The processes underlying two frequent casual speech phenomena in Dutch: A production experiment

Iris Hanique[1,2], *Mirjam Ernestus*[1,2]

[1]CLS, Radboud University Nijmegen, the Netherlands
[2]Max Planck Institute for Psycholinguistics, the Netherlands
iris.hanique@mpi.nl, m.ernestus@let.ru.nl

Abstract

This study investigated whether a shadowing task can provide insights in the nature of reduction processes that are typical of casual speech. We focused on the shortening and presence versus absence of schwa and /t/ in Dutch past participles. Results showed that the absence of these segments was affected by the same variables as their shortening, suggesting that absence mostly resulted from extreme gradient shortening. This contrasts with results based on recordings of spontaneous conversations. We hypothesize that this difference is due to non-casual fast speech elicited by a shadowing task.

Index Terms: pronunciation variation, acoustic reduction, shadowing task, experimental methodology

1. Introduction

One main characteristic of casual speech is that many words are not produced in their full forms, but in reduced pronunciation variants. For example, the English word *hilarious* may be produced as /hlɛrɛs/ instead of /hɪlɛriʌs/ [1]. This production study investigates the processes underlying two of these reduction phenomena in Dutch and compares the results to those of a corpus study. This paper contributes to our knowledge of speech reduction and shows the advantages and disadvantages of two very different research methods for the study of casual speech phenomena.

Processes underlying speech reduction may be gradient. The altered pronunciation or even absence of a segment then originates from gradient overlap and decrease in magnitude of articulatory gestures [2]. For instance, /t/ in *must be* may be partly or even completely hidden by the closure of the following bilabial stop, which results in /mʌsbi/.

In addition, the absence of segments may result from categorical deletion processes, which may be phonological rules or the selection of a reduced pronunciation variant from the variants stored in the mental lexicon. So far, two types of reduction phenomena have been found to be categorical: processes that affect only highly frequent words or word combinations (e.g., /e/-deletion in French *c'était* /setɛ/ 'it was' [3]), and phenomena that not only occur in casual connected speech, but also in words produced in isolation or formal speech (e.g., word-internal schwa deletion in French [4]).

So far, only one study has investigated the nature of reduction phenomena that affect many different words and are restricted to connected informal speech [5]. It examined schwa and /t/ reduction in past participles in two speech corpora of Dutch. The results showed that the presence and duration of /t/ are affected by roughly the same phonetic variables, suggesting that absence of /t/ results from the same gradient process as its shortening. Also the presence and duration of schwa were mainly influenced by phonetic variables, but the presence of schwa was affected by more and different variables than its duration. The authors therefore argued that schwa reduction can result from gradient as well as categorical processes.

A disadvantage of corpus studies is that they are often restricted to highly frequent word types, and that these are represented by a widely varying number of tokens. Also, the segmental context of units under study cannot be well-controlled for. These disadvantages do not apply to controlled production experiments, but it is currently unclear which experimental task can elicit casual speech.

We conducted a shadowing experiment, in which, like [5], we focused on schwa and /t/ reduction in Dutch past participles. These words usually consist of /xə/, the verbal base, and /t/ (e.g., *gedanst* /xə+dɑns+t/ 'danced'). We examined whether this controlled experiment produces results similar to those of the corpus study mentioned above.

2. Method

2.1. Participants

We tested 35 Dutch native speakers aged between 18 and 27 (mean 20 years).

2.2. Materials

Our experiment consisted of 180 target past participles starting with *ge* (140 end in /t/) and 100 filler past participles starting with *ver* or *be* (88 end in /t/). These past participles consisted of two or three syllables and the schwa in the initial syllable was followed by a consonant. They spanned the entire range of frequency of occurrence, which was based on the Spoken Dutch Corpus [6].

Each past participle was embedded in the middle of a sentence (which on average consisted of 10 words). Sentence accent was never on the past participle. Also, past participles were never preceded by a fricative, and those ending in /t/ were never followed by /t/ or /d/. Whereas we created one sentence for each filler, we created two sentences for each target: In one sentence the past participle was followed by a vowel, and in the other by a consonant. For example, the sentences for the target *getankt* /xətɛnkt/ 'refueled' were (1) *Ze had per ongeluk diesel getankt in plaats van benzine* 'She accidentally refueled diesel instead of gas', and (2) *Hij heeft voornamelijk getankt waar de brandstof goedkoop is* 'He mainly refueled where fuel was cheap'. All sentences were recorded by a native Dutch female speaker in a casual (average sentence duration: 2019 ms) and careful way (average sentence duration: 2208 ms). On the basis of au-

tomatically generated transcriptions (using the same procedure as described below), we observed that schwa was absent in 125 of the 180 casually produced sentences, while it was never absent in the carefully produced sentences. The average durations of the present schwas were 28.1 ms (std: 9.7 ms) in the casual and 49.2 ms (std: 12.9 ms) in the careful condition. Participants heard every past participle only once, that is, only in the casual or in the careful condition, and followed either by a consonant or a vowel.

The experiment started with a practice block of 10 filler trials followed by four experimental blocks, each consisting of 45 target and 25 filler sentences. The first block after the practice trials started with seven fillers, while the other blocks started with three fillers. Within one block, we presented either casual or careful sentences. If block one and two contained casual sentences, blocks three and four consisted of careful sentences, and vice versa.

We created three pseudo-randomizations of all stimuli, in which no more than three target sentences occurred in succession. On the basis of each randomization, we created four lists. In each list, half of the target past participles were followed by a vowel, and the other half by a consonant. Moreover, half of the stimuli in a list had been produced in a casual way, and the other half in a careful way. Together all four lists contained all sentences in both speech styles.

2.3. Procedure

Each participant was tested individually in a sound-attenuated booth. We presented sentences via headphones, and asked participants to repeat these sentences as quickly and accurately as possible, and to start repeating as soon as possible. We recorded responses on an R-09 Edirol recorder. Each trial started with a fixation point shown for 500 ms on a computer screen, and after an interval of 100 ms the stimulus was presented. The next trial started 1500 ms after the end of this stimulus. Each session lasted approximately 30 minutes.

2.4. Data processing

For all target sentences produced, we created automatic broad phonetic transcriptions by means of forced alignment as described in [5]. The automatic system selected the variant for each word that best matched the speech signal from a lexicon that contained, among others, pronunciation variants in which schwa and /t/ were present or absent. Since the acoustic models consisted of three emitting states and the frame shift was 5 ms, the system assigned to each segment a duration of at least 15 ms. If a segment was in reality shorter than 15 ms, its boundaries were placed within the neighboring segments (which were consequently assigned a shorter duration than they really had). We validated the automatic transcriptions by manually transcribing 100 target schwas, and found that the agreement on the presence of schwa between two transcribers (90%) was very similar to those between each transcriber and the forced alignment (85% and 87%). The average differences in duration were smaller than 13 ms, which is usually considered as acceptable for this type of speech [7]. As expected, given the automatic method used, the durations assigned by the automatic system were generally greater than the durations assigned by the human transcribers.

We excluded 35% of the 6300 trials in two steps. First, 1866 trials were excluded on the basis of the transcriptions. We excluded transcriptions that indicated a silence directly before or after the target word, since in these trials the target words are not embedded in stretches of connected speech. In addition, we excluded transcriptions that were likely to be incorrect. We identified these incorrect transcriptions by means of a set of criteria, which we determined by checking 200 automatically generated transcriptions. Transcriptions were excluded that contained three or more segments of 15 ms in the preceding, target, or following word, which typically indicates that the sentence was produced incompletely or with a wrong word order, or that the forced alignment system had selected a non-suitable pronunciation variant. Further, we also excluded trials if the longest phoneme was the schwa or shorter than 50 ms, which are highly likely to be transcription errors. Finally, we excluded those targets in which the longest phoneme was suspiciously long, as this often indicates that multiple phonemes are transcribed as one long phoneme. We set the boundary for suspiciously long plosives at 175 ms, for vowels and fricatives at 165 ms, and for other consonants at 155 ms. In the second step, we listened to all remaining trials, and excluded from further analyses those 339 trials in which the speaker had not produced the target word or the directly preceding or following word fluently. The fact that we discarded 2205 out of 6300 trials shows that the task was difficult. Interestingly, the number of disregarded trials varied among participants from 23 to 104 trials.

2.5. Predictors

We tested the influence of several predictors on the presence and duration of schwa and /t/. Three of these predictors were defined on the basis of the experimental design. The first predictor is the register (careful or casual) in which the stimulus was presented. The second predictor tested whether there were differences between the blocks of the experiment. The third predictor is the duration of the sentence presented to the participant.

Further, we added other predictors to our statistical models that the literature has shown to be relevant. One of these is speech rate, since segments tend to be more reduced in faster speech (e.g., [8]). We defined speech rate as the number of syllables per second in the whole sentence produced by the participant. In addition, we tested a factor word length, which indicates whether the past participle consisted of two or three syllables, as segments are often shorter if they are followed by more syllables [9].

We also examined three measures of word predictability, since words tend to be more reduced if they are more predictable (e.g., [10]). One measure was the log-transformed word's frequency of occurrence. The second and third measures were the conditional probabilities (CP) of the target word (w_{target}) given the preceding word ($w_{preceding}$) or the following word ($w_{following}$), which were calculated with formulae 1 and 2, respectively. The frequencies used for these predictability measures were based on all components of the Spoken Dutch Corpus [6].

$$log_2(\frac{Frequency(w_{preceding}, w_{target}) + 1}{Frequency(w_{preceding}) + 1}) \quad (1)$$

$$log_2(\frac{Frequency(w_{target}, w_{following}) + 1}{(w_{following}) + 1}) \quad (2)$$

Finally, the surrounding segments may affect reduction (e.g., [8]). For schwa, we therefore tested the log-transformed durations of the preceding and following consonants, the place and manner of articulation of the following consonant, as well as its voicing, and whether it was velar. For /t/, we examined the place and manner of articulation of the preceding segment,

the log-transformed durations of the preceding and following segments, and whether the following segment was a vowel or a consonant (henceforth *type*).

2.6. Analyses

We used mixed effects regression models with contrast coding (i.e., for factors, one level is placed on the intercept, and all other levels are compared to this default level). In order to account for differences between individual stimuli and participants, the models contained *target word* and *participant* as crossed random effects. Each predictor was added individually to a model, and only remained in that model if it was significant *and* improved the AIC value. Duration analyses were based on present segments only.

3. Results and discussion for schwa

Table 1 presents the two final statistical models for the presence and duration of schwa. As expected, schwa was more often absent and shorter if the stimulus was produced in a casual (29.8%; 42 ms) compared to a careful way (22.9%; 45 ms). However, this difference in duration was much smaller than the difference between the stimuli in the two conditions.

Replicating earlier findings [10], schwa was more often absent and shorter in past participles with higher frequencies of occurrence. Highly frequent words are produced more often, and their production has therefore become more automatized and efficient. This typically results in more overlapping gestures.

Also, the consonants surrounding schwa affected its reduction. The main effect of the manner of articulation of the following consonant showed that schwa was significantly longer if this consonant was a plosive (47 ms), and shorter if it was a fricative (38 ms; other segments: 44 ms). We hypothesize that schwa can more easily be co-articulated with a following fricative than a following plosive, since fricatives are continuants.

In addition, we found an effects of the duration of the following consonant. Since the duration of the following consonant correlated with its manner of articulation, we orthogonalized these variables: We replaced this duration by the residuals of a model that predicted duration as a function of manner of articulation. Schwa tended to be shorter if the preceding and following consonants were longer. Schwa was also more likely to be absent if the preceding consonant was longer, especially in block 1 and 3. Moreover, this vowel was more often absent if the following consonant was longer, especially if this consonant was voiced. The effect of the duration of the surrounding consonants can also be explained by co-articulation: Schwa appears shortened or completely hidden by the preceding or following consonants, which are then assigned longer durations. Schwa may often be absent especially before long consonants that are voiced, since it is more difficult to observe (for both humans and automatic speech recognizers) a short, co-articulated, voiced vowel next to a voiced rather than a voiceless consonant.

Finally, speech rate correlated with stimulus register, which we had therefore orthogonalized. Schwa was more likely to be shorter at higher speech rates, but only in bi-syllabic past participles. A possible explanation is that vowels tend to be longer in the initial syllables of bi-syllabic than tri-syllabic words [9]. Consequently, schwa in bi-syllabic words can show more variation in its duration (as shortening is less likely to result in deletion), and may therefore be more easily affected by gradient reduction processes like speech rate.

Presence of schwa (N = 4095)			
Fixed effects	F	df	$p <$
Stimulus register	26.79	1,3864	.0001
Word frequency	6.51	1,3864	.05
Duration $_{\text{following C}}$	36.16	1,3864	.0001
Duration $_{\text{preceding C}}$	12.80	1,3864	.001
Voicing $_{\text{following C}}$	40.33	1,3864	.0001
Block	2.76	3,3864	.05
Duration $_{\text{following C}}$ × Voicing $_{\text{following C}}$	18.45	1,3864	.0001
Duration $_{\text{preceding C}}$ × Block	3.22	3,3864	.05
Random effects	Word	Participant	
Intercept	0.46	0.94	
Duration of schwa (N = 3016)			
Fixed effects	F	df	$p <$
Stimulus register	43.16	1,2678	.0001
Word frequency	12.27	1,2678	.0005
Duration $_{\text{following C}}$	117.72	1,2678	.0001
Duration $_{\text{preceding C}}$	100.37	1,2678	.0001
Manner $_{\text{following C}}$	32.05	2,2678	.0001
Speech rate	4.67	1,2678	.05
Speech rate × Word length	7.49	1,2678	.01
Random effects	Word	Participant	
Intercept	14.73	4.74	
Duration $_{\text{preceding C}}$	2.43		

Table 1: *Results for schwa: those for its presence are above the double line, and those for its duration are below the double line. For the factors Stimulus register, Voicing $_{\text{following C}}$, and Block, the levels on the intercept are casual, voiced, and block 4, respectively. For the random effects, the table reports the estimated standard deviations (in ms for duration).*

4. Results and discussion for /t/

The final statistical models for the presence and duration of /t/ are presented in Table 2. Since the duration and type (vowel versus consonant) of the following segment were correlated, we had orthogonalized them. First, as expected, the models show that /t/ was shorter at higher speech rates. In addition, /t/ was more likely to be absent and shortened if it was followed by a consonant (28.3%; 58 ms) than a vowel (10.5%; 66 ms). The articulatory gestures of /t/ are more similar to those of other consonants than to those of vowels, since vowels require a relatively open vocal tract whereas consonants typically involve a (almost) closed one. Therefore, /t/s may more easily overlap with and be hidden by other consonants. If so, they may be difficult to distinguish from these overlapping consonants, and appear acoustically shortened or even absent.

Word-final /t/ was also more likely to be absent and shorter if the following segment was longer, but only if this was a consonant. An explanation may be that if /t/ overlaps with a following consonant, (part of) its duration may be attributed to this following consonant. For the presence of /t/, the effects of the following consonant were also greater if the word was more predictable given the following word. Word combinations that are often used together are more automatized, can thus be produced more easily, and are consequently more likely to show effects of co-articulation.

Further, we investigated the roles of the duration of the preceding segment and its manner of articulation. Since these two predictors were correlated, they were orthogonalized (following the method for orthogonalization described above). Word-final

Presence of /t/ (N = 3133)			
Fixed effects	F	df	$p <$
Duration $_{\text{following segm}}$	28.87	1,3123	.0001
Type $_{\text{following segm}}$	149.12	1,3123	.0001
CP $_{\text{following word}}$	7.85	1,3123	.05
Duration $_{\text{following segm}}$ × Duration $_{\text{preceding segm}}$	13.80	1,3123	.001
Duration $_{\text{following segm}}$ × Type $_{\text{following segm}}$	22.66	1,3123	.0001
Duration $_{\text{following segm}}$ × CP $_{\text{following word}}$	30.94	1,3123	.0001
Duration $_{\text{following segm}}$ × Type $_{\text{following segm}}$ × CP $_{\text{following word}}$	29.33	1,3123	.0001
Random effects	Word	Participant	
Intercept	1.92	0.84	
Duration $_{\text{following segm}}$	0.79	0.39	
Duration $_{\text{preceding segm}}$	0.04		
Duration of /t/ (N = 2472)			
Fixed effects	F	df	$p <$
Duration $_{\text{following segm}}$	42.83	1,2389	.0001
Type $_{\text{following segm}}$	109.36	1,2389	.0001
Duration $_{\text{preceding segm}}$	21.53	1,2389	.0001
Manner $_{\text{preceding segm}}$	13.01	4,2389	.0001
CP $_{\text{following word}}$	24.04	1,2389	.0001
Speech rate	28.40	1,2389	.0001
Duration $_{\text{following segm}}$ × Type $_{\text{following segm}}$	70.10	1,2389	.0001
Duration $_{\text{following segm}}$ × Duration $_{\text{preceding segm}}$	7.81	1,2389	.01
Duration $_{\text{preceding segm}}$ × Manner $_{\text{preceding segm}}$	5.37	1,2389	.001
Random effects	Word	Participant	
Intercept	5.83	6.87	
Duration $_{\text{following segm}}$	5.58	2.07	

Table 2: *Results for /t/: those results for its presence are above the double line, and those for its duration are below the double line. For the factors Type $_{\text{following segm}}$ and Manner $_{\text{preceding segm}}$, the levels on the intercept are consonant and fricative, respectively. For the random effects, the table reports the estimated standard deviations (in ms for duration).*

/t/ tended to be shorter if the preceding segment was longer, especially if this segment was a fricative or nasal. If the gestures of /t/ overlap with a preceding fricative or nasal, its closure may be incomplete, and /t/ may therefore be hard to distinguish from this fricative or nasal, which then appears longer.

In addition, the interaction of the durations of the preceding and following segments showed that /t/ was more likely to be absent and shorter if either the preceding or following segment is longer, especially if the other immediately neighboring segment is shorter. This suggests that gestural overlap of /t/ with an adjacent segment is larger if it overlaps less with the other adjacent segment.

Finally, /t/ tended to be shorter and more often absent in words that are more predicable given the following word. These predictability effects are likely the result of more sloppy pronunciations of more often repeated and thus more automatized words or word sequences.

5. General discussion and conclusions

We demonstrated that the shortening and absence of schwa and /t/ show patterns that can easily be interpreted as resulting from co-articulation. Furthermore, we found that their presence and duration were conditioned by similar variables, suggesting that the absence of these segments is the extreme result of their shortening, and thus of a gradient underlying process. Note that we did find slightly more effects for the duration measures, as expected, since analyses of a continuous variable have generally more statistical power than analyses of a factor.

We expected participants to repeat the pronunciation variants that were presented, and thus that many more schwas would be absent in the casual than in the careful condition. However, these percentages were relatively low in both conditions (29.8% and 22.9%, respectively). Our results thus suggest that participants did not aim at repeating the variant they heard, but at producing the word's full form. This would explain why we did not find evidence for categorical absence of schwa, as reported in the Corpus study [5]. Our shadowing task did elicit reduction phenomena resulting from co-articulation. Apparently, this task evokes non-casual fast speech. This hypothesis is supported by the fact that several words which are often drastically reduced in casual speech (e.g., *eigenlijk* /eixələk/ is often reduced to /eik/, *allemaal* /aləmal/ to /ɑm/, and *helemaal* /heləmal/ to /heməl/) are never produced in these extremely reduced forms by the participants. We therefore recommend, when studying natural speech, to always use this task in combination with another research method.

In conclusion, our production experiment shows that the shadowing task elicits non-casual fast speech, in which reduction of schwa and /t/ in Dutch past participles is only affected by gradient co-articulation.

6. Acknowledgments

This work was partly funded by an European Young Investigator Award and by an ERC starting grant (284108) to the second author. Further, the authors like to thank Marjoleine Sloos for her contribution to the experimental method, and Lou Boves for his comments on a previous version of this article.

7. References

[1] K. Johnson, "Massive reduction in conversational American English," in *Proc. of the workshop on Spontaneous Speech: Data and Analysis*, Tokyo, Japan, 2004, pp. 29–54.

[2] C. Browman and L. Goldstein, "Articulatory Phonology: An Overview," *Phonetica*, vol. 49, pp. 155–180, 1992.

[3] F. Torreira and M. Ernestus, "Vowel elision in casual French: the case of vowel /e/ in the word c´était," *J. Phonetics*, vol. 39, pp. 50–58, 2011.

[4] A. Bürki, M. Ernestus, and U. Frauenfelder, "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," *J. Mem. Lang.*, vol. 62, pp. 421–437, 2010.

[5] I. Hanique, M. Ernestus, and B. Schuppler, "The processes underlying two casual speech phenomena in Dutch," submitted.

[6] N. Oostdijk, "The design of the Spoken Dutch Corpus," in *New Frontiers of Corpus Research*, P. Peters, P. Collins, and A. Smith, Eds. Amsterdam: Rodopi, 2002, pp. 105–112.

[7] M. Wesenick and A. Kipp, "Estimating the quality of phonetic transcriptions and segmentations of speech signals," in *Proc. of ICSLP*, Philadelphia, USA, 1996, pp. 129–132.

[8] J. Dalby, "Phonetic structure of fast speech in American English," Ph.D. dissertation, Indiana University, 1984.

[9] S. Nooteboom, "Production and perception of vowel duration: a study of durational properties of vowels in Dutch," Ph.D. dissertation, University of Utrecht, 1972.

[10] A. Bell, J. Brenier, M. Gregory, C. Girand, and D. Jurafsky, "Predictability effects on durations of content and function words in conversational English," *J. Mem. Lang.*, vol. 60, pp. 92–111, 2009.