SAME OR DIFFERENT? SCHWAS IN NATURALLY SPOKEN GERMAN

Frank Zimmerer & Henning Reetz

Goethe University, Frankfurt (Main), Germany
zimmerer@em.uni-frankfurt.de; reetz@em.uni-frankfurt.de

ABSTRACT

Linguistic research has found that in conversational speech variation in general, and reduction in particular, are ubiquitous processes affecting pronunciation of words. They are shortened, syllables are not produced canonically, segments get lenited. These processes can have quite dramatic effects on the pronunciation of words, which can have detrimental effects for word recognition (e.g. [13, 17]). While deletion of a vowel can remove syllabic information, reduction has a less severe effect on the syllabic structure of words. However, the exact amount of reduction in conversational speech in German is still not examined satisfactorily. This paper investigates in how far in conversational German, "reduced Schwas", being the product of reduction or neutralization of other vowels, are comparable to underlying Schwas. An analysis of the Kiel corpus of spontaneous Speech [7] sheds light on this question. The results show that Schwas being the product of reduction are different from underlying Schwas in some respects, but these differences are rather small and are not identical for female and male speakers. The results are compared to prior research and we discuss repercussions for modeling reduction.

Keywords: speech production, reduction, conversational speech, variation

1. INTRODUCTION

Linguistic research has only recently become aware of the importance of natural speech for modeling and theory building. For a long time, many of the processes occurring in conversational speech have been considered to be fast speech phenomena only, and have not been regarded as being part of the linguistic competence of speakers (cf. [3, 8, 17]). Pronunciation of words can differ considerably from "perfect" or canonical speech. For instance, in conversational speech, speakers tend to utter the words more "sloppy" (c.f. "Hypospeech", [11]) than in careful speech. The vowel Schwa is an excellent example of the reduction of segments. In conversational English, for instance, vowels in unstressed positions are regularly reduced to Schwa, or, more drastically, they can be (almost) deleted (e.g. [2, 12]). This is also true for German. While the number of massive reduction in conversational German is comparatively small (e.g. [17]), the reduction of vowels to Schwa is a process that does occur, leading to the question in how far it is even correct to characterize the outcome of vowel reduction as regular Schwa, with respect to Schwas that have not been altered by speech production processes.

2. SCHWA

The vowel Schwa has been used as a descriptive label for many different acoustic realizations in many different languages of the world (e.g. [15]). In the IPA chart, the symbol [ə] denotes the most central vowel in the overall system. Being the most unmarked and central vowel, with a kind of neutral jaw position for vowel production, makes Schwa an ideal vowel as outcome of neutralization or reduction processes[1]. For instance, Dutch as well as English show reduction processes that have Schwa as target (e.g. [15] and references therein). From a phonetic point of view, Schwa is harder to grasp, due to its huge amount of variation depending on its context, as will be discussed in the next section.

2.1. Phonological description

In phonological terms, Schwa is often characterized by the lack of its (featural) specification. Thus, this minimally specified vowel is an ideal target in vowel reduction, deletion but also insertion processes ([15], and references therein). The phonological description does not take into account the acoustic nature and differences of Schwa. As a matter of fact, the acoustic nature of this vowel has been found to be extremely variable, and not easy to subsume under a single label (e.g. [4, 15]). Furthermore, the

phonemic status of Schwa in German is debatable [9, 16]. Thus, before studying the nature of reduced Schwa in comparison to underlying Schwa in German, a phonetic description is warranted.

2.2. Phonetic description

From the point of view of a phonetician, Schwa is harder to describe than for a phonologist. Schwa can be variable across contexts. In English, for instance, word medial, unstressed vowels can reduce to Schwa, such as in *suppose*, or *police*. At the same time, there is also vowel neutralization in cases like *China* or *comma*. However, acoustic analyses have shown, that the two vowels are rather different. While word-internal Schwa is very variable and depending on context, final Schwa is not [5]. This finding led to a proposal of different Schwas, whose transcription as a single symbol [ə] is misleading [4], and probably driven by phonological considerations than based on phonetic analysis. We wanted to extend and replicate this data in German, also a Germanic language. In German, vowels do not get reduced to Schwa as regularly as in English (e.g. [17]), but speakers do produce this kind of reduction. Yet, it is not clear in how far there are acoustic differences between Schwas that are the result of reduction processes on the one hand and Schwas that are underlying Schwas on the other hand. It is also not clear whether they are comparable to the differences found in English. In the following section, we present data of an analysis of a corpus of conversational German comparing acoustic properties of reduced Schwas with the acoustic properties of underlying Schwas.

3. EXPERIMENTAL DATA: CORPUS ANALYSIS

3.1. Methods and materials

For the phonetic analysis, we chose the Kiel Corpus of spontaneous speech [7], the largest phonetically annotated corpus of conversational German. Due to its architecture, it provides us with ideal data for examining Schwas in conversational German. The corpus consists of dialogues from 42 German speakers (18 female), summing up to about 4 hours of speech. Two speakers were engaged in making appointments with each other but they had conflicting time-tables, thus they had to negotiate dates for meetings, and did so quite naturally. Furthermore, the Kiel Corpus offers high quality recordings, with all dialogues being transcribed and labeled by trained phoneticians using spectrogram and oscillogram information as well as listening to the conversations [10]. The corpus consists of three different transcriptions. Firstly, there is an orthographic transcription of the dialogues. Then, there is a canonical phonetic transcription. And finally, and most importantly, the corpus has also a phonetic transcription of what was actually produced by the speakers. This allows for a comparison of an idealized pronunciation (i.e. canonical transcription) with the actual (i.e. phonetically transcribed) pronunciation. The idealized canonical transcription denotes how speakers should utter the words if they were talking in accordance with a careful dictionary-like pronunciation. The nature of the task and the fact that all the pairs of talkers had to make the same appointments restricted the vocabulary on the one hand and let to a large number of utterances for other words, on the other hand. For instance, days of the week as well as dates and times occur very often. Nevertheless, the conversations were very natural and the corpus meets all the requirements that were asked for as a basis for an analysis of the processes that occur in natural speech.

All acoustic measurements were taken at the vowel midpoints using Praat [1]. For the analysis, only those vowels were chosen as underlying Schwa, where, according to a canonical pronunciation, Schwa was the actual target vowel, and was produced as such. As reduced Schwas, we considered those instances where any vowel (or diphthong) should have been the target vowel, but the actual realization was transcribed as Schwa. We did not include epenthetic Schwas or cases where underlying Schwas were produced as another vowel. Statistical analyses were calculated in JMP [14].

3.2. Results

The number of vowels that are reduced to Schwa in conversational German is rather small (see Table 1 and Table 2). In the corpus, of the 53433 vowels that are produced, 3685 vowels are underlying Schwas and 432 (1%) are reduced to Schwa. This shows that, reduction to Schwa is not very common in conversational German, but it occurs. Also, not all vowels are equally likely to be reduced to Schwa. Mid-vowels account for the vast majority of the sources for Schwa (258 – 60%), whereas low (21%) and high vowels (7%) are reduced less often to Schwa. The missing 12% of

the vowels are underlyingly diphthongs. Front vowels are the basis for Schwa in 229 productions (53%), whereas back vowels get centralized in 153 cases (35% - diphthongs 12%). Long vowels (and diphthongs) sum up to 198 instances of reduced Schwas (46%), short vowels to 234 (54%). When they are reduced, long vowels tend to be even shorter than underlying short vowels (37.3ms (SD 20.23) vs. 45.7ms, (SD 15.9).

Table 1: Number of reduced vowels, compared to canonical pronunciation, by height categories and overall, Underlying Schwas are not included. Percentages are in relation to overall numbers and to the number of schwa reduction.

	High	Mid	Low	Dipth.	Overall
Reduced	32	258	92	50	432
Overall	13220	11460	14257	10787	49724
Percent	0.2/7	2.2/60	0.6/21	0.4/12	0.9/100

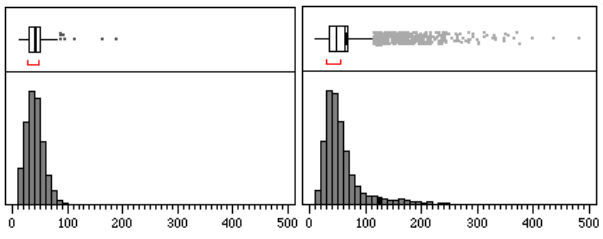
Table 2: Number of reduced vowels, compared to canonical pronunciation, by front/back categories and overall, Underlying Schwas are not included. Percentages are in relation to overall numbers and to the number of schwa reductions.

	Front	Back	Dipth.	Overall
Reduced	229	153	50	432
Overall	17434	21503	10787	49724
Percent	1/53	0.7/35	0.4/12	0.9/100

Men reduced 257 vowels to Schwa, women accounted for 175 instances of vowel reduction. This difference is not significant, since they overall contributed different amounts of segments in the corpus production. Next, a statistical analysis was carried out to investigate in how far the reduced Schwas are like underlying Schwas. The results show that indeed, reduced Schwas are significantly different from underlying Schwas.

First, concerning vowel duration, reduced Schwas are significantly shorter than their underlying counterparts ($F(1,4115)=83.16$, $p<0.001$, see also Figure 1).[2]

Figure 1: Histograms of vowel duration (in ms) of reduced (left) and unreduced (right) Schwas, with Box Plots.



We analyzed the formant values for men and women separately and found further evidence that reduced Schwas differ from underlying Schwas. Women produced (i.e. underlying) unreduced Schwas with a mean F1 of 510 Hz (SD 143), their reduced Schwas were produced with a mean F1 of 478 Hz (SD 154). For F2, unreduced Schwas showed a mean F2 of 1793 Hz (SD 339), reduced Schwas had a mean F2 of 1716 Hz (SD 217). Men showed the following pattern. Their unreduced Schwas had a mean F1 of 509 Hz (SD 266), whereas reduced Schwas were produced with a mean F1 of 517 Hz (SD 281). For F2, unreduced Schwas had an average of 1712 Hz (SD 338); reduced Schwas 1654 Hz (SD 348).

An ANOVA with REDUCTION (red/unred), and SPEAKER (as random effect) was calculated separately for F1 and F2. The results for female speakers showed that F1 differed significantly ($F(1,1682)=6.49$; $p<0.01$) between reduced and unreduced Schwas. The analysis of F2 was not significant. Neither were the differences between the formants of reduced and unreduced Schwas for male speakers.

4. CONCLUSIONS

First, the statistical analysis did not provide clear-cut results for all the acoustic cues. While reduced Schwas are different from underlying Schwas in some respects, not all the cues (duration, F1, F2) behave in parallel. The factors that influence the significant differences between the factors differ also between male and female speakers. Therefore, no clear factor can be given that separates unreduced (i.e. underlying) Schwas from reduced ones. The results are thus not identical with results from conversational English [4, 5]. In those studies, the results supported the idea of having different Schwa categories. The fact that reduced Schwas are significantly different in length from underlying Schwas and in mean F1 (32 Hz) for female speakers is not enough evidence favoring different categories, the F1 difference for female speakers is also rather small. Furthermore, also within the underlying Schwa segments, a huge variation can be observed. For those segments, one would then also have to posit different categories, which would reduce explanatory power in turn. The question that arises is on what factors such categories would be based. Length, for instance, would not be enough. As can be seen in Figure 1 there is a great amount of overlap in Schwa length. A simple formula for deciding on a category cannot be given. Speech perception tasks with speakers of different language backgrounds could

shed more light on the question in how far it is correct to treat different Schwas as one single category. For instance, a comparison of languages with Schwa (e.g. German) and languages without Schwa in their inventory (e.g. Turkish) could be revealing.

Next, an extra-linguistic factor that has been found in some instances to have an effect on reduction is gender (e.g. [2, 17]). In this study, we also observed some gender effects. However, they are not completely in line with prior research. In this study, men and women reduced similarly often, and if at all, woman reduced more drastically (as can be seen in the significant F1 difference), which is different from results reported in prior research.

The results presented in this paper are further evidence that despite the massive variation in conversational speech, the tendency of German speakers to neutralize contrasts is not very high. Compared to English, the number of completely neutralized vowels is rather small. From prior research we know that in cases of complete neutralization, or massive reductions, listeners need the context in order to recognize the intended meaning [17], which can be interpreted as additional recognition cost, since listeners have to wait for and include more context for recognition. It seems that German speakers refrain from putting additional memory load on listeners for succeeding in speech recognition.

Thus, successful linguistic modeling has to include factors to determine in how far speakers reduce in their native languages, and why speakers from different language backgrounds reduce differently.

5. REFERENCES

[1] Boersma, P., Weenink, D. 2010. PRAAT, V.5.1.33. *http://www.praat.org/*

[2] Byrd, D. 1994. Relations of sex and dialect to reduction. *Speech Communication* 15, 39-54.

[3] Ernestus, M. 2000. *Voice Assimilation and Segment Reduction in Casual Dutch*. Utrecht: LOT.

[4] Flemming, E. 2007. *The Phonetics of Schwa Vowel*. Unpublished manuscript. *http://web.mit.edu/flemming/www/paper/schwaphonetics.pdf*

[5] Flemming, E., Johnson, S. 2007. Rosa's roses: Reduced vowels in American English. *Journal of the International Phonetic Association* 37, 83-96.

[6] Gick, B. 2003. Articulatory correlates of ambisyllabicity in English glides and liquids. In Local, J., et al. (eds.), *Papers in Laboratory Phonology 6: Constraints on Phonetic Interpretation*. Cambridge: Cambridge University Press, 223-236.

[7] IPDS - Institut Für Phonetik und Digitale Sprachverarbeitung. 1994. *The Kiel Corpus of Spontaneous Speech*. Kiel.

[8] Johnson, K. 2004. *Massive Reduction in Conversational American English*. *Proceedings of the 1st Session of the 10th International Symposium* Tokyo.

[9] Kohler, K.J. 1995. *Einführung in die Phonetik des Deutschen*. Berlin: Erich Schmidt Verlag.

[10] Kohler, K.J., Pätzold, M., Simpson, A.P. (eds.). 1995. *From Scenario to Segment - The Controlled Elicitation, Transcription, Segmentation and Labelling of Spontaneous Speech*. AIPUK, IPDS, Kiel, vol. 29.

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

[12] Manuel, S.Y. 1991. Recovery of "Deleted" Schwa. Paper presented at *Phonetic Experimental Research at the Institute of Linguistics* (PERILUS XIV), Stockholm.

[13] Pickett, J.M., Pollack, I. 1963. Intelligibility of excerpts from fluent speech: Effects of rate of utterance and duration of excerpt. *Language and Speech*, 6, 151-164.

[14] SAS, 2003. JMP. 5. 0.12.

[15] Silverman, D. 2011. Schwa. In van Oostendorp, M., Ewen, C., Hume, E., Rice, K. (eds.). *Companion to Phonology*. Chichester, UK: Wiley-Blackwell.

[16] Wiese, R. 1996. *The Phonology of German: The Phonology of the World's Languages*. Oxford: Oxford University Press.

[17] Zimmerer, F. 2009. *Reduction in Natural Speech*. Ph.D. thesis, Goethe-Universität, Frankfurt (M). *http://publikationen.ub.uni-frankfurt.de/volltexte/2010/7332/*

[1] The notion of neutral Schwa is challenged (e.g. [6])

[2] All vowels were analyzed, irrespective of possible (hesitational) lengthening by speakers.