

# LEVELS OF REDUCTION FOR GERMAN TENSE VOWELS

*Christina Widera and Thomas Portele*

Institut für Kommunikationsforschung und Phonetik (IKP), Universität Bonn  
Poppelsdorfer Allee 47, 53115 Bonn, Germany  
{cwi, tpo}@ikp.uni-bonn.de

## ABSTRACT

In natural speech there are differences in the realisation of vowels. Numerous factors such as speaking style, prosody, or word class can cause vowel reductions. It is investigated whether vowel reductions can be described using discrete levels and, if yes, how many levels can be reliably perceived. The reduction of a vowel was judged by matching stimuli to representatives of reduction levels (prototypes). The results were investigated on the basis of inter-subject agreement. The resulting prototypes were evaluated by further perception experiments as well as artificial neural networks. The transferability of the reduction levels to other speakers was also investigated. The experiments show that listeners can reliably discriminate 3 to 5 reduction levels depending on the vowel. They use the prototypes speaker-independently, while neural networks trained with the material from one speaker are not applicable to other speakers.

Lastly, the relationships between reduction levels and prosodic factors (lexical word stress, pitch accent, prominence) as well as word class (content words vs. function words) were investigated.

## 1. INTRODUCTION

In natural speech, we find a lot of inter- and intrasubjective variation in the realisation of vowels. Vowels spoken in isolation or in a neutral context are considered to be ideal vowel realisations with regard to vowel quality. Vowels differing from the ideal vowel are described as reduced. From an articulatory point of view, vowel reduction is explained by articulators not reaching the canonical target position (target undershoot, [1]). From an acoustic point of view, vowel reduction is described by smaller spectral distances between the sounds. Perceptually, reduced vowels sound like the neutral vowel ('schwa').

Numerous factors such as speaking style, prosody, or word class can cause vowel reductions. Speakers mark important parts of an utterance by accentuation and they pronounce these parts more carefully and clearly ([2], [3], [4]). Stressed syllables are important for lexical access, they contain less reduced vowels than unstressed ones [3]. Function words are less important and have a high frequency of occurrence. Reduced vowels are found more frequently in function words than in content words [3].

On the continuum from unreduced vowels to 'schwa', listeners are able to hear a lot of differences with regard to vowel quality. However, results of perception experiments where subjects had to classify vowels according to their vowel quality into 2 (full vowel or 'schwa' [5]) or 3 groups (without any duration information [6]) show that the inter-subject agreement is quite low. The question is whether vowel reductions can be

described using discrete levels and if yes, how many levels can be reliably perceived. A subdivision of reduction in levels allows an automatic labelling of reduction.

## 2. DATABASE

The database consists of isolated sentences, question and answer pairs, and short stories read by 3 speakers [7]. The utterances were labelled by hand (SAMPA [8]). For each vowel, the frequencies of the first 3 formants were computed every 5 ms [9]. The values of each formant for each vowel were estimated by a 3rd order polynomial function. The polynomial fits the formant trajectory. The frequency of each formant is defined here as the value in the middle of the vowel [10]. Each vowel has also information about its normalized energy values of 4 frequency bands (0-1kHz, 1-2kHz, 2-4kHz, 4-8kHz), its mean fundamental frequency (F0), and its duration. Furthermore, each syllable has been labelled with the perceived prominence (median of 3 subjects' judgements scaled from 0 to 30 [7]).

## 3. EXPERIMENTS

### 3.1 Pre-test

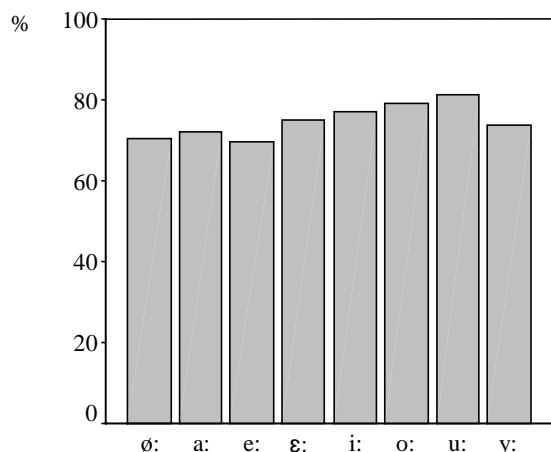
The first two formant frequencies (F1, F2) are assumed to be the main factors determining vowel quality [11]. Their values were mel-scaled and standardized (z-scores), F1 and F2 values of the 8 German tense vowels of one speaker (Speaker1) were clustered by mean cluster analysis. The number of clusters varied from 2 to 7.

In a pre-test, the strength of the reduction of vowels with the same label in the same phonetic context was judged perceptually by a single subject (open answer form). The comparison of the perceived reduction levels and vowel qualities with the groups of the different cluster analysis shows a better agreement between judgements and the cluster analysis with 7 groups for [i:], [y:], [a:], [u:], and [o:]. For [e:], [ɛ:], and [ø:] the judgements are better described by the cluster analysis with 6 groups.

From each cluster, one prototype was determined whose formant values are the closest to the cluster centre. Within a cluster, the distances between the formant values and the cluster centre were computed by:

$$d = (ccF1 - F1)^2 + (ccF2 - F2)^2$$

where ccF1 stands for mean F1 value of the vowels of the cluster; F1 is the F1 value of a vowel of the same cluster; ccF2 stands for mean F2 value of the vowels of the same cluster; F2 is the F2 value of a vowel of the same cluster.



**Fig. 1:** Average agreement between individual judgements and overall reduction level for each vowel.

These prototypes are supposed to be representative for different reduction levels; this hypothesis is tested in the following perception experiment.

### 3.2 Perception experiment

#### Method

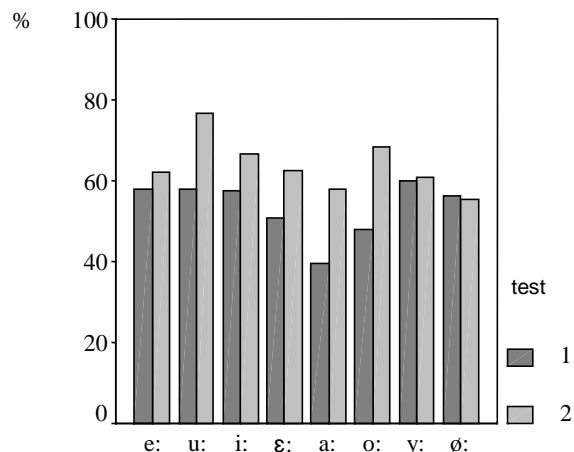
The experiments were carried out for each of the 8 tense vowels separately. 9 subjects participated in the first perception experiment. All subjects are experienced in labelling speech. The prototypes were presented on the computer via headphones. The subjects could listen to each prototype as often as they wanted. The task was to arrange the prototypes by strength of reduction from unreduced to reduced. The reduction level of each prototype was defined by the modal value of the subjects' judgements.

Furthermore subjects had to classify stimuli based on their perceived qualitative similarity to these prototypes. Six vowels from each cluster (if available) whose acoustical values are maximally different as well as the prototypes were used as stimuli. The test material contains each stimuli twice ([i:], [o:], [u:] n=66; [a:] n=84; [e:] n=64; [y:] n=48; [ε:] n=40; [ø:] n=36). Each stimulus was presented over headphones together with the prototypes as labels on the computer screen. The subjects could hear the stimuli within and outside their syllabic context and could compare each prototype with the stimulus as often as they wanted. Assuming that a stimulus shares its reduction level with the pertinent prototype, each stimulus received the reduction level of its prototype. The overall reduction level (ORL) of each judged stimulus was determined by the modal value of the reduction levels of the individual judgements.

#### Results

Prototypes stimuli were assigned to the prototypes correctly in most of the cases (average value of all subjects and vowels: 93.6%). 65.4 % of all stimuli (average value of all subjects and vowels) were assigned to the same prototype in the repeated presentation. The results indicate that the subjects are able to assign the stimuli more or less consistently to the prototypes, but it is a difficult task due to the large number of prototypes.

The relevance of a prototype for the classification of vowels was determined on the basis of a confusion matrix. The



**Fig. 2:** Agreement (%) between individual judgements and overall reduction level with respect to the number of prototypes of the first (1) and second (2) experiment for each vowel.

prototypes themselves were excluded from the analysis. If individual judgements and ORL agree in more than 50% and more than one stimulus is assigned to the prototype, the prototype might represent a reduction level. Depending on the vowel, 5 prototypes were found for [i:], [u:] as well as for [e:], and 3 prototypes for the other vowels. The resulting prototypes were evaluated in further experiments with the same design as used before.

### 3.3 Evaluation of prototypes

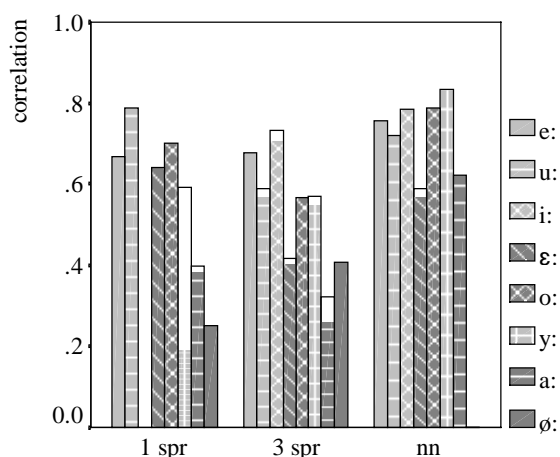
8 subjects were asked to arrange the prototypes with respect to their reduction and to transcribe them narrowly using the IPA system. Then they had to classify the stimuli using the prototypes. Stimuli were vowels with maximally different syllabic context. Each stimulus was presented twice in the test material ([i:] n=82; [o:] n=63; [u:] n=44; [a:] n=84; [e:] n=68; [y:] n=52; [ε:] n=34; [ø:] n=30).

The average agreement between individual judgements and ORL (stimuli with 2 modal values were excluded) is equal or greater than 70% for most vowels. For [i:] it is found that two prototypes are frequently confused. Since the prototypes sound very similar one of them was excluded. No separate tests were carried out, the resulting prototypes were evaluated in the next experiment (cf. section 3.4).

$\chi^2$ -tests show a significant relation between the judgements of any two subjects for most vowels ([i:], [u:], [e:], [o:], [y:]  $\alpha < .01$ ; [a:]  $\alpha < .02$ ; [ε:]  $\alpha < .05$ ). Only for [ø:], 9 non-significant ( $\alpha > .05$ ) inter-subject judgements are found, most of them (6) due to the judgement of one subject. Fig. 1 shows the agreement between individual judgements and ORL for each vowel.

To test whether the agreement is improved because the prototypes are good representatives of reduction levels or only because of the decrease in their number, the agreement between individual judgements and ORL are computed with respect to the number of prototypes [12]:

$$agreement(pa) = n(ra) - \frac{n(wa)}{n(pa) - 1}$$



**Fig. 3:** Correlation for each vowel grouped by experiments. Correlation between subjects of the test with 1 speaker (1 spr) and with 3 speakers (3 spr); correlation between NNs and subjects (nn).

where  $n(ra)$  is the number of right answers;  $n(wa)$  is the number of wrong answers;  $n(pa)$  is the number of possible answers.

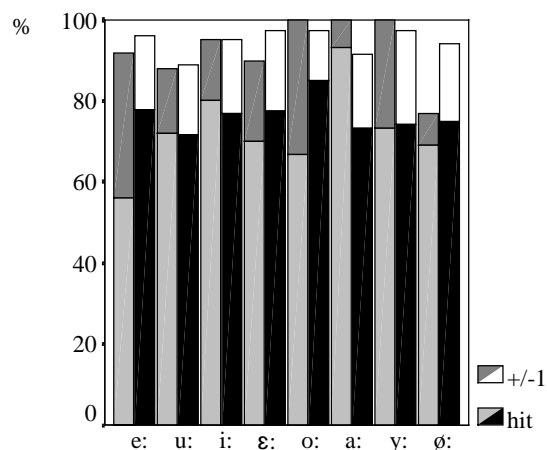
In comparison to the agreement between individual judgements and ORL in the first experiment, the results are actually improved (Fig. 2). We assume that the agreement between individual judgements and ORL is sufficiently good. The prototypes represent reduction levels, and the assigned stimuli can be regarded as classified with respect to their reduction. This is supported by the inter-subject judgements of stimuli for most vowels. The average correlation between any two subjects is significant at the .01-level for the vowels [i:], [e:], [u:], [o:], [y:] and at the .04-level for [ε:]. For [a:] and [ø:], the inter-subjective correlation is low but significant at the .02 or at the .05-level, respectively (Fig. 3; value of [i:] is shown in the 2<sup>nd</sup> group (3 spr)).

### 3.4 Speaker-independent reduction levels

In a further experiment it was investigated whether the reduction levels and their prototypes can be transferred to other speakers, and whether artificial neural networks (NNs) can reliably predict subjects' judgements. For each vowel the judged stimuli (ORL) from Speaker1 were used for training a NN (feed-forward, with 2 hidden layers). Fundamental and formant frequencies, normalized energy in frequency bands, and duration (all values standardized by z-scores) were used as input. Output of the NNs was the pertinent reduction level (5 reduction levels for [u:] and [e:]; 4 for [i:]; 3 for the other vowels). 2830 vowels of 3 speakers (Speaker1 and 2 additional speakers) were classified by the NNs. For the evaluation, subjects had to judge 5 stimuli per speaker and per reduction level classified by the NNs. The same experimental design as in the other perception experiments was used.

The comparison of individual judgements and ORL shows that independently of the speaker, the average agreement between these values are quite similar (76.4% for Speaker1; 73.1% for Speaker2; 76.5% for Speaker3). There is also little difference between the average confusion rate of related reduction levels (+/- 1 level) for each speaker (18.4% for Speaker1; 20.8% for Speaker2; 18.3 for Speaker3; Fig. 3). The correlation of any two subjects' judgements is comparable to the correlation of the

last perception experiments, only the correlation of the inter-



**Fig. 4:** Hit rate and confusion rate (%) of related reduction levels (+/-1 level) of NN (grey colours) and of subjects (black/white) for each vowel of Speaker1.

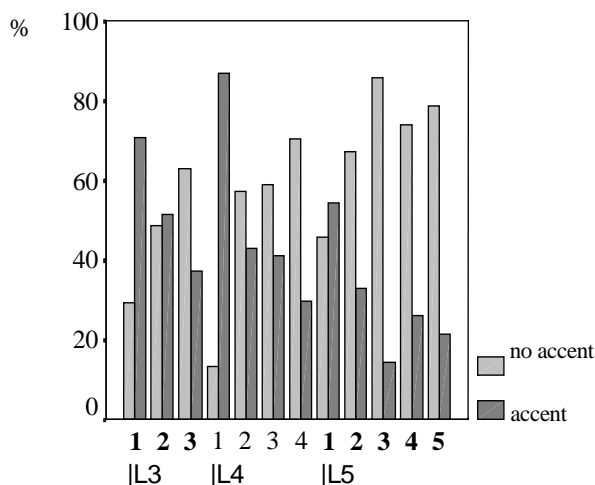
subject judgements of the two back vowels [u:] and [o:] as well as for [ε:] decreases (Fig. 3). We conclude that the subjects can compensate speaker differences by matching stimuli and prototypes.

A comparison of the performance of the NNs and the subjects' judgements (ORL) indicates that the classifications of the NNs are only comparable with the subjects' judgements for Speaker1 whose material was used for training (average hit rate: 72.6% for Speaker1; 52.6% for Speaker2; 50.9% for Speaker3 and average confusion rate of related reduction levels: 20.2% for Speaker1; 37.6% for Speaker2; 30.4% for Speaker3). For the vowels [e:] and [o:] of Speaker1, the hit rate and confusion rate of related levels of the NN differ from those of the subjects' judgements, but the cumulative sum of these two rates is quite similar to those of the subjects' judgements (Fig. 4). Therefore, the misclassification can be regarded as negligible. For [ø:] ( $n=13$ ) the hit rate is comparable to the subjects' agreement, but 30 % are not correctly classified by the NN. This is explainable by the low correlation of the inter-subject judgements. The results indicate that the classification of NNs is comparable to the deviation of the judgements between any two subjects. The comparison of the correlation between perceived reduction levels (ORL) and the levels classified by the NNs with the average inter-subject correlation supports this view. Significant correlation ( $\alpha < .01$ ) between perceived reduction levels and the levels classified by the NNs is found for all vowels, except for [ø:] (Fig. 3).

## 4. PROSODY, WORD CLASS, AND REDUCTION LEVELS

The relationships between reduction levels classified by the NNs (only the material of Speaker1, except classified [ø:]) and prosodic factors (lexical word stress, pitch accent, prominence) as well as word class (content words vs. function words) were also investigated.

For the analyses, the vowels are grouped by their maximal number of reduction levels ([e:] and [u:] with 5 levels (L5), [i:] with 4 levels (L4), the other vowels with 3 levels (L3)). No significant relation between lexical word stress and the reduction levels can be found. However, in accented syllables vowels



**Fig. 5:** Distribution of reduction levels (from unreduced vowels (1) to strongly reduced one (3, 4 or 5)) in accented and unaccented syllables grouped by maximal reduction levels (L3, L4 or L5; s. text).

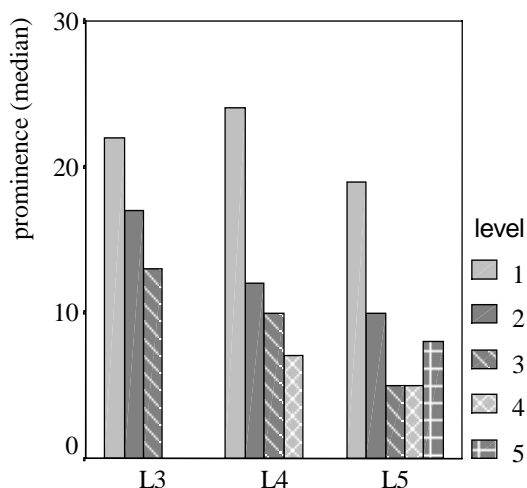
less reduced than in unaccented ones are observed (Fig. 5). The effect of syllable accentuation and reduction level is significant (L3:  $\chi^2=22.63$ ;  $df=2$ ;  $\alpha<.01$ ; L4:  $\chi^2=15.07$ ;  $df=3$ ;  $\alpha<.01$ ; L5:  $\chi^2=17.91$ ;  $df=4$ ;  $\alpha<.01$ ). There is only a significant effect of the dichotomy function vs. content words for L4 ( $\chi^2=12.83$ ;  $df=3$ ;  $\alpha<.01$ ) L5 ( $\chi^2=24.70$ ;  $df=4$ ;  $\alpha<.01$ ). In comparison to content words, function words tend to have more strongly reduced vowels. Examining the relation between perceived prominence of syllables and the reduction levels of vowels, it is found that a syllable is perceived to be less prominent if its vowel is more reduced (for L5 this is not so obvious; Fig. 6). The investigation of reduction level depending on prominence and word class shows that function words with unreduced vowels are perceived to be as prominent (median=20,  $n=75$ ) as content words (median=22,  $n=91$ ) with unreduced vowels.

## 5. DISCUSSION

The aim was to investigate a method for labelling vowel reduction in terms of levels. The reduction of a vowel was judged by matching stimuli to prototypes according to their qualitative similarity. The assumption is that vowel realisations have the same reduction level as their chosen prototypes. The results were investigated according to the inter-subject agreement and compared with the performance of artificial neural networks. Furthermore, the transferability of the reduction levels to other speakers was tested.

The experiments show that subjects can reliably assign stimuli to the prototypes of most vowels. They use these prototypes speaker-independently. NNs trained with the material from one speaker can reliably reflect subjects' judgements of this speaker. The classification is comparable to the deviation of the judgements between any two subjects. Listeners compensate speaker-characteristic variations of vowel reduction, but for NNs a training run with material of other speakers is necessary.

Analyses of the relation between reduction level and prosody as well as word class indicate that the more reduced a vowel is the more its prominence decreases. Accented syllables have more unreduced vowels than unaccented ones. There are more strongly reduced vowels in function words than in



**Fig. 6:** Relation between prominence and reduction level (from unreduced (1) to strongly reduced vowels (3, 4 or 5)) grouped by maximal reduction levels (L3, L4 or L5; s. text).

content words, but function words with unreduced vowels are perceived to be as prominent as content words.

The results show that a description of reduction in terms of levels is possible, and indicate the feasibility of automatic labelling of vowel reductions.

## REFERENCES

- [1] Lindblom, B. (1963). Spectrographic study of vowel reduction. *Journal of the Acoustical Society of America*, 35, 1773-1781.
- [2] Bolinger, D. (1972). Accent is predictable (if you're a mind-reader). *Language*, 48, 633-644.
- [3] Van Bergem, D. R. (1993). Acoustic vowel reduction as a function of sentence accent, word stress, and word class. *Speech Communication*, 12, 1-23.
- [4] Rietveld, A. C. M. and Koopmans-van Beinum, F. J. (1987). Vowel reduction and stress. *Speech Communication*, 6, 217-229.
- [5] Van Bergem, D. R. (1995). Perceptual and acoustic aspects of lexical vowel reduction, a sound change in progress. *Speech Communication*, 16, 329-358.
- [6] Aylett, M. and Turk, A. (1998). Vowel quality in spontaneous speech: What makes a good vowel? *Proc. of the 5<sup>th</sup> International Conference on Spoken Language Processing*, Sydney Australia
- [7] Heuft, B.; Portele, T.; Höfer, F.; Krämer, J. Meyer, H.; Rauth, M. and Sonntag, G. (1995). Parametric description of F0-contours in a prosodic database. *Proceedings of the XIIIth International Congress of Phonetic Sciences*, 378-381, Stockholm: KTH.
- [8] <http://www.phon.ucl.uk/home/sampa/german.htm>
- [9] ESPS (Version 5.0), Entropic Research Laboratory
- [10] Stöber, K.-H. (1997). Unpublished software.
- [11] Pols, L. C. W.; van der Kamp, L. J. T. and Plomp, R. (1969). Perceptual and physical space of vowel sounds. *Journal of the Acoustical Society of America*, 46, 458-467.
- [12] Lienert, G. A. and Raats, U. (1994) *Testaufbau und Testanalyse*, Psychologie Verlags Union: Weinheim, 5. edition

This work is funded by the Deutsche Forschungsgemeinschaft (DFG) under grant HE 1019/9-1.