Rate Dependent Vowel Reduction in German

Benjamin Weiss
Humboldt University Berlin

Abstract

The relationship between local speaking rate and acoustic vowel reduction is examined for a corpus of spontaneous speech. Therefore, the variation in speaking rate was not given by instruction but reflects speakers' choice. Automatically measured first and second formant frequencies of German monophthongs are statistically analysed with regard to vowel duration as well as another local measure of speaking rate mainly based on syllable rate. Both measures show comparable effects.

For most of the treated vowels, significant centralization occurs with higher rates. This result applies to intra-speaker rate variation. It cannot be explained by stress or differences between function words or content words, as these factors have been controlled. It can be shown that this vowel reduction is an effect of increased co-articulation by reanalysing data in one condition. In that way, one effect of decentralization is explained.

There is also significant spectral vowel reduction by fast speakers compared to slow speakers, but not to a degree comparable to intra-speaker effects. Both results cannot be compared properly, because rate variation between speakers is much lower than within speakers.

1 Introduction

There are many studies concerned with the relationship between vowel formant frequencies and vowel duration as a measure of local speaking rate. At a first glance, there is a tendency, that with shorter durations, formant values measured in the middle of vowels are more centralized. A side effect of this result is a smaller acoustic space in the F1-F2 plane. However, there is strong evidence, that this tendency is caused by averaging over different contextual conditions. In fact, with higher rates co-articulation increases. Most striking, there are some contextual conditions, where this increased co-articulation effectively leads to decentralization [1, 2]. But such effects do not become visible in averaged data. By analysing formant transitions instead of steady states it is confirmed that higher rates lead to increased contextual assimilation, albeit shorter vowels also show steeper transitions [3].

To explain this increased co-articulation linguistic factors like stress or paralinguistic factors like speaking style are given [4, 1, 5]. Unstressed vowels and those uttered in a less formal style are more reduced and shorter. This effect has to be distinguished from phonological reduction like schwa realizations of unstressed vowels.

But there are conflicting results regarding rate dependent vowel reduction within such conditions as stress or style. In some studies strong vowel reduction is found [2, 1, 6, 7], while results of others show no or only minimal effects [4, 8, 9, 10].

Recent studies could not solve this contradiction. Data from large corpora of French and German journalistic speech show strong centralization for both languages [11]. The authors present evidence for French, showing that this centralization is a result of contextual assimilation. However, stress is not considered there. In another study, there are only small changes in vowel steady state formants of one female speaker due to rate or consonantal context, although these factors systematically influenced segment duration. Instead, a significant target undershoot was measured for read sentences compared to isolated words [12].

To conclude, it is still unclear if within different linguistic conditions local changes in speaking rate co-occur with vowel reduction. The aim of this study is to clarify this question. The effect of speaking rate on vowel articulation does not seem to be obligatory [13]. Therefore, it is important to see if such an effect appears within natural rate variation instead of induced speaking rate variation.

Additionally, there is evidence, that rate variation within speakers has to be separated from variation between speakers, as there seems to be no clear effect for the latter [14]. Therefore, both sources of rate variation are considered here.

2 Material

Recordings from 32 speakers (14 women, 18 men) are analysed [15]. The corpus contains transcribed dialogues of a made up appointment scenario resulting in conversations with a semi-spontaneous speaking style.

Vowel segment identifiers are collected with *Emu* [16], along with corresponding information about speaker and gender, formant frequencies (automatically measured with *Praat*), stress, part of speech (function vs. content words), local speaking rate and duration. Formant frequencies are extracted at 40%, 50% and 60% of the vowel duration, averaged, and transformed to Bark [17]. To capture rate information, the Perceived Local Speaking Rate (PLSR [18]) is used. This is a linear combination of Hann windowed reciprocal syllable and phone durations with more emphasis on syllable rate. The PLSR was chosen because it highly correlates ($R = 0.9$) with user judgments in a listening test for German speech.

Only for content words stress is annotated. Items with secondary stress are not included, because the total number of those syllables is low. Items directly neighboring pauses and hesitations are also excluded to avoid possible systematical influence in these positions. The same is done for items marked as nasal or creaky voiced, because automatic formant tracking might fail here. Altogether, more than 25000 items are taken into account.

3 Results

Both measures of local speaking rate, vowel duration and PLSR, do not differ in their significant results. Of course, the exact values and the sign of the estimates are different. To save space, only exact values for PLSR are given here.

3.1 Intra-personal rate effects

This section deals with intra-speaker effects. Do formant values differ for relatively fast produced vowels compared to relatively slow produced ones of one speaker? The average natural occurring rate variation is about 1.35 syllables per second (standard deviation (SD)). This is quite high, as it reflects only rate variation within different conditions of vowel class, stress and part of speech.

Both, formant frequencies and rate information are Z-transformed for every speaker and vowel class. Gender effects are also tested. Separate regression analyses for every vowel class, formant, gender, stress and part of speech show significant formant changes for higher rates on a 1% level for most of the vowels considered. For [ø] and [yː], there are not enough items within each condition, [ɪ], [ʊ], [ʏ] show no rate dependent changes at all.

The results are shown in Fig. 1 to Fig. 3. Instead of Z-values, which are difficult to present, results are given in Hz. Means of subject means of formant frequencies for slow vs. fast items are printed. Slow and fast items are < -1 SD and $> +1$ SD of relative rate, respectively. Hence, the figures only illustrate the results with meaningful values. Labels are in SAMPA, with added "f" for "fast" and "s" for "slow". As only significant results are printed, many fast vs. slow items differ only in one formant.

The directions of formant frequency changes for higher rate point in most cases towards the center of the vowel space. Exceptions are central vowels [ə] and [ɐ] (Fig. 2), which show lower F1 values. One effect of decentralization appears in F2 for stressed [ɔ] for both gender (Fig. 1), accompanied by centralization in F1.

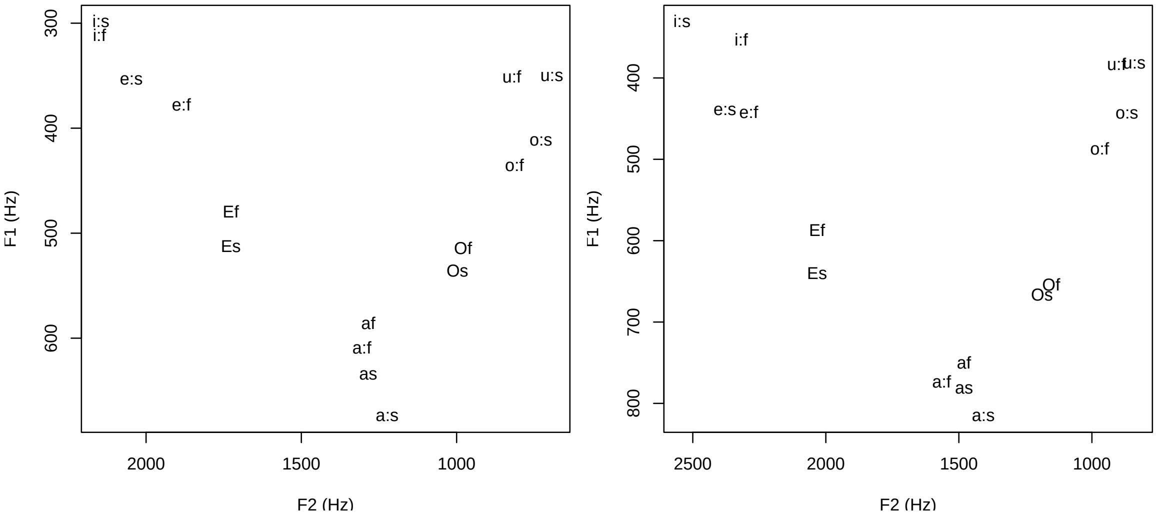

Figure 1: stressed vowels; men (left), women (right)

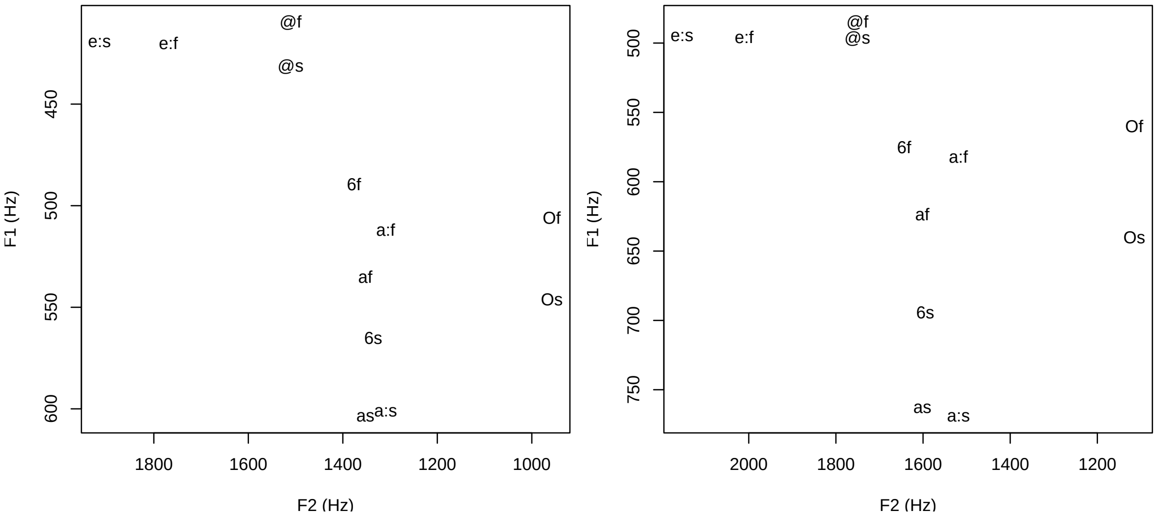

Figure 2: unstressed vowels; men (left), women (right)

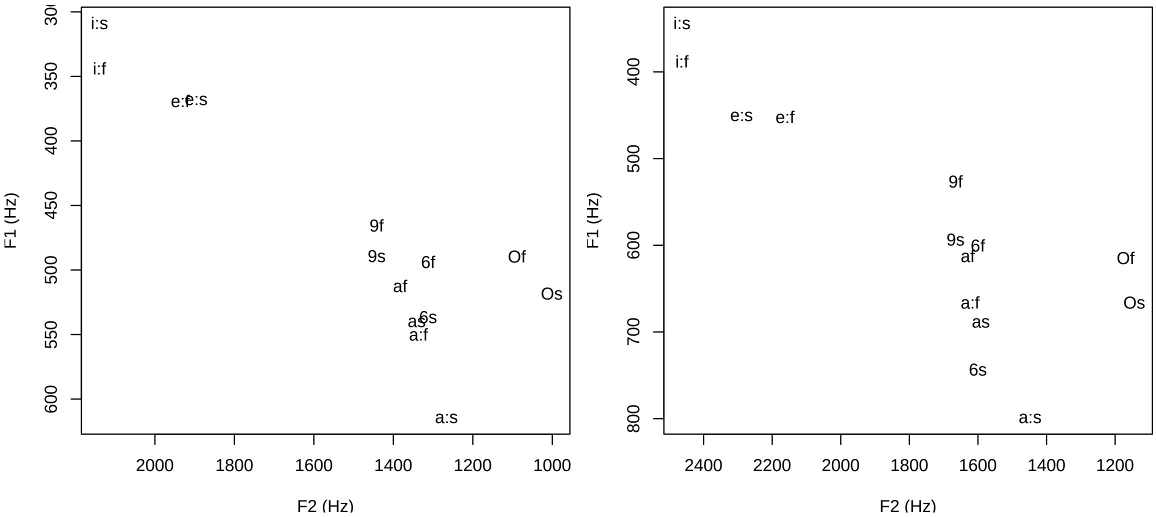

Figure 3: vowels in function words; men (left), women (right)

3.2 Centralization or increased co-articulation?

The results presented in the previous section (3.1) do not reflect differences in consonantal context. In order to test if the general tendency to centralize with higher relative rates results from averaging over these contexts, as mentioned in the introduction, one condition is examined more closely.

F2 of [ɔ] in stressed content words is chosen, because there is an effect of decentralization, that has not been explained yet (see Fig. 1). As F2 for [ɔ] has already extreme values on average (about 1000 Hz for men, 1150 Hz for women), systematic differences in the way F2 changes with rate for different consonantal contexts could be regarded as strong cue for interpreting the results in the previous section as increased co-articulation instead of centralization.

Those items with alveolar contexts (which should increase F2 because of a higher F2 locus [19]) are compared to labial and velar contexts, as pure labial contexts do not occur in the data.

Those items with positive Z-values are considered as fast and compared to slow ones (negative Z-values), because the low number of items in these conditions. Separate one sided Mann-Whitney tests for both contexts show significant changes (5% level) in F2 in Bark: For the non-alveolar context (222 items for men, 153 for women) fast items have lower F2 as slow ones. The opposite effect, an F2 increase for fast items is true for alveolar contexts. This is significant for women (106 items), but there is only a tendency for men ($p = 0.09$).

3.3 Inter-personal rate effects

This section deals with the question, whether spectral differences do not only occur for rate variation within speakers, but also between naturally slow and fast speaking subjects.

Differences in mean speaking rate between speakers in this corpus are small (SD is about 0.8 syllables per second). Therefore, only the 10 slowest and 10 fastest speakers are compared. Even then, these groups differ in their rates on average only 1.22 syllables per second; the mean SD within both groups is 0.64. Compared to within-speaker rate variation, the difference between fast and slow items is about ½ (compared to 2 times SD of 1.35).

As local speaking rate varies strongly within one speaker (see section 3.1), this variability has to be minimized for comparing speaker specific (global) rates. Only items within $\pm$ 1 SD of rate about the individual mean in each condition are used. Additionally, speaking rate is used as control variable to exclude effects from relative PLSR. Speakers are modelled as random factor. As there are few speakers in every group, they are not separated in different tests for gender, but this variable is included into the analyses. Gender is of course highly significant for formant values, but results show an interaction with another variable in only one condition ([aː] in function words).

On a 1% significance level, there are significant differences between the slow and fast group for 9 conditions. These are presented in Table 1. The formant for which the significant effect occurs is given (F1 or F2), as well as the condition examined: Content words are separated for stressed (s) and unstressed (u), function words (F) are not. The estimate is in Bark because in these tests men and women are not treated separately. It shows the deviation in the mean formant frequency of the fast group from the slow group. The above mentioned interaction of factor gender means, that the significant effect for F1 for [aː] in function words is only valid for women.

Most of the significant results are centralizations for the fast group of speakers compared to the slow group. The exception is [ə], that shows lower F1 as in section 3.1. The significant changes in F2 for [ə] cannot clearly be classified because the absolute values of both groups are located in the area of the center of the vowel space.

4 Discussion

The presented analyses show, that formant frequencies differ significantly and systematically for variation in rate. More importantly, these effects are not forced by given rate variation as

Table 1: Significant changes in formant frequencies between slow and fast speakers.

vowel	condition	F1/F2	estimate	F(DF)
aː	u	F1	−0.20	21.95(16)
	F	F1	−0.53	19.02(7)[a]
a	s	F1	−0.08	9.69(16)
	F	F1	−0.07	11.66(16)
eː	s	F2	−0.40	25.70(16)
ɪ	u	F2	−0.15	23.08(16)
œ	F	F2	−0.50	19.84(14)
ə	u	F2	−0.33	11.84(16)
	F	F1	−0.38	17.63(16)

[a] Only for women.

it is done in some experiments, but occur naturally, for one specific speaking style and within linguistic conditions of stress and part of speech. Data in other studies are more precise because of phonologically controlled items, but are produced for example in citation style and therefore are not directly conferrable to communicative situations.

Considering within-speaker effects first, it can be shown that most of the vowels examined are affected in at least one condition. Both, vowel duration and PLSR do account for the presented effects.

The results, averaged over different consonantal contexts, are mostly centralizations for higher rates. Another measure of vowel reduction, the so called vowel space of cardinal vowels, was not tested for, as its shrinking for higher rates is eminent with these results. Exceptions to this centralization are both central vowels [ə] and [ɐ] as well as [ɔ] in one condition. The latter effect could be allocated to consonantal context, despite one non-significant result for men. The additional analysis (section 3.2) also confirms in line with other studies mentioned in the introduction, that in fact the tendency to centralize has to be taken for increased co-articulation.

On this basis, the first mentioned effects – lowering in F1 for both central vowels – do not have to be considered as exceptions. However, more detailed analyses would be desirable to examine consonantal context more closely. This has not been done here, as the focus of this study lies on the question, if there are rate effects at all within the above mentioned categories for naturally produced rate variation.

By comparing within-speaker and between-speaker effects, it becomes visible, that inter-personal rate variation is much less than intra-personal variation (about ½). The estimates are not comparable, as they vary for each vowel and condition. Because of the small difference in speaking rate between both groups of speakers, these results do not provide a clear picture for inter-personal rate effects. It can be stated that for the kind of speaking style examined, there are some speaker specific vowel reduction with rate. This does at least not support the results in [14].

One can argue, that the presented results are not rate induced, which has not been tested here. They might be explained by other factors that influence both, local rate and spectral reduction, like positional effect, as found for consonants [20].

Also, it is not tested yet, if these rate dependent vowel reductions are relevant for perception. Indeed, some of the results are rather small in amount and may therefore have no effect on identification, but perhaps on the coherence of natural speech.

References

[1] S. J. Moon and B. Lindblom. Interaction between duration, context, and speaking style in English stressed vowels. *JASA*, 96(1):40–55, 1994.

[2] B. Lindblom. Spectrographic study of vowel reduction. *JASA*, 35(11):1773–1781, 1963.

[3] M. Pitermann. Effect of speaking rate and contrastive stress on formant dynamics and vowel perception. *JASA*, 107(6):3425–3437, 2000.

[4] T. Gay. Effect of speaking rate on vowel formant movements. *JASA*, 63(1):223–230, 1978.

[5] D. R. van Bergem. Acoustic vowel reduction as a function of sentence accent, word stress, and word class. *Speech Communication*, 12:1–23, 1993.

[6] O. Engstrand and D. Krull. Determinants of spectral variation in spontaneous speech. In *Proc. of Speech Research, Budapest*, pages 84–87, 1989.

[7] G. S. Turner, K. Tjaden, and G. Weismer. The influence of speaking rate on vowel space and speech intelligibility for individuals with amyotrophic lateral sclerosis. *Journal of Speech and Hearing Research*, 38:1001–1003, 1995.

[8] L. C. W. Pols and R. J. J. H. van Son. Acoustics and perception of dynamic vowel segments. *Speech Communication*, 13:135–147, 1993.

[9] M. Fourakis. Tempo, Stress, and Vowel Reduction in American English. *JASA*, 90(4):1816–1827, 1991.

[10] L. C. W. Pols and R. J. J. H. van Son. Formant frequencies of dutch vowels in a text, read at normal and fast rate. *JASA*, 88(4):1683–1693, 1990.

[11] C. Gendrot and M. Adda-Decker. Impact of Duration on F1/F2 Formant Values of Oral Vowels: An Automatic Analysis of Large Broadcast News Corpora in French and German. In *9th European Conference on Speech Communication and Technology Lisbon*, pages 2453–2456, 2005.

[12] J. W. Stack, W. Strange, J. J. Jenkins, W. D. Clarke III, and S. A. Trent. Perceptual invariance of coarticulated vowels over variations in speaking rate. *JASA*, 119(4):2394–2405, 2006.

[13] J. E. Flege. Effects of speaking rate on tongue position and velocity of movement in vowel production. *JASA*, 84(3):901–916, 1988.

[14] Y.-C. Tsao and G. Weismer. Interspeaker variation in habitual speaking rate: Evidence for a neuromuscular component. *Journal of Speech and Hearing Research*, 40:858–866, 1997.

[15] IPDS. *Kiel Corpus of Spontaneous Speech*. CDROMs #IPDS, Kiel, 1995–1997. vols. 1–3.

[16] J. Harrington and S. Cassidy. Multi-level annotation in the emu speech database management system speech communication. *Speech Communication*, 33:61–77, 2001.

[17] H. Traunmüller. Analytical expression for the tonotopic sensory scale. *JASA*, 88(1):97–100, 1989.

[18] H. R. Pfitzinger. Local Speech Rate Perception in German Speech. In *Proc. of the 14th ICPhS, San Francisco*, pages 893–896, 1999.

[19] P. Delattre, A. M. Liberman, M. Alvin, and F. S. Cooper. Acoustic loci and transitional cues for consonants. *JASA*, 27(4):769–773, 1955.

[20] R. J. J. H. van Son and L. C. W. Pols. An acoustic description of consonant reduction. *Speech Communication*, 28:125–140, 1999.