




Listeners' processing of a given reduced word pronunciation variant directly reflects their exposure to this variant: evidence from native listeners and learners of French

Sophie Brand & Mirjam Ernestus

To cite this article: Sophie Brand & Mirjam Ernestus (2017): Listeners' processing of a given reduced word pronunciation variant directly reflects their exposure to this variant: evidence from native listeners and learners of French, The Quarterly Journal of Experimental Psychology, DOI: [10.1080/17470218.2017.1313282](https://doi.org/10.1080/17470218.2017.1313282)

To link to this article: <http://dx.doi.org/10.1080/17470218.2017.1313282>




View supplementary material 



Accepted author version posted online: 04 Apr 2017.



Submit your article to this journal 



Article views: 1



View related articles 



View Crossmark data 

Publisher: Taylor & Francis & The Experimental Psychology Society

Journal: *The Quarterly Journal of Experimental Psychology*

DOI: 10.1080/17470218.2017.1313282

**Listeners' processing of a given reduced word pronunciation variant directly reflects their
exposure to this variant: evidence from native listeners and learners of French**

Sophie Brand

Radboud University, Nijmegen, the Netherlands

&

Mirjam Ernestus

Radboud University, Nijmegen, the Netherlands

and Max Planck Institute for Psycholinguistics, Nijmegen, the Netherlands

Correspondence concerning this article should be addressed to Sophie Brand,

Radboud University, Nijmegen. Email: s.brand@let.ru.nl

Abstract

In casual conversations, words often lack segments. This study investigates whether listeners rely on their experience with reduced word pronunciation variants during the processing of single segment reduction. We tested three groups of listeners in a lexical decision experiment with French words produced either with or without word-medial schwa (e.g., /rəvy/ and /rvy/ for *revue*). Participants also rated the relative frequencies of the two pronunciation variants of the words. If the recognition accuracy and reaction times for a given listener group correlate best with the frequencies of occurrence holding for that given listener group, recognition is influenced by listeners' exposure to these variants. Native listeners' relative frequency ratings correlated well with their accuracy scores and RTs. Dutch advanced learners' accuracy scores and RTs were best predicted by their own ratings. In contrast, the accuracy and RTs from Dutch beginner learners of French could not be predicted by any relative frequency rating; the rating task was probably too difficult for them. The participant groups showed behaviour reflecting their difference in experience with the pronunciation variants. Our results strongly suggest that listeners store the frequencies of occurrence of pronunciation variants, and consequently the variants themselves.

Keywords: word recognition, pronunciation variation, lexical representations, language learners, frequency estimations

Introduction

In casual speech, native speakers do not pronounce all sounds. In American English, the word *yesterday* /jɛstədəɪ/, for instance, may be pronounced as something like /jɛfɛɪ/. Johnson (2004) studied 88,000 word tokens produced by 40 native speakers of American English in interviews (part of the Buckeye corpus of Conversational Speech, Pitt et al., 2007) and found that 25% of the content words lacked at least a single sound. This phenomenon of reduction, whereby words are produced with fewer segments compared to their citation forms, is also highly frequent in Germanic languages, such as Dutch (e.g., Ernestus, 2000) and German (e.g., Kohler, 1990), and in non-Germanic languages such as French (e.g., Adda-Decker, Boula de Mareüil, Adda & Lamel, 2005) and Finnish (Lennes, Alaroty & Vainio, 2001). Despite the ubiquity of reduction in everyday speech, little is known about how listeners comprehend reduced words. This study contributes to filling this gap by investigating whether native listeners and learners of a language rely on lexical representations of reduced word pronunciation variants during the processing of single segment reduction.

To date, there has been little agreement on the processes that underlie the comprehension of reduced word pronunciation variants, and on which word forms are stored in the mental lexicon. We can distinguish two main accounts. The first account assumes that only unreduced word pronunciation variants are stored in the mental lexicon. Reduced words (e.g., French /rvy/ *revue* 'magazine') are matched with the representations for the unreduced variants (e.g., /rɔvy/), by means of general processes (e.g., schwa-insertion), each applying to several words. This account is compatible with most

psycholinguistic models of word recognition, such as TRACE (McClelland & Elman, 1986), the Neighborhood Activation Model (NAM: Luce & Pisoni, 1998), PARSYN (Luce, Goldinger, Auer & Vitevitch, 2000), and Shortlist B (Norris & McQueen, 2008). The second account assumes that both unreduced and reduced word pronunciation variants are stored in the mental lexicon. A French speaker would recognize reduced /rvy/ by matching the acoustic signal with the lexical representation /rvy/. This account requires that the word recognition models mentioned above extend their lexicons with additional word pronunciation variants.

The debate on the roles of storage versus computation is not restricted to the domain of word pronunciation variants. There have, for instance, also been extensive debates about whether regular morphologically complex words are stored in the mental lexicon or whether they are computed from their morphemes every time they are processed (e.g., Baayen, McQueen, Dijkstra & Schreuder, 2003; Alegre & Gordon, 1999; Clashen, 1999; Sereno & Jongman, 1997; Stemberger & MacWhinney, 1988). Many researchers argue that a class of regular complex words (e.g., plurals) is stored if the members with high frequencies of occurrence are processed more quickly than members of lower frequencies (given similar lemma frequencies).

The debate on storage versus computation of word pronunciation variants also focusses on the role of the variant's frequency of occurrence in processing (e.g., Ranbom & Connine 2007; Pitt, Dilley & Tat, 2011). Ranbom and Connine studied how quickly American English listeners recognize words that can be pronounced with both /nt/ and a nasal flap (e.g., *gentle*) and found that native listeners recognize the flap variant more quickly if it is the word's most frequent variant (as is the case for *gentle*) than if it is not the most frequent variant (as for *lantern*). Pitt and colleagues reported similar results for the recognition of American English words in which word-medial /t/ occurs in one of four phonological contexts, each favouring one of four realisations of /t/ (/t/, /ɾ/, /ɹ/ or a deleted variant). Pitt and

colleagues documented a recognition benefit for the variant that is the most frequent in the given context.

These frequency effects reported by Ranbom and Connine (2007) and by Pitt and colleagues (2011) can easily be explained in models that allow for storage of more than one pronunciation variant per word (i.e., the second account discussed above): the variants are stored together with their a-priori probabilities, which determine how easily they are processed (e.g., since they co-determine the variants' resting activation levels).

Frequency effects can also easily be accommodated in models that assume that word pronunciation variants are never stored but always computed (i.e., the first account, e.g., Roelofs, 1997), but only in speech production. During production, the frequency of occurrence of a reduced variant may modulate the ease of application of the rule that transforms the stored, unreduced variant into the corresponding reduced variant. It can be argued that a similar process may account for frequency effects in speech comprehension. Prelexical reconstruction rules may be sensitive to the frequencies with which they are applied to each word. That is, when hearing a word, a listener may apply a reconstruction rule and consider the resulting prelexical representation as likely for the given acoustic input as indicated in the description of the rule. For instance, when a native listener of French hears /rvy/ *revue* 'magazine', (s)he may apply schwa-insertion (resulting in /rəvy/) and keep the two prelexical representations active until the word is identified. The recognition process would favor the prelexical representation that is most likely given the likelihood that the word is produced with schwa, as indicated by the rule. However, the word-specific probability of an insertion rule can only become available after the word has been identified. If information about the probability of an insertion rule only becomes available after the word has been identified, how then can this probability affect the ease of word recognition?

The assumption that variants are stored together with their frequencies of occurrence predicts that listeners with different experience with the variants react differently on these variants. For instance, if one listener tends to hear one word in its reduced variant and a second word in its unreduced variant while another listener tends to hear the former word in its unreduced variant and the latter word in its reduced variant, we expect that the former listener recognizes the former word best in its reduced variant and the latter word in its unreduced variant while the opposite pattern holds for the latter listener. The aim of this study is to investigate whether this expectation is correct, that is, whether the recognition accuracy and reaction times for a given listener group correlate best with the frequencies of occurrence of the variants holding for that given listener group. This would indicate that the pronunciation variants are lexically stored.

We focus on schwa in the initial syllable of French words, which can be absent even if the words are produced in isolation and in a formal situation (e.g., /fənɛtʁ/ 'window', which can be pronounced as /fnɛtʁ/; Bürki, Ernestus & Frauenfelder, 2010). The absence of schwa is often complete, leaving no traces in the acoustic signal (e.g., Bürki, Fougeron & Gendrot, 2007). This study focusses on French as spoken in the north of France, where schwa is more often absent than in the south of France (e.g., Durand & Eychenne, 2004; Coquillon, 2007). In the north of France, schwa is absent in more than 50% of the word tokens in conversational speech (e.g., Fougeron, Goldman & Frauenfelder, 2001; Hansen, 1994). Sociolinguistic factors influence the likelihood of the absence of schwa in the north of France (e.g., Hansen, 2000) as they do in the south of France (e.g., Eychenne & Pustka, 2007). This study focusses on Northern French as produced by a young adult, highly educated speaker.

Our study, which concentrates on vowels, thus differs from those of Ranbom and Connine (2007) and Pitt and colleagues (2011), who examined variations in the realisation of consonants.

Several studies have shown that listeners process vowels differently from consonants. For instance,

Pisoni (1973) and Stevens, Liberman, Studdert-Kennedy, and Öhman (1969) showed that listeners perceive finer distinctions in vowels than in consonants. Cutler, Sebastián-Gallés, Soler Vilageliu, and Van Ooijen (2000) reported that participants are more likely to change vowels than consonants if they are asked to turn pseudo words into real words. Given this processing difference between consonants and vowels, findings on consonants may not be generalized to vowels. The question thus arises whether listeners are also sensitive to frequencies of word pronunciation variants that differ from their respective full variants only in their vowels.

Two studies (Bürki, Ernestus, Gendrot, Fougeron & Frauenfelder, 2011; Bürki, Ernestus & Frauenfelder, 2010) investigated whether reduced pronunciation variants of French words with schwa are stored in the mental lexicon, both focussing on speech production. The corpus study by Bürki and colleagues (2011) showed that the duration of schwa in a word like *revue* 'magazine' is physiologically conditioned (e.g., by the voicing specifications of the consonant following schwa), whereas the presence of schwa is particularly affected by prosodic factors. The limited overlap between the predictors for variant choice (i.e., with or without schwa) and for schwa duration, combined with the nature of these variables, suggests that selection of the variant to be pronounced occurs before phonetic implementation, which determines schwa duration. As such, Bürki and colleagues assume that both pronunciation variants of a word, like /RƏvy/ and /Rvy/ for *revue*, are lexically stored. Bürki and colleagues (2010) showed that native speakers of French have shorter production latencies for a given variant of a schwa-word if they think this variant occurs more often. The authors argue that speakers store the relative frequencies of the different pronunciation variants of a word in their mental lexicons and therefore these pronunciation variants are themselves also stored. As explained above, however, a mental lexicon that contains only unreduced variants can also account for frequency effects in production: frequency of occurrence of a variant may modulate the ease of application of a rule that

transforms the stored variant into the variant to be produced. A study on frequency effects in speech production is therefore not highly informative for answering the question whether word pronunciation variants are stored.

We found one study that investigated the *comprehension* of word pronunciation variants without schwa in French (Racine, Bürki & Spinelli, 2014). The authors studied how the recognition of French reduced schwa words by French native children is affected by spelling. The results show that both readers and pre-readers recognize the variant with schwa (e.g., /ʁənɑʁ/ *renard* 'fox') more quickly than the variant without schwa (e.g., /ʁnɑʁ/) for words that are most frequently produced with schwa. In contrast, readers and pre-readers react differently to words that are spelled with schwa but never pronounced with this vowel (e.g., *bracelet* /braslɛ/ 'bracelet'): whereas pre-readers recognize the schwa variant more slowly than the non-schwa variant, readers recognize the two variants equally quickly. Racine and colleagues conclude that spelling influences readers' word recognition and that pre-readers' word recognition is only influenced by frequency of occurrence.

We investigated in a more direct way whether listeners are sensitive to variants' frequencies. We first asked participants to perform an auditory lexical decision task in which French schwa words with optional schwa (e.g., *revue*, which can be pronounced as /ʁəvy/ or /ʁvy/) were presented with or without schwa. We then asked them to perform a relative frequency estimation task where they, for each visually presented word, were instructed to indicate the relative frequencies of the two variants based on what they hear in everyday life (following e.g., Bürki et al., 2010; Racine, 2008; Ranbom & Connine, 2007).

We tested native speakers of French (Experiment 1) as well as learners of French (Experiments 2, 3, and 3a in Supplemental Material C), who we expect to have encountered reduced pronunciation variants, of at least some words, less often than natives. It has been shown that there is a mismatch

between what tends to be presented to language learners and the actual use of that language outside the classroom (see e.g., Jones & Ono, 2000; McCarthy & Carter 1995; see for French Askildson, 2008; Fonseca-Greber & Waugh, 2003; Waugh & Fonseca-Greber, 2002; O' Connor di Vito, 1991). As a consequence, the relative frequency of the unreduced and the reduced variant of a given word is likely to be different for language learners than for native listeners. We hypothesize that learners' accuracies and reaction times (henceforth RTs) are better predicted by their own relative frequency ratings of the unreduced and reduced variants of a given word than by the natives' ratings and that is especially true for learners who have had some exposure to both variants.

Advanced learners are likely to have encountered the reduced variants of a (large) number of words. If they have encountered a given reduced variant, we hypothesize that they have stored this variant together with its frequency. Importantly, this frequency is expected to be different from the frequency stored by natives, as advanced learners have had less and different exposure to casual speech. As a consequence, the relative frequencies of the unreduced and reduced variants that hold for native speakers may not correlate well with advanced learners' accuracies and RTs. These dependent variables are predicted to correlate better with the frequency ratings provided by advanced learners themselves.

Beginner learners are only likely to have encountered the reduced variants of a small number of words. Moreover, the frequencies of occurrence of the variants in their input differ from those stored by natives. Thus, we hypothesize that beginner learners' response accuracies and reaction times also do not correlate with natives' relative frequency ratings. In contrast, because beginner learners and advanced learners received somewhat similar input, their performance may show a (weak) correlation with advanced learners' relative frequency ratings.

It is difficult to predict whether beginner learners' performance will correlate with their own relative frequency ratings. We might expect this correlation because these ratings should reflect

beginner learners' input. However, beginner learners' ratings may be unreliable for many words because these learners have very little experience with these words' variants. They may therefore be just guessing. Their relative frequency ratings then do not correlate with their response accuracies and RTs.

We rely on subjective frequency ratings instead of objective frequency measures. Some researchers assume that subjective frequency ratings better predict lexical processing than objective frequency ratings do (e.g., Balota, Pilotti & Cortese, 2001). Others claim the opposite, at least for word frequency. For instance, Ghyselinck, Lewis, and Brysbaert (2004) claim that subjective ratings of word frequency are likely to be influenced by the age the word was acquired. They state that objective word frequency measures are to be preferred above subjective frequency measures if the word frequency measures can be based on reliable large data bases. This statement is supported by Ernestus and Cutler (2015), who showed that a subjective word frequency measure predicts RTs in a Dutch lexical decision task less accurately than several measures of objective word frequency.

We nevertheless relied on subjective relative frequency measures for several reasons. First of all, no objective measures for word pronunciation variants without schwa are available. This holds for French native speakers and above all for language learners. Second, Racine (2008) obtained frequency estimations from Swiss francophone students that correlated well with the variants' frequencies in the productions of 16 different Swiss francophone students who were asked to summarize stories ($r = .44$, $p < .01$). This result demonstrates that native speakers of French are able to provide reliable estimations of the frequencies of occurrence of the schwa and non-schwa variants of French words.

While the effect of relative frequency can thus only be tested with subjective ratings as a predictor for language behaviour, the effect of word frequency can be tested with predictors representing objective counts taken from several databases (e.g., Lexique 3.80, New, Pallier, Ferrand &

Matos, 2001). This objective word frequency information may, however, only be representative for the language user group from whose speech these counts were extracted (typically native adult speakers). For other groups, subjective word frequency information may outperform objective word frequency measures in predicting language processing. This study also tests this hypothesis.

We tested learners of French who are native speakers of Dutch. The Dutch phonological system is highly comparable to that of French: the two phoneme inventories are about the same, and Dutch also shows schwa reduction. We would therefore expect that Dutch learners have little difficulties identifying the phonemes in the French stimuli of the experiment and that it will therefore be easy to test the effect of reduction and the role of relative frequency.

We tested native listeners in Experiment 1 to see whether their accuracies and RTs were predicted by their relative frequency ratings. Experiment 2 tested Dutch advanced learners of French (C1-C2 levels according to the Common European Framework of Reference for Languages, CEFR, Council of Europe, 2011), while Experiment 3 tested Dutch beginner learners (B1-B2 levels according to CEFR). We investigated whether the accuracies and RTs in the lexical decision task of the learner groups were better predicted by their own relative frequency ratings of the variants or by those provided by one of the other listener groups. Moreover, because we expect a discrepancy between the words' frequencies as listed in traditional databases, such as Lexique 3.80 (New, Pallier, Ferrand & Matos, 2001), and those experienced by beginner learners, we also tested in Experiment 3, whether the objective word frequency holding for native listeners or the subjective frequency measure holding for the specific listener group better predicts their accuracies and RTs.

Experiment 1

Method

Participants

Thirty-six native speakers of French from Paris (three males), aged between 19 and 30 years, were paid to participate in the experiment. They were born and raised in the north of France, and their parents were also native French speakers. Like all participants tested for this study, they did not report any hearing problems and did not guess the purpose of the experiment.

Materials

We selected 44 French, morphologically simple, bisyllabic target words with a schwa in the first syllable (see Supplemental Material A for the complete list). The target words were selected from vocabulary lists in teaching methods used at Dutch secondary schools. In standard French, these words can be produced with and without schwa in the initial syllable. The absence of schwa results in an illegal consonant cluster in 38 of the 44 words (e.g., /ʁn/ in /ʁnaʁ/ 'fox'). All words tested in this study were presented with their definite determiners.

We added several types of filler words (520 in total, see Table B.1 in Supplemental Material B). First, we created 44 bisyllabic pseudo words with schwa in the first syllable. These pseudo words consisted of real French syllables and did not closely resemble real Dutch words. We determined their definite determiners, *le* or *la*, by adhering to the broad trend that French nouns ending in *-esse*, *-ie*, *-ite*, *-ine*, *-té*, *-ure* are feminine, while nouns ending in *-age*, *-aire*, *-al*, *-at*, *-eau*, *-euil*, *-in*, *-is* are masculine.

Second, we added 149 bisyllabic real words without schwa, in order not to draw the participants' attention to the target words. Lexique 3.80 (New, Pallier, Ferrand & Matos, 2001), a list of 135,000 French words, shows that words with schwa in their initial syllable form only 2.8% of the bisyllabic French words. We further added 149 corresponding bisyllabic pseudo words without schwa. These pseudo words were constructed in the same way as the 44 pseudo words with schwa.

Finally, we selected 89 monosyllabic real words without schwa and we created 89 monosyllabic pseudo words. Due to these filler words, only a few (3.9%) of the monosyllabic real words presented to a given participant resulted from the absence of schwa.

A female French native speaker, aged 21 years, recorded all words in a sound-attenuated booth. She produced the words containing a schwa once with this schwa (unreduced) and once without (reduced). We used Adobe Audition 1.5 and a Sennheiser ME 64 microphone. The stimuli were digitised at a sampling rate of 44.1 kHz, a 16-bit quantization, and were scaled to an average intensity of 70 dB. The unreduced variants of the target words and their determiners (e.g., /larəvy/ *la revue* ‘the fox’) had a total mean duration of 771.70 ms and the word-medial schwa had a mean duration of 107.07 ms (range: 61 ms - 158 ms). The reduced variants of the target words and their determiners (e.g., /larvy/ *la revue* ‘the fox’) had a total mean duration of 725.11 ms. In these variants, word-medial schwa was absent, except in the word *seconde*, which contained a short schwa (12.86 ms). Due to absence of schwa, the onset of the noun formed the syllable coda of the determiner (e.g., reduced *la revue* /la. RƏ.vy/ was syllabified as /lar.vy/). The schwa in the determiner *le* had a mean duration of 118.34 ms if followed by an unreduced noun versus 115.74 ms if followed by a reduced noun.

For the auditory lexical decision experiment, we created four different randomised lists, each containing all 564 words. Each list contained half of the 44 target words and half of the 44 bisyllabic pseudo words in their unreduced variants (i.e., with schwa), and the other halves in their reduced variants. Subsequently, we created the mirror images of these four lists by replacing the reduced variants by the corresponding unreduced variants and vice versa. This resulted in eight lists. Each participant heard one of these lists. The 564 tokens in a list were divided in three equal blocks, which were separated by short breaks in the experiment.

Each list was preceded by eight trials which familiarised the participants with their task. These

trials presented three real schwa words, of which two were reduced and one unreduced, and three pseudo words, one with and two without schwa, and two real words without schwa. None of these words occurred in the main experiment. The familiarisation trials were presented in one of two different orders.

The rating experiment only contained the target words, which were presented in alphabetical order. This task also started with a familiarization trial, presenting a schwa word not occurring in the experiment.

Procedure

Experiment 1 took place at the *Laboratoire Charles Bruneau* of the *Institut de Linguistique et Phonétique Générales et Appliquées* (ILPGA) in Paris. Participants were tested individually in a sound-attenuated room. A session consisted of both the auditory lexical decision task and the relative frequency estimation task, always in that order. The two tasks were presented in E-prime 2.0 (Schneider, Eschman & Zuccolotto, 2007) from a laptop. The entire session lasted approximately 45 minutes. In the auditory lexical decision task, the participants heard the stimuli via Sennheiser HD 215 headphones. They were instructed to indicate as quickly as possible for each stimulus whether it was a real French word or not. Right handed participants responded by pressing the key “m” for “yes”, and the key “z” for “no”, on the keyboard. For left-handed participants, the “yes” and “no” keys were reversed. Each trial started with an asterisk shown for 250 ms in the center of the screen. Then, the stimulus was presented auditorily. Participants had to respond within 3000 ms from word onset. After the participant had responded, or after 3000 ms if no answer was given, the screen remained blank for another 300 ms and then the asterisk appeared announcing the next trial. After each block of approximately 10 minutes, the participants took a short break.

For the unspeeded relative frequency estimation task, participants were instructed to rate, for

each word, the frequency of the reduced variant compared to the unreduced variant, based on what they commonly hear in daily life. They were asked to choose a value on the Likert scale presented on the computer screen by pressing a number between 1 and 6 on the keyboard. A score of 1 meant that the word was always produced with schwa, whereas a score of 6 meant that it was always produced without schwa. The two pronunciation variants were visually presented at the extreme left end (unreduced variant) and extreme right end (reduced variant) of the scale. The words occurred out of context, but with their definite determiners, as in the lexical decision task.

Results

Relative frequency estimation task

We first examined the validity of the relative frequency estimations (a total of 1568 due to 16 missing ratings; 99.0% of the trials). For each participant, we calculated a frequency value for each variant of every word by attributing the value entered by the participant to the reduced variant of the word, and subtracting this value from seven to calculate the value for the unreduced variant of that word. For example, if a participant selected 2 on the keyboard for the word *le chemin* 'the way' (meaning that the word was almost always produced with schwa), the reduced variant was assigned the value 2 and the unreduced variant the value 5. The interrater agreement, calculated on the basis of the ratings for the reduced variants, appeared to be slight ($k_{\text{fleiss kappa}}: .09, p < .001$; Fleiss, Levin & Paik, 2003). The average by word relative frequency ratings ranged between 1.17 and 5.83. We compared our ratings to those provided by participants from Nantes obtained by Racine (2008) in a similar rating experiment. As illustrated in Figure 1, the correlation between the by word average relative frequency ratings for reduced variants given by our French participants and by Racine's participants' is high ($r = .83, p < .0001$). This shows that two groups of speakers of more or less the same variant of French produce ratings that are very similar, which strongly suggests that speakers are able to perform the task.

"(Figure 1 about here)"

Lexical decision accuracy

We examined the accuracy of the responses to the target words in the lexical decision task and investigated whether relative frequency rating was a significant predictor. Here and for the other analyses in this paper, we only report models including relative frequency ratings averaged over participants and not models including the individual ratings as predictor because nearly all these latter models had higher AIC values. We excluded trials where the computer had not registered responses as well as trials with words for which the participant had not entered frequency ratings. The number of responses left for analysis totalled 1558 (98.4% of the data, of which 91.9% correct and 8.1% incorrect).

In Figure 2, the natives' mean accuracy score per word variant is plotted against their average relative frequency rating for that variant. The plot shows that the higher the relative frequency rating for the reduced word variant, the higher the French natives' accuracy score for this variant.

“(Figure 2 about here)”

This effect of relative frequency on accuracy is supported by statistical analyses. We analysed the accuracy of the answers by means of logistic mixed-effects models (Faraway, 2006) in R (R Development Core Team, 2015), including the package lme4 (Bates, Maechler, Bolker & Walker, 2015). We included word and participant as crossed random effects, and tested variant relative frequency averaged over participants and variant type (reduced versus unreduced) as the main fixed predictors. We also added several control predictors in order to reduce the variance in the data: trial

number and word frequency as listed in Lexique 3.80 for subtitles of films (New et al., 2001). Random slopes were tested for all fixed predictors. We only retained those predictors in the model that were statistically significant or figured in statistically significant interactions. The final model is summarized in Table B.2 (see Supplemental Material B).

French native listeners made fewer errors for variants with higher average relative frequency ratings. Furthermore, they made more errors when responding to reduced than to unreduced variants (see the upper panel of Figure 8). As shown by the random slope, the effect of variant type on accuracy differed per word. While the plot suggests that the effect of average relative frequency rating is restricted to the processing of reduced variants, the statistical analysis does not show an interaction between variant type and average relative frequency rating, possibly because of lack of statistical power.

Lexical decision RTs

We analysed the RTs measured from word onset for the trials in the dataset used for the accuracy analysis, discarding trials where the target words had incorrectly been classified as pseudo words (this left 1456 trials, 93.5%). RTs deviating from the grand mean (1214.22 ms) by more than 2.5 times the standard deviation (282.79 ms) were considered to be outliers and removed. The number of observations left for analysis totalled 1390 (87.8% of the data). In Figure 3, the natives' average RT per word variant is plotted against the relative frequency rating averaged over participants for that variant. We see that the higher the relative frequency rating for the (reduced) word variant, the faster the French natives reacted.

“(Figure 3 around here)”

We performed statistical analyses to see whether the effect of the average relative frequency rating is statistically significant. We analysed the log-transformed RTs by means of mixed-effects regression models, with word and participant as crossed random effects and with the same fixed predictors as those used to analyse accuracy. We added two more control predictors: the log-transformed duration of the target word and the log-transformed RT to the previous stimulus. Since the unreduced pronunciation variants were longer (mean duration of the determiner - word combination: 771.70 ms) than the reduced pronunciation variants (mean: 725.11 ms), we orthogonalised target duration and variant type by replacing target duration with the residuals of a linear model predicting target duration by variant type. Random slopes were again tested for all fixed predictors. We only retained those predictors in the model that were statistically significant or figured in statistically significant interactions. Finally, we discarded all RTs that deviated more than 2.5 times the standard error from the values predicted by the best statistical model and refitted the model. The final model is summarized in Table B.3 in Supplemental Material B.

The French native listeners responded significantly faster to shorter words, to stimuli presented later in the experiment, and if they had also responded quickly to the previous trial. More interestingly, we found a significant effect of variant type on RT: the French native listeners took more time to respond to a reduced variant than to an unreduced variant (see the lower panel of Figure 8). This effect of variant type was larger for some words than for others, as shown by the significant random slope of variant type for word.

Most importantly with regard to our research question, we found a main effect of relative frequency rating. The French native listeners responded faster to variants that according to this group's average relative frequency ratings occur more often.

Summary and discussion

The average relative frequency ratings provided by the French natives predict both their accuracy and RTs in a lexical decision task. These results support earlier findings suggesting that listeners show sensitivity to the frequency of occurrence of a given pronunciation variant (e.g., Pitt et al., 2011; Ranbom & Connine, 2007). These frequencies must then be stored, which strongly suggests that the pronunciation variants are lexically stored. Furthermore, Experiment 1 provides additional evidence for a privileged status for unreduced variants (e.g., Pitt et al., 2011; Ranbom & Connine, 2007; Tucker & Warner, 2007; Ernestus & Baayen, 2007; Janse, Nootboom & Quené, 2007): the native listeners made fewer mistakes and responded faster to unreduced variants than to reduced variants.

In order to discover whether participant groups with different experience with the reduced and unreduced variants of a word show behaviour reflecting this difference in experience, we replicated Experiment 1 with Dutch advanced learners of French. Dutch advanced learners of French have less experience with the reduced variants of most words presented in our experiment. As a consequence, we expect that their recognition accuracy and reaction times better correlate with their own variant frequency ratings than with those provided by the French natives.

The learners performed two additional tests to reveal their proficiencies. The first is the visual lexical decision task LexTALE (Brysbaert, 2013; Lemhöfer & Broersma, 2012), which provides a measure of general lexical proficiency. The second was a dictation task, which provides a good indication of learners' comprehension of connected speech (Kennedy & Blanchet, 2013).

Experiment 2

Method

Participants

Forty-seven Dutch undergraduate students of French (aged 19-30 years; eleven males) participated in

the experiment. All were born and raised in the Netherlands, had taken French classes for five or six years at secondary school, and had studied French at university for at least seven months and at most three years and seven months. Their CEFR-levels roughly corresponded to C1-C2 level (CEFR, Council of Europe, 2011).

Materials

The materials for the auditory lexical decision task and the relative frequency estimation task were the same as those used in Experiment 1.

The LexTALE task, the first proficiency test that we administered, includes 56 real French words and 28 phonotactically legal pseudo words.

The dictation task consisted of 10 sentences (138 words in total), extracted from an informal conversation between two men in The Nijmegen Corpus of Casual French (Torreira, Adda-Decker & Ernestus, 2010). The sentences were produced at an average speech rate of 5.57 syllables per second (excluding pauses) and contained high-frequency words, in which many schwas were absent.

Procedure

Dutch advanced learners of French were tested in sound-attenuated booths at the Max Planck Institute for Psycholinguistics in Nijmegen and at Leiden University. The procedures of the lexical decision experiment and the rating experiment were the same as in Experiment 1. These two experiments were followed by the two proficiency tests.

The visual lexical decision task LexTALE was run in E-prime, as were the two main experiments. Each trial started with a blank screen shown for 250 ms and once the stimulus was visually presented (in an 18-point Courier New font size), participants had to respond within 3000 ms. As soon as the participant had responded (or after 3000 ms), a blank screen appeared for 100 ms, after which the next trial started.

The dictation task was presented via WebExp2, a web-based program (Keller, Gunasekharan, Mayo & Corley, 2009). Participants listened to the speech to be transcribed orthographically, sentence by sentence. They could replay a sentence as often as necessary. However, after a certain amount of time (range: 10-120 seconds), determined on the assumption that one minute of speech takes approximately ten minutes to transcribe, participants were forced to start on the next sentence. The entire experimental session lasted approximately 60 minutes.

Results

Relative frequency estimation task

We obtained 2068 relative frequency ratings, which showed a slight interrater agreement (on the basis of the ratings for the reduced variants $k\text{ fleiss kappa: } .01, p < .001$). The average by word relative frequency ratings ranged between 1.89 and 4.06 and showed a correlation of .40 ($p < .01$) with the French participants' average ratings obtained by Racine (2008), and of .48 ($p < .001$) with those obtained in Experiment 1 (Figure 4).

"(Figure 4 about here)"

Interestingly, the advanced learners' relative frequency ratings averaged by word correlated with the word frequencies from Lexique: the more often a word occurs according to the Lexique database, the higher the average relative frequency rating for its reduced word pronunciation variant (and thus the lower the relative frequency rating for its unreduced variant). Possibly, the advanced learners applied the strategy to partly base their relative frequency ratings on the words' frequencies.

Lexical decision accuracy

We obtained responses for 2039 trials (98.6% of the data). In the left panel of Figure 5, the Dutch advanced learners' mean accuracy score per word variant is plotted against their mean relative frequency rating for that variant. Like for the French natives, for the Dutch advanced learners, the accuracy scores for the (reduced) word variants increase with the group's average relative frequency ratings.

“(Figure 5 about here)”

We also plotted the Dutch advanced learners' accuracy score per word variant against the French natives' average rating for that variant (right panel of Figure 5). We see that the accuracy score for the word variant also increases with this average relative frequency rating for the variant, although the correlation seems smaller in the right panel than in the left panel of Figure 5.

We tested whether the correlation between accuracy score and average relative frequency rating visible in the two figures is statistically significant. We first analysed the effect of the average relative frequency ratings provided by the advanced learners themselves. We analysed response accuracy (76.7% of the responses were correct, 23.3% incorrect) by means of logistic mixed-effects models using the same fitting procedure and using the same random and other fixed predictors (i.e., variant type, trial number, word frequency) as in the analysis of the accuracy data of Experiment 1. In addition, we included the percentage of correct words in the orthographic dictation task as a predictor indicating proficiency. We did not include a predictor for proficiency based on the LexTALE task because due to a technical error the data from this task were not available for 21 participants. Because we are interested in variation in the advanced learners' ratings that is not based on their knowledge of word frequency but on exposure to reduced pronunciation variants, we removed the variation in the

frequency ratings that can be explained by the variation in word frequency. That is, we replaced the average ratings provided by the advanced learners with the residuals of a linear model predicting average ratings by variant type, the word frequencies from Lexique, and their interaction. Table B.4 (see Supplemental Material B) shows the final model.

The advanced learners made fewer errors to words that occur more often and if they performed better on the dictation task. Like the participants in Experiment 1, they made more errors when responding to reduced variants relative to unreduced variants (see the upper panel in Figure 8). As shown by the random slope, this effect was not equally large for all participants.

Importantly, the accuracy scores were also predicted by the average relative frequency ratings from the advanced learners: the higher the average rating for a given variant, the fewer errors they made. In contrast, the accuracy scores could not be predicted by the average ratings provided by the French native listeners, in contrast to what the right panel of Figure 5 suggests.

We directly compared whether the effect of relative frequency rating differed between the advanced learners and the native listeners. We pooled the data from the two participant groups and investigated the effect of the natives' average relative frequency ratings and of the advanced learners' average relative frequency ratings. The model with the lowest AIC (2257.8 versus 2260.8; Akaike, 1973) contained the French participants' average ratings as a simple effect and in interaction with participant group (main effect: $\beta = -.826$, $z = -3.853$, $p < .001$; interaction: $\beta = .865$, $z = 4.271$, $p < .001$). The interaction shows that the native average relative frequency ratings only predict the natives' accuracies, and not those of the advanced learners. There were no other significant interactions with participant group.

Lexical decision RTs

The analyses of the lexical decision RTs were based on correct responses only (1564 trials, 76.7% of

the data), and on RTs that were within 2.5 standard deviations (335.91 ms) of the grand mean (1386.53 ms). The number of observations left for analysis totalled 1473 (72.2% of the data).

The left panel of Figure 6 presents the Dutch advanced learners' mean RTs per word variant as a function of the relative frequency ratings averaged over participants for that variant. The higher the average relative frequency ratings for the word variants, the faster the advanced learners reacted.

“(Figure 6 around here)”

The right panel of Figure 6 plots the average advanced learners' RTs against the French natives' average relative frequency ratings. The plot does not show a clear relationship between the two.

We performed statistical analyses to further investigate these observations. We analysed the log-transformed RTs by means of mixed-effects regression models, following the same fitting procedure as in Experiment 1. We tested the same predictors as for the accuracy analysis of this experiment, in addition to log-transformed target word duration and log-transformed previous RT. The final statistical model is summarized in Table B.5 (see Supplemental Material B).

The Dutch advanced learners responded faster to words with higher frequencies of occurrence. Like the French natives in Experiment 1, they speeded up over the course of the experiment, they were faster if they had responded more quickly to the previous trial, and, more interestingly, responded faster to unreduced than to reduced variants (see the lower panel in Figure 8). The random slope of variant type by word shows that this latter effect was larger for some words than for others. Finally, the advanced learners responded faster to word variants that, according to the average of the whole group of students, occur more often. A different model with the ratings provided by the French native listeners as predictor showed that this predictor does not reach significance.

We then combined the advanced learners' RTs with those provided by the native listeners to investigate whether the predictors showed different effects for the advanced learners and the natives. The best model included a main effect of the average relative frequency ratings by the French participants ($AIC = -2514.6$). This model also contains two interactions with participant group. The first is expected, given the main effects in the separate models; it is the interaction with word frequency (main effect: $\beta = -.00305$, $t = -.614$; interaction: $\beta = -.0265$, $t = -3.474$), which shows that the word frequency effect is substantially larger for the advanced learners than for the French natives. The second interaction is less expected. We found a group by target duration interaction, which shows that target duration effects were larger for the natives than for the learners (main effect: $\beta = .235$, $t = 6.207$; interaction: $\beta = -.143$, $t = -2.538$).

Summary and discussion

The advanced learners' average relative frequency ratings predict both their accuracy and RTs. Importantly, the learners' accuracies were not predicted by the natives' average ratings. This strongly suggests that how well a listener recognizes a pronunciation variant is determined by how much exposure this listener has had with the given variant and thus that advanced learners have lexical representations for these variants specified for their variants' frequency in their own exposure.

In Experiment 3, we tested beginner learners. These learners may have encountered the reduced variants of only a few words, and with different frequencies than native listeners. We therefore expect that they have only lexically stored a few reduced variants. Moreover, the variants that they have stored are specified for these variants' frequencies of occurrence in the learners' own input rather than in the natives' input.

We therefore do not expect a correlation between the learners' performance and the natives' relative frequency ratings, as these ratings reflect the variant frequencies in the natives' input rather

than in the beginner learners' input. The advanced learners ratings may better reflect the beginner learners' input, and there may therefore be a (weak) correlation between the beginner learners' performance and the advanced learners' variant frequency ratings.

We also tested whether the ratings of the relative frequencies of the unreduced and reduced variants provided by the beginner learners themselves predict their performance. On the one hand, following the rationale of Experiments 1 and 2, we might expect that this is the case since these ratings would reflect the beginner learners' input. On the other hand, the beginner learners' ratings may be unreliable for many word variants because they have very little experience with these variants. These learners may therefore be just guessing. Their relative frequency ratings then do not correlate with their response accuracies and RTs.

Finally, we investigated in Experiment 3 which measure of word frequency best predicts beginner learners' performance. Since their input is very different from the input that a native listener receives, the word frequencies listed in Lexique 3.80 (New et al., 2001) for film subtitles probably do not well reflect the frequencies with which these learners encounter words. The learners' ratings of these word frequencies may therefore better predict their performance than the word frequencies extracted from Lexique.

Experiment 3

We performed two experiments with Dutch beginner learners. One experiment was an exact replication of Experiment 2 (Supplemental Material C; Experiment 3a). In the other experiment (Experiment 3, reported below), we changed the instruction for the relative frequency estimation task. The participants were explicitly asked to rate the variants' frequencies in their own language input instead of in everyday French. This change in instruction, however, did not substantially affect the results. We

therefore only report here the results of Experiment 3 and describe Experiment 3a in Supplemental Material C, focusing on the differences between the two experiments¹. Experiment 3 also incorporated a word frequency rating task.

Method

Participants

Fifty-six Dutch university students (aged 18-26 years; ten males) were paid to participate in the experiment. All were born and raised in the Netherlands and had taken a maximum of 2 hours of French classes a week, for a maximum of six years at secondary school. They had spent no more than 15 days a year in France, and were not exposed to French films, books, poems or music on a weekly basis. Their proficiency levels roughly corresponded to the B1-B2 level (CEFR, Council of Europe, 2011).

To verify whether these Dutch beginner learners are able to perceive the difference between the reduced and unreduced variants of French schwa words, we asked them to perform an AXB task. In this task, the participant heard three different tokens of the same word in a sequence (in every trial a different word) and had to say whether the second (X) token was more similar to the first (A) or to the third (B) token. If the first token was unreduced, the third token was reduced and vice versa. There were 17 trials: ten trials with reduced tokens and seven trials with unreduced tokens in the middle of the sequence. Eighteen participants gave incorrect responses to all trials; it is very likely that they reversed the response buttons. We therefore interpreted their responses as correct responses. The complete dataset (96.8% correct) shows that these Dutch beginner learners of French are able to distinguish reduced from unreduced pronunciation variants of French schwa words.

¹ We also conducted all analyses comparing the beginners with the other listener groups pooling the data from Experiments 3 and 3a, instead of with just the data from Experiment 3, in order to increase statistical power. The obtained results were similar to those reported in this section.

Materials

The materials for the lexical decision task and the relative frequency estimation task were the same as in Experiment 2.

The word frequency rating task comprised all target words and 28 real word fillers.

Procedure

The experiment was conducted in sound-attenuated booths at the Max Planck Institute for Psycholinguistics in Nijmegen. The procedure was the same as for Experiment 2, except for the question asked with regards to the relative frequency rating task. We adapted our question to the Dutch translation of "How often have you heard the two pronunciation variants?".

After the relative frequency estimation task, we asked our participants to perform the word frequency rating task. This rating task started with two high frequency filler words (*le matin* 'the morning' and *la vie* 'the life') and two low frequency filler words (*le péage* 'the toll (gate)' and *la gendarmerie* 'the gendarmerie') to give the participant the opportunity to interpret the extreme values of the scale. The fillers were followed by six target words and then by four new fillers with extreme frequencies of occurrence. This combination of target and filler words was repeated until all target words were presented. To rate the frequency of each word, participants were asked to choose a value on the Likert scale presented on the computer screen by pressing a number between 1 and 7 on the keyboard. A score of 1 reflected that the word had a very low frequency, whereas a score of 7 reflected that the word had a very high frequency.

Results

Relative frequency estimation task

Due to a technical error the relative frequency ratings for the target word *la vedette* were not registered.

We obtained 1151 ratings, which showed a very poor interrater agreement in terms of *k fleiss* (on the

basis of the ratings for the reduced variants $k_{fleiss\ kappa}: .002, p > .1$). Although we again used a six-point Likert scale, all average relative frequency ratings were between 3.39 and 4.50. These average relative frequency ratings did not correlate with any of the other relative frequency ratings obtained in this study ($ps > .05$). Surprisingly, the reduced variants were rated as occurring more frequently than the corresponding unreduced variants (see the left panel of Figure 7).

Lexical decision accuracy

In the left panel of Figure 7, the beginner learners' mean accuracy score per word variant is plotted against their average relative frequency rating for that variant. The plot shows that the beginner learners' ratings do not predict the learners' accuracy score.

“(Figure 7 around here)”

We further examined whether the beginner learners' mean accuracy scores could be predicted by the advanced learners' (middle panel) or the French natives' (right panel) average relative frequency ratings. The middle and right panels of Figure 7 suggest that if there is an effect of average relative frequency ratings of either group on beginner learners' mean accuracy, this effect is small.

We performed statistical analyses to see whether the effect of any of the different average relative frequency ratings is, nevertheless, statistically significant. We analysed the 2464 responses (of which 48.8% were correct) by means of logistic mixed-effects models, using the same random predictors as in the analysis of the accuracy data of Experiment 2. We created different models with the relative frequency ratings provided by the native listeners, the advanced learners, and the beginner learners tested in this experiment. In addition, we included as fixed predictors variant type, trial number, dictation task performance, and the Ghent score (the number of correctly classified real words minus

twice the number of pseudo words classified as real words in the LexTALE task by the given participant, Brysbaert, 2013). Furthermore, we tested separate models in which we replaced the word frequencies taken from Lexique by the average word frequency ratings as indicated by the participants on a scale from one to seven. The average word frequency ratings did not correlate with the word frequencies in Lexique as reported for subtitles (Lexique 3.80, New et al., 2001, $p > .05$).

Table B.6 (see Supplemental Material B) shows a summary of the best statistical model. Participants made more errors for reduced than for unreduced variants (see the upper panel of Figure 8), and again the effect of variant type was larger for some words than for others, as shown by the significant random slope of variant type for word.

“(Figure 8 around here)”

Moreover, participants who were better at the dictation task made fewer errors. Participants made fewer errors for words that occur more frequently. Interestingly, we find this latter effect with both predictors reflecting word frequency (the one derived from Lexique and the one derived from the participants' ratings), but the predictor resulting in the best statistical model is based on the word frequency ratings provided by the participants. Most importantly, we did not find an effect of the average relative frequency rating of any participant group.

We combined the data from Experiment 3 with the data from Experiment 1 to investigate whether the beginner learners and the natives showed statistically significant differences. The best model contained an interaction between learner group and the relative frequency ratings averaged over the natives, demonstrating that only the natives reacted significantly more accurately to variants that, according to them, on average, had higher relative frequencies (main effect: $\beta = 0.143$, $z = 2.183$;

interaction: $\beta = -0.217$, $z = -2.629$).

We further combined the data from Experiment 3 with the data from Experiment 2 in order to see whether there were statistically significant differences between the beginner and advanced learners. The best model contained an interaction between the relative frequency ratings averaged over the advanced learners and variant type, demonstrating that the advanced learners' ratings could only well predict the beginner and advanced learners' accuracy scores for the reduced variants (simple effect of relative frequency for unreduced variants: $\beta = -0.571$, $z = -1.818$; $p > .05$; interaction with pronunciation variant: $\beta = 1.188$, $z = 2.0194$, $p < .05$).

Lexical decision RTs

We restricted the analyses of the RTs to those trials where the participants had provided the correct responses (1202 trials, 48.8% of the data). RTs deviating from the grand mean (1442.57 ms) by more than 2.5 times the standard deviation (364.48 ms) were considered to be outliers and removed, which left 1145 observations (46.5% of the data). In the left panel of Figure 9, the beginner learners' mean RTs per word variant are plotted against their average relative frequency ratings for that variant. There seems to be no clear correlation between the two variables.

“(Figure 9 around here)”

We also plotted the beginner learners' mean RTs per word variant against the advanced learners' (middle panel of Figure 9) and natives' (right panel of Figure 9) average relative frequency ratings per word variant. Again we see no clear relationship between the relative frequency ratings and the beginner learners' mean RTs.

This is confirmed by statistical analyses. We analysed the log-transformed RTs by means of

mixed-effects regression models, following the same fitting procedure as in Experiment 2. We tested the same predictors as for the accuracy analysis of this experiment, in addition to log-transformed target word duration and previous log RT. Table B.7 (see Supplemental Material B) summarizes the best statistical model.

Participants responded more slowly to longer words and if they had responded slowly in the preceding trial. More interestingly, they also responded more slowly to reduced variants relative to unreduced variants (see the lower panel of Figure 8). As shown by the random slope, this effect of variant type was not equally large for all participants. No predictor reflecting the frequencies of the words emerged as significant. Importantly, none of the variant relative frequency ratings (neither the participants' own average ratings, nor the average ratings from the other groups tested in this study) predicted the participants' RTs.

We combined the data from Experiment 3 with the data from Experiment 1 to investigate whether the predictors showed different effects for the beginner learners and the natives. The best model contains an interaction between participant group and the average relative frequency ratings provided by the natives, demonstrating that only the natives reacted significantly faster to variants that according to them, on average, had higher relative frequencies (main effect: $\beta = -0.0194$, $t = -4.048$; interaction: $\beta = 0.0288$, $t = 4.174$).

We further combined the data from Experiment 3 with the data from Experiment 2 in order to see whether there were statistically significant differences between the beginner and advanced learners. The best model contains an interaction between the average ratings provided by the advanced learners and learner group, showing that only the advanced learners reacted significantly faster to variants that according to them, on average, had higher relative frequencies (main effect: $\beta = -0.0301$, $t = -3.236$; interaction: $\beta = 0.0274$, $t = 2.769$).

Summary and discussion

The beginner learners' accuracy scores and RTs can neither be predicted by the advanced learners' nor by the natives' average relative frequency ratings if the data from the beginners are analysed just by themselves. However, when we combine the data from the beginner and advanced learners, the advanced learners' ratings appear to predict the beginner learners' accuracy. In contrast, the natives' ratings do not predict the beginner learners' accuracy in a combined dataset. This suggests that the frequencies for the reduced pronunciation variants stored in their mental lexicons differ between beginner learners and natives, while there may be some similarity between beginner and advanced learners.

Furthermore, we found that variant relative frequency ratings provided by beginner learners do not correlate with their own word recognition accuracy, nor with their RTs. There may be no variant frequency effect because the relative frequency rating task is too hard for beginner learners. Due to their low exposure to reduced pronunciation variants, they may not be able to provide different ratings for the different words, and may often just be guessing.

The participants in Experiment 3 also rated how often they encountered the target words (independently of the variant). This word frequency measure outperformed the word frequency measure based on Lexique in the accuracy analysis. This suggests that word frequency ratings better reflect learners' stored word frequencies than lexical databases based on language produced by native speakers.

General discussion

The aim of this study was to further investigate the hypothesis that listeners rely on lexical representations of reduced word pronunciation variants. These lexical representations would be

specified for the variants' frequencies of occurrence. If so, how easily a listener recognizes a given word pronunciation variant should directly reflect the listener's exposure to this variant. We investigated this by exploring whether the recognition accuracy and reaction times for a word pronunciation variant in the lexical decision task correlate best with the frequencies of occurrence of that variant holding for the group to which the listener belongs.

We conducted an experiment with several groups of listeners. First, participants performed an auditory lexical decision task with schwa words that were pronounced with schwa (e.g., /larɔvy/ *la revue* 'the magazine') or without schwa (e.g., /larvy/). Second, they indicated the relative frequencies of the two pronunciation variants of each word by choosing a value on a scale from 1 to 6. We examined whether the recognition accuracy and reaction times (RTs) in the lexical decision task correlated with the participants' relative frequency ratings. We tested native listeners of French (Experiment 1) and beginner and advanced learners of French (Experiments 2, 3, and 3a in Supplemental Material C). The learners of French are expected to have encountered reduced word pronunciation variants less often than the natives have, and the unreduced and reduced variant of a word may therefore have different relative frequencies for them than for natives.

The natives in Experiment 1 provided relative frequency ratings that, on average, correlated well ($r = .83$) with other francophone participants' ratings collected by Racine (2008). This shows that two groups of speakers of more or less the same variant of French produce ratings that are very similar, which suggests that speakers are able to indicate their exposure to the two variants of a schwa word.

The Dutch advanced students' ratings from Experiment 2 correlated with those provided by native listeners in Racine (2008) and those obtained in our Experiment 1, but these correlations were substantially lower ($r = .40, .48$, respectively) than the correlation between our

group of native listeners and the natives in Racine (2008; $r = .83$). This suggests that advanced learners' exposure to reduced variants is different from that of native listeners.

Interestingly, the advanced learners' relative frequency ratings correlated with the word frequencies of the words listed in Lexique. A higher frequency in Lexique correlated with a lower average relative frequency rating for the unreduced variants. This correlation is in line with the general finding that more frequent words tend to be more often reduced and to a greater extent (e.g., Pluymaekers, Ernestus & Baayen, 2005; Jurafsky, Bell, Gregory & Raymond, 2000; Bybee, 1998). Possibly, the advanced learners are aware of this generalisation and applied it when providing their frequency ratings.

Experiment 3 tested beginner learners of French. It could be expected that these learners indicated that they knew most words only in their unreduced variants. This, however, was not the case: their average relative frequency ratings on a scale from 1 (the unreduced variant is the most frequent variant) to 6 (the unreduced variant hardly ever occurs) range between 3.39 and 4.50 (i.e., a range of 1.1). The ranges were larger for the advanced learners' ratings (2.17) and for the natives' ratings (4.66).

The fact that the beginner learners mostly chose values in the middle of the continuum indicates that these beginner learners hardly differentiated among words and may just have been guessing. They may have adopted a guessing strategy because they had encountered the vast majority of words (nearly) only in their unreduced variants and were thus unfamiliar with the reduced ones. They will not have lexically stored the reduced pronunciation variants that they had not encountered yet. The frequencies for the variants that they have encountered and stored will hardly differ among the words.

In the lexical decision task, the French native listeners responded faster and more accurately to (reduced) pronunciation variants with higher relative frequency ratings. The advanced learners' accuracies and RTs only correlated with the average frequency ratings provided by these learners themselves. These results suggest that advanced learners are sensitive to relative frequencies during word recognition, but that these frequencies differ from those for the native listeners. These findings support the hypothesis that the recognition of French reduced pronunciation variants is influenced by the listener's exposure to these variants.

The beginner learners' accuracies and RTs were neither predicted by the natives' nor by the advanced learners' ratings. Since beginner learners receive input that is very different from the input to natives, and that also deviates from the input to advanced learners, this finding is also in line with our hypothesis that listeners lexically store the variants' frequencies reflecting their own input.

The beginner learners' performance could neither be predicted by their own relative frequency ratings. We believe that this is the case because the beginner learners were just guessing in the relative frequency estimation task, as explained above, and that these ratings were thus unreliable. We assume that when reduced pronunciation variants are not stored at all or stored with only low frequencies, as is the case for the beginner learners, listeners may not or hardly use these lexical representations during recognition but use other mechanisms and accept, for instance, imperfect matches of the acoustic signal with the lexically stored unreduced variants.

Thus, we find differences between the listener groups in which relative frequency rating (if any) best predicts their performance in the lexical decision experiment. These differences are supported by the pairwise comparisons of the data from the different groups, which show

statistically significant differences between natives and beginner learners in the effects of the ratings provided by the different groups on their accuracy and RTs, between beginner and advanced learners on their RTs, and between advanced learners and natives on their accuracy.

The beginner and advanced learners did not differ significantly in their sensitivity to relative frequency ratings provided by the advanced learners in the accuracy analysis. Similarly, the advanced learners and the natives did not differ in their sensitivity to relative frequency ratings provided by the natives in the RT analysis. The lack of significant differences in sensitivity to relative frequency ratings in the combined datasets have to be considered with care. There may simply be power issues, especially because the learner groups show large within-group variation (see, e.g., the error bars in Figure 8): some learners are better (or worse) than suggested by their group's average proficiency level, and therefore performed more similarly to participants in the higher (or lower) proficiency level group.

Our finding that native listeners and advanced learners are sensitive to frequencies of occurrence of reduced variants for French schwa words adds to previous research on the comprehension of pronunciation variants (e.g., Pitt et al., 2011; Ranbom & Connine, 2007) in several respects. First, while Ranbom and Connine documented listeners' sensitivity to the frequencies of word pronunciations resulting from the substitution of one segment by another, we show that this may also hold for variants from which a segment is lacking. Second, while Ranbom and Connine and Pitt and colleagues reported evidence concerning different pronunciations for a consonant, we provide evidence concerning the presence versus absence of a vowel. Frequency effects in comprehension are thus not limited to words in which a consonant is reduced; they can also occur in words in which a vowel is reduced. Third, our data show that listeners not only rely on the frequencies of different pronunciation variants of words in

American English, but also in French. Fourth, our study shows that advanced learners of a language may also show sensitivity to variants' frequencies during speech comprehension. Like natives, learners keep frequency counts for pronunciation variants and use them during recognition. In sum, we showed that listeners of advanced proficiency levels are sensitive to the frequency with which a given word lacks a vowel in their own speech input.

Our results thus show that exposure to a given variant of a given word predicts how well a listener processes that variant. Listeners appear to store the frequencies with which they hear each word form. This makes them very efficient listeners to the type of speech they are trained on, but less efficient listeners to other speech registers.

We obtained these results even though we only had subjective measures of the relative frequencies of the variants. Objective measures may better reflect language exposure, but in the absence of objective measures, subjective measures may also provide insight into speech processing.

Suppose we had only tested native listeners. We could then have explained the relative frequency effect in their processing in a different way. Previous studies have shown that more frequent words tend to be more reduced and to show stronger co-articulation than less frequent words (e.g., Wright, 2004; Jurafsky, Bell, Gregory & Raymond, 2000; Bybee, 1998; Lindblom 1990). These acoustic differences among words and word pronunciation variants as a function of frequency could then have explained the attested relative frequency effects in processing: listeners could have recognized more frequent variants more easily because these variants tend to show more co-articulation, which could have facilitated recognition (e.g., Salverda, Kleinschmidt & Tanenhaus, 2014; Mattys, White & Melhorn, 2005). However, if the frequency effects had resulted from the words' acoustic properties and had thus been *indirect* consequences of the variants' frequencies of occurrence, all

listener groups should have been sensitive to similar relative frequencies, which is contrary to fact. Our results show that the relative frequency effects that we observe do not result from the acoustic properties of the words' variants. Groups that differ in their experience with the language show sensitivity reflecting their own exposure.

In Experiment 3, we collected subjective word frequency ratings and investigated whether these or a word frequency measure derived from a general database for French (e.g., Lexique, New et al., 2001) better predict beginner learners' behaviour. We found that the subjective word frequency ratings outperformed the objective frequency counts for predicting learners' accuracy. In the analysis of the RTs, neither word frequency measure showed a statistically significant effect. The subjective word frequency ratings probably outperformed the objective word frequency measure in predicting accuracy because the former measure better reflects the learners' experience with the words. This experience may be very different from the experience that native listeners generally have and that is reflected in the frequency database that we consulted. Objective frequency measures may generally outperform subjective frequency measures, as claimed by Ghyselinck and colleagues (2004), but only if the objective frequency measure faithfully reflects the given participants' exposure.

Our data further contribute to the question of whether unreduced word pronunciation variants have a special status in word recognition (e.g., Ernestus & Baayen, 2007; Janse, Nöteboom & Quené, 2007; Tucker & Warner, 2007). The native listeners in Experiment 1 performed well in the lexical decision task, but nevertheless showed a clear processing advantage for the unreduced variants. They processed the variants with schwa more accurately (99.0% correct) and more quickly (mean RT: 1153.96 ms) than the variants without schwa (88.4% correct; mean RT: 1217.99 ms). This privileged status of the unreduced variant may be due to the role of orthography or of context. In our experiment,

context may have indeed favoured the unreduced variant, as the unreduced variant is more likely than the reduced variant when the word is only preceded by its determiner. Moreover, the duration of the vowel of the determiner was typical of a determiner followed by an unreduced word (mean /ə/ and /a/ duration in the determiner was 119 and 118 ms, respectively, if followed by an unreduced variant and 115 and 116 ms if followed by a reduced variant). Findings by Bürki and colleagues (2010) suggest that unreduced variants may lose their privileged status in speech production when presented in more natural context. Future research has to show whether this also holds for speech comprehension.

Since learners have less experience with reduced word variants than natives, these learners are expected to show a greater bias for the unreduced variants than the natives. This is exactly what we found. While the difference in lexical decision accuracy for the unreduced and reduced variants was 10.6% for the native listeners, it was 24.9% for the advanced learners and 29.3% for the beginner learners. This shows that the focus on unreduced variants in the classroom (e.g., Fonseca-Greber & Waugh, 2003) is harmful for learners' comprehension of casual speech. Interestingly, the beginners categorised more than half of the reduced variants (63.3%) as pseudo words, possibly because the majority of these variants have illegal consonant clusters (e.g., /RV/ in reduced /RVy/ 'magazine').

Furthermore, the privileged status of the unreduced variant appears from the interaction between variant type and the relative frequency rating provided by the advanced learners in the combined dataset of the beginner and advanced learners (see accuracy analysis). The unreduced variant is well recognized, independent of its relative frequency. Nearly all accuracy and RT plots seem to show this interaction but it does not surface as statistically significant in the other analyses (possibly because of lack of statistical power).

The question arises how to account for these results in models of word recognition. As explained in the Introduction, a rule-based model of word recognition cannot easily account for effects

of relative frequency for reduced word pronunciation variants, since in this type of model, word-specific variant frequencies can only become available (and thus affect the word recognition process) after the word has been identified. In contrast, relative frequency effects can easily be accounted for in models assuming lexical storage of multiple word pronunciation variants. For instance, if Shortlist B (Norris & McQueen, 2008) is adapted in such a way that it stores these variants, the relative frequencies can co-determine the variants' priors and thus how easily these variants are processed.

In line with this, we assume that French natives and advanced learners of French store the relative frequencies of pronunciation variants together with the lexical representations of these variants in the mental lexicon, for instance, in the form of resting activations or in the form of priors. Similarly, beginner learners also store the reduced variants they have encountered, together with their (low) frequency counts. The variants stored and their frequencies change as a function of the type and amount of exposure. We leave open the possibility that variants with very short schwas are also stored, which we view as a topic for future research.

An alternative account for the processing of segment reduction is *naive discriminative learning* as proposed by Baayen (2010). Although this model can account for reduction, it cannot easily simulate the observed relative frequency effects. In the *naive discriminative learning* model, a two-layer network with symbolic representations for a word's form and a word's semantics, each input unit (unigram, bigram, uniphone, biphone or triphone) is linked to each meaning, and a weight is associated with each link. Effects of the relative frequencies of the two pronunciation variants of a word have to be expressed in the weights of the connections between the biphones or triphones of the two variants of the word and the word's meaning. These weights also have to indicate the likelihoods that these input units represent different word types. That is, word-specific relative frequencies cannot be separately represented from other (relative) frequencies, and therefore these relative frequency effects cannot

simply emerge. The *naive discriminative learning* model cannot easily explain the observed relative frequency rating effect.

Finally, the two tests added to the experiment to determine the learners' proficiency levels provide interesting insights in the relation between different types of proficiencies. Learners who processed reduced variants more accurately were also better at the dictation task, while they were not necessarily better at visual lexical decision (the LexTALE task). This result suggests that the ease with which learners process spontaneous speech is not strongly correlated with their vocabulary knowledge, but more specifically with their listening skills.

In conclusion, this study provides evidence that listeners lexically store reduced pronunciation variants with their relative frequencies. We compared listener groups differing in their proficiency levels and demonstrated that their processing of unreduced and reduced pronunciation variants is best predicted by their exposure to these variants, as captured by subjective frequency ratings. Together, these results reveal that listeners easily comprehend the speech that they hear in daily life because they are fine-tuned to the frequencies of occurrence of reduced pronunciation variants in their own daily life input.

Acknowledgments

This research was supported by a Consolidator grant (284108) from the European Research Council awarded to the second author.

The authors thank Cécile Fougeron and Willemijn Heeren, who made it possible to conduct the experiments in Paris and Leiden. We also express our appreciation to Evelyne Vos-Fruit, Janine Berns, Mattie Lagarrigue and Jenny Doetjes, who facilitated the data collection.

Supplemental Material

The Supplemental Material can be found at the address (online address to be filled in)

References

Adda-Decker, M., Boula de Mareüil, P., Adda, G., & Lamel, L. (2005). Investigating syllabic structures and their variation in spontaneous French. *Speech Communication*, 46, 119-139.

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. *Proceedings of isit-1973*, 267-281. Budapest, Hungary.

Alegre, M., & Gordon, P. (1999). Frequency effects and the representational status of regular inflections. *Journal of Memory and Language*, 40, 41-61.

Askildson, V. (2008). What do teachers and students want from a foreign language textbook? *Dissertation abstract international*, 69(3), 274A.

Baayen, R. H. (2010). Assessing the processing consequences of segment reduction in Dutch with naive discriminative learning. *Lingue & Linguaggio*, 9(2), 95-112.

Baayen, R. H., McQueen, J., Dijkstra, A., & Schreuder, R. (2003). Frequency effects in regular inflectional morphology: Revisiting Dutch plurals. In R.H. Baayen, & R. Schreuder (Eds.), *Morphological Structure in Language Processing* (pp. 355-390). Berlin: Mouton de Gruyter (Trends in

linguistics. Studies and monographs, 151).

Balota, D.A., Pilotti, M., & Cortese, M.J. (2001). Subjective frequency estimates for 2,938 monosyllabic words. *Memory & Cognition*, 29, 639-647.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67(1), 1-48. doi:10.18637/jss.v067.i01.

Brysbaert, M. (2013). LEXTALE_FR: A fast, free, and efficient test to measure language proficiency in French. *Psychologica Belgica*, 53, 23-37.

Brysbaert, M., & Cortese, M.J. (2010). Do the effects of subjective frequency and age of acquisition survive better word frequency norms? *The Quarterly Journal of Experimental Psychology*, 64(3), 545–559.

Bürki, A., Ernestus, M., & Frauenfelder, U.H. (2010). Is there only one "fenêtre" in the production lexicon? On-line evidence on the nature of phonological representations of pronunciation variants for French schwa words. *Journal of Memory and Language*, 62, 421-437.

Bürki, A., Ernestus, M., Gendrot, C., Fougerson, C., & Frauenfelder, U.H. (2011). What affects the presence versus absence of schwa and its duration: a corpus analysis of French connected speech. *Journal of the Acoustical Society of America*, 130, 3980-3991.

Bürki, A., Fougeron, C., & Gendrot, C. (2007). On the categorical nature of the process involved in schwa elision in French. *Interspeech*, 1026-1029. Antwerp, Belgium.

Bybee, J. (1998). *The Phonology of the lexicon: Evidence from lexical diffusion. Usage-based models of language*, M. Barlow & S. Kemmer (Eds.). Stanford: CSLI.

Clashen, H. (1999). Lexical entries and rules of language: A multidisciplinary study of German inflection. *Behavioral and Brain Sciences*, 22, 991-1060.

Clopper, C. G. (2014). Sound change in the individual: Effects of exposure on cross-dialect speech processing. *Laboratory Phonology*, 5, 69-90.

Council of Europe (2011). *Common European Framework of Reference for Languages: Learning, Teaching, Assessment*. Council of Europe. Retrieved from http://www.coe.int/t/dg4/linguistic/cadre1_en.asp

Coquillon, A. (2007). Le français parlé à Marseille : exemple d'un locuteur PFC. In S. Detey & D. Nouveau (Eds.), *PFC: enjeux descriptifs, théoriques et didactiques* (Bulletin PFC 7), 145-156.

Cutler, A., Sebastián-Gallés, N., Soler Vilageliu, O., & van Ooijen, B. (2000). Constraints of vowels and consonants on lexical selection: Cross-linguistic comparisons. *Memory & Cognition*, 28, 746-755.

Durand, J., & Eychenne, J. (2004). Le schwa en français : pourquoi des corpus ? *Corpus*, 3, 311-356.

Ernestus, M. (2000). *Voice assimilation and segment reduction in casual Dutch. A corpus-based study of the phonology-phonetics interface* (Doctoral dissertation, LOT, Utrecht). Retrieved from http://www.mirjamernestus.nl/Ernestus/publications/Ernestus_2000_Dissertation.pdf/

Ernestus, M., & Baayen, R. H. (2007). The comprehension of acoustically reduced morphologically complex words: The roles of deletion, duration and frequency of occurrence. *Proceedings of the 16th International Congress of Phonetic Sciences*, 773-776. Saarbrücken, Germany.

Ernestus, M., & Cutler, A. (2015). BALDEY: A database of auditory lexical decisions. *Quarterly Journal of Experimental Psychology*, 68, 1469-1488.

Eychenne, J., & Putska, E. (2007). The initial position in Southern French: elision, suppletion, emergence. In J.-P. Angoujard, & O. Crouzet (Eds.), *Proceedings of JEL 2007*, 199-204. Nantes, France.

Faraway, J. (2006). *Extending linear models with R: generalized linear mixed effects and nonparametric regression models*. Boca Raton, FL: Chapman and Hall/CRC.

Fleiss, J. L., Levin, B., & Paik, M. C. (2003). *Statistical Methods for Rates and Proportions*, 3rd Edition. New York: John Wiley & Sons.

Fonseca-Greber, B., & Waugh, L. (2003). On the radical difference between the subject personal

pronouns in written and spoken European French. In P. Leistyna, & C. Meyer (Eds.), *Corpus Analysis: Language Structure and Language Use* (pp. 225-240). Editions Rodopi B.V., Amsterdam.

Fougeron, C., Goldman, J.-P., & Frauenfelder, U. H. (2001). Liaison and schwa deletion in French: an effect of lexical frequency and competition? *Proceedings of Eurospeech 2001*, 639-642. Aalborg, Denmark.

Foulkes, P. (2010). Exploring socio-indexical knowledge: A long past but a short history. *Journal of Laboratory Phonology*, 1, 5-39.

Ghyselinck, M., Lewis, M. B., & Brysbaert, M. (2004). Age of acquisition and the cumulative-frequency hypothesis: A review of the literature and a new multi-task investigation. *Acta Psychologica*, 115, 43-67.

Hansen, A. B. (1994). Etude du E caduc - Stabilisation en cours et variations lexicales. *Journal of French Language Studies*, 4, 25-54.

Hansen, A. B. (2000). Le E caduc interconsonnantique en tant que variable sociolinguistique - une étude en région parisienne. *LINX*, 42(1), 45-48.

Janse, E., Nootboom, 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. *Proceedings of the 10th international symposium on spontaneous speech: Data and analysis*, 29-54. Tokyo, Japan.

Jones, K., & Ono, T. (2000). Reconciling textbook dialogues and naturally occurring talk:

What we think we do is not what we do. *Arizona Working Papers in SLAT*.

Jurafsky, D., Bell, A., Gregory, M., & Raymond, W. D. (2000). Probabilistic relations between words: Evidence from reduction in lexical production. In: J. Bybee, & P. Hopper (Eds.), *Frequency and the emergence of linguistic structure*. Amsterdam: John Benjamins.

Keller, F., Gunasekharan, S., Mayo, N., & Corley, M. (2009). Timing Accuracy of Web Experiments: A Case Study Using the WebExp Software Package. *Behavior Research Methods*, 41(1), 1-12.

Kennedy, S., & Blanchet, J. (2013). Language awareness and perception of connected speech in a second language. *Language awareness*, 23(1-2), 92-106. doi: 10.1080/09658416.2013.863904

Kohler, K. J. (1990). Segmental reduction in connected speech in German: Phonological facts and phonetic explanations. In W.J. Hardcastle, & A. Marchal (Eds.), *Speech production and speech modelling* (pp. 21-33). Dordrecht: Kluwer Academic Publishers.

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

Lennes, M., Alaroty, N., & Vainio, M. (2001). Is the phonetic quality of unaccented words unpredictable? An example from spontaneous Finnish. *Journal of the International Phonetic Association*, 31, 127-138.

Lindblom, B. (1990). Explaining phonetic variation: A sketch of the H & H theory. In W. J. Hardcastle, & A. Marchal (Eds.), *Speech Production and speech modelling*. Dordrecht: Kluwer Academic Publishers, 403-439.

Luce, P. A., & Pisoni, D. B. (1998). Recognizing spoken words: The neighborhood activation model. *Ear and Hearing*, 19, 1-36.

Luce, P. A., Goldinger S. D., Auer, E. T., Vitevitch, M. S. (2000). Phonetic priming, neighborhood activation, and PARSYN. *Perception & Psychophysics*, 62, 615–625.

Mattys, S. L., White, L., & Melhorn, J. F. (2005). Integration of multiple speech segmentation cues: A hierarchical framework. *Journal of Experimental Psychology*, 134 (4), 477–500.

McCarthy, M., & Carter, R. (1995). Spoken grammar: What is it and how can we teach it? *ELT Journal* 49(3), 207-218.

McClelland, J. L., & Elman, J. L. (1986). The TRACE Model of Speech Perception. *Cognitive Psychology*, 18, 1-86.

New, B., Pallier, C., Ferrand, L., & Matos, R. (2001). Une base de données lexicales du français contemporain sur internet: LEXIQUE 3.80. *L'Année Psychologique*, 101, 447-462. Retrieved from <http://www.lexique.org>

Norris, D., & McQueen, J. M. (2008). Shortlist B: A Bayesian model of continuous speech recognition. *Psychological Review*, 115(2), 357-395.

O'Connor Di Vito, N. (1991). Incorporating native speaker norms in second language materials. *Applied Linguistics*, 12(4), 383-396.

Pisoni, D.B. (1973). Auditory and phonetic memory codes in the discrimination of consonants and vowels. *Perception and Psychophysics*, 13, 253-260.

Pitt, M., 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).

Pitt, M., Dilley, L., & Tat, M. (2011). Exploring the role of exposure frequency in recognizing pronunciation variants. *Journal of Phonetics*, 39, 304-311.

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

R Development Core Team (2015). *R: A language and environment for statistical computing* [<http://www.R-project.org/>]. Vienna: R Foundation for Statistical Computing.

Racine, I. (2008). *Les effets de l'effacement du schwa sur la perception et la production de la parole en français*. (Doctoral dissertation, Université de Genève, Genève, no. L. 652). Retrieved from <http://www.unige.ch/cyberdocuments/theses2008/RacineI/meta.html/>

Racine, I., Bürki, A., & Spinelli, E. (2014). The implication of spelling and frequency in the recognition of phonological variants: evidence from pre-readers and readers. *Language, Cognition and Neuroscience*, 29(7), 893-898.

Ranbom, L.J., & Connine, C.M. (2007). Lexical representation of phonological variation in spoken word recognition. *Journal of Memory and Language*, 57, 273-298.

Roelofs, A. (1997). The WEAVER model of word-form encoding in speech production. *Cognition*, 64(3), 249-284.

Salverda, A. P., Kleinschmidt, D., & Tanenhaus, M. K. (2014). Immediate effects of anticipatory coarticulation in spoken-word recognition. *Journal of Memory and Language*, 71(1), 145–163.

Schneider, W., Eschman, A., & Zuccolotto, A., (2007). *E-Prime: User's Guide, Version 2.0*. Psychology Software Tools.

Sereno, J. A., & Jongman, A. (1997). Processing of English inflectional morphology. *Memory & Cognition*, 25, 425-437.

Stemberger, J. P., & MacWhinney, B. (1988). Are inflected forms stored in the lexicon? In: M. Hammond & M. Noonan (Eds.). *Theoretical morphology: Approaches in modern linguistics* (pp. 101-116). San Diego, CA, Academic Press.

Stevens, K. N., Liberman, A. M., Studdert-Kennedy, M., & Öhman, S. (1969). Crosslanguage study of vowel perception. *Language and Speech*, 12, 1-23.

Torreira, F., Adda-Decker, M., & Ernestus, M. (2010). The Nijmegen corpus of casual French. *Speech Communication*, 52, 201-212. doi:10.1016/j.specom.2009.10.004.

Tucker, B. V., & Warner, N. (2007). Inhibition of processing due to reduction of the American English flap. *Proceedings of the 16th International Congress of Phonetic Sciences*, 1949-1952. Saarbrücken, Germany.

Waugh, L. R., & Fonseca-Greber, B. (2002). Authentic Materials for Everyday Spoken French: Corpus Linguistics vs. French Textbooks. *Arizona Working Papers in SLAT*, 9, 114-127.

Wright, R. (2004). A review of perceptual cues and cue robustness. In: B. Hayes, R. Kirchner, & D. Steriade (Eds.). *Phonetically-based phonology*, 34-57. Cambridge: Cambridge University Press.

Figure Captions

Figure 1. The average relative frequency ratings for the reduced variants provided by the native listeners in Experiment 1 plotted against the average relative frequency ratings provided by Racine's (2008) participants. A score of 1 meant that the reduced variant hardly ever occurs, while a score of 6 (for our participants) or 7 (for Racine's participants) meant that the word was always produced without schwa. Every dot represents a word.

Figure 2. The French natives' average relative frequency ratings for the reduced variants (open circles) and for the unreduced variants (asterisks) plotted against the French natives' mean accuracy scores.

Figure 3. The French natives' average relative frequency ratings for the reduced variants (open circles) and for the unreduced variants (asterisks) plotted against the French natives' mean reaction times.

Figure 4. The average relative frequency ratings for the reduced variants provided by the advanced learners plotted against the average relative frequency ratings provided by Racine's participants (2008, left panel) and by the native listeners from Experiment 1 (right panel). A score of 1 meant that the reduced variant hardly ever occurs, while a score of 6 (for our participants) or 7 (for Racine's participants) meant that the word was always produced without schwa. Every dot represents a word.

Figure 5. The advanced learners' mean accuracy scores plotted against the average relative frequency ratings for the reduced variants (open circles) and for the unreduced variants (asterisks) as provided by the advanced learners (left panel) and by the French natives (right panel).

Figure 6. The advanced learners' mean reaction times plotted against the average relative frequency ratings for the reduced variants (open circles) and for the unreduced variants (asterisks) as provided by the advanced learners (left panel) and by the French natives (right panel).

Figure 7. The beginner learners' mean accuracy scores plotted against the average relative frequency ratings for the reduced variants (open circles) and for the unreduced variants (asterisks) as provided by

the beginner learners (left panel), by the advanced learners (middle panel) and by the French natives (right panel).

Figure 8. The top figure shows the lexical decision accuracy to unreduced and reduced words by French native participants (Experiment 1), by advanced learners (Experiment 2) and by beginner learners (Experiments 3). The error bars show standard deviations. The bottom figure shows the mean reaction times to unreduced and reduced words by French native participants (Experiment 1), by advanced learners (Experiment 2) and by beginner learners (Experiments 3). The error bars show standard deviations.

Figure 9. The beginner learners' mean reaction times plotted against the average relative frequency ratings for the reduced variants (open circles) and for the unreduced variants (asterisks) as provided by the beginner learners (left panel), by the advanced learners (middle panel) and by the French natives (right panel).

Table B.1

Fillers in the lexical decision experiment.

	bisyllabic	monosyllabic
real words	149 (e.g., <i>la boisson</i>)	89 (e.g., <i>le lit</i>)
pseudo words	44 with schwa: one of two variants (e.g., <i>le beseuil</i> , <i>le b'seuil</i>) 149 without schwa (e.g., <i>la bimise</i>)	89 (e.g., <i>le tuit</i>)

Table B.2

Summary of *the regression model predicting the logit of an incorrect answer by the French native listeners. The intercept represents an unreduced variant.*

Fixed effects	β	z	$p <$
Average relative frequency rating	-1.818	-5.769	.001
Variant type (reduced)	6.151	2.638	.01
Random effects			SD
Word	intercept		6.986
	variant type (reduced)		7.155
Participant	intercept		.995

Table B.3

Summary of the linear mixed-effects model fitted on the log-transformed lexical decision times from the French native listeners. The intercept represents an unreduced variant.

Fixed effects	β	t
Target duration	.221	5.748
Trial number	-.0000545	-2.396
Previous RT	.0824	4.576
Variant type (reduced)	.0627	6.0477
Average relative frequency rating	-.0132	-2.743
Random effects		SD
Word	intercept	0.0265
	variant type (reduced)	0.0485
Participant	intercept	0.0715
Residual		0.128

Table B.4

Summary of the regression model predicting the logit of an incorrect answer by the advanced learners.

The intercept represents an unreduced variant.

Fixed effects	β	z	$p <$
Word frequency from Lexique	-.425	-3.482	.001
Dictation task performance	-.0188	-2.336	.05
Variant type (reduced)	1.935	10.160	.001
Average rating by the advanced learners	-.408	-2.272	.05
Random effects			<i>SD</i>
Word	intercept		0.921
Participant	intercept		0.814
	variant type (reduced)		0.762

Table B.5

Summary of the linear mixed-effects model fitted on the log-transformed lexical decision times of the advanced learners. The intercept represents an unreduced variant.

Fixed effects	β	t
Trial number	-.0000496	-2.0211
Word frequency from Lexique	-.0284	-4.595
Previous RT	.0724	4.611
Variant type (reduced)	.0616	5.484
Average rating by the advanced learners	-.0421	-3.102
Random effects		SD
Word	intercept	0.0467
	variant type (reduced)	0.0543
Participant	intercept	0.0601
Residual		0.135

Table B.6

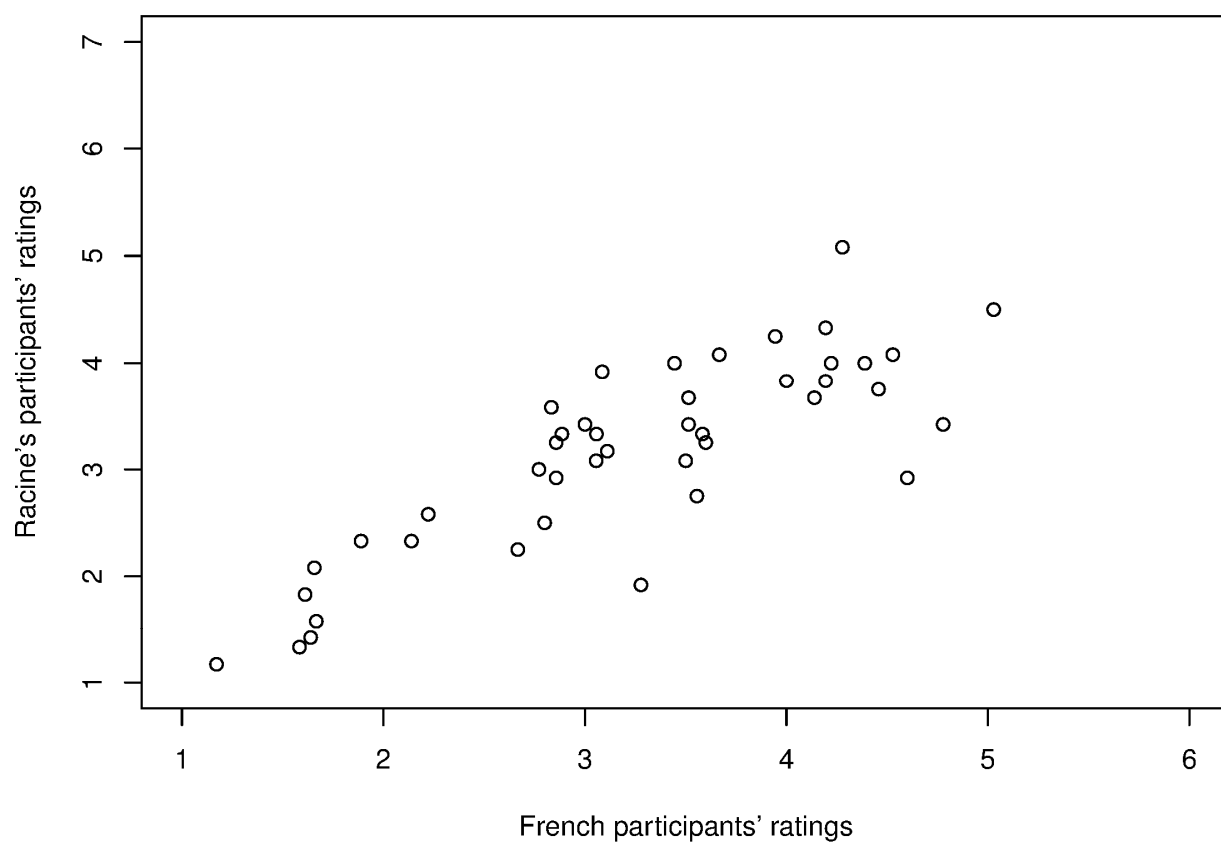
Summary of the regression model predicting the logit of an incorrect answer by the group of beginner learners in Experiment 3. The intercept represents an unreduced variant.

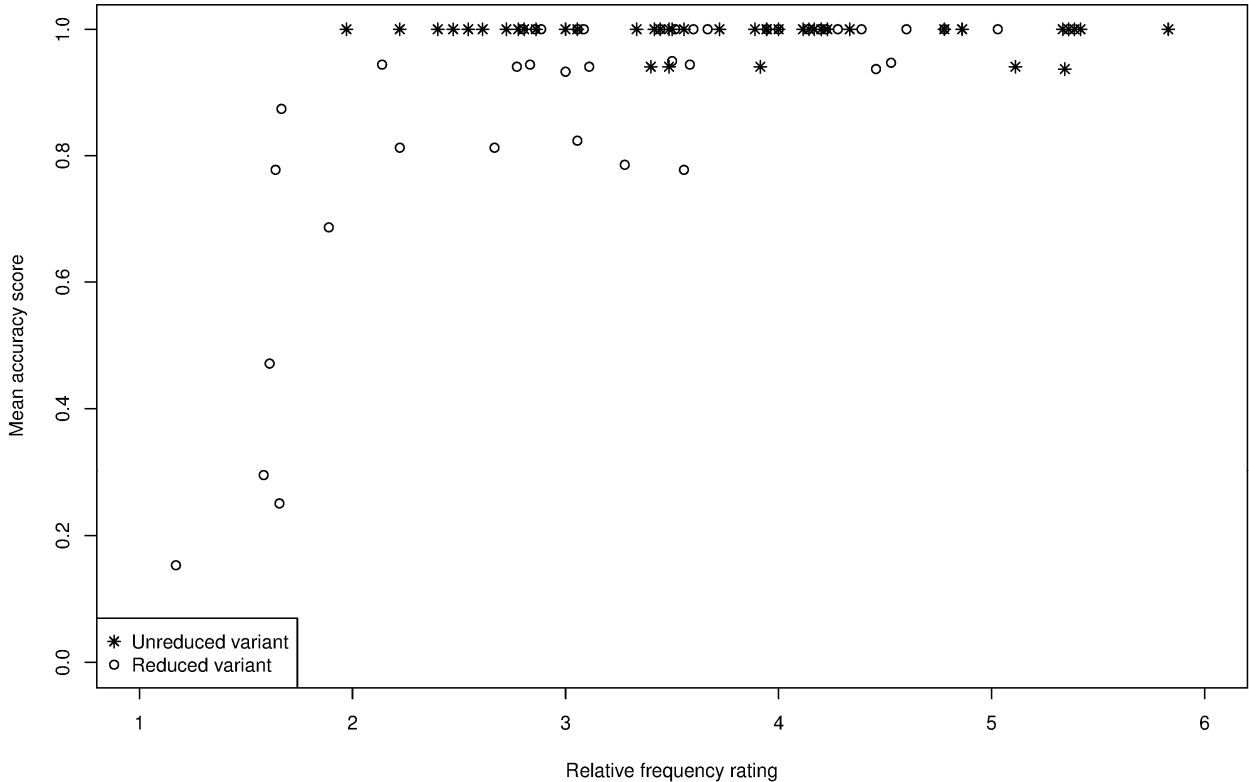
Fixed effects	β	z	$p <$
Variant type (reduced)	1.474	9.580	.001
Word frequency	-0.118	-4.298	.001
Dictation task performance	-0.0289	-2.460	.05
Random effects	SD		
Word	intercept		0.984
	variant type (reduced)		0.757
Participant	intercept		0.671

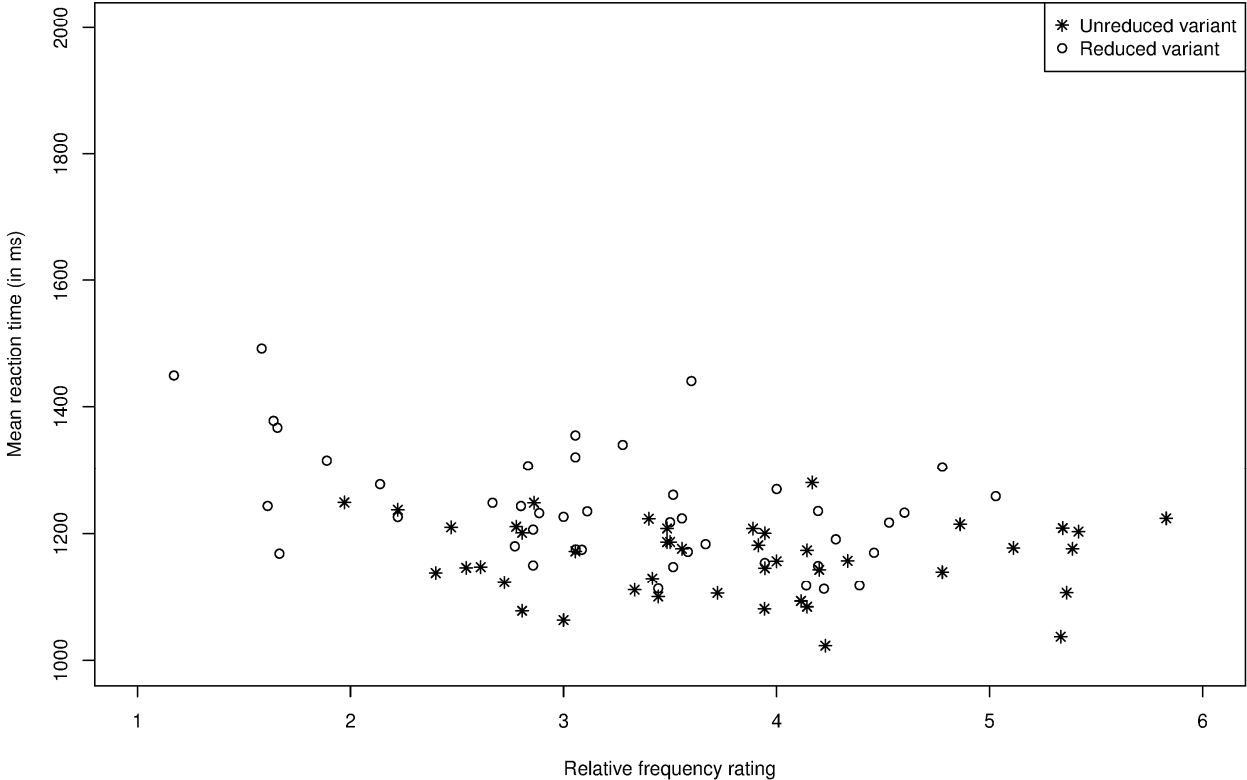
Table B.7

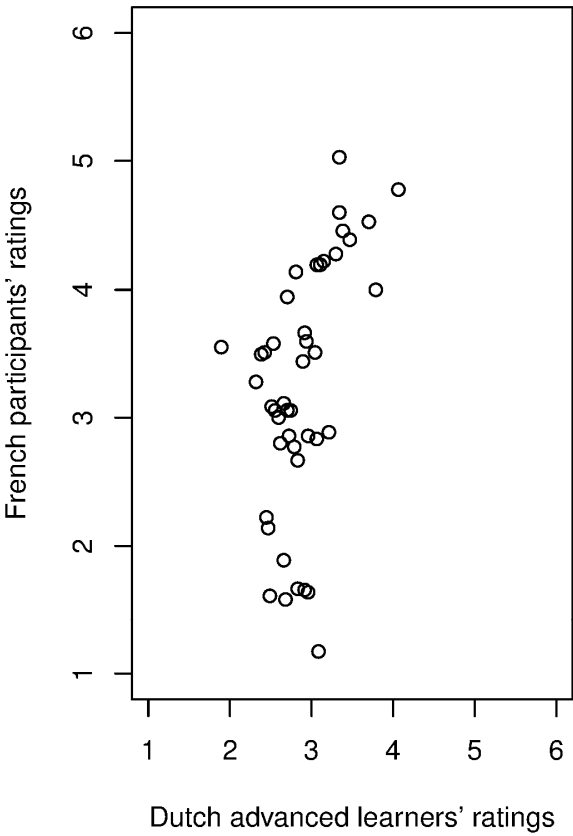
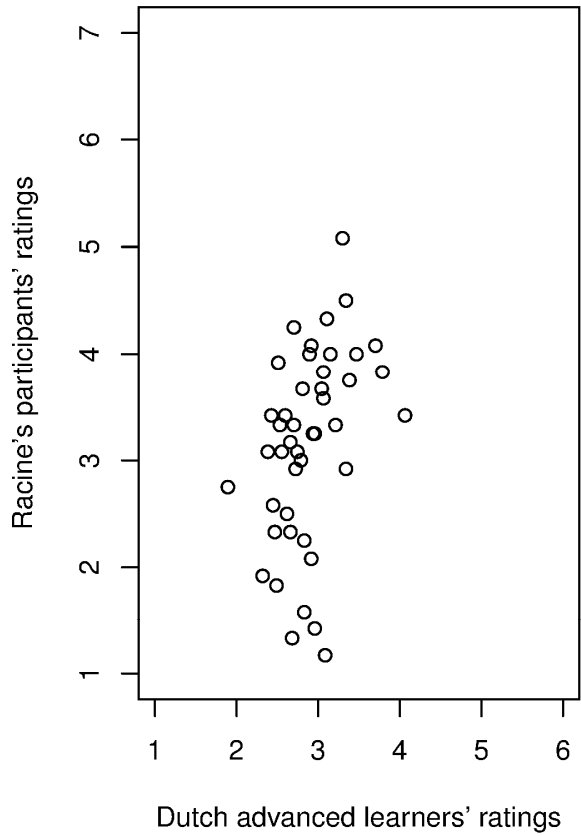
Summary of the linear mixed-effects model fitted on the log-transformed lexical decision times of the group of beginner learners in Experiment 3. The intercept represents an unreduced variant.

Fixed effects	β	t
Target duration	0.167	5.758
Previous RT	0.196	11.808
Variant type (reduced)	0.0164	1.961
Random effects	SD	
Word	intercept	0.0196
Participant	intercept	0.0951
	variant type (reduced)	0.0333
Residual		0.164

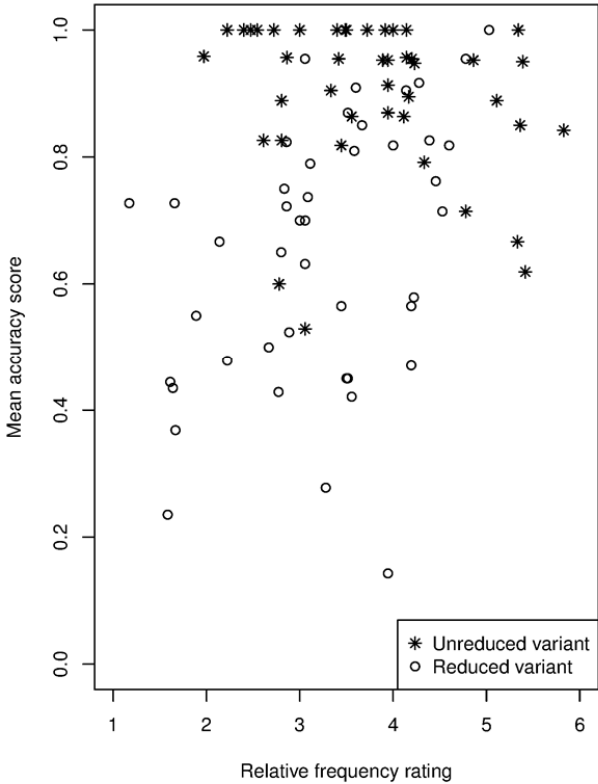
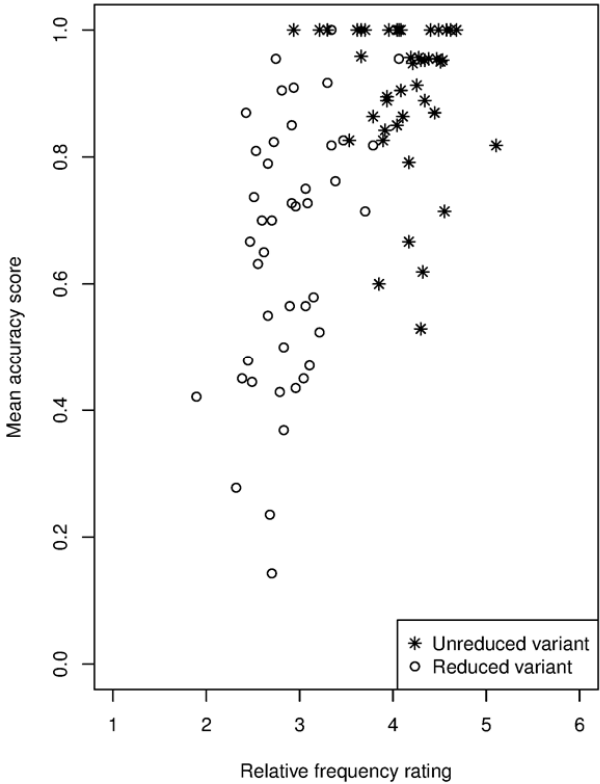


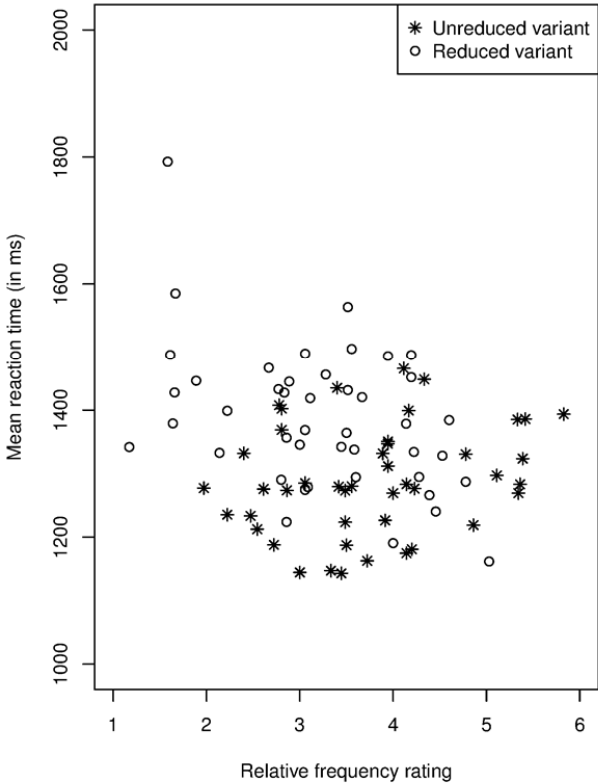
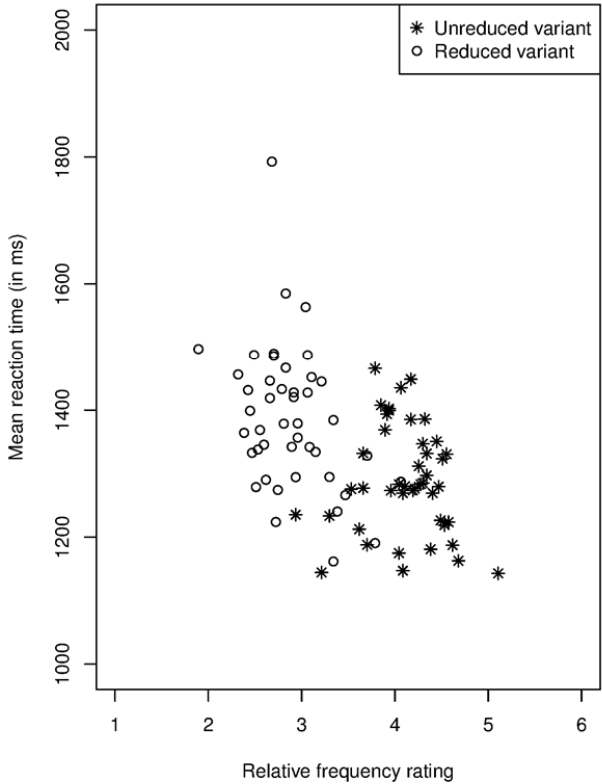






ACCEPTED





ACCEPTED

