

Beat Siebenhaar, Martin Forst, Eric Keller

Prosody of Bernese and Zurich German. What the development of a dialectal speech synthesis system tells us about it.

Abstract

In this article, we motivate the use of speech synthesis as a tool for research on dialectal prosody. The general architecture of our speech synthesis systems (LAIPTTS) is presented and the choice of data for building a dialectal system is justified. Here we present an analysis of a speaker of the Zurich dialect and of a speaker of Bernese origin. These analyses show that differences in timing cannot be related to a different phoneme distribution in the two dialects. We thus have good reasons to assume that rhythm and intonation do differ between these dialects. In our data, differences of segment duration affect mostly vowels. Our data also indicate that accentuation as well as the grammatical vs. lexical status of the word in which the segment occurs affect the realisation of a segment differently for the speaker of Bernese and Zurich German. Another salient difference can be noted in the way phrase boundaries are marked in the two dialects. The speaker of Bernese German marks phrase boundaries mainly by a lengthening of the elements of phrase-final syllables, whereas the speaker from Zurich shows less lengthening of phrase-final syllables, but greater lengthening of elements in phrase-initial syllables. Despite these first results that suggest truly prosodic differences between Bernese vs. Zurich German, further research is necessary in order to clearly identify dialectal features and to distinguish them from idiolectal or stylistic ones.

1. Introduction

In this article, we present first insights into Bernese and Zurich German prosody that have been obtained in a Swiss National Fund study on dialectal prosody. These two dialects are each centres of a wider dialect area with clearly distinct dialectal features on the segmental level, in morphology, in syntax and in the lexicon, which are described in the Linguistic Atlas of German speaking Switzerland (SDS 1962–1997). With respect to prosody, there is only one description of Bernese intonation compared to northern German intonation patterns (Fitzpatrick-Cole 1999). The aim of our study is to lay the foundations for a systematic investigation of the contribution of different prosodic aspects to the peculiarity of a dialect. Initially, it consists of building two modular synthesis systems, one for Bernese and one for Zurich German. On the one hand, these systems will let us compare globally the prosody of the two dialects. On the other hand, they will let us change (synthesised) utterances in a consistent fashion to investigate dialectal variation of phrasing, timing and into-

nation. We are currently still in the analysis phase of the overall project, but we wish to present some initial results here.

The article is organised as follows: Section 2 motivates the use of speech synthesis as a tool for research on dialectal prosody. Section 3 introduces the general architecture of the speech synthesis systems developed at the LAIP (Laboratoire d'Analyse Informatique de la Parole), which will also be the base for the dialectal speech synthesis. In section 4 we justify the choice of the data for our dialectal speech synthesis. Section 5 presents a first analysis of timing aspects of these Bernese and Zurich German data. Finally, section 6 concludes the paper with an outlook concerning the use of the emerging dialectal speech synthesis system in linguistic experiments.

2. Speech synthesis as a tool for linguistic research

The traditional scientific procedure for obtaining information consists of analysing data and explaining the findings concerning a certain phenomenon. The advantage as well as the disadvantage of this procedure is that apparently non-relevant phenomena are excluded. For instance, most intonation research puts aside timing phenomena on the phonetic level. As a consequence, it deals mainly with f0 contours, without taking possible interactions between rhythm and intonation into account. Specific contours are selected from a corpus; they are described and compared to other contours in similar and different environments. They are classified into highs and lows, early and late highs, and boundary tones. A sophisticated description of a specified aspect of intonation is obtained, but quite often the repercussions that phonetic rhythmic or segmental phenomena may have on such f0 contours are neglected. For example it is not taken into account that a different duration of the same phoneme in two dialects may have a direct influence on the f0 contour, just because a longer sound allows a greater modification of f0.

Using synthesis as a scientific method allows us, indeed forces us, to model all aspects of the domain, as well as interactions between the different subsystems that are typically set aside in an exclusively analytic procedure. For the investigation of prosody, this means that the design of a synthesis system has to take into account rhythmical phenomena just as much as intonation, and that it has to reflect the interrelation between the different linguistic levels. Therefore a speech synthesis system can be used to test the interplay of segmental and suprasegmental information, of phrasing, timing and intonation.

It is in this sense that we wish to use speech synthesis for conducting research on dialectal prosody. In a synthesis system, each prosodic parameter can be modified independently of all others, while maintaining the interactions between different aspects of prosody. This allows us to test hypotheses concerning the relevance of these different aspects for the naturalness of speech as well as claims as to their language-specific or language-independent nature. Sound examples can be generated with different models at each linguistic level, and their importance can be evaluated by submitting them one by one to perception tests.

3. Architecture of the LAIPTTS systems

As the design of our dialectal synthesis systems will follow the 'LAIP tradition' of speech synthesis concerning major theoretical decisions, we will briefly present the basics of the psycholinguistically and statistically motivated models of phrasing, timing, intonation and their interplay that Keller and Zellner elaborated for French (Keller and Zellner 1996) and that was subsequently adapted to German (Siebenhaar et al. 2001). They are for the most part implemented in the French speech synthesis system LAIPTTS_F and in its German counterpart LAIPTTS_D. Since the first adaptation from French to German was reasonably successful, we are now building the Bernese and Zurich German synthesis systems along the same lines.

The input to our prosodic module is the phonetic chain, annotated for word and syllable boundaries, as well as for the grammatical or lexical status of words. This phonetic chain is split up into prosodic phrases. Augmented with information concerning phrase boundaries, it is the basis for the calculation of the duration of the single segments. As a final step, the f_0 contour is calculated on the basis of the phonetic material, the corresponding durations and phrase boundaries.

Thus it is assumed in our systems that the phrasing component precedes the temporal calculation component, and that the temporal calculation component precedes the intonation component. This conception is based on the fact that any human action process is necessarily embedded in a temporal structure. Furthermore it has been demonstrated that the durational domain is subject to more rigid constraints than the f_0 domain. For example, it was shown by Keller (1994, based on data by Caelen-Haumont 1991) that within a given syllable, duration correlates much more between speakers than f_0 .

Due to this linear succession of speech generation modules, our speech synthesis models can represent influences of phrasing on segment duration, and influences of timing on intonation. But there is no influence from intonation to phrasing or segment duration in this architecture. The model is therefore a linear model without recursivity. The different components of the model will now be explained in detail.

3.1 Phrasing

Speaking is constrained by human cognitive and physiological abilities. For this reason, human beings structure utterances into intelligible parts that can be handled by these abilities. The most evident structuring elements in speech are pauses and they can also be marked in the temporal structure of the utterance by the speeding up or slowing down of syllables. Psycholinguistic tests for English and French (Gee and Grosjean 1983; Keller et al. 1993; Zellner 1994, 1997a, 1997b, Zellner Keller 2002) have shown that phrasing can only partially be predicted from syntax. A more adequate prediction is achieved with a psycholinguistically motivated model that splits up sentences into rhythmically balanced phrases. This means that phrases tend to be of similar length and can hardly ever be longer than a certain number of syllables, which seems quite a plausible assumption, considering cognitive and articulatory constraints. For French, this number rarely exceeds 12 (Zellner

1997b, 1998), and according to our data the same holds for German and the Swiss German dialects (cf. fig. 9). That does not mean that syntactic and psycholinguistic aspects are in complete disaccord, but where the principles are in conflict, the psycholinguistic model is often more adequate (Zellner 1997a).

Based on statistical tests of these psycholinguistic principles, Keller and Zellner Keller developed a word grouping algorithm for French (Keller et al. 1993; Keller and Zellner 1996), which was refined for different speech rates (Zellner 1998) and adapted for German by Siebenhaar (Siebenhaar et al. 2001).

This algorithm developed for read speech was adapted for spontaneously spoken language. The breaks generated are mostly reasonable, but they often do not fit the data of the dialects as there are much more pauses and breaks. Therefore the algorithm was a valuable basis for further work on spontaneous speech but again it had to be adopted. This time it seems less to be an adaptation to a new language but more an adaptation to a new style.

3.2 Temporal calculation

For the calculation of the segment durations, we rely mainly on statistical models built on manually labelled corpora. These are general linear models (GLM) that use input parameters such as the durational class of the current segment and the surrounding segments, the structure of the syllable the segment occurs in, the grammatical status of the word, the position of the segment within the syllable, within the word and within the prosodic phrase (Keller et al. 1993; Keller and Zellner 1996; Zellner 1998, Siebenhaar et al. 2001). Depending on the language these parameters may differ somewhat. A statistical analysis first has to select these parameters among all variables that are supposed to have a potential influence on segment durations. Part of such an analysis for the dialectal synthesis system is presented in section 5.

3.3 Intonation

F0 contours are calculated with a superpositional Fujisaki model (first presented in Fujisaki and Hirose 1982). The model implements relatively slow phrase commands to determine the general intonation contour in a prosodic phrase on the one hand, and relatively fast accent commands on the other. The resulting curves of both kinds of commands are summed up and result in the final f0 contour. For the analysis, the position, duration and slope of these commands provide a mathematical description of concrete f0 contours. In a next step these factors are correlated to the linguistic structures, and the results are implemented into the model. The output of the model – the generation of a concrete f0 contour on the base of a linguistic description – is then compared to the original contours, which gives the background for a refinement of the model.

4. Building a speech synthesis system for dialectal variants

These models are now being adapted to a synthesis of Bernese and Zurich German. As these Swiss German dialects are not written languages like standard French and standard German, the new model will simulate quite a different style of speech. The ‘neutral’ news reading style modelled by the French and the German system would not be natural in these varieties. Hence, not only are we dealing with new varieties, but we also have to model a different style of speech, which probably has a considerably greater inherent variation in prosodic terms than the ‘neutral’ news reading style.

The style we have decided to analyze in order to build our synthesis models for Bernese and Zurich German is the one of public interviews. Its main advantage is to be naturally embedded in a communication situation while at the same time being characterized by a certain degree of formality that prevents excessively ‘exotic’ prosodic patterns. In our opinion, this style thus combines naturalness and a certain degree of formal control of the language.

More precisely, the data we are now using to build our models are an interview in Bernese of about 20 minutes and two interviews of about 50 minutes altogether with a speaker of Zurich German. All three interviews were recorded in a studio, which resulted in a style that was spontaneous but still formal. Nevertheless, the two speakers differ stylistically in that the speaker from Zurich talks with many pauses, but with few self-corrections, while the speaker from Berne formulates less carefully and has many self-corrections.¹

We are aware of the fact that it is problematic to rely on a single speaker for each dialect. Prosody, even within a single variety, is characterised by a great variability, whose causes can be inter- or intraindividual. In analytic work, researchers mostly investigate a representative sample of speakers in order to counter-balance those factors. Speech synthesis, on the other hand, usually relies on one or a few typical cases. This corresponds to the reference speaker method of traditional dialectology, which excludes interindividual variation.

For any work based on a single source, the selection of the source is one of the most delicate issues. As long as we have only little data on the prosody of Bernese and Zurich German, let alone securely representative data, the choice of the source can hardly be justified. Consequently we have simply taken two speakers whose dialects can clearly be assigned to one of the two dialectal areas on the segmental level.

¹ This differences result in a lower speech rate of the speaker from Zurich compared to the speaker from Berne. This observation is in contrast to the stereotype that the Berne dialect is a slower dialect than the Zurich dialect, a stereotype that is supported by duration measurements by Löffler (1984) of texts recorded in the 1940s (Der sprechende Atlas, published again: Phonogrammarchiv 2000).

5. Bernese and Zurich German timing – Comparison of timing aspects

This section shall present first results from the analysis of two speakers of Bernese (BE) and Zurich (ZH) dialect. The results of these analyses will be the basis for the dialectal speech synthesis system. They show which aspects have to be implemented in a dialectal speech synthesis. The analyses also show differences between the two speakers that can give hints to the aspects that may be important to distinguish between the two dialects. The data for the analysis are the recordings mentioned above. These recordings were labelled with the help of an automatic aligning programme and were then corrected manually. They comprise – excluding pauses – 7847 segments for the Bernese part and 15,017 segments for the Zurich German part.

5.1 Syllable structure

The structure of the syllable in which a segment occurs has an influence on its duration. For this reason, we first present the distribution of different syllable types in our data. Table 1 and 2 show the percentages of C-initial vs. V-initial syllables and of open vs. closed syllables, respectively.

Table 1: Distribution of CV(C) / V(C)-syllables

	Total percent	BE percent	ZH percent
CV(C)	89.50	87.88	90.39
V(C)	10.50	12.12	9.61
Total	100.00	100.00	100.00

The Bernese recording shows a slightly higher proportion of syllables beginning with vowels (Table 1). The difference is weak, but statistically significant². Swiss German does not show the glottal stop. Therefore, in contrast to standard German, VC-syllables are possible, even if relatively rare because of a general tendency towards CV-syllables. Every so often, an originally syllable-final consonant is ‘pulled over’ to the next originally vowel-initial syllable, where it becomes the syllable onset and thereby turns it into a CV(C) syllable. This principle of syllable structures is of higher priority than in standard German. In standard German, the morphological or even word structure is mostly maintained while in Swiss German, the tendency to CV-syllables quite often breaks the morphological and word structure (Nübling/Schrambke i. pr.).

² Contingency coefficient = .039, Chi-square-p = .0008.

Table 2: Distribution of (C)VC / (C)V-syllables

	Total percent	BE percent	ZH percent
(C)VC	44.15	42.19	45.24
(C)V	55.85	57.81	54.76
Total	100.00	100.00	100.00

Table 2 shows that open syllables are somewhat more frequent in Bernese German than in Zurich German, a difference that is not significant, however. Consequently, we do not expect it to have an influence on segment durations.

Table 3: Distribution of vowels/consonants

	Total percent	BE percent	ZH percent
Consonants	66.74	65.34	67.47
Vowels	33.26	34.66	32.53
Total	100.0	100.00	100.00

Table 3 shows the distribution of consonants and vowels in both recordings. Bernese German has a slightly higher proportion of vowels than Zurich German. The correlation is significant, but very weak.³

Table 4: Frequency distribution of the single vowels

	Total %	BE %	ZH %	Frequency relation (=BE/ZH; 100 = no difference)
ə	21.841	19.180	23.320	82.25
i / ɪ	20.824	21.729	20.321	106.93
a	15.674	18.293	14.218	128.66
o / ɔ	8.781	10.384	7.890	131.61
u / ʊ	7.368	6.430	7.890	81.50
e/ ε	7.131	7.280	7.047	103.31
æ	5.559	5.580	5.548	100.58
y / ʏ	3.631	3.215	3.863	83.23
Diphthongs	3.565	4.398	3.103	141.73
Falling diphthongs	2.483	1.035	3.287	31.49
ø/ œ	2.113	1.220	2.609	46.76
Syllabic consonants	.832	.998	.740	134.86
Nasalized vowels	.198	.259	.164	157.93
Total	100.000	100.000	100.000	

Table 4 shows that the individual vowel classes in both dialects are similarly distributed. The table does not take into account differences in phonological systems. Because of the different distribution of open and closed vowels in the two dialects, the variants of the vowels concerned are taken together. The most remarkable difference between the dialects appears with the schwa. Schwa occurs approximately 4 % more frequently in Zurich than in

³ Contingency coefficient = .02, Chi-square-p = .0012.

Bernese German and it is hardly ever compensated by syllabic consonants. Other differences show up only in the smaller classes further down in the table containing diphthongs, falling diphthongs, with $\emptyset/\text{œ}$, the syllabic consonants and nasal vowels.

With consonants, differences in the frequency distribution are even smaller (Table 5). This reflects the relative stability of the consonant system for all Swiss German dialects⁴. Only the category of the semivowels shows an important difference. This difference is due to the vocalisation of /l/, a typical feature of most western Swiss German dialects, which the Bernese dialect belongs to. Besides, Bernese German shows a higher proportion of fortis fricatives.

Table 5: Frequency distribution of the single consonant classes

	Total %	BE %	ZH %	Frequency relation (=BE/ZH; 100 = no difference)
Fortis fricatives	18.526	20.259	17.651	114.78
Lenis plosives burst	14.176	14.885	13.817	107.73
Lenis plosives occlusion	13.169	12.963	13.273	97.66
Nasals	12.366	11.924	12.589	94.72
Fortis plosives/affricates occlusion	12.287	11.120	12.876	86.36
Fortis plosives burst	8.424	8.119	8.578	94.65
r	9.181	8.688	9.429	92.14
Affricates	4.014	3.412	4.319	79.00
l	3.659	3.216	3.883	82.82
Lenis fricatives	2.869	2.765	2.922	94.63
Semivowels	1.270	2.589	.604	428.64
Aspirated plosives	.059	.059	.059	100.00
Total	100.000	100.000	100.000	

Thus it becomes clear that differences in prosody cannot be attributed to a fundamentally different relationship of the individual sounds. Such differences do exist, but they are only of marginal importance. Hence, prosodic differences must be explained by other aspects, as a different control of timing and intonation.

5.2 Timing

In order to represent the differences in the segment duration between the dialects, analyses of variance (ANOVA) were carried out. These require normally distributed data. The following histograms of the durations of the individual segments, however, show a strongly left-skewed distribution (Figure 1, left panels).

⁴ The wordbook of Swiss German dialects (Schweizerisches Idiotikon 1882 ff.) does not enter the words in alphabetical order because of the many vowel differences between the dialects. The words are ordered following their consonant skeleton, that is much more similar among the dialects. The atlas of Swiss German dialects (SDS 1962–1997) shows differences in the vowel system in one and a half volumes while differences in the consonant system only cover half a volume.

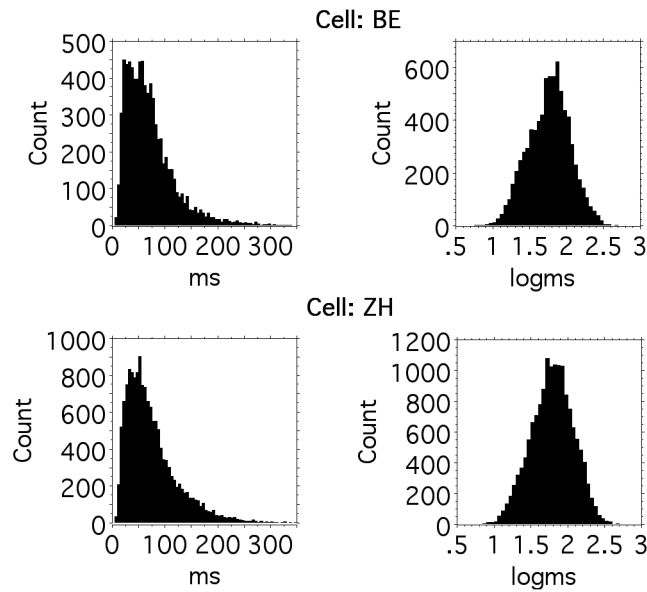


Figure 1: Histograms of segment durations in ms (left panels) and in log ms (right panels) for Bernese (top, $n = 7847$) and Zurich data (bottom, $n = 15017$).

Many studies have shown that segment durations are approximately normally distributed in the logarithmic space (cf. Zellner 1998; Riedi 1998; van Santen 1998). Figure 1 (right panels) shows that this is also the case with our data. We thus continue with the analysis of the logarithmic values of our data.

Our analysis shows that intrinsic segment duration can explain a great proportion of general timing variation. First we will present the duration distribution of some selected phoneme classes and then continue with influences of suprasegmental aspects on syllable, word and phrase level. At this place we refrain from presenting the multiple factorial ANOVA because of the multiple interactions of the ten different factors that will make a presentation confusing. Instead we present results of mostly only three-dimensional analyses, which show differences between the selected phonetic factors and between the speakers of the two dialects.

5.2.1 Segmental level

We have already pointed out that the subject from Zurich speaks more carefully than the Bernese speaker. So, against the stereotype, the segment durations of the Zurich German recording are significantly longer than the ones of the Bernese recording. However, the difference does not affect all segments to the same extent. Figure 2 shows that differences

occur particularly between vowels, which are significantly longer for the Zurich speaker, while the consonants do not show any significant differences between speakers⁵.

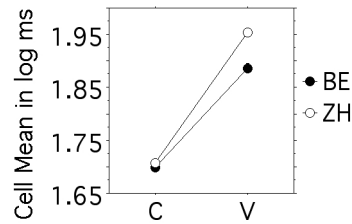


Figure 2: Cell mean and 95% confidence interval (almost not visible) in log ms: Segment duration by sound class (vowel/consonant) and by dialect (BE/ZH)

Figure 3 shows the duration of the plosives. Lenes and fortes are depicted separately and for both, occlusion and burst are separated. Both lenes and fortes have an occlusion that is longer than the burst. The Zurich speaker differs from the Bernese speaker in all aspects except of the burst of the lenes (Le-Pl-Bu). In spite of the generally longer segment durations of the Zurich speaker, his occlusions of fortes and lenes are significantly shorter than those of the Bernese speaker (Fo-Pl-Ok/Le-Pl-Ok), while the burst of the fortes is shorter for the Bernese speaker (Fo-Pl-Bu).⁶ A particularly interesting thing to notice is the relationship between the bursts of fortes and lenes. The Zurich speaker distinguishes fortes and lenes not only in the occlusion, but also in the burst. For the Bernese speaker, bursts of lenes and fortes coincide but, on the other hand, the difference in the duration of the occlusion is clearer.

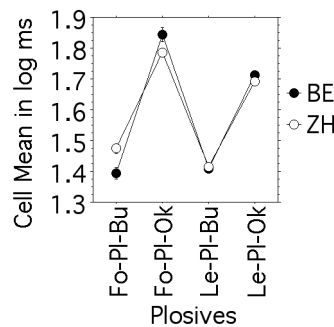


Figure 3: Cell mean and 95% confidence interval in log ms: Duration of the plosives by fortes (Fo-)/lenes (Le-) plosives (-Pl-) and each burst (-Bu) and occlusion (-Ok) by dialect (BE/ZH)

⁵ The difference between the two speakers is not significant for consonants (t-test: $p = .052$), while it is highly significant for vowels (t-test: $p < .001$).

⁶ For fortis and lenis in Zurich German see Willi (1996).

Figure 4 shows the distribution of the vowel durations by phonological length. The Zurich speaker shows four significantly different classes, while the Bernese speaker shows no difference between diphthongs and long vowels⁷. Comparing the speakers, the differences are highly significant for short vowels and diphthongs. Schwa-vowels and long vowels do not differ significantly⁸. So the slower speech rate of the Zurich speaker is mainly due to longer short vowels and diphthongs.

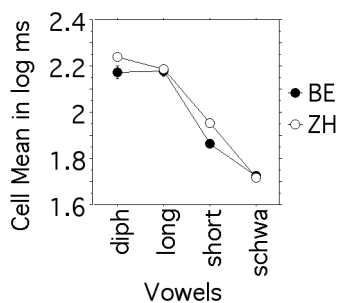


Figure 4: Cell mean and 95% confidence interval in log ms: Segment duration of vowels by phonological length and by dialect (BE/ZH)

5.2.2 Syllable level

Figure 5 shows the differences in the realization of consonants and vowels by accentuation. Preliminary examination has shown that it is useful to distinguish between syllables with schwa, non-accented syllables with full vowels, and accented syllables with full vowels. These diagrams point out that the differences between the Bernese and the Zurich speaker are not very important concerning consonants. Both speakers realize consonants in schwa syllables significantly shorter than in syllables with full vowels. Consonants in accented and non-accented syllables are not differentiated. The situation looks different with regard to the vowels: While schwas are equally long for both speakers, the durations between accented and non-accented syllables differ significantly. In both cases, the realisations of the Zurich speaker are significantly longer than those of the Bernese speaker.⁹

⁷ For the Zurich speaker all differences between single classes are highly significant ($p < .0001$). For the Bernese speaker long vowels and diphthongs show no significant difference, while the other classes also show highly significant differences ($p < .0001$).

⁸ Differences between two speakers for diphthongs: $p < .001$, for short vowels: $p < .001$. For long vowels ($p = .4517$) and for schwa ($p = .5322$), there is no significant difference.

⁹ t-test for difference between Bernese and Zurich speaker for consonant duration in schwa-syllables: $p = .60$; in unaccented syllables $p = .0490$; in accented-syllables: $p = .0090$. t-test for difference between Bernese and Zurich speaker for vowel duration in schwa-syllables: $p = .54$; in unaccented syllables: $p < .0001$; in accented-syllables: $p < .0001$.

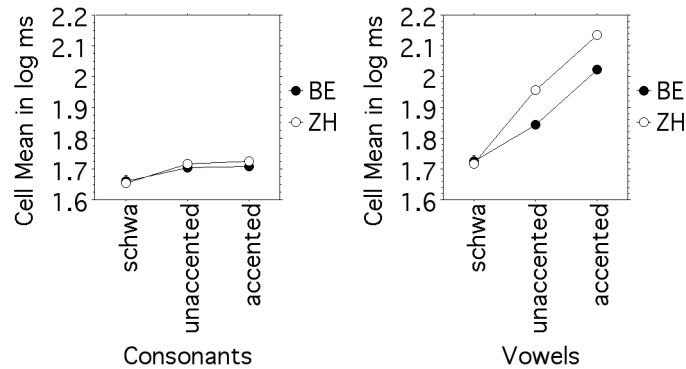


Figure 5: Cell mean and 95% confidence interval in log ms: Segment duration by accentuation and by dialect (BE/ZH) for consonants (left) and vowels (right)

Figure 6 shows the length of the consonants by position in the syllable. As in other languages and varieties (French: Keller & Zellner 1996, German: Riedi 1998, 52; Siebenhaar et al. 2001, 169), consonants in Swiss German are shorter in syllable onsets than in syllable codas. The figure reveals that this difference is much stronger for the Zurich speaker than for the Bernese speaker. The two speakers do not differ with respect to the consonants in the onset, the difference in the coda, however, is highly significant.

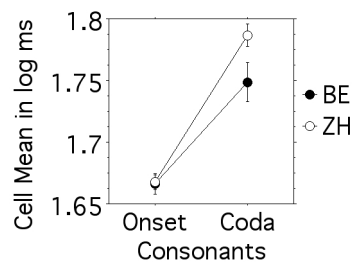


Figure 6: Cell mean and 95% confidence interval in log ms: Segment duration of the consonants by position in the syllable (Onset/Coda) and by dialect (BE/ZH)

5.2.3 Word level

In most languages, the distinction between lexical and function or grammatical words is an important factor for segment duration (cf. Riedi 1998 for German, Zellner 1998 for French, van Santen 1998, 137 ff. for English, Mandarin Chinese, French and German). This is also true for the two dialects examined: As usual, segments are shorter in grammatical words than in lexical words. Figure 7, which only considers vowel duration, confirms the difference between vowels in lexical and grammatical words for both dialects.

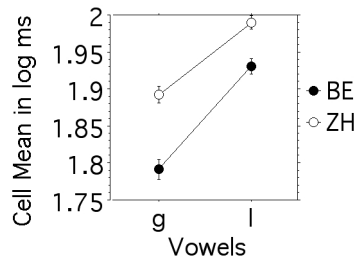


Figure 7: Cell mean and 95% confidence interval in log ms: Segment duration of vowels by grammatical/lexical words (g/l) and by dialect (BE/ZH)

In addition, the diagram suggests that the Bernese speaker makes a greater difference between vowels in lexical and grammatical words than the Zurich speaker. Beyond the general tendencies, finer differences appear when the accentuation of the syllables is taken into account. Figures 8a-c show that, depending on word accent, both speakers arrange the relation of the vowel durations in lexical and grammatical words differently.

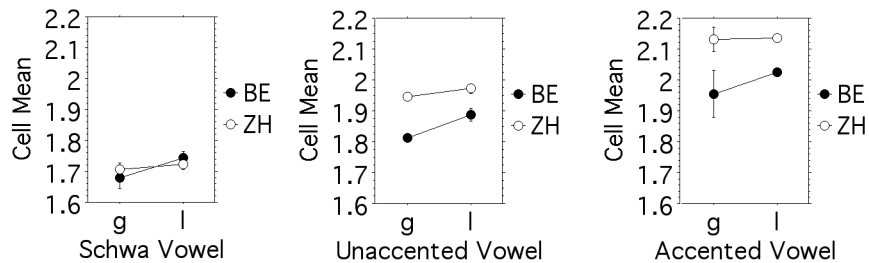


Figure 8: Cell mean and 95% confidence interval in log ms: a) Segment duration of schwa-vowels (left), b) non-accented vowels (centre), and c) accented vowels (right) by grammatical/lexical words (g/l) and by dialect (BE/ZH)

The Bernese speaker shows a greater difference of segment duration between the two word classes than the Zurich speaker. For the Bernese speaker, these differences are highly significant for schwa syllables and non-accented syllables and significant for stressed syllables. For the Zurich speaker, the difference is significant only for non-accented syllables. The fact that the values collapse for the Zurich speaker can be explained by his more careful articulation. This careful articulation puts up a resistance to a stronger reduction of syllables also in the grammatical words.

This comparison exemplifies the complex interaction of the different factors. Any of them shows significant differences as main effect, and the interaction between the single linguistic factors and dialect is also significant, while the interaction between grammatical status and accent, as well as the interaction of all three factors are not significant.

5.2.4 Phrase level

Prosodic phrasing substantially contributes to the global impression of the prosody of a language. Here we only look at major phrases. For our study we defined these phrases perceptually on the basis of perceived pauses and of the resetting of the intonation curve. Table 6 shows that the Zurich speaker has significantly shorter phrases than the speaker from Berne.¹⁰

Table 6: Phrases: Count, mean number of syllables, standard deviation, and standard error.

	Count	Mean nr. of syllables	Std. Dev.	Std. Err.
BE	609	4.48	2.81	.114
ZH	1265	3.83	2.55	.072

The distribution of the phrase length shows some differences (Figure 9). The modal value for both speakers is at 2 syllables. The steeper slope of the Zurich distribution conforms with the data of Table 6. They show that the Zurich speaker has fewer longer phrases and therefore a lower mean of phrase duration.

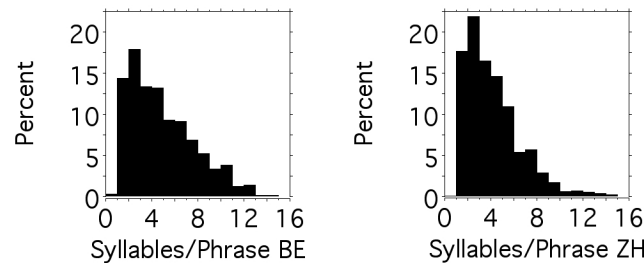


Figure 9: Histograms of the number of syllables per phrase for the Bernese (left panel) and the Zurich speaker (right panel).

Phrasing influences segment durations, as syllables at the end of a phrase normally are lengthened, which may be a language-universal phenomenon (Maddieson 1997, 631 f.). So one could expect that the differences shown up to now are due to the fact that the Zurich speaker has shorter phrases and therefore more syllables with phrase-final lengthening. Comparing vowel length with respect to the syllable position in the phrase shows that the differences between the two speakers remain. Figure 10 shows that in any position, i. e. one-syllable phrases, first, middle, penultimate and last syllables of a phrase, the vowels of the Zurich speakers are longer than those of the Bernese speaker. In the penultimate position, these differences are significant, while they are even highly significant in any other position. So we can state that the Zurich speaker really has a slower speech rate – a fact that does not reflect the stereotypes. The picture, however, also reflects that both speakers show a similar behaviour in vowel lengthening. Compared to the phrase-middle syllables, both the ultimate and penultimate syllables are lengthened. For both speakers the differences of these three positions are significant. For the Zurich speaker the change between the three

¹⁰ Unpaired t-test: Mean Difference = .722, DF = 1808; t-value = 5.537; $p < .0001$.

positions is similar. For the Bernese speaker the durations of vowels in the next to last syllable are closer to those of the last syllable.

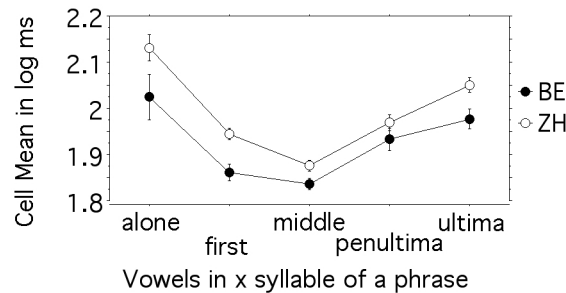


Figure 10: Cell mean and 95% confidence interval in log ms: Segment duration of the vowels by position in the phrase and by dialect (BE/ZH)

In phrase-initial syllables, vowels are also lengthened. The difference to the vowels in phrase-medial syllables is clearer for the Zurich speaker than for the Bernese speaker. For the Zurich speaker, vowels in phrase-initial syllables are almost as long as vowels in the penultimate syllables. For the Bernese speaker, vowels in phrase-initial syllables are clearly shorter than vowels in penultimate syllables.

As single-syllable phrases show the longest vowels, it seems that both final and initial lengthening are added. These differences do affect all vowels regardless of their accentuation. But for accented syllables alone, the interaction between dialect and syllable position is not significant.

In consonants, a slightly different pattern emerges (cf. Figure 11). Compared to the Bernese speaker, the consonants of the Zurich speaker are longer only in one-syllable phrases and in the first syllable of a longer phrase, while in other positions they are even shorter than those of the Bernese speaker. The differences are significant for the single-syllable phrases and the penultimate syllable¹¹.

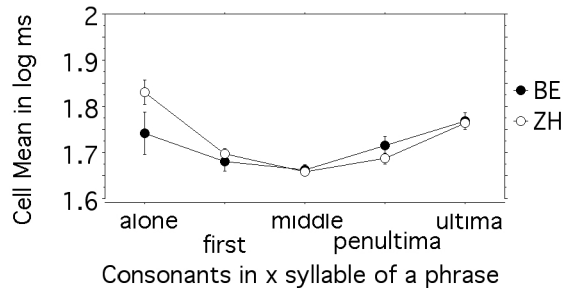


Figure 11: Cell mean and 95% confidence interval in log ms: Segment duration of the consonants by position in the phrase and by dialect (BE/ZH)

¹¹ t-test for difference of consonant duration between Bernese and Zurich speaker: in single-syllable phrases $p = .0011$, in penultimate syllables $p > .0001$.

The difference in consonant length cannot be explained by the position of the syllable in the phrase alone. When taking into account the position of the consonants in the syllable as well, it appears that the consonants in syllable coda position are quite similar between the two speakers (cf. Figure 12). In fact, there is no statistical difference between the speakers.

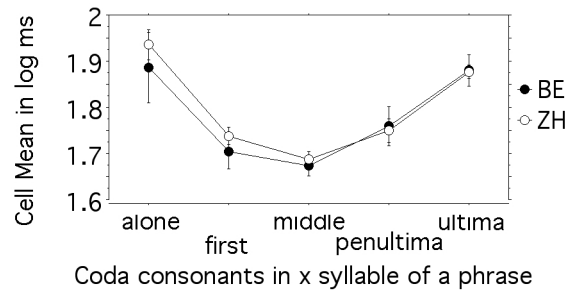


Figure 12: Cell mean and 95% confidence interval in log ms: Segment duration of syllable coda consonants by position in the phrase and by dialect (BE/ZH)

With respect to syllable-initial consonants, however, the two speakers behave differently (Figure 13). The Bernese speaker has no significant lengthening of consonants in the syllable onset of phrase initial syllables. The Zurich speaker shows a clear lengthening in these positions. On the other hand, the final lengthening of syllables affects Bernese syllable initial consonants more than those of the Zurich speaker. The 'crossing of the lines' in Figure 13 shows that there is an interaction between dialect and position in the phrase. This interaction is significant.

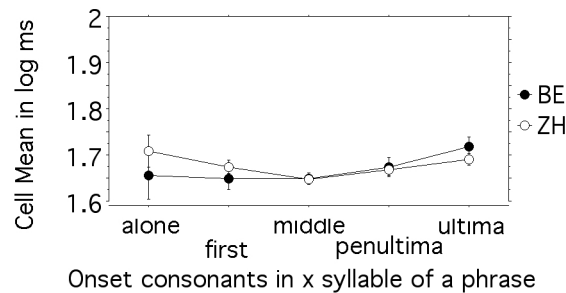


Figure 13: Cell mean and 95% confidence interval in log ms: Segment duration of syllable onset consonants by position in the phrase and by dialect (BE/ZH)

So the two speakers mark phrase boundaries in a slightly different way. They have a similar behaviour for the syllable nucleus and especially for syllable final consonants. At both edges of phrase boundaries, both speakers lengthen these elements. We do not only find a final lengthening but also a phrase-initial lengthening. This result is in contrast to findings on Swiss High German (Siebenhaar et al. 2001) and to findings on French (Zellner Keller 2002), where phrase-initial syllables are reported to be accelerated. Byrd and Saltzman

(1998) observe articulatory lengthening for both phrase-final and phrase-initial positions. For French there are also reports for articulatory initial lengthening of certain, mostly consonantal segments (Fougeron 2001). For Korean (Cho/Jun 2000) domain-initial strengthening is observed for consonants. These results are all based on read speech. The difference of the results may therefore be a stylistic difference of read and spontaneous speech that may have some psycholinguistic basis in the sense of increased semantic and syntactic processing times required by spontaneous utterances.

The longest syllables in our data are those in single-syllable phrases. It seems that here phrase-final and phrase-initial lengthening coincide.

One main difference between the speakers lies in the syllable onset consonants, which are generally shorter than syllable-coda consonants for both speakers. Moreover, the syllable-onset consonants remain short in phrase-initial position for the Bernese speaker. At this position, the Zurich speaker shows the same lengthening as for vowels and syllable coda consonants. On the other hand, we have a contrary behaviour at phrase final syllables. Here, the Bernese speaker shows the 'normal' lengthening of syllable onset consonants, while for the Zurich speaker they remain rather short.

The other main difference in marking phrase boundaries concerns vowel lengthening. While the Bernese speaker has a longer lengthening of the penultimate syllable, the lengthening of the first syllable is smaller. The Zurich speaker, on the other hand, has a clear lengthening of the first syllable nucleus, while the lengthening of the penultimate nucleus is less marked.

These results show that the Bernese speaker tends to mark phrase boundaries more clearly at the end of the phrase, while the Zurich speaker marks them more clearly at the beginning.

6. Conclusion

Because the data analysis is still incomplete, these results are to be considered as preliminary and should mainly point out some aspects of the analysis that have to be taken into consideration when building a speech synthesis system. It has been shown that for many linguistic aspects the two speakers behave similarly, the general tendencies are the same but the distance covered is different. Therefore, all these aspects implemented in the models for read speech also have to be implemented in the synthesis system for dialects, i. e. into the synthesis of a spoken speech variety. On the basis of the analyses shown above, we are now building a statistical general linear model (GLM) for each speaker. As has been shown, the same levels for each factor can be chosen, but the values connected to these levels are different for the two speakers. Therefore, the statistical parameters of these models represent the individual speakers who are – in the classical dialectological sense – representatives for their dialect area.

However similar the levels of these factors are, the varying values of the parameters demonstrate that the temporal organisation of speech of the two speakers shows differences. Generally, the differences are greater for vowels than for consonants, be it in the distinction of lexical and functional words, be it in the distinct accentuation or be it in the distinct dura-

tion regarding the position in a phrase. But even for consonants we could demonstrate some differences between the two speakers; for example, the different relations of occlusion and burst in plosives or the distinct duration of onset consonants in respect to the position of the syllable in the phrase. One of the most remarkable results concerns the marking of phrase boundaries. While the speaker of Bernese German mainly marks the phrase boundary with a lengthening of the elements of phrase-final syllables, the speaker from Zurich shows a less distinct lengthening of these phrase-final syllables but a stronger lengthening of elements in phrase-initial syllables. Very often these differences make up only a few milliseconds. However, some informal tests with French vowels in the LAIP have shown that systematic temporal changes of as little as 2% can be perceived, so the sum of these differences may well constitute a part of the audible difference between the speakers of Berne and Zurich German.

With the data of two speakers presented here it is not yet possible to generalize the results and to make universally valid statements about the difference between two dialects. These results characterise two speakers who have a different dialectal background. Therefore it is not yet possible to distinguish between idiolectal and dialectal features of prosody. Further data and further extended analyses will make such statements possible and it will be possible to build the models for an automated generation of dialectal prosody.

In many articles of the last 10 years, Brigitte Zellner Keller has pointed to the importance of temporal organisation for a natural sounding speech synthesis. From our preliminary results on Berne and Zurich German it has hopefully become clear that also for research on dialectal prosody both intonation and the temporal organisation of speech are important.

As for the ongoing work in view of a dialectal synthesis system, the analyses presented here are to be refined and will lead into a (statistical) model of segment durations for each dialect. Moreover, models for phrasing and intonation are currently being built. The procedure corresponds to the procedure presented here. For intonation, a mathematical description of the f_0 -curve following the approach of Fujisaki is generated (Fujisaki & Hirose 1982). The parameters of this description are then correlated with the linguistic description. For example, the two aspects of a phrase command, its position and its magnitude, are correlated with the linguistic description of the phrase, its length, the length of the previous phrase and other factors. Analyses of variance reveal the important factors and the distinct levels that have to be implemented in a model. After an implementation of these algorithms and the completion of the dialectal diphone database, the computer will be able to read dialectal texts. This will let us verify the adequacy of our models by comparing the output to original data and with perception tests, and we will be able to compare the prosody models of one dialect to the models of the other dialect.

References

- Byrd, Dani/Saltzman, Elliot (1998): "Intragestural dynamics of multiple phrasal boundaries." – In: *Journal of Phonetics* 26, 173–199.

- Caelen-Haumont, Geneviève (1991): *Stratégies des locuteurs et consignes de lecture d'un texte: Analyse des interactions entre modèles syntaxiques, sémantiques, pragmatique et paramètres prosodique*. – Thèse de doctorat d'état: Université d'Aix-en-Provence (unpublished).
- Cho, Taehong/Jun, Sun-Ah (2000): "Domain initial strengthening as enhancement of laryngeal features: Aerodynamic evidence from Korean." – In: *Chicago Linguistics Society* 36, 31–44.
- Fitzpatrick-Cole, Jennifer (1999): "The alpine intonation of Bern Swiss German." – In: J. Ohala (ed.): *Proceedings of the XIVth International Congress of Phonetic Sciences (ICPhS)*, 941–944. San Francisco.
- Fougeron, Cécile (2001): "Articulatory properties of initial segments in several prosodic constituents in French." – In: *Journal of Phonetics* 29, 109–135.
- Fujisaki, Hiroya/Hirose, Keikichi (1982): "Modelling the dynamic characteristics of voice fundamental frequency with applications to analysis and synthesis of intonation." – In: *Preprints of the Working Group on Intonation, 13th International Congress of Linguists*, 57–70. Tokyo.
- Gee, James P./Grosjean, François (1983): "Performance structures: A psycholinguistic and linguistic appraisal." – In: *Cognitive Psychology* 15, 411–458.
- Keller, Eric (1994): "Fundamentals of phonetic science." – In: E. Keller (ed.): *Fundamentals of Speech Synthesis and Speech Recognition: Basic Concepts, State of the Art, and Future Challenges*, 5–21. Chichester: John Wiley.
- Keller, Eric/Zellner, Brigitte/Werner, Stefan/Blanchoud, Nicole (1993): "The prediction of prosodic timing: Rules for final syllable lengthening in French." – In: D. House, P. Touati (eds.): *Proceedings ESCA Workshop on Prosody, September 27–29*, 212–215. Lund, Sweden: Department of Linguistics and Phonetics.
- Keller, Eric/Zellner, Brigitte (1996): "A timing model for fast French." – In: *York Papers in Linguistics* 17, 53–75. University of York.
- Löffler, Heinrich (1984): "Sprechtempo – ein Merkmal der Sprache oder der Sprecher? Beobachtungen und Überlegungen zu einem vernachlässigten Problem." – In: W. Besch, K. Hufeland, V. Schupp, and P. Volker (eds.): *Festschrift für Siegfried Grosse zum 60. Geburtstag*, 111–141. Göttingen: Kümmerle.
- Maddieson, Ian (1997): "Phonetic Universals." – In: W. J. Hardcastle and J. Laver (eds.): *The Handbook of Phonetic Sciences*, 619–639. Cambridge: Blackwell.
- Nübling, Damaris/Schrambke, Renate (i. pr.): "Silben- versus akzentsprachliche Züge in germanischen Sprachen und im Alemannischen." – In: E. Glaser, P. Ott and R. Schwarzenbach (eds.): *Alemannisch im Sprachvergleich*. Stuttgart: Steiner.
- Phonogrammarchiv der Universität Zürich (ed.) (2000): *Der sprechende Atlas. "Gespräch am Neujahrstag" in 24 Dialekten (CD)*. – Zürich: Phonogrammarchiv der Universität Zürich.
- Riedi, Marcel P. (1998): *Controlling Segmental Duration in Speech Synthesis Systems*. – Zürich: TIK. *Schweizerisches Idiotikon. Wörterbuch der schweizerdeutschen Sprache. Gesammelt auf Veranstaltung der Antiquarischen Gesellschaft in Zürich unter Beihilfe aus allen Kreisen des Schweizervolks. Herausgegeben mit Unterstützung des Bundes und der Kantone*. Begonnen von Friedrich Staub und Ludwig Tobler und fortgesetzt unter der Leitung von Albert Bachmann, Otto Gröger, Hans Wanner, Peter Dalcher, Peter Ott. (1881 ff.) – Frauenfeld: Huber.
- Siebenhaar Beat/Zellner Keller, Brigitte/Keller, Eric (2001): "Phonetic and Timing Considerations in a Swiss High German TTS System." – In: E. Keller/G. Bailly/A. Monaghan/J. Terken/M. Huckvale (eds.): *Improvements in Speech Synthesis*, 165–175. Chichester: John Wiley.
- SDS = *Sprachatlas der deutschen Schweiz*. Begründet von Heinrich Baumgartner und Rudolf Hotzenköcherle. In Zusammenarbeit mit Konrad Lobeck, Robert Schläpfer, Rudolf Trüb und unter Mitwirkung von Paul Zinsli herausgegeben von Rudolf Hotzenköcherle. (1962–1997). – Bern: Francke, vol. VII and VIII Basel: Francke.
- van Santen, Jan (1998): "Timing." – In: R. Sproat (ed.): *Multilingual Text-to-Speech Synthesis: The Bell Labs Approach*, 115–139. Dordrecht, Boston, London: Kluwer Academic Publishers.
- Willi, Urs (1996): *Die segmentale Dauer als phonetischer Parameter von "fortis" und "lenis" bei Plosiven im Zürichdeutschen. Eine akustische und perzeptorische Untersuchung*. – Stuttgart: Steiner.
- Zellner, Brigitte (1994): "Pauses and the temporal structure of speech." – In: E. Keller (ed.): *Fundamentals of speech synthesis and speech recognition*, 41–62. Chichester: John Wiley.

- (1997a): “Improving Speech Fluency in French through Psycholinguistic Principles.” – In: F. L. Borchardt/E.M.T. Johnson (eds.): *14th CALICO Annual Symposium, [CD-ROM]*. Durham: CALICO.
- (1997b): “Fluidité en synthèse de la parole.” – In: E. Keller/B. Zellner (eds.): *Les défis actuels en synthèse de la parole*, 47–78. Lausanne: Université de Lausanne.
- (1998): *Caractérisation et prédiction du débit de parole en français. Une étude de cas.* – Thèse de Doctorat. Faculté des Lettres, Université de Lausanne. (http://www.unil.ch/imm/docs/LAIP/pdf.files/Zellner_Dissertation.pdf)
- (2002): “Revisiting the Status of Speech Rhythm.” – In: B. Bel and I. Marlien (eds.): *Proceedings of the Speech Prosody 2002 conference, 11–13 April 2002*, 727–730. Aix-en-Provence: Laboratoire Parole et Langage.