

PDF hosted at the Radboud Repository of the Radboud University Nijmegen

The following full text is a publisher's version.

For additional information about this publication click this link.

<http://hdl.handle.net/2066/194916>

Please be advised that this information was generated on 2019-12-08 and may be subject to change.

The role of segmental and durational cues in the processing of reduced words

Marco van de Ven

Centre for Language Studies, Radboud University Nijmegen, the Netherlands;
Max Planck Institute for Psycholinguistics, the Netherlands

Mirjam Ernestus

Centre for Language Studies, Radboud University Nijmegen, the Netherlands;
Max Planck Institute for Psycholinguistics, the Netherlands

Language and Speech
2018, Vol. 61(3) 358–383
© The Author(s) 2017



Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/0023830917727774
journals.sagepub.com/home/las



Abstract

In natural conversations, words are generally shorter and they often lack segments. It is unclear to what extent such durational and segmental reductions affect word recognition. The present study investigates to what extent reduction in the initial syllable hinders word comprehension, which types of segments listeners mostly rely on, and whether listeners use word duration as a cue in word recognition. We conducted three experiments in Dutch, in which we adapted the gating paradigm to study the comprehension of spontaneously uttered conversational speech by aligning the gates with the edges of consonant clusters or vowels. Participants heard the context and some segmental and/or durational information from reduced target words with unstressed initial syllables. The initial syllable varied in its degree of reduction, and in half of the stimuli the vowel was not clearly present. Participants gave too short answers if they were only provided with durational information from the target words, which shows that listeners are unaware of the reductions that can occur in spontaneous speech. More importantly, listeners required fewer segments to recognize target words if the vowel in the initial syllable was absent. This result strongly suggests that this vowel hardly plays a role in word comprehension, and that its presence may even delay this process. More important are the consonants and the stressed vowel.

Keywords

Acoustic reduction, word recognition, speech perception, gating, phonetic detail

Introduction

Research on speech comprehension has focused on the comprehension of carefully pronounced, laboratory speech. In everyday conversations, however, words are generally realized much shorter

Corresponding author:

Marco van de Ven, Centre for Language Studies, Radboud University Nijmegen, P.O. Box 9104, 6500 HD Nijmegen, The Netherlands.

Email: Marco.vandeVen@pwo.ru.nl

and with less articulatory effort than in laboratory speech (an introduction to the phenomenon of acoustic reduction is provided by Ernestus & Warner, 2011). For example, the English word “ordinary” can be pronounced like [ˈɔnri] and, likewise, the Dutch word *natuurlijk* “of course” may be reduced to [ˈtyk]. Reduced pronunciations are ubiquitous in spontaneous speech. To illustrate, Johnson (2004) found that, in American English, segments are changed or missing in 25% and complete syllables are missing in 6% of the word tokens. Similarly, in Dutch, segments are changed or missing in 48% of the word tokens and complete syllables are missing in approximately 19% of the word tokens (Schuppler, Ernestus, Scharenborg, & Boves, 2011).¹ The present study investigated how reduced word pronunciation variants are recognized and whether this can be assessed by means of an adapted version of the gating paradigm.

Several studies have already investigated how listeners recognize reduced word pronunciation variants. Research by Pollack and Pickett (1964) was the first to show that the intelligibility of words excised from fluent speech is increased by adding surrounding context. In line with this, Ernestus, Baayen, and Schreuder (2002) found that listeners had difficulty recognizing highly reduced pronunciation variants out of context (ca. 50% correct) and when these variants were presented together with minimal phonetic context (the neighboring vowels and intervening consonants; ca. 70% correct). Within sentence context, listeners did not have any difficulty recognizing these reduced variants (more than 90% correct). These findings indicate that listeners need some information from the sentence context to recognize highly reduced word pronunciation variants. Consequently, experiments investigating how listeners recognize reduced word pronunciation variants can only yield ecologically valid results if they present the variants in their context.

Listeners can base their predictions of omitted reduced words on the preceding context as well as on the following context. Van de Ven, Ernestus and Schreuder (2012) showed that participants can better guess the identity of an omitted reduced word if they are presented with both the preceding and following context rather than just the preceding context. The relevant semantic/syntactic information is not restricted to the meanings of directly surrounding words, but may also include the larger (discourse) context (e.g., Nieuwland & Van Berkum, 2006). Furthermore, the context may contain informative acoustic cues, as was also shown by Van de Ven and colleagues. They found that participants better predicted omitted reduced words if they heard rather than read the context.

Context alone, however, is insufficient to recognize reduced variants, as shown by Janse and Ernestus (2011) and van de Ven et al. (2012). Janse and Ernestus (2011) presented participants only with orthographic transcriptions of the preceding and following context of reduced word pronunciation variants, or participants also heard the reduced variants (in a separate experiment; the context was again presented visually). Listeners could not identify most target words on the basis of the written context alone (only 13% of the items were guessed correctly by at least a third of the participants), but the auditory presentation of the target words significantly increased participants' performance (90% correct). Apparently, context only becomes highly informative once listeners have heard the reduced variants. This raises the question, which acoustic information from the reduced variants is, above all, informative.

Many studies suggest that even if listeners hear reduced words in their natural context, their recognition is slower than the recognition of well-articulated words. Nearly all these studies present reduced variants in isolation (e.g., Ernestus & Baayen, 2007; Ranbom & Connine, 2007; Tucker, 2011; Tucker & Warner, 2007; van de Ven, Tucker, & Ernestus, 2011). There are, however, two clear exceptions. Results obtained by Brouwer, Mitterer, and Huettig (2012) came from several eye-tracking experiments in which participants heard fragments of conversational speech and saw orthographic representations of words on a computer screen (i.e., the printed words version of the visual world paradigm). Participants were instructed to click on the printed word that matched a word in the fragment; if they did not hear any of the words on the screen (which was the case for all target trials) they

had to click in the middle of the screen. The results suggest the recognition of reduced pronunciation variants is inhibited compared to the recognition of unreduced variants. These findings are unexpected since everyday conversations are full of reduced words. This raises the question whether the printed words version of the visual world paradigm can be used for investigating the comprehension of reduced words. The words' orthographic forms represent their full pronunciations, and participants may therefore expect these pronunciations. As a consequence, they may recognize words more slowly when they are realized as reduced variants. Further, presenting orthographic information while listeners hear (casual) speech also leads to questions concerning ecological validity because listeners are normally not presented with orthographic transcriptions of what they will hear.

An EEG study by Drijvers, Mulder, and Ernestus (2016) shows that gamma oscillations only increase when listeners hear reduced rather than unreduced word pronunciation variants in mid-sentence positions. The authors interpret this result as suggesting that it is more difficult for listeners to activate the semantic network when hearing reduced instead of unreduced pronunciation variants (in line with van de Ven et al., 2011). The target words were presented in read-aloud sentences, and were cross-spliced. The effect of reduction might have been absent if the reduced words had been presented in their natural contexts.

The present study aims at contributing to the understanding of how listeners identify reduced words in their natural contexts. We do so by focusing on three questions. First of all, we investigated which segments are used by listeners to recognize reduced word pronunciation variants. Second, we assessed to what extent word token duration contributes to the recognition of reduced pronunciation variants. The third question of our study was whether the gating paradigm (Grosjean, 1980) can be adapted for studying how listeners understand reduced pronunciation variants in their context.

The present study focused on the recognition of reduced words with unstressed initial syllables, which are likely to be reduced (e.g., the Dutch verb form *verlaten* [fər'latən]² “leave” may be realized like [ˈflatə]). Since this reduction is located (far) before the word's uniqueness point, it may increase uncertainty about the word's identity during the word recognition process. For example, the reduced variant ([ˈflatə]) of the Dutch word *verlaten* is initially very similar to the Dutch word *flater* “blunder,” which may be realized like [ˈflatə]. One may therefore predict that listeners are better at recognizing words with unreduced rather than reduced initial syllables, in line with the literature showing that reductions hinder comprehension (see above). On the other hand, however, if the first unstressed vowel is missing, listeners hear more segments known to be especially relevant for word recognition: they hear subsequent consonants and the stressed vowel earlier. The absence of the first vowel may consequently increase the relevance of the following segments as cues to recognize reduced pronunciation variants.

Segments may be completely absent or may leave acoustic traces that listeners can pick up on. For example, Manuel (1992) showed that listeners can distinguish between English “*sport*” and “*support*” pronounced without the schwa, based on the duration of the aspiration of the following /p/.³ Another example is provided by Zimmerer and Reetz (2014), who found that, in German, if the word-final /t/ is missing in final /st/ clusters, the duration of the preceding /s/ tends to be longer (Zimmerer, Scharinger, & Reetz, 2011, 2014), and listeners use this subsegmental cue to reconstruct the missing /t/. Likewise, the absence of the initial vowel in reduced pronunciation variants may leave beneficial cues for the listener. This would be another reason why listeners may not be hindered by such reductions, and these reductions may even enhance the recognition of these variants. In fact, reduction in these cases actually leads to more information in the same stretch of time.

Another potential cue for the word's identity is its duration. Listeners may use the duration of a reduced word, relative to the durations of (segments in) surrounding words (to estimate speech rate; Nootboom & Doodeman, 1980), to deduce its number of syllables/segments. If the listener is (unconsciously) aware of the possible pronunciation variants of a word (for instance, because

some of them are lexically stored, e.g., Ranbom & Connine, 2007), the duration of the word may thus form a cue to the intended word. Previous research has shown that listeners take word duration into account and that they build expectations that even influence the number of words and word boundaries they perceive (e.g., Dilley & Pitt, 2010). Because of all these cues, listeners may not be hindered by reductions in initial syllables in words presented in context, in contrast to what has been found so far for words presented in isolation, or in experiments that are not ecologically valid for some other reason.

We tested the recognition of reduced word pronunciation variants in a gating task. In a typical gating task, participants hear incremental portions of a target word (i.e., the gates), and for each gate (usually 50 ms longer than the previous one) they need to identify the target. Using this technique, Grosjean (1980) has shown that listeners can recognize carefully pronounced words already before their acoustic offsets and, in many cases, even before their uniqueness points. Furthermore, when these words are embedded in context, listeners need even less acoustic information.

We expect that the gating task is highly suitable for investigating the processing of spontaneous speech. Although some authors criticized the gating paradigm for not being a true on-line paradigm, Tyler and Wessels (1985) showed that this paradigm is equally sensitive to the real-time processes involved in spoken word recognition as other on-line paradigms. Moreover, Bruno, Manis, Keating, Sperling, Nakamoto, and Seidenberg (2007) suggested that the gating task is highly suitable for measuring phonological processing because it is independent of a phonemic level of representation, as is the case in other tasks (e.g., categorization or phonological awareness tasks). Further, the task can indicate how much acoustic information is required to recognize a word (e.g., Grosjean, 1996).

We are not the first to use the gating paradigm with spontaneous speech instead of connected, laboratory speech. Bard, Shillcock, and Altmann (1988) presented participants with utterances extracted from a corpus of spontaneous speech that were gated in increments of one word. They found that for 21% of the words listeners did not only need the preceding context and the word itself, but also the following context to recognize the word. Apparently, listeners also need the following context to recognize words when they are presented within spontaneous rather than laboratory speech (see also van de Ven et al., 2012, discussed above). The findings of Bard et al. may (partly) be due to the frequent occurrence of reductions in spontaneous speech.

We created a version of the gating paradigm where the gates are aligned with the edges of consonant clusters or vowels. This approach is highly suitable for studying the contributions of the different segments in the word to the recognition of reduced pronunciation variants because we could control the segments participants heard in each gate. We placed gate boundaries (1) at word onset; (2) at the end of the first realized consonant (cluster); (3) at the end of the first realized vowel; and (4) after the second realized consonant (cluster; see Cutler & Otake (1999) for a similar approach, using the gating paradigm to study the role of pitch-accent information in spoken word recognition). Note that gate 2 may not only contain more segments if the initial unstressed vowel is absent but may also be longer. We address this multicollinearity with statistical modeling, as we explain in Experiment 1.

We report three auditory gating experiments, in Dutch. Listeners were presented with the natural preceding and following context (since both are relevant, see Bard et al., 1988) of reduced word pronunciation variants (henceforth “target words”), and some acoustic information from these variants themselves (except for the baseline condition). The materials were extracted from a corpus of spontaneous speech.

In Experiment 1, we investigated the role of the first realized consonant or consonant cluster (henceforth “consonant cluster,” for the sake of convenience). The experiment consisted of two parts. In part one (gate 1), participants heard the preceding and following contexts, separated by a

square wave. In the second part of the experiment (gate 2), participants heard the preceding contexts and the initial consonant clusters of the target words, followed by a square wave and the following contexts. Each part contained half of the target sentences, and each sentence only occurred once throughout the experiment (the same holds for subsequent experiments reported in this study).

The initial consonant cluster of a given target word consisted of only the onset consonants from the citation form if the first unstressed vowel was present, whereas it consisted also of consonants from the coda and/or the onset of the following stressed syllable from the citation form if the first unstressed vowel was absent (henceforth “merged clusters”). For example, the Dutch word *principe* “principle” with the citation form [pɾɪn'sipə] was realized like [pə'sipə] in one token and like [psipə] in a different token from the experiment, and the participants in the second half of Experiment 1 either heard the segments [p] or [ps] of these target words, depending on which pronunciation variant they heard (each token was only presented once throughout the experiment). Almost half of the target words contained merged initial clusters. As illustrated in the example (where [r] appears missing), in many initial unstressed syllables with reduced vowels, consonants were also reduced.

This experimental design allowed us to make two comparisons. First, we could compare the conditions with and without the initial consonant cluster (gate 1 vs. gate 2), which would show the contribution of this consonant cluster to the recognition of reduced pronunciation variants. Second, we could compare tokens with simple initial consonant clusters (e.g., [p] from [pə'sipə]) to tokens with merged initial consonant clusters (e.g., [ps] from [psipə]), which allowed us to investigate the effects of missing vowels on the word recognition process.

In Experiment 2, we investigated whether listeners can make use of word duration as a cue to word identity. This experiment was identical to Experiment 1, except that the duration of the square wave (combined with the duration of the initial consonants in the second half of the experiment) now equaled the duration of the target word.

Finally, Experiment 3 investigated the role of the consonants and vowels from the second, stressed syllable in the recognition of reduced target words. This experiment also consisted of two parts, and the duration of the square wave was fixed. In part one (gate 3), participants heard the context, and the reduced target word up to and including the first vowel. This vowel was either the vowel from the first, unstressed syllable (e.g., the first schwa in [pə'sipə]) or the vowel from the second, stressed syllable (e.g., [i] in [psipə]). This part allowed us to compare the contribution of the vowel and consonants from the unstressed initial syllable with the contribution of the initial consonants and stressed vowel in the absence of the unstressed vowel.

In part two (gate 4), listeners heard the context and the target words up to and including the consonant cluster immediately following the first vowel. For example, for the Dutch word *principe* “principle” listeners heard [pə's] and [psip] for the realizations [pə'sipə] and [psipə], respectively. This part shows to what extent hearing these additional consonants influences participants' performance.

For all experiments, we also investigated how the acoustic information from reduced realizations of words interacts with the contextual predictabilities of these words given their context. Van de Ven et al. (2012) observed that contextual predictability as indicated by word trigram frequency becomes less important when more acoustic cues are present. We hypothesize that the contribution of contextual predictability becomes smaller if a larger portion of the reduced word is presented.

In short, we report a series of experiments using an adapted version of the gating paradigm that allows us to investigate the contribution of segmental and durational information to the recognition of reduced pronunciation variants in their natural context (rather than in clearly articulated laboratory speech). We compared reduced pronunciation variants with and without the initial unstressed vowel being present. The segmental information and the average durations of the segment

Table 1. An overview of the segments provided in gates 1–4 (top line), and their average durations (bottom two lines), for tokens in which the first unstressed vowel was acoustically present or absent (exemplified by two tokens for the target word *principe*: [pə'sipə] and [p'sipə]).

Stimulus type	Gate 1 Baseline	Gate 2 C (ms)	Gate 3 CV (ms)	Gate 4 CVC (ms)
Vowel present	∅	74.33 ([p])	109.31 ([pə])	176.37 ([pə's])
Vowel absent	∅	128.84 ([p's])**	209.48 ([p'si])**	271.64 ([p'sip])**

C, consonant cluster; V, vowel. Significance values were obtained by applying *t*-tests comparing the durations of the target words with and without first vowel presence. * = $p < 0.05$ ** = $p < 0.01$ *** = $p < 0.001$.

sequences provided in gates 1–4 for tokens with and without the first unstressed vowel being present are summarized in Table 1.

2 Materials and methods

2.1 Experiment 1

2.1.1 Participants. Twenty native speakers of Dutch were paid to take part in the experiment. They did not report any hearing loss, and most of them were undergraduate students (the same holds for all subsequent experiments).

2.1.2 Materials. The materials were extracted from the Ernestus Corpus of Spontaneous Dutch (Ernestus, 2000), which consists of casual conversations between 10 pairs of Dutch native speakers, recorded in a soundproof booth. We selected as our target stimuli 38 high-frequency multisyllabic Dutch word types with unstressed initial syllables, all starting with a consonant in their citation form. Many of these word types were content words, or they at least contributed substantially to the meaning of the utterance. In addition, we selected 20 different Dutch word types, including words with word-initial stress and monosyllabic words, as filler items, to introduce more variation in the experiment.

For each target word type, we selected two tokens on average (one token for 23 word types, two tokens for nine word types, three tokens for two word types, and four tokens for four word types). The stimuli were produced by 20 different speakers in total; the distribution of tokens across speakers is shown in Table 2.

If the first unstressed vowel was present, all consonants in the initial (but not the coda) consonant cluster were nearly always present, too. We tried to select as many tokens with simple as with merged initial consonant clusters (i.e., clusters consisting of more than the onset consonants from the full forms). Since, for most word types, we could not find a token with a simple cluster and a token with a merged cluster, we varied these two cluster types across (rather than within) word types. Further, we selected 1.5 tokens for each filler word type on average.

We extracted these tokens embedded in their prosodic phrases (mean preceding context: 5.46 words, range: 2 to 18 words; mean following context: 4.12 words, range: 1 to 15 words). None of the extracted speech fragments contained overlapping speech or loud background noises.

We verified the intelligibility of the resulting 73 possible target and 30 possible filler tokens, embedded in their contexts, in a control experiment, because we only wanted to include tokens that could easily be recognized in context. Following the procedure described in van de Ven et al. (2012), we presented 20 native speakers of Dutch with the full sentence fragment (e.g.,

Table 2. An overview of the distribution of tokens across speakers for the stimuli used in this study.

Number of tokens	Number of speakers
1	2
2	5
3	7
4	3
5	2
8	1

Kan je op verschillende ['fsχɪln] manieren doen. “You can do that in various ways.”), followed by the reduced target word and its two preceding and following words (e.g., *je op verschillende ['fsχɪln] manieren doen.* “you do that in various ways.”). The participants were instructed to orthographically transcribe this shorter fragment (i.e., consisting of five words in total). This experiment (as well as all subsequent experiments reported in the present study) was carried out in a sound-attenuated booth, with E-prime 1.2 (Schneider, Eschman, & Zuccolotto, 2002). The experiment consisted of 20 blocks, and each block contained the materials of one of the 20 speakers from the corpus. The blocks and trials within blocks were randomized across participants. Each block was preceded by a short monologue (on average 21.46 s) by the speaker, which allowed the participants to get used to the speaker’s (voice) characteristics. Further, two filler tokens preceded the target tokens in each block. We found that most, but not all, of our stimuli were relatively easy to understand in their contexts (93.72% correct, range: 16.67%–100% correct).

For the main experiments, we selected those stimuli that were easy to understand in their contexts (more than 75% correct in the control experiment). In total, the main experiments contained 63 target tokens (again representing 38 word types) and 30 fillers, produced by 20 speakers. We include the orthographic transcriptions of the 63 target tokens in the Appendix.

Subsequently, we carried out a second control experiment to assess how easily the filler and target tokens could be recognized in isolation. This experiment was identical to Control Experiment 1, except that the words were presented in isolation. Participants (who did not take part in Control Experiment 1) recognized the target tokens in 69.24% of the trials on average (range: 0%–100%), which indicates that listeners require context to recognize these reduced pronunciation variants, in line with previous research (e.g., Bard et al., 1988; Ernestus et al., 2002; van de Ven et al., 2012).

Two transcribers, naive to the purpose of the experiments, determined which segments were present in the speech signal. They disagreed on the presence/absence of consonants in the first syllable and on the presence/absence of vowels in the first syllable in 12.7% and 15.87% of the target tokens, respectively. Whenever there was a difference between the two transcriptions, a third transcriber (the first author) determined the correct transcription. A phonetic transcription of the materials, which provides insight into the degree of reduction of the target and filler tokens, is provided in the Appendix.

The descriptive statistics for the reduction in the initial consonant cluster are shown in Table 3. Vowels were missing in the initial syllable (and thus in the initial consonant cluster) in 39 target tokens (61.90%), for instance, in [pɾɪn'sipə] realized like ['psipə]. Spectrograms and transcriptions of two tokens of *principe*, one realized with and one without the first unstressed vowel, are provided in Figure 1. Missing vowels lead to phonotactically illegal consonant clusters in 11 target tokens (17.46%, e.g., [vər'kopt] realized like ['fkopt]). We included more tokens with legal than

Table 3. Absolute numbers (and percentages) of target words with different types of reduction in the initial consonant cluster, broken for the phonotactic well-formedness of this cluster.

Missing segments	Phonotactically legal	Phonotactically illegal
None	13 (20.63%)	6 (9.52%)
Vowel only	10 (15.87%)	2 (3.17%)
Consonants only	5 (7.94%)	–
Vowel + consonants	18 (28.57%)	9 (14.29%)

with illegal initial consonant clusters because legal initial consonant clusters may incur higher processing costs for the listener, as suggested by Spinelli and Gros-Balthazard (2007).

We determined the gates for our main experiment based on the locations of the segment boundaries in the phonetic transcription. Whenever there was a difference between these locations set by the two transcribers, the same third transcriber determined the correct boundary location. The average discrepancy between the locations of the segment boundaries of two transcribers equaled 2.11 ms.

Experiment 1 presented two different gates. Gate 1 consisted of only the preceding and following context of the experimental items, separated by a square wave. Gate 2 also contained the initial consonant cluster (which was the merged cluster in 62% of the tokens). Here, nine target tokens contained two consonants (14.29% of all target tokens), 14 target tokens contained three consonants (22.22% of all target tokens), and eight target tokens contained four consonants (12.70% of all target tokens).

Truncated speech sounds highly unnatural and may lead listeners to perceive an inserted labial or plosive consonant (Pols & Schouten, 1978), especially when the truncated speech is followed by silence. This is less the case if the truncated speech is followed by a square wave (Warner, 1998), and in our experiments we therefore used a square wave (rather than silence) to indicate the original location of the target word. We used a 500 Hz square wave, which consisted of an onset of 5 ms with gradually increasing amplitude and 500 ms with a fixed amplitude of 52 dB. The intensity of the sound fragments (without the square wave) was normalized to 70 dB.

As for the control experiments, the experiment consisted of 20 blocks. Each block contained the speech materials of one of the 20 speakers and was preceded by the same familiarization phase as in the control experiments. The blocks and trials within blocks were again randomized across participants, and each speaker block started with two filler tokens. Participants heard the materials of a particular speaker in either gate 1 or gate 2. After 47 of the 93 trials, the current speaker block was completed with gate 1, and the trials of the subsequent speakers were presented with gate 2. As a consequence, part one contained more target tokens than part two (33 vs. 30 target tokens on average).

2.1.3 Procedure. In both parts, participants were instructed to orthographically transcribe the target words while seated in a sound-attenuated booth, and while wearing headphones. The experiment was self-paced.

2.2 Experiment 2

2.2.1 Participants. Twenty native speakers of Dutch were paid to take part in the experiment. These participants did not take part in any of the other experiments.

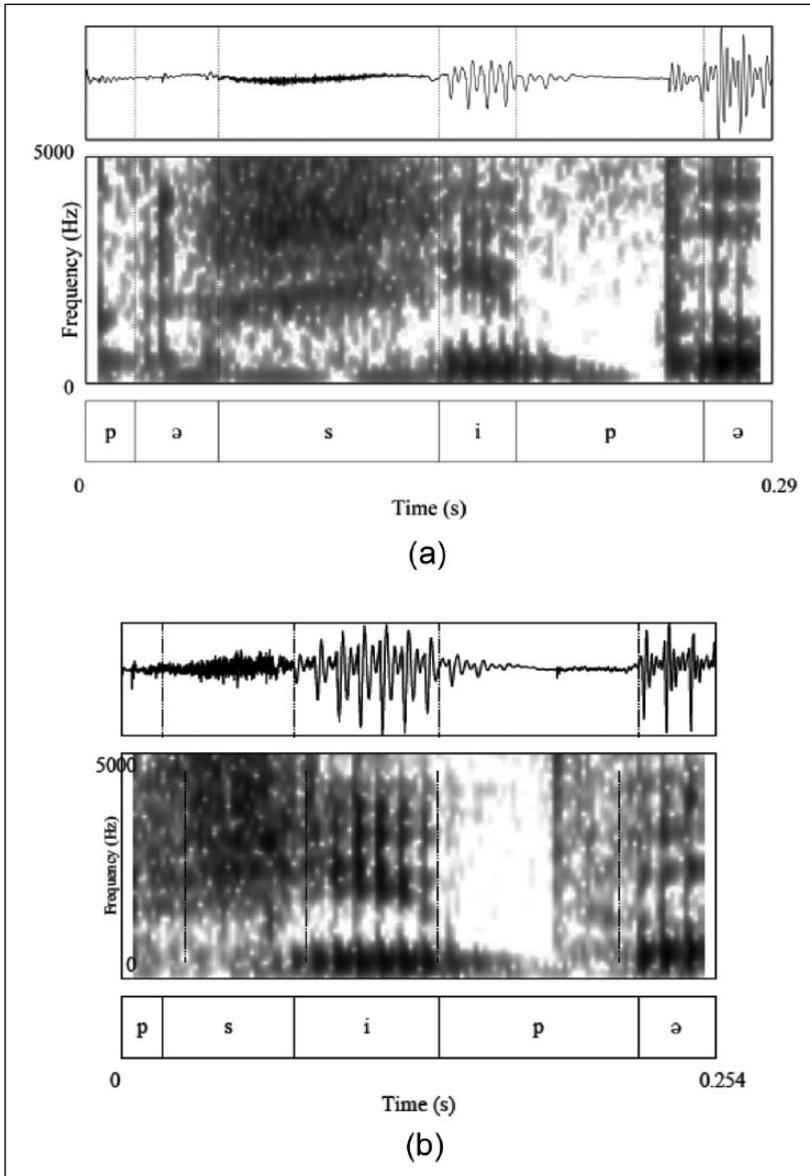


Figure 1. Spectrograms and transcriptions for two tokens of the word *principe* "principle." (a) with the first unstressed vowel; (b) without the first unstressed vowel.

2.2.2 Materials. The materials were identical to those of Experiment 1, except that the duration of the square wave now equaled the duration of the reduced word (in gate 1) or the duration of the word minus the duration of the initial consonant cluster (in gate 2). We used a minimum duration of 20 ms because a pilot experiment indicated that for shorter durations listeners have difficulty locating the square wave. The minimum duration of 20 ms meant that, in gate 2, the combined duration of the square wave and the initial consonant cluster for three fillers was longer than these reduced filler tokens themselves.

2.2.3 Procedure. The experimental procedure was identical to that of Experiment 1, except that participants were now told that the duration of the square wave equaled that of the missing word in gate 1, and of the part that was missing in gate 2.

2.3 Experiment 3

2.3.1 Participants. Twenty native speakers of Dutch were paid to take part in the experiment. These participants did not take part in any of the other experiments.

2.3.2 Materials. Each stimulus used in Experiment 1 was extended to include the first realized vowel for gate 3 and this vowel as well as the second consonant cluster for gate 4. For example, for the target word *principe* “principle” pronounced like [pə'sipə], participants heard [pə] in gate 3 and [pəs] in gate 4. On the other hand, for a different token of this target word, pronounced like ['psipə], participants heard ['psi] in gate 3 and ['psip] in gate 4. This meant that two tokens of the target word *manier* (both realized like ['mni]) were presented in full in both gates 3 and 4 (3.17% of the trials), while an additional 22 target words (38.10%) were presented in full in gate 4 (e.g., ['mir] for the target word *manier* “manner” and ['χat] for the target word *gehad* “had”).

Our phonetic transcriptions showed that 13 target stimuli contained merged second consonant clusters (20.63%). For example, the Dutch word *verschillende* [vər'sχɪləndə] “different” was realized like ['fsχɪln] and in gate 4 participants then also heard the consonants immediately following the second unstressed vowel in the word’s citation form. In 40 target stimuli (63.49%) consonants were missing in the second consonant cluster. For example, the Dutch word *vanzelf* “by itself” [vɑn'zɛlf] was realized like [və'zɛlf] and [n] was missing in [vəz] in gate 4. Since Experiment 2 showed that participants were misled by durational information, we used the same square wave with a fixed duration as in Experiment 1.

2.3.3 Procedure. The experimental procedure was identical to those of Experiments 1 and 2.

3 Results and discussion

3.1 Experiment 1

A transcriber labeled participants’ responses as correct or incorrect.⁴ Participants produced 430 correct and 830 incorrect responses for the target words. Descriptive statistics indicated that listeners experienced difficulty guessing the target words on the basis of just the context (26.02% correct) or the context combined with the first consonant cluster (43.19% correct). Nevertheless, listeners performed better when presented with some acoustic information about the target words (average correctness increased by 17.17%). More acoustic information for these target words was required to correctly identify these words.

We analyzed the correctness of participants’ responses (*correctness*; in the analysis of the present and subsequent experiments) by means of generalized linear mixed-effects regression with the logit link function (Jaeger, 2008), using the statistical package lme4 (Bates, Maechler, Bolker, & Walker, 2015), and starting with the maximal random effects structure (Barr, Levy, Scheepers, & Tily, 2013). We included random effects for *participant*, *target type* (e.g., the Dutch word *principe* or *manier*), and *target token* (e.g., the first or second token of the Dutch word *principe*). We tested the significance of the random intercepts, random slopes, fixed effects and interactions of fixed effects by means of χ -squared tests comparing nested models. For the fixed effects, we also examined the

Table 4. Results for the statistical analysis for Experiment 1.

Predictor	β	z	p	Variance explained	χ^2	p
Fixed effects						
Intercept	3.122	6.21	<.0001			
Gate (gate 2)	-1.978	-6.81	<.0001			
Preceding bigram frequency	-0.004	-2.29	<.05			
Random effects						
Participant				0.178	6.59	<.05
Target token				9.085	519.48	<.0001
Target token: gate				1.953	13.98	<.001

p-values obtained from the model summary, and relied on the most conservative *p*-value if there was a difference between the two. Variables were removed if they did not attain significance at the 5% level. We first determined the fixed-effects structure, and subsequently whether the inclusion of random slopes improved the model fit. In a first analysis, we investigated whether participants gave significantly more correct responses after hearing the initial consonant cluster.

We entered several fixed predictors. Most importantly, we included *gate* (gate 1 vs. gate 2). We also incorporated as predictors the likelihoods of the target words given the preceding and following words in the sentences. We determined the words' bigram frequencies with their preceding or following words on the basis of the Spoken Dutch Corpus (Oostdijk, 2002).⁵

Finally, we included the control variable *trial number* (trial number within each gate), in order to capture effects due to learning or fatigue. The results of the mixed effects model are shown in Table 4.

We found the main effects of *gate* and *preceding bigram frequency*, indicating that participants performed better if they heard the initial part of the target word (gate 2) and if the target word had a higher bigram frequency with the preceding word. The random slope of the variable *gate* for *token type* further indicated that the effect of adding the initial consonant cluster varied significantly across tokens. More specifically, we found that seven target tokens had strong positive slopes ($\beta > 1$) for gate 2 and thus listeners did not benefit much from hearing the first realized consonants and vowels for these target tokens. Four of these tokens also had strong positive intercepts ($\beta > 1$), indicating that these tokens were harder than the other tokens. On the basis of a visual inspection, it appears that the surrounding context of these target tokens allowed a relatively large number of semantically and syntactically legal alternatives, although this did not imply a relatively low bigram frequency of the preceding word with the target word or the target word with the following word. For example, *de manier waarop* "the way in which" contains a highly frequent bigram (*de manier* "the way"), but there are also many alternative answers for *de... waarop* resulting in highly frequent bigrams. On the other hand, there were three target tokens with strong negative slopes ($\beta < -1$), and for these tokens participants apparently benefited greatly from hearing the initial part of the target word in gate 2. Subsequently, we conducted a post hoc analysis to investigate the effects of the presence of the vowel in the initial syllable in gate 2. As mentioned in the Materials section, there was a relation between the presence of the first vowel and the presence of the complete initial consonant cluster; the same holds for segments in subsequent syllables, which is relevant for later experiments. Detailed analyses (for all main experiments reported in this study) showed, however, that the presence of the first unstressed vowel was a better predictor of

the correctness of the responses than was the presence of the first complete consonant cluster of the unstressed syllable. Importantly, the presence of the vowel in the initial syllable affected the duration of the second gate (and subsequent gates; see Experiment 3) in the experiment. To make sure that the effects of vowel presence were not merely due to differences in duration between the stimuli presented with simple and merged consonant clusters, we included *stimulus duration* as a variable in the analysis.

Given that *first vowel presence* and *stimulus duration* were highly correlated, we had to orthogonalize these predictors before we could proceed with the analysis. Since orthogonalization reduces the predictive power of residualized predictors (Wurm & FisiCaro, 2014), the to-be-residualized predictor, which is least predictive, had to be determined in an objective manner. We used the following procedure to determine which of the two predictors was to be residualized (in all models that included both predictors in this paper). First of all, we fitted a separate regression model on participants' responses with *stimulus duration* as predictor and another model with *first vowel presence* as predictor. These models included any other (uncorrelated) significant predictors (in this case *preceding bigram frequency*) and interactions. Importantly, we did this separately for each gate, because *stimulus duration* increased incrementally with each gate, whereas *first vowel presence* was always constant. We then compared the Akaike Information Criteria (AIC) of the two models (i.e., for *stimulus duration* and *preceding vowel presence*) and ranked them accordingly. This ranking determined which predictor was residualized. For all experiments reported in this study, we found a better fit for the model containing *first vowel presence* than for the one containing *stimulus duration*. Hence, for all models that contained both predictors, *stimulus duration* was residualized from *first vowel presence*, and *stimulus duration*_{resid} was used to replace *stimulus duration* in the analyses.

Importantly, we found a significant effect of *first vowel presence*, $t(1,592) = -3.37$, $p < .001$, but not of *stimulus duration*_{resid}. This finding shows that participants benefited especially from hearing the consonants following unstressed vowels in the full forms rather than from hearing just longer stretches of speech.

We also investigated participants' incorrect responses. For this purpose, the same transcriber first marked whether the incorrect response was contextually appropriate, that is, if it could fit within the syntactic structure of the sentence, and whether the resulting sentence made any sense. For example, the response *papier* "paper" was labeled contextually inappropriate for the sentence *Met die slaapzakken ook in het verleden wel problemen gehad eigenlijk*, "With those sleeping bags in the past also had problems actually," because if we replace the target word *problemen* with *papier*, then the sentence becomes semantically uninterpretable.

Further, the transcriber marked the correctness of the word's first segment, second segment (if the first segment was correct), third segment (if the first and second segment were correct), the word-final segment, and the number of syllables. A segment was labeled as correct if its pronunciation matched that of the word's citation form(s). For example, in certain regions of the Netherlands, voiced fricatives are frequently pronounced as voiceless, and therefore if a participant's answer for the target word [vər'kɔpt] *verkoopt* started with an "f," the first segment of this answer was labeled as "correct."

The descriptive statistics for the incorrect responses are provided in Table 5 (the total number of incorrect responses equals 100%). These descriptives suggest that, in the case of an error, participants could better identify the first segments of the word's citation form in gate 2 than in gate 1, which is as expected, since participants heard these segments in gate 2, whereas they did not in gate 1. Moreover, participants could better identify the first segments of the word's citation form in gate 2 if they heard a merged consonant cluster. In both cases, however, listeners could not always recognize the initial consonants or did not always use this segmental information.

Further, participants provided more contextually appropriate responses for target words with simple than with merged consonant clusters, and this difference was smaller in gate 2 than in gate

Table 5. Percentage of contextually appropriate incorrect responses, and percentages of correct first segments, final segments, and number of syllables for the incorrect responses in Experiment 1, broken down by the gate and whether the initial unstressed vowel was realized.

Gate	Vowel realized	Contextually appropriate	1st segment	1st and 2nd segment	1st–3rd segment	Final segment	Number of syllables
Gate 1	No	34.56%	8.05%	1.68%	0.34%	19.13%	11.74%
Gate 1	Yes	59.79%	6.70%	1.03%	0.00%	22.68%	14.43%
Gate 2	No	32.48%	36.31%	14.01%	7.01%	15.29%	12.74%
Gate 2	Yes	41.44%	34.25%	6.08%	2.76%	25.41%	20.44%

1. Possibly, participants had more difficulties understanding the contexts of target tokens with merged consonant clusters, since highly reduced word tokens tend to occur in acoustically reduced contexts. This effect was smaller in gate 2, probably because any effects of reduction in the context become smaller as participants hear more acoustic information from the target words, for instance due to compensation for coarticulation.

Finally, 22.49% of the incorrect responses were semantically and syntactically possible and shared their first segments with those of the reduced target words. Apparently, Dutch allows multiple word candidates on the basis of the context and the word's first segments. For example, for the sentence *Ik had vandaag weer een auto geleend*. "Today I borrowed a car again," two participants answered *vanochtend* "this morning," which shares the initial consonant with the target word and is contextually appropriate. On the other hand, participants produced semantically and/or syntactically incorrect responses in 41.05% of all trials in gate 1 (including the correct trials), and in 30.42% of all trials in gate 2. Participants clearly require more segments than the initial ones to recognize reduced words.

To summarize, our results show that listeners had difficulties guessing the target word on the basis of the context alone or on the basis of the context and the initial consonant cluster. Importantly, performance was better if words started with merged consonant clusters, in gate 2, and this effect was not simply due to longer durations of these consonant clusters compared to simple consonant clusters. This finding indicates that hearing additional consonants outweighs the absence of the first unstressed vowel in word recognition.

Listeners apparently need more information from reduced pronunciation variants than the initial consonants, which could be more segmental information, or perhaps the durations of the words, as mentioned in the Introduction. Experiment 2 investigated whether listeners are able to use word duration to recognize words more easily.

3.2 Experiment 2

A transcriber labeled the responses, using the same criteria as for Experiment 1. Participants produced 335 correct and 925 incorrect responses for the target words (see Table 6).

We fitted a regression model for the combined data set of Experiments 1 and 2, so that we could compare their results. We included the same random and fixed variables as for Experiment 1, in addition to *experiment* (Experiment 1 vs. Experiment 2). The results are provided in Table 7.

Importantly, we found a main effect of *experiment*, indicating that participants who heard the durations of the target words performed worse than those who did not. Durational information thus did not make the task easier, but appeared misleading.

Interestingly, we also found an interaction between *experiment* and *preceding bigram frequency*. This result indicates that the effect of *preceding bigram frequency* was restricted to Experiment 1. Apparently, participants focused less on context when also provided with durational information.

Table 6. The percentages correct for target words in Experiment 2, broken down by whether the initial unstressed vowel was realized, and by gate.

Gate	Vowel realized	Vowel absent
Gate 1	18.99%	18.91%
Gate 2	25.66%	45.02%

Table 7. Results for the statistical analysis comparing Experiments 1 and 2.

Predictor	β	z	p	Variance explained	χ^2	P
Fixed effects						
Intercept	3.884	6.66	<.0001			
Experiment (experiment 2)	0.561	2.51	<.05			
Gate (gate 2)	-2.288	-7.51	<.0001			
Preceding bigram frequency	-0.004	-1.97	<.05			
Experiment (experiment 2): preceding bigram frequency	0.001	2.32	<.05			
Random effects						
Participant				0.244	20.02	<.0001
Target type				3.835	4.87	<.05
Target token				6.941	336.79	<.0001
Target token: gate				3.081	67.57	<.0001

Subsequently, we fitted an additional model for gate 2 of both experiments, in which we included *first vowel presence* and *stimulus duration_{resid}* as predictors, following the same procedure as for Experiment 1.

Importantly, we found an effect of *first vowel presence* $t(1,1186) = -2.05, p < .05$, yet no effect of *stimulus duration_{resid}*. Thus, this effect of *first vowel presence* was not simply due to durational differences between merged and simple initial consonant clusters.

Why did participants perform worse if they were provided with additional, durational information to rely on? Listeners appear generally to be unaware of the reductions that occur in spontaneous speech (e.g., Kemps, Ernestus, Schreuder, & Baayen, 2004) and consequently participants may have tried to match the durations of the square waves to the durations of words' citation forms. Since the target words were all segmentally and durationally reduced, participants may consequently have preferred candidates that are shorter than the citation forms of the target words.

We converted participants' orthographic responses into phoneme sequences, and subsequently compared the lengths of the responses in phonemes (henceforth *response length*) in Experiments 1 and 2. Long vowels were counted as one phoneme (a decision that did not affect the outcome of the comparison). Since the lengths were not distributed normally, we converted *response length* into a binary variable by applying a median split: We labeled responses as "long" if they contained more than five phonemes; otherwise we labeled them as "short." We then fitted a generalized linear mixed-effects regression model with the logit link function for the dependent variable *response length*, including the same random variables as for the analysis of the correctness of the responses. We found a significant main effect of *experiment* $\beta = 0.477, F(1, 2518) = 8.20, p < .01$, indicating that participants provided shorter responses in Experiment 2 than in Experiment 1 (4.92 vs. 5.31 phonemes on average).

Table 8. Percentage of contextually appropriate incorrect responses, and percentages of correct first segments, final segments, and number of syllables for the incorrect responses in Experiment 2, broken down by the gate and whether the initial unstressed vowel was realized.

Gate	Vowel realized	Contextually appropriate	1st segment	1st and 2nd segment	1st–3rd segment	Final segment	Number of syllables
Gate 1	No	25.90%	7.19%	2.88%	2.16%	18.71%	8.99%
Gate 1	Yes	39.85%	8.81%	2.30%	0.00%	16.86%	17.24%
Gate 2	No	28.24%	30.59%	12.94%	11.18%	19.41%	8.82%
Gate 2	Yes	43.98%	27.78%	6.48%	1.85%	23.15%	15.28%

The descriptive statistics for the incorrect responses are provided in Table 8 (the total number of incorrect responses equals 100%). These descriptives were largely similar to those of Experiment 1. One important difference may be noted, however. The difference between simple and merged consonant clusters in terms of participants' recognition of the first segments in incorrect responses appeared smaller than in Experiment 1.

To conclude, listeners are misled by the durational information from reduced pronunciation variants if this durational information is provided separately from other acoustic information. These results are in line with the hypothesis that listeners are unaware of the reductions in spontaneous speech, and therefore cannot use word duration by itself to recognize these reduced pronunciation variants.

So far, we have established the contribution of the initial consonant cluster and of word duration to the recognition of reduced pronunciation variants. In Experiment 3, we investigated the contributions of the first realized vowel and of the subsequent consonant or consonant cluster (henceforth the "second consonant cluster,") to the recognition of these variants.

3.3 Experiment 3

Participants produced 745 correct responses and 515 incorrect responses for the target words (see Table 9). First, we investigated the contribution of the first vowel to the recognition of reduced pronunciation variants by comparing the results for gate 2 from Experiment 1 to those for gate 3 from Experiment 3, with a regression model, including the dependent variable *correctness* and the same fixed and random effects as for Experiment 1. The results are provided in Table 10.

Importantly, we found a main effect of *first vowel presence* and two-way interactions between *preceding bigram frequency* and *gate* and between *preceding bigram frequency* and *first vowel presence*. These interactions indicated that listeners better recognized target words if they heard the vowel in addition to the initial consonant (i.e., in gate 3), and if they heard a merged consonant cluster (in both gates) and a vowel from the stressed rather than the unstressed syllable (in gate 3), and these effects were larger for target words with low bigram frequencies with their preceding words. There was no effect of *stimulus duration*_{resid.}

The random slope of the factor *gate* for *target token* indicated that the main effect of *gate*, which had a beta estimate of -1.20 , was reversed or completely absent for 13 target tokens (20.63%); thus, for these target tokens, the beta estimates were maximally 1.95 (range of beta estimates for these tokens: 1.21 to 3.15).

Subsequently, we fitted a regression model for the complete data set of Experiment 3 in order to determine the effect of the consonant cluster following the first vowel in the stimulus. We included the same random and fixed variables as for Experiment 1, in addition to the fixed variable *complete*

Table 9. The percentages correct for tokens in which listeners either heard the unstressed or stressed vowel from the target words in gates 3 and 4.

Gate	Unstressed vowel	Stressed vowel
Gate 3	40.19%	61.03%
Gate 4	55.59%	81.10%

Table 10. Statistical results for gate 2 and gate 3 in Experiments 1 and 3 respectively.

Predictor	β	z	p	Variance explained	χ^2	P
Fixed effects						
Intercept	2.530	4.46	<.001			
Gate (gate 3)	-1.199	-3.20	<.01 ⁶			
Preceding bigram frequency	-0.008	-3.23	<.01 ¹			
First vowel presence (no.)	-2.308	-3.70	<.001			
Gate (gate 3): preceding bigram frequency	0.004	2.93	<.01			
Preceding bigram frequency: first vowel presence (no.)	0.006	2.43	<.05			
Random effects						
Participant				0.209	8.43	<.01
Target token				7.173	339.54	<.0001
Target token: gate				3.071	39.74	<.0001

auditory form (whether the reduced target word was presented in full). Given that *complete auditory form* was correlated with *gate* (participants heard more acoustic information in gate 4 than in gate 3), and neither of the two was numeric, two separate models were fitted containing only one of these two predictors. Subsequently, we selected the best model on the basis of the AIC and the Bayesian Information Criterion (BIC). This model comparison showed a substantially better fit for the model containing *gate* (AIC difference = 58.60, BIC difference = 58.60).

The results of the analysis are provided in Table 10. We found a three-way interaction between *gate*, *first vowel presence*, and *preceding bigram frequency*. In addition, the random slope of the factor *gate* for *target token* showed that the main effect of *gate* was absent for one target token (i.e., the slope estimate for this token equaled 1.47), and the random slope of the factor *first vowel presence* for *participant* indicated that the main effect of *first vowel presence* was, despite a significant amount of variation, not completely absent for any of the participants. In order to interpret the three-way interaction, we split the data by *gate*. Post hoc analyses (for the two gates separately) revealed an interaction between *first vowel presence* and *preceding bigram frequency* for gate 3 $t(1,659) = 2.44, p < .05$, yet no such interaction for gate 4 (only main effects of *first vowel presence* and *preceding bigram frequency*).

Hence, while participants' benefit from hearing a merged consonant cluster in the absence of the first vowel depended on the target words' bigram frequencies with their preceding words (gate 3), there was no such dependency in gate 4. Since we did not find any effects of *stimulus duration_{resid}*, this main effect of *first vowel presence* cannot purely be attributed to the durations of the gates.

Finally, we investigated participants' incorrect responses in Experiment 3. The descriptive statistics are provided in Table 11. The results show that listeners often did not recognize the first

Table 11. Statistical results for Experiment 3.

Predictor	β	z	p	Variance explained	χ^2	P
Fixed effects						
Intercept	1.280	2.55	<.01			
Gate (gate 4)	-1.409	-3.89	<.001			
First vowel presence (no.)	-2.368	-3.61	<.001			
Preceding bigram frequency	-0.004	-1.94	n.s.			
Gate (gate 4): first vowel presence	0.781	1.56	n.s.			
Gate (gate 4): preceding bigram frequency	0.002	0.93	n.s.			
First vowel presence (no.): preceding bigram frequency	0.006	2.45	<.05			
Gate (gate 4): first vowel presence (no.): preceding bigram frequency	-0.006	-2.63	<.01			
Random effects						
Participant				0.015	8.45	<.01
Participant: first vowel presence				0.294	8.27	<.05
Target token				3.740	229.90	<.0001
Target token: gate				1.198	13.69	<.01

three segments of the reduced target word at all. Further, unlike the incorrect answers for Experiments 1 and 2, participants' incorrect responses more frequently contained the correct initial segment if the first unstressed vowel was present. Apparently, listeners could identify more segments after hearing more following segments, but they nevertheless could not identify the target words in these cases (Table 12).

To summarize, listeners better recognized target words if they heard the vowel from the stressed syllable and the first, unstressed vowel was missing than if they heard the vowel from the initial, unstressed syllable. This suggests again that the possibly disturbing absence of a vowel may be compensated for by information from the stressed vowel and additional consonants becoming more readily available. In addition, phonetic residues from the unstressed vowel may play a role. This effect was larger for words with low bigram frequencies with their preceding words. Finally, we found that the role of bigram frequency information decreased as listeners heard more segments of the target words.

4 Discussion

Listeners need both the context and acoustic information from reduced word pronunciation variants to recognize these variants (e.g., Janse & Ernestus, 2011; van de Ven et al., 2012). The present study investigates which types of acoustic information listeners rely on most. We addressed three questions, namely: (1) which segments are especially important for listeners to recognize reduced word pronunciation variants; (2) what is the contribution of word token duration to the recognition of reduced pronunciation variants; and (3) whether the gating paradigm (Grosjean, 1980) can be adapted for studying how listeners understand reduced pronunciation variants in their context. We focused on target words with reduced unstressed initial syllables because missing vowels in the initial syllables are likely to create ambiguity and increase uncertainty during the recognition process (e.g., the Dutch words *verlaten* "leave" and *flater* "blunder" with the citation forms [fər'latən]⁶ and [flatər] may both sound like [flatə]).

Table 12. Percentage of contextually appropriate incorrect responses, and percentages of correct first segments, final segments, and number of syllables for the incorrect responses in Experiment 3, broken down by the gate and whether the initial unstressed vowel was realized.

Gate	Vowel realized	Contextually appropriate	1st segment	1st and 2nd segment	1st–3rd segment	Final segment	Number of syllables
Gate 3	No	14.53%	20.51%	9.40%	4.27%	21.37%	13.68%
Gate 3	Yes	48.08%	35.10%	11.54%	7.21%	21.15%	22.12%
Gate 4	No	18.87%	13.21%	13.21%	11.32%	30.19%	15.09%
Gate 4	Yes	43.07%	32.12%	14.60%	8.76%	25.55%	18.98%

In an adapted version of the gating paradigm, participants heard fragments of spontaneous speech always consisting of the context preceding the reduced target word, some segments of this target word (except for the baseline condition, gate 1, in which listeners heard only the context), a square wave, and the following context. By aligning the gates with the boundaries of consonant clusters (rather than using gates with fixed durations), we controlled the types of segments that participants heard in each gate. This allowed us to investigate the role of these segments. Importantly, by comparing simple and merged consonant clusters we could investigate the role of the first unstressed vowel in the recognition of reduced pronunciation variants. Merged consonant clusters contained more segments and could contain subphonemic cues signaling the missing vowels. Hence, the question arises whether listeners are hindered (or, on the contrary, aided) by the absence of the initial unstressed vowel, if we take into account the durational differences.

Each participant heard two out of four gates. They only heard the context in gate 1. In addition to the context, they heard the initial consonant cluster of the target word in gate 2, the initial consonant cluster and the first realized vowel of the target word in gate 3, and the initial consonant cluster, the first realized vowel, and the second consonant cluster of the target word in gate 4.

We found that participants' performance improved with every gate (percentages correct for gate 1: 26.02%; gate 2: 43.19%; gate 3: 51.13%; gate 4: 68.07%). Importantly, the performance for gates 2–4 was higher for merged than for simple consonant clusters. This shows that the full presence of unstressed vowel is less important than the presence of additional consonants. This result may partially be explained by research indicating that, at least in carefully pronounced speech, consonants play a larger role in word recognition than vowels (e.g., Bontatti, Peña, Nespór, & Mehler, 2005; Cutler, Sebastián-Gallés, Soler-Vilageliu, & van Ooijen, 2000; Mehler, Peña, Nespór, & Bonatti, 2006).

Importantly, our study is the first to indicate that reductions may actually benefit the listener. This result contrasts with previous findings suggesting that reductions inhibit word recognition (e.g., Ernestus & Baayen, 2007; Ranbom & Connine, 2007; Tucker, 2011; Tucker & Warner, 2007; van de Ven et al., 2011), lead to relatively high cognitive demands (Drijvers et al., 2016), and delay spreading of activation to semantically related words (e.g., Drijvers et al., 2016; van de Ven et al., 2011). These previous findings nearly all come from experiments testing listeners' comprehension of reduction in read-aloud isolated words or in words embedded in short (e.g., Ernestus & Baayen, 2007; Ranbom & Connine, 2007; Tucker, 2011; Tucker & Warner, 2007; van de Ven et al., 2011) or more elaborate (Drijvers et al., 2016) read-aloud sentences. This may explain these divergent findings, especially since previous research has shown the importance of natural contexts (e.g., Ernestus et al., 2002; Janse & Ernestus, 2011). Only Brouwer et al. (2012) tested the comprehension of reduced words in their natural contexts, as in the present study. They used a printed words version of the visual world paradigm, which may have activated the words' citation forms. Possibly,

these orthographic representations are responsible for the inhibition that these authors found for reduced forms.

The present study also investigated the role of durational information in the recognition of reduced words. In Experiment 2, the duration of the square wave (gate 1), or its duration combined with the duration of the initial consonant cluster (gate 2), equaled that of the reduced target word. Surprisingly, listeners found this durational information misleading, and they made more errors and gave shorter words as responses in Experiment 2 than in Experiment 1, where the duration of the square wave was fixed. In line with Kemps et al. (2004), this finding shows that listeners are unaware of the reductions that occur in spontaneous speech, and, because the target words were short, they therefore expected them to contain few segments in their citation forms.

In all three experiments, we tested the contribution of local semantic/syntactic contextual information, operationalized as bigram frequencies, to the recognition of the reduced target words. Theoretical models of word recognition predict that listeners can use contextual information to narrow down their lexical search space (e.g., van Berkum, Brown, Zwitserlood, Kooijman, & Hagoort (2005) or enhance semantic integration (van Petten & Kutas, 1990)). We found a gradually decreasing effect of preceding bigram frequency as a function of how much participants heard of the target words. This finding shows that listeners rely less heavily on probabilistic information based on the context to recognize these reduced variants if more acoustic information from the word is available, even for a reduced word (with e.g., shorter segment durations, spectral reduction). Hence, reduced segmental information seems to outweigh probabilistic contextual information in recognizing reduced pronunciation variants, in line with van de Ven et al. (2012).

These results are expected. They show that listeners predominantly rely on their acoustic input. Contextual information mostly facilitates the word recognition process. It only determines the outcome if insufficient acoustic information is available. If contextual information played a larger role, listeners would not be able to understand unexpected words/information.

We obtained these results by operationalizing local semantic/syntactic contextual information as bigram frequencies. We could have operationalized contextual probability differently, for instance by means of a visual cloze task. We believe that a different operationalization would have produced the same result because bigram frequency well reflects semantic/syntactic contextual information and because the result reflects the fact that listeners are able to understand unexpected information.

Participants' incorrect responses also provide information about the recognition process. These responses mainly show that, when provided with just the initial consonant cluster, participants could better identify the segments of the cluster when it was merged, as a result of vowel reduction, than when it was a simple cluster. However, since merged clusters also typically contained more segments and were therefore probably more noticeable, it is difficult to draw strong conclusions based on this finding. Moreover, participants could frequently come up with contextually appropriate alternatives for our target words with the same initial segments, which testifies the importance of hearing the complete realization of reduced pronunciation variants.

Finally, this study demonstrates that the gating paradigm (Grosjean, 1980), designed for studying the comprehension of laboratory speech, can also be used for studying the comprehension of highly reduced pronunciation variants in conversational speech. In our version of the gating paradigm, we placed gates at the end of segment boundaries, thereby controlling for the number of vowels and consonant clusters listeners heard. Since we statistically controlled for the confound between vowel reduction and gate duration, we could use the gating paradigm to study the influence of vowel reduction on the recognition of reduced words.

Preferably, future studies follow up on our study in order to investigate whether the same results are also found with different experimental paradigms. This holds for all studies using only one

experimental paradigm. Furthermore, one disadvantage of our version of the gating experiment is that (part of) the target word is replaced by noise, which decreases the task's ecological validity.

5 Conclusions

The present study shows that the gating paradigm can be effectively adapted to investigate the effects of initial vowel reduction on the recognition of reduced pronunciation variants embedded in natural contexts. The results show that acoustic cues in reduced words override probabilistic cues based on preceding context, and that reductions may enhance word recognition if this means that subsequent segments from the stressed syllable become more readily available.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was funded by a European Young Investigator Award from the European Science Foundation and by an ERC starting grant (284108), both to the second author.

Notes

1. Descriptive differences between the amount of reduction found by Johnson (2004) and Schuppler et al. (2011) are probably due to the different procedures used to elicit and measure natural speech.
2. Many speakers of Northern Standard Dutch pronounce word initial /v/ as /f/ (e.g., van de Velde, Gerritsen, & van Hout, 1996)
3. Note that this finding does not generalize to languages, including Dutch, where the duration of the aspiration hardly depends on whether or not the /p/ is the only consonant in the onset of a stressed syllable.
4. Ten percent of the trials of all experiments were also labeled by a second transcriber; the interrater reliability was >95% for all measures.
5. We measured the contextual probabilities of the target words in our study by means of bigram frequencies, instead of trigram frequencies (used in a previous study by van de Ven et al., 2012), because trigram frequencies resulted in a larger number of zero counts, and therefore minimal discriminatory value.
6. This effect was not significant in the analysis of variance results.

References

- Bard, E. G., Shillcock, R. C., & Altmann, G. T. M. (1988). The recognition of words after their acoustic offsets in spontaneous speech: Effects of subsequent context. *Perception & Psychophysics*, *44*, 395–408.
- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*, 255–278.
- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*, 1–48.
- Bonatti, L. L., Peña, M., Nespor, M., & Mehler, J. (2005). Linguistic constraints on statistical computations: The role of consonants and vowels in continuous speech processing. *Psychological Science*, *16*, 451–459.
- Brouwer, S., Mitterer, H., & Huettig, F. (2012). Can hearing puter activate pupil? Phonological competition and the processing of reduced spoken words in spontaneous conversations. *The Quarterly Journal of Experimental Psychology*, *65*, 2193–2220.
- Bruno, J. L., Manis, F. R., Keating, P., Sperling, A. J., Nakamoto, J., & Seidenberg, M. S. (2007). Auditory word identification in dyslexic and normally achieving readers. *Journal of Experimental Child Psychology*, *97*, 183–204.
- Cutler, A., & Otake, T. (1999). Pitch accent in spoken-word recognition in Japanese. *The Journal of the Acoustical Society of America*, *105*, 1877–1888.

- 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.
- Dilley, L. C., & Pitt, M. A. (2010). Altering context speech rate can cause words to appear or disappear. *Psychological Science*, 21, 1664–1670.
- Drijvers, L., Mulder, K., & Ernestus, M. (2016). Alpha and gamma band oscillations index differential processing of acoustically reduced and full forms. *Brain and Language*, 153–154, 27–37.
- Ernestus, M. (2000). *Voice assimilation and segment reduction in casual Dutch, a corpus-based study of the phonology-phonetics interface*. Utrecht, Netherlands: LOT.
- Ernestus, M., & Baayen, H. (2007). Paradigmatic effects in auditory word recognition: The case of alternating voice in Dutch. *Language and Cognitive Processes*, 22, 1–24.
- Ernestus, M., Baayen, H., & Schreuder, R. (2002). The recognition of reduced word forms. *Brain and Language*, 81, 162–173.
- Ernestus, M., & Warner, N. (2011). An introduction to reduced pronunciation variants. *Journal of Phonetics*, 39, 253–260.
- Grosjean, F. (1980). Spoken word recognition processes and the gating paradigm. *Perception & Psychophysics*, 28, 267–283.
- Grosjean, F. (1996). Gating. *Language and cognitive processes*, 11, 597–604.
- Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, 59, 434–446.
- Janse, E., & Ernestus, M. (2011). The roles of bottom-up and top-down information in the recognition of reduced speech: Evidence from listeners with normal and impaired hearing. *Journal of Phonetics*, 39, 330–343.
- Johnson, K. (2004). Massive reduction in conversational American English. In K. Yoneyama & K. Maekawa (Eds.), *Proceedings of the first session of the 10th international symposium on spontaneous speech: Data and analysis* (pp. 29–54). Tokyo, Japan: The National International Institute for Japanese Language.
- Kemps, R., Ernestus, M., Schreuder, R., & Baayen, H. (2004). Processing reduced word forms: The suffix restoration effect. *Brain and Language*, 90, 117–127.
- Manuel, S. Y. (1992). Vowel reduction and perceptual recovery in casual speech. *The Journal of the Acoustical Society of America*, 91, 2388–2388.
- Mehler, J., Peña, M., Nespore, M., & Bonatti, L. (2006). The “soul” of language does not use statistics: Reflections on vowels and consonants. *Cortex*, 42, 846–854.
- Nieuwland, M. S., & Van Berkum, J. J. A. (2006). When peanuts fall in love: N400 evidence for the power of discourse. *Journal of Cognitive Neuroscience*, 18, 1098–1111.
- Nooteboom, S. G., & Doodeman, G. J. N. (1980). Production and perception of vowel length in spoken sentences. *The Journal of the Acoustical Society of America*, 67, 276–287.
- Oostdijk, N. (2002). The design of the Spoken Dutch Corpus. In Peters, P., Collins, P., & Smith, A. (Eds.), *New frontiers of corpus research* (pp. 105–112). Amsterdam, Netherlands: Rodopi.
- Pollack, I., & Pickett, J. M. (1964). Intelligibility of excerpts from fluent speech: Auditory vs. structural context. *Journal of Verbal Learning and Verbal Behavior*, 3, 79–84.
- Pols, L. C. W., & Schouten, M. E. H. (1978). Identification of deleted consonants. *The Journal of the Acoustical Society of America*, 64, 1333–1337.
- Ranbom, L., & Connine, M. (2007). Lexical representation of phonological variation in spoken word recognition. *Journal of Memory and Language*, 57, 273–298.
- Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-prime user's guide, psychology software tools*. Pittsburgh: Psychology Software Tools Inc.
- Schuppler, B., Ernestus, M., Scharenborg, O., & Boves, L. (2011). Acoustic reduction in conversational Dutch: A quantitative analysis based on automatically generated segmental transcriptions. *Journal of Phonetics*, 39, 96–109.
- Spinelli, E., & Gros-Balthazard, F. (2007). Phonotactic constraints help to overcome effects of schwa deletion in French. *Cognition*, 104, 397–406.

- Tucker, B. V. (2011). The effect of reduction on the processing of flaps and /g/ in isolated words. *Journal of Phonetics*, 39, 312–318.
- Tucker, B. V., & Warner, N. (2007). Inhibition of processing due to reduction of the American English flap. In Trouvain, J., & Barry, W. J. (Eds.), *Proceedings of the 16th International Congress of Phonetic Sciences. Saarbrücken, Germany*, (pp. 1949–1952). Saarbrücken, Germany: Pirot.
- Tyler, L. K., & Wessels, J. (1985). Is gating an on-line task? Evidence from naming latency data. *Attention, Perception, & Psychophysics*, 38, 217–222.
- Van Berkum, J. J. A., Brown, C. M., Zwitserlood, P., Kooijman, V., & Hagoort, P. (2005). Anticipating upcoming words in discourse: Evidence from ERPs and reading times. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31, 443–467.
- Van de Velde, H., Gerritsen, M., & van Hout, R. (1996). The devoicing of fricatives in Standard Dutch: A real-time study based on radio recordings. *Language Variation and Change*, 8, 149–175.
- Van de Ven, M., Ernestus, M., & Schreuder, R. (2012). Predicting acoustically reduced words in spontaneous speech: The role of semantic/syntactic and acoustic cues in context. *Laboratory Phonology*, 3, 455–481.
- Van de Ven, M., Tucker, B. V., & Ernestus, M. (2011). Semantic context effects in the comprehension of reduced pronunciation variants. *Memory & Cognition*, 39, 1301–1316.
- Van Petten, C., & Kutas, M. (1990). Interactions between sentence context and word frequency in event-related brain potentials. *Memory & Cognition*, 18, 380–393.
- Warner, N. (1998). *The role of dynamic cues in speech perception, spoken word recognition, and phonological universals*. PhD Dissertation, University of California, Berkeley.
- Wurm, L. H., & Fisičaro, S. A. (2014). What residualizing predictors in regression analyses does (and what it does not do). *Journal of Memory and Language*, 72, 37–48.
- Zimmerer, F., & Reetz, H. (2014). Do listeners recover “deleted” final /t/ in German? *Frontiers in Psychology*, 5, 73–103.
- Zimmerer, F., Scharinger, M., & Reetz, H. (2011). When BEAT becomes HOUSE: Factors of word final /t/-deletion in German. *Speech Communication*, 53, 941–954.
- Zimmerer, F., Scharinger, M., & Reetz, H. (2014). Phonological and morphological constraints on German /t/-deletions. *Journal of Phonetics*, 45, 64–75.

Appendix

This Appendix contains orthographic transcriptions of the materials and phonetic transcriptions of the target words used in the present study. We have underlined the target words in the orthographic transcriptions.

Daarna [dar'na] “subsequently”

Ze hadden ons gevraagd of wij de allerlaatste keer in die boot wilden roeien en daarna [nə'na] zou die in stukken gehakt worden.

“They had asked us to row that boat for the last time and subsequently it would be cut up in pieces.”

En een jaar daarna [nə'na] ben jij erbij gekomen.

“And the year after that you joined us.”

Hij heeft daarna [tə'na] helemaal opnieuw leren praten.

“After that he had to learn to talk again from square one.”

dezelfde [də'zɛlvdə] “the same”

Het was precies dezelfde ['tɛsələ] tijd.

“It was exactly the same time.”

Familie [fa'mili] “family”

En daar zit nu ook de hele familie ['fmili] weer bij, of niet?

“And the whole family will join once again, right?”

Ja, bij jullie familie ['fmili] zijn jullie echt snel.

“Yes, in your family they are really quick.”

Het ging dan meer om familie [ˈfmili]-bezoek dus het hoefde niet.

“It was then more like a family visit, so it did not have to.”

Gegevens [χəˈχevəns] “data”

En dan de gegevens [ˈχevəs] aan te vullen.

“And then update the data.”

Gehad [χəˈhat] “had”

Ik heb een tijd gehad [ˈχat] dat ik veel naar eh naar Derrick keek.

“I have had a period in which I frequently watched eh Derrick.”

Of heb jij ook te maken gehad [ˈχad] met eh ambtelijke teksten zeg maar?”

“Or have you also had to deal with eh so-called official texts?”

Ik heb ook een periode van een jaar ofzo gehad [ˈχat] dat ik één keer gereden had.

“I have also had a period of one year or so, in which I drove only once.”

Want ik heb nooit het idee gehad [ˈχat] dat de organisatie een probleem was.

“Because I have never had the feeling that the organization was a problem.”

Gesproken [χəˈsprokən] “speaking”

De mensen van wie je normaal gesproken [ˈsprokə] veel vuurwerk ziet.

“The people who normally speaking show fireworks.”

Goedkoop [χutˈkop] “cheap”

Een grote partij in te slaan en dan heel goedkoop [χəˈkop] aan te bieden.

“Stock a large amount and then offer them at a very low price.”

Goedkope [χutˈkopə] “cheap”

Straks staan ze allemaal tegen die [χəˈkop] tenten aan te loeren.

“Soon they will all be looking at those cheap tents.”

Hetzelfde [hetˈzɛlvdə] “the same”

Dat was de tweede keer dat we op hetzelfde [ˈsɛldə] instituut zaten.

“That was the second time that we were at the same institute.”

Je betaalt exact hetzelfde [ˈsɛldə] bedrag als vorig jaar.

“You pay exactly the same amount as last year.”

Kunstmatige [kʏnstˈmatɪχə] “artificial”

Waarom we voor een kunstmatige [kəsˈmatɪχ] taal hebben gekozen.

“Why we opted for an artificial language.”

Manier [maˈnir] “manner”

Nee, maar het is toch de manier [ˈmir] waarop het gebouw gemaakt is.

“No, but it is still the way the building was constructed.”

Maar dat was al op die manier [ˈmni] gegarandeerd.

“But that was already guaranteed in that way.”

Maar goed, dan wordt het toch op een of andere manier [ˈmni] vastgelegd.

“But well, that will still be recorded in some way.”

Moet natuurlijk dat geld op de een of andere manier [ˈmir] beheren.

“Of course that money had to be administered in a certain way.”

Moment [moˈment] “moment”

Maar hij leest op dit moment [ˈmen] meer kinderboeken dan ik.

“But at the moment he reads more children’s books than me.”

In Amsterdam duurt het op dit moment [ˈmet] heel lang.

“In Amsterdam it takes very long at this moment.”

Normaal [nɔrˈmal] “normally”

Maar waar wordt dit normaal [nə'mal] voor gebruikt?

“But what is this normally used for?”

Partij [par'tei] “batch”

Als ik iets koop dan moet het maximaal een partij [pə'tei] van 75 stuks zijn.

“If I buy anything then it has to be maximally a batch of 75 pieces.”

Partijen [par'teɪən] “batches”

In het verleden heb je vrij forse partijen [pə'tei] afgenomen.

“In the past you bought quite large batches.”

Principe [prɪn'sipə] “principle”

Ik kan eh in principe [pə'sipə] gaan wanneer ik wil.

“I can eh in principle go whenever I want to.”

De dingen die je meet zijn in principe [pə'sipə] makkelijker.

“The things that you measure are in principle easier.”

Ik voel me daar in principe [ˈpsip] ook helemaal niet bij thuis.

“In principle I really do not feel comfortable there.”

Boeken die er in principe [ˈpsipə] hadden kunnen zijn.

“Books that in principle could have been there.”

Problemen [proˈblemən] “problems”

Met die slaapzakken ook in het verleden wel problemen [ˈplemə] gehad eigenlijk.

“With those sleeping bags in the past there were also problems actually.”

Procent [proˈsent] “percent”

Nee dan wil ik toch echt 25 procent [ˈpsent] korting op die eerste prijs van je hebben.

“No then I really want to have a 25 percent discount on that first price of yours.”

Dan kan ik daar wel eh twintig procent [pəˈsent] afkrijgen denk ik.

“I think I can get eh a 20 percent discount.”

Programma [proˈɣrama] “programme”

Maar dat vind ik een slecht programma [pəˈɣama] eigenlijk.

“But I consider that a bad programme actually.”

Project [proˈjekt] “project”

Hij is weer met een ander Europees project [pəˈjek] bezig.

“He is working on a different European project again.”

Vakantie [vɑˈkɑnsi] “holiday”

Echt het idee van op vakantie [ˈfkɑnt] misschien een auto huren ofzo.

“Really the idea of maybe renting a car during the holidays or something.”

Vandaag [vɑnˈdɑχ] “today”

Ik had vandaag [fəˈdɑ] weer een auto geleend.

“Today I borrowed a car again.”

Vanzelf [vɑnˈzɛlf] “by itself”

Dat het opeens vanzelf [vəˈzɛlf] gaat.

“That suddenly it goes automatically.”

Verdieping [vərˈdipiŋ] “floor”

Op die verdieping [fəˈniːpiŋ] ergens op de Keizersgracht.

“On that floor somewhere along the Keizersgracht.”

Verhaal [vərˈhal] “story”

Ik zal het verhaal [ˈfal] vertellen ja.

“I will tell the story, yes.”

Het moraal van het verhaal [ˈfal] kwam er voor mij op neer van

“The moral of the story to me was that”

Verjaardag [vər'jɑrdɑχ] “birthday”

Jij was niet op de verjaardag [fə'jad] van Jet, toch?

“You were not present at Jet’s birthday, were you?”

Ik vind een verjaardag [fə'jar] is nog wel leuk om te doen.

“I think a birthday is still enjoyable to do.”

Verkeerd [vər'kert] “wrong”

Hij had toch wel eh een verkeerd [f'kɪt] tentje of iets dergelijks.

“He did have a eh wrong tent or something.”

Verkeerde [vər'kerdə] “wrong”

Dat ze een grote kans hebben om eh het verkeerde [fə'kɪdə] pad op te gaan.

“That they run a larger risk to eh go off the track.”

Verkoopt [vər'kopt] “sell”

Want jij verkoopt [f'kopt] er tenslotte meer.

“Because after all you sell more.”

Verleden [vər'ledən] “past”

Mijn oma heeft verleden [f'lej] jaar voor het eerst in januari haar verjaardag gevierd.

“Last year, my grandmother celebrated her birthday in January for the first time.”

Maar dit jaar ga ik niet het risico lopen, want verleden [f'led] jaar ben ik het schip in gegaan.

“However, this year I will not run that risk, because last year I was financially disadvantaged.”

Verloopt [vər'lopt] “elapse”

Van hoe hoe dat afscheid verloopt [f'lopt] van een vakgroep.

“Of how how one takes leave of a research group.”

Verschillende [vər'sɣɪləndə] “different”

Kan je op verschillende [f'sɣɪln] manieren doen.

“You can do that in various ways.”

Corpus dat bestaat uit materiaal van verschillende [f'sɣɪlə] taalfasen, toch?

“Corpus that consists of materials from various phases of language development, right?”

Ik vind wel een heleboel verschillende [f'sɣɪləndə] dingen wat dat betreft.

“I like a lot of different things as far as that is concerned.”

Steekproef te nemen van verschillende [f'sɣɪlnə] vakgebieden.

“To take a sample of different research fields.”

Vertellen [vər'tɛlən] “tell”

Dus ik kan meer vertellen [f'tɛlə] wat ik wel leuk vind.

“So I can better tell you what I do like.”

Vervelend [vər'velənt] “annoying”

Gewoon het idee dat je niet af en toe even kan praten over je werk vond ik heel vervelend [f'velənt], want dat had ik dus erg weinig vond ik zelf.

“Simply the thought that you cannot occasionally talk about your work, I considered very annoying, because I thought I had very little opportunity to do that.”

Ik moet ze ook netjes houden, want anders is het voor jou vervelend [f'vent] als ik ze

“I also need to keep them tidy, because otherwise it is very annoying for you if I”

Verzamelt [vər'zaməlt] “collects”

Nou hij verzamelt [fə'zamət] al heel lang kinderboeken.

“Well he has been collecting children’s books for a very long time.”

Voorname [vɔr'namələk] “mainly”

Het ging die ene persoon dan ook voornamelijk [f'vnamək] om het programma.

“It concerned that one person who mainly for the programme.”

Je komt weleens langs en voornamelijk [ˈvnamə] zit je in de kroeg.

“You occasionally pass by and you are mainly spending time in the pub.”

Waarschijnlijk [warˈsɣeinlək] “probably”

En jullie hebben waarschijnlijk [wəˈsɣeik] alleen al het oud-Engelse deel eruit gevist.

“And you have probably only extracted the Old English part.”

Zoals [zoˈals] “such as”

Net zoals [ˈzəz] wat ik een keertje bij de Albert Heijn had.

“Just like what I once had at the Albert Heijn.”

Te vieren zoals [ˈzɔz] dat gebruikelijk is.

“To celebrate it as usual.”