersome for native English listeners.

The native Spanish listeners had more experience with Spanish accented English than the other listener groups. Nevertheless, they did not outperform the native listeners of English or of Dutch on the stimuli produced by native speakers of Spanish. This pattern of results shows that the native Spanish listeners did not benefit much from their larger experience with Spanish accented English. Like the native English listeners, the native Dutch listeners seemed able to quickly adapt to the Spanish accent in the English stimuli produced by the native speakers of Spanish. This observation seems to support earlier evidence that learners of a foreign language can quickly adapt to a foreign accent in their foreign language (e.g., Weber et al., 2014). Our results contribute to this line of research by showing that this quick adaptation also takes place when non-native listeners are presented with stretches of conversational speech rather than clear speech.

The listeners' proficiencies in English seem to have played a minor role in their identification accuracies. The average proficiency level in English was slightly higher in the group of native Dutch listeners than in the group of native Spanish listeners (as indicated by the LexTale test, Lemhöfer & Broersma, 2012). However, if the analysis of the comprehension results is restricted to only a subset of the participants in which the native Spanish listeners have a higher average proficiency level than the native Dutch listeners, the results are the same. Proficiency level in English cannot be the only explanation for the differences in comprehension patterns between the different groups of non-native listeners.

Similarly, the role of the non-native listener's experience with the variant of English presented in the experiment seems not to have played an important role. The native Dutch listeners probably had more exposure to American English than the native Spanish listeners, since American English is abundant on Dutch television and radio. Accordingly, the native Dutch listeners outperformed the native Spanish listeners. However, the native Mandarin listeners had also probably more exposure to North American English than the native Spanish listeners because they were enrolled in a university program taught in

North American English and lived in a country where North American English was spoken. Nevertheless, they did not outperform the native Spanish listeners. It is possible that listeners' ability to quickly adapt to speech produced by non-native speakers (Weber et al., 2014) has reduced the effect of their experience with the specific variant of English presented in the experiment.

The native Spanish listeners produced comprehension results very similar to those produced by the native Mandarin listeners, while the comprehension results from the native Dutch listeners pattern with those from the native English listeners. This pairing of the native Spanish listeners with the native Mandarin listeners and of the native English listeners with the native Dutch listeners suggests that the major comprehension patterns result from the phonologies of the listeners' native languages.

The native Mandarin listeners showed a clear direct effect of their native language's ban on /nt/ clusters: they more often misidentified unreduced *can't* as *can* (in 35% of all the unreduced *can't* stimuli) than they misinterpreted unreduced *can* as *can't* (in approximately 20% of all the unreduced *can* stimuli). Native Mandarin listeners thus not only tend to omit segments (e.g., the /t/ of word-final /nt/) in a foreign language such that the words comply to the phonotactic constraints of their native language (e.g., Chang, 2001), but they also tend to ignore these segments when listening to that foreign language. Added to the finding that language learners tend to alter segments that do not occur in their native languages in the same way in production and in perception (Moreton, 2002), our finding indicates that learners of a language generally treat phoneme sequences that are phonotactically illegal in their native languages similarly in the production and comprehension of the foreign language: in both, a segment is changed (Moreton, 2002) or omitted / ignored (as shown in the present study).

We argue that native Mandarin listeners, who reduce word-final /t/ in English because of the phonotactic constraints in their native language, do not compensate in perception for their reduction patterns, while native listeners do, because the mechanisms underlying /t/ reduction in the two speaker groups are different. Native speakers of Mandarin mostly delete /t/ from word-final /nt/ in English due to phonological processes (based on the phonotactic constraints of their native language) affecting the mental representations of words, which both the production and comprehension processes operate on. In contrast, native speakers of English may reduce /t/ primarily during the articulation process.

In contrast to the native Mandarin listeners, the native Spanish listeners showed no clear bias for either *can* or *can't* when hearing unreduced tokens of these words. They thus showed no clear direct effect of the ban on word-final /nt/ in their native language. Nevertheless, these listeners produced significantly more errors for the unreduced tokens of *can't* and *can* than the native listeners of English or of Dutch. A likely explanation is that the same processes are at work as in the native Mandarin listeners, impeding the native Spanish listeners' ability to interpret the final /t/ of unreduced *can't*, while, simultaneously, the native Spanish listeners are aware of their difficulties with interpreting word-final [t] and (consciously) (over)compensate for these difficulties. The data from the native Spanish listeners therefore also lend some support for a direct effect of phonotactic constraints in the listener's native language on the interpretation of casual speech in a foreign language.

We think that the native Spanish listeners showed a smaller direct effect of their native language's phonotactics than the native Mandarin listeners because the native Spanish listeners had, on average, a higher proficiency level in English. This suggests that the direct effect of the native language's phonotactic constraints role decreases with proficiency level. It may be the case that at beginning stages of proficiency, non-native listeners are more strongly affected by direct effects of the phonotactic constraints of their native languages. The effects may change with the learning process as listeners advance to higher proficiency levels.

Comparison of the listener groups' comprehension patterns also shows evidence for indirect effects of the phonotactic constraints in their native languages on the comprehension of foreign casual speech. The native listeners of Spanish and Mandarin resemble the native listeners of English and of Dutch in the number of errors for the reduced tokens of *can't* produced by native speakers of Spanish (the number of errors produced by the native Mandarin listeners was slightly higher, showing their bias for *can*). In contrast, the native Spanish and Mandarin listeners produced significantly more errors than the native English and Dutch listeners for reduced tokens of *can't* produced by native speakers of American

English. This combination of results indicates that the native Spanish and Mandarin listeners could not benefit as much as the native English and Dutch listeners from the subsegmental information in the native American English stimuli. This result supports the hypothesis that non-native listeners may show indirect effects of their native languages' phonotactic constraints by not taking full advantage of the subsegmental information in the speech signal. Moreover, the indirect effect hardly depends on the listeners' proficiency level, since we do not see a difference between the native Spanish and Mandarin listeners.

The question arises of how to account for this smaller sensitivity to subsegmental cues in models of speech comprehension. Some models, like Trace (McClelland & Elman, 1986), neither specify how native listeners extract phoneme strings from the speech signal nor do they account for the role of subsegmental details of the speech signal in this process. Shortlist B (Norris & McQueen, 2008) postulates that subsegmental information in the speech signal may affect the listener's estimation of the probability that a given phoneme is present in this signal. This model can account for non-native listeners' smaller sensitivity to (some) subsegmental information by assuming that the probability of a phoneme given the speech signal depends on the listener's sensitivity to the subsegmental information in the signal.

Exemplar theory (e.g., Goldinger, 1998; Johnson, 2004) assumes that many tokens of a word ever produced or perceived by the language user are mentally stored with all their acoustic detail, and thus including the subsegmental cues. Exemplar theory can thus explain the role of subsegmental cues in speech comprehension. Without additional assumptions, however, the theory cannot account for why some listener groups do not rely on some acoustic cues whereas other groups do, depending on their native languages: all subsegmental cues are equally relevant, for all listener groups. Our findings clearly show that the theory has to be adapted in this respect. The mechanisms computing the fit between the speech signal and the stored exemplars should put more weight on some properties of the acoustic signal than on others, depending on the listener's sensitivity to the different types of subsegmental information. Such a mechanism has never been proposed so far within exemplar-based theories. Another possibility is that listeners only store the acoustic details they are sensitive to. This possibility has also not been explored before.

Although the native Spanish and Mandarin listeners produced many errors for reduced *can't* tokens, they did not consistently classify them as *can*. Importantly, they classified fewer reduced *can't* tokens than *can* tokens as *can*, which implies that they were able to rely on (some) subsegmental cues to *can't* to some extent. Thus, these non-native listeners have acquired sensitivity to cues that they are not sensitive to in their native languages (e.g., exact vowel quality) or they have learnt to assign more functions to the acoustic details of the speech signal that they were already sensitive to from their native languages, such that these cues now also signal word-final /t/ of *can't*. For instance, they may have learnt that acoustic traces of /t/ may also cue /t/ in word-final /nt/. This finding is in line with the perceptual learning literature showing that listeners can adjust their interpretation of a sound in their native and in a non-native language on the basis of recent speech input (e.g., Sjerps & McQueen, 2010). Our study contributes to this line of research by showing that non-native listeners can also learn to interpret subsegmental properties of the acoustic signal cuing the presence of a segment, and that the learning may be long lasting.

The many errors produced by the native Spanish and Mandarin listeners for reduced *can't* also show, however, that learning to interpret an acoustic cue under natural learning conditions is not as easy as learning a new sound or the correct interpretation of an ambiguous sound in some perceptual learning experiments in the laboratory. Future research has to investigate whether the difference in results between our study and the results of some sound learning laboratory experiments (e.g., Sjerps & McQueen, 2010) is due to the learning conditions (natural learning conditions versus perceptual learning experiments conducted in the laboratory) as claimed by Pallier et al. (1997), or whether the learners have to acquire a new sound or adjust their phoneme boundaries, versus learning to interpret a subsegmental cue for overcoming a segment reduction.

Finally, in our analysis of the comprehension results, we investigated the role of the learner's estimations of the probabilities of *can* and *can't* in the presented phrases, which we had gathered in a frequency rating experiment after the comprehension experiment. Listeners may rely on these estimations, in addition to the acoustic signal. As expected, the individual ratings predicted the responses in the comprehension experiment. The statistical analyses also showed, however, that the predictive power of the ratings was small for all participant groups, including the native listeners (see the low coefficients for Relative Frequency Rating in Tables 5 and 7, and the absence of an effect of Relative Frequency Rating in Table 6). This is possibly due to participants being often confronted with infrequent combinations of *can* / *can't* and an infinitive, which may have discouraged them from taking their expectations into account.

In conclusion, this study has shed more light on how adult non-native listeners process casual conversations in a foreign language. We presented different listener groups with spontaneously uttered phrases, each containing a reduced /t/ whose presence is crucial to the meaning of the phrase. We observed a crucial role for the phonotactic constraints of the listeners' native languages. These constraints may have a direct effect, especially on non-native listeners of intermediate proficiency levels, and make these listeners ignore the segments that occur in illegal segmental contexts according to the phonotactic constraints in their native languages. In addition, these constraints may have more indirect effects, even on listeners of advanced proficiency levels, causing the non-native listeners not to take full advantage of the subsegmental cues in the speech signal.

4.0 Acknowledgements

The authors would like to thank Benjamin V. Tucker and his colleagues at the Department of Linguistics of the University of Alberta for conducting the experiment with native listeners of English and of Mandarin. We would also like to thank Natasha Warner for her valuable input. Finally, we would like to thank José Manuel Pardo of the Grupo de Tecnología del Habla at the Escuela Técnica Superior de Ingenieros de Telecomunicación of the Universidad Politécnica de Madrid for allowing us to use a sound-attenuated booth in the group's laboratory to conduct the experiment with native Spanish listeners, and for his assistance in the recruitment of participants. This work was partly funded by an ERC starting grant (284108) to the first author.

References

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

Bent, T., Bradlow, A. R., & Smith, B. L. (2007). Segmental errors in different word positions and their effects on intelligibility of non-native speech. In O. S. Bohn, & M. J. Munro (Eds.), *Language experience in second language speech learning* (pp. 331-347). Amsterdam: John Benjamins Publishing.

Bergem, D. R. van (1993). Acoustic vowel reduction as a function of sentence accent, word stress, and word class. *Speech communication*, *12*(1), 1-23.

Bosch, L. ten, Giezenaar G., Boves, L. & Ernestus, M. (2016). Modeling language-learners' errors in understanding casual speech. In G. Adda, V. Barbu Mititelu, J. Mariani, D. Tufiş, & I. Vasilescu (Eds.), *Errors by humans and machines in multimedia, multimodal, multilingual data processing. Proceedings of Errare 2015* (pp. 07-121). Bucharest: Editura Academiei Române.

Chang, J. (2001). Chinese speakers. In M. Swan, & B. Smith (Eds.), *Learner English: A teacher's guide to interference and other problems* (2nd ed., pp. 310-324). Cambridge: Cambridge University Press.

Clarke, C. M., & Garrett, M. F. (2004). Rapid adaptation to foreign-accented English. *The Journal of the Acoustical Society of America*, *116*(6), 3647-3658.

Cobb, K., & Simonet, M. (2015). Adult second language learning of Spanish vowels. *Hispania*, *98*(1), 47-60.

Coe, N. (2001). Speakers of Spanish and Catalan. In M. Swan, & B. Smith (Eds.), *Learner English: A teacher's guide to interference and other problems* (2nd ed., pp. 90-112). Cambridge: Cambridge University Press.

Council of Europe (2011). *Common European Framework of Reference for Languages: Learning, Teaching, Assessment.* Council of Europe.

Dilley, L. C., & Pitt, M.A. (2010). Altering context speech rate can cause words to appear or disappear. *Psychological Science*, 21, 1664-1670.

Ernestus, M., & Warner, N. (2011). An introduction to reduced pronunciation variants. *Journal of Phonetics*, *39*, 253-260.

Flynn, N. & Foulkes, P. (2011). Comparing vowel formant normalization procedures. In Lee, W. & Zee, E. (Eds.), *Proceedings of the 17th International Conference on Phonetic Sciences* (pp. 83-686). Hong Kong.

Fox, J., & Weisberg, S. (2011). *An R Companion to Applied Regression* (2nd ed). Thousand Oaks CA: Sage. http://socserv.socsci.mcmaster.ca/jfox/Books/Companion.

Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review*, 105, 251-279.

Gut, U. (2003). Non-native speech rhythm in German. In *Proceedings of the ICPhS conference* (pp. 2437-2440).

Guy, G. R. (1991). Contextual conditioning in variable lexical phonology. *Language variation and change*, *3*, 223-239.

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.

Johnson, K. (2004). Massive reduction in conversational American English. In *Proceedings of the 10th international symposium on spontaneous speech: Data and analysis* (pp. 29-54). Tokyo, Japan.

Kouwenhoven, H., Ernestus, M., & Van Mulken, M. (to appear). Register variation by Spanish users of English: The Nijmegen Corpus of Spanish English. *Corpus Linguistics and Linguistic Theory*. doi: 10.1515/cllt-2013-0054.

Labov, W., Ash, S., & Boberg, C. (2005). *The atlas of American English: Phonetics, phonology and sound change.* Walter de Gruyter.

Labov, W. (1972). *Sociolinguistic patterns*. Philadelphia: University of Pennsylvania Press.

Lemhöfer, K., & Broersma, M. (2012). Introducing LexTALE: A quick and valid Lexical Test for Advanced Learners of English. *Behavior Research Methods, 44,* 325-343.

Manuel S. Y. (1991). Recovery of "deleted" schwa. In *Proceedings Phonetic Experimental Research at the Institute of Linguistics University of Stockholm* (pp. 115-118). Stockholm, Sweden.

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

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., & Tuinman, A. (2012) The role of native-language knowledge in the perception of casual speech in a second language. *Frontiers in psychology*. doi: 10.3389/fpsyg.2012.00249.

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.

Moreton, E. (2002). Structural constraints in the perception of English stop-sonorant clusters. *Cognition*, 84, 55-71.

Nouveau, D. (2012). Limites perceptives de l'e caduc chez des apprenants néerlandophones. *Revue Canadienne de linguistiquea Appliquée*, 15, 60-78.

Norris, D., & McQueen, J. M. (2008). Shortlist B: A Bayesian model of continuous speech recognition. Psychological Review, 115(2), 357-395. doi:10.1037/0033-295X.115.2.357.

Pallier, C., Bosch, L., & Sebastián-Gallés, N. (1997). A limit on behavioral plasticity in speech perception. *Cognition*, *64*(3), B9-B17.

Pisoni, D. B., Lively, S. E., & Logan, J. S. (1994). Perceptual learning of nonnative speech contrasts: Implications for theories of speech perception. In Goodman, J. C. & Nusbaum, H.C. (Eds), *The development of speech perception: The transition from speech sounds to spoken words* (pp. 121-166). Cambridge, MA, US: The MIT Press.

Pitt, M. A. (2009). How are pronunciation variants of spoken words recognized? A test of generalization to newly learned words. *Journal of Memory and Language*, *61*, 19-36.

Pitt, M. A., Johnson, K., Hume, E., Kiesling, S., & Raymond, W. (2005). The Buckeye corpus of conversational speech: Labeling conventions and a test of transcriber reliability. *Speech Communication*, *45*, 89-95.

Pitt, M.A., Dilley, L., Johnson, K., Kiesling, S., Raymond, W., Hume, E., & Fosler-Lussier, E. (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, R. H. (2006). Effects of word frequency on the acoustic durations of affixes. *Proceedings of Interspeech 2006*, 953-956.

R Core Team (2014). *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing. *http://www.R-project.org/.*

Schuppler, B., Ernestus, M., Scharenborg, O., & Boves, L. (2011). Acoustic reduction in conversational Dutch: A quantitative analysis based on automatically generated segmental transcriptions. *Journal of Phonetics*, *39*, 96-109.

Sjerps, M. J., & McQueen, J. M. (2010). The bounds on flexibility in speech perception. *Journal of Experimental Psychology: Human Perception and Performance*, *36*(1), 195.

Sjölander, K. (2001). Automatic alignment of phonetic segments. *Lund University, Department of Linguistics Working Papers*, *49,* 140-143.

Spinelli, E., & Gros-Balthazard, F. (2007). Phonotactic constraints help to overcome effects of schwa deletion in French. *Cognition*, *104*(2), 397-406.

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., Mitterer, H., & Cutler, A. (2014). Use of syntax in perceptual compensation for phonological reduction. *Language and Speech*, 57, 68-85.

Ven, M. van de, Ernestus, M., & Schreuder, R. (2012). Predicting acoustically reduced words in spontaneous speech: The role of semantic/syntactic and acoustic cues in context. *Laboratory Phonology,* 3, 455-481.

Ven, M. van de, Tucker, B.V. & Ernestus, M. (2010). Semantic facilitation in bilingual everyday speech comprehension. In *Proceedings of the 11th Annual Conference of the International Speech Communication Association (*pp.1245-1248). Makuhari, Japan.

Ven, M. van de, Tucker, B.V. & Ernestus, E. (2011). Semantic context effects in the comprehension of reduced pronunciation variants. *Memory and Cognition*, 39, 1301-1316.

Viebahn, M., Ernestus, M. & McQueen, J. (2015). Syntactic predictability in the recognition of carefully and casually produced speech. *Journal of Experimental Psychology: Learning, Memory, and Cognition,*41 (6), 1684-1702.

Vorstermans, A., Martens, J. P., & Van Coile, B. (1996). Automatic segmentation and labelling of multi-lingual speech data. *Speech Communication*, *19*, 271-293.

Weber, A., Di Betta, A. M., & McQueen, J. M. (2014). Treack or trit: Adaptation to genuine and arbitrary foreign accents by monolingual and bilingual listeners. *Journal of phonetics*, *46*, 34-51.

Witteman, M. J., Weber, A., & McQueen, J. M. (2013). Foreign accent strength and listener familiarity with an accent codetermine speed of perceptual adaptation. *Attention, Perception, & Psychophysics*, 75(3), 537-556.

Wong, S. W., Mok, P. P., Chung, K. K. H., Leung, V. W., Bishop, D. V., & Chow, B. W. Y. (2015). Perception of native English reduced forms in Chinese learners: Its role in listening comprehension and its phonological correlates. *TESOL Quarterly*.

Young, S., Evermann, G., Gales, M., Hain, T., Kershaw, D., Liu, X. A., Moore, G., Odell, J., Ollason, D., Povey, D., Valtchev, V., & Woodland, P. (2006). *The HTK book* (for HTK version 3.4). Cambridge: Cambridge University Engineering Department.

Zimmerer, F., Scharinger, M., & Reetz, H. (2011). When BEAT becomes HOUSE: Factors of word final/t/-deletion in German. *Speech Communication*, *53*(6), 941-954.

Zimmerer, F., & Reetz, H. (2014). Do listeners recover “deleted” final/t/ in German?. *Frontiers in Psychology*, *5*.

Appendix. Automatic generation of the phonetic transcription of the NCSE

The speech in the Nijmegen Corpus of Spanish English (NCSE, Kouwenhoven et al., to appear) is orthographically transcribed. This transcription is aligned with the speech signal at the level of chunks of approximately two seconds, which reach from one natural pause in the speech signal to the next. As no phonetic transcription of the NCSE was available, we used the automatic speech recognition (ASR) system HTK (Hidden Markov Model Toolkit; Young et al., 2006) to generate broad phonetic transcriptions of the chunks containing *can* or *can't* following a forced alignment procedure similar to the one described by Schuppler et al. (2011).

Forced alignment uses one acoustic model for each phone in the language. Since the speech in the NCSE is heavily accented, phone models trained on native English speech were considered inaccurate for this corpus. We therefore trained our own phone models on the NCSE. The input for the training phase consisted of the wave files of all chunks of speech containing *can* or *can't* tokens, and a pronunciation lexicon with the standard pronunciations (see also Vorstermans et al., 1996) of all words in these chunks. We took the standard pronunciations from Celex (Baayen et al., 1995) or created them manually for those words not in Celex. We excluded the chunks with Spanish words from the training materials. This procedure resulted in a training set of 919 chunks of speech, with a total duration of approximately 38 minutes.

We trained 49 32-Gaussian tri-state monophone models, including four models for non-speech sounds (laughter, breath-taking, clicks produced by the speakers' mouths, and sounds resulting from microphone touches). We are aware that models cannot reliably be trained for non-speech sounds, but we are confident that by including these models, the ASR can more accurately place the boundaries of the speech sounds. The models were trained at a frame rate of 10 ms and a window length of 25 ms. For each frame, 13 MFCCs (the mel-scaled cepstral coefficients C0-C12) and their first and second order derivatives (39 features in total) were calculated.

For the forced alignment procedure, we created a pronunciation dictionary that included two pronunciation variants of *can* with two different phones for the vowel, and four pronunciation variants of *can't* with the same two vowel options and with or without /t/. The ASR determined for each token of *can* and *can't* which pronunciation variant was present in the speech signal.

We validated the phonetic transcriptions by comparing the transcriptions of a subset of 79 *can* and 51 *can't* tokens with two human-made transcriptions. We compared the mean differences between the positions of the phone boundaries (in ms) and the percentages of differences smaller than 20 ms, a widely used accuracy measure (see e.g., Vorstermans et al., 1996; Sjölander, 2001; Pluymaekers et al., 2006). The agreement between the two human transcribers was high (see Table A1). In contrast, a first comparison of the automatically generated transcriptions and the two human-made transcriptions showed that the ASR consistently placed the boundaries for the start and the end of /n/ too early. We resolved this issue by shifting all /n/ boundaries 25 ms to the right (see also Pluymaekers et al., 2006). After this adjustment, the agreement between both human transcribers and the ASR was high (see Table A1).

Table A.1: ***Comparison of the automatic (A) and human-made (H1 and H2) phonetic transcriptions. The number of tokens (N) in each comparison is given in the left column*** [a]**.**

	Mean difference between boundaries			Percentage of boundaries within 20 ms		
Boundary	A – H1	A – H2	H1 – H2	A – H1	A – H2	H1 – H2
Start /k/ (N = 130)	8.71 ms	8.81 ms	4.42 ms	90.77 %	88.46 %	97.70 %
Start /a/ (N = 130)	7.00 ms	9.97 ms	6.08 ms	95.38 %	85.38 %	95.38 %
Start /n/ (N = 130)	11.36 ms	13.29 ms	4.63 ms	85.15 %	82.30 %	96.15 %
End /n/ (N = 130)	18.18 ms	19.97 ms	10.23 ms	72.31 %	62.31 %	84.62 %
Start /t/[a] (N = 16-20)	11.29 ms	8.99 ms	6.99 ms	88.24 %	87.50 %	90.00 %
End /t/[a] (N = 16-20)	22.65 ms	14.21 ms	13.69 ms	70.59 %	93.75 %	75.00 %

[a] In each pairwise comparison, the two transcribers could disagree about whether a /t/ was present. If the two transcribers disagreed (i.e., one transcribed a /t/, but the other did not), no comparison for /t/ boundaries could be made, which explains the variation in the Number of tokens for start and end of /t/.

Since the presence versus absence of /t/ in *can't* is the main focus of the present study, we also compared the three transcriptions of the 51 tokens of *can't* in this respect. The agreement on the presence or absence of /t/ was high: in 47 cases (92.2%), the ASR agreed with at least one human transcriber, and only in four cases (7.8%) did the two human transcribers both differ from the ASR. As the ASR provides consistent phonetic transcriptions relatively quickly, we accepted the validity of the automatically generated phonetic transcriptions for the present study.