INTONATIONAL AND TEMPORAL FEATURES OF SWISS GERMAN

Adrian Leemann and Beat Siebenhaar

University of Berne, Switzerland
adrian.leemann@isw.unibe.ch, beat.siebenhaar@isw.unibe.ch

ABSTRACT

The study examines the timing of 17 speakers and the intonation of 6 speakers of two Swiss German dialects. Results show that the relative mean duration of segments and final lengthening are only similar in the two dialects observed. A crucial difference is that Valais speakers generally speak at a faster rate. In terms of intonation, the Valais produce more accent commands than the Bernese; largely due to stressing more lexical words than the Valais. Phrase accents are fairly weak as opposed to standard German. The study shows phonetically motivated differences in dialectal prosody.

1. INTRODUCTION

Swiss German dialects have been thoroughly investigated over the past decades with regard to the sound structure, morphology, and the lexicon [9], [8]. Yet, prosodic features of Swiss German have largely been ignored. Two typological descriptions of Bernese and Zurich prosody show a general difference to standard German. Fitzpatrick [2] and Fleischer/Schmid [3] claim that the default pitch accent in both dialects consists of a low-rising contour (L*+H) compared to the Standard German falling accent (H*+L). Earcatching differences between the dialects are not described. This is where our project pitches in.

In recording two dialects we work out a gross linguistic model that is geared at revealing temporal and intonational features of the dialects. The comparison of the informants from each recording location allows for a distinction to be made between region-specific and individual prosodic characteristics.

2. DATA SET & METHODS

It is the aim of this project to study spontaneously produced language of Bernese and Valais Swiss German (Zurich and Chur will follow) on a suprasegmental level. In an interview setting, 41 Gymnasium students were confronted with spontaneous questions largely regarding the subjects' plans after school.

For the study at hand, 17 speakers were recorded up to now; 9 from the Midland dialect of Berne (henceforth BE) and 8 from Brig, representing the Alpine, Valais (henceforth WS) dialect. The sound files were subsequently labelled on a segmental level in PRAAT [1] and annotated with factors that are known to influence the duration of these segments. The database consists of 54'787 articulatory segments without pauses.

For the intonational component 3 WS and 3 BE speakers werde analyzed, pitch values were extracted, and pitch contours were subsequently smoothened. The Fujisaki model [4] is used for the analysis. It describes intonation as superposition of a phrase component (henceforth Ap), representing the temporally stretched F0 alterations, and of an accent component (henceforth Aa), often word accents, which can, however, span over several syllables. We opted for this approach because the model accurately generates and reproduces any given F0 contour through a mathematical formulation. This parametrization is beneficial in that it enables quantification and modeling of the F0 representations. The Fujisaki-parameters were extracted with a tool provided by Mixdorff [6]. The parameters were tied to syllables, which were again manually annotated.

Analyses are executed at various levels. First, ANOVAs reveal regionally different or similar influences. Secondly, generalized linear models, which relate the distribution of a variable to predictor factors with a linear link function, are used to model the distribution of the timing and the Fujisaki-parameters. The models for each individual are then used to cluster the speakers to groups that show a similar prosodic behaviour.

3. RESULTS – TIMING

The following figures show some of the general differences of the dialects. All of these differences are found for all speakers; yet, for the single speaker the differences are not necessarily significant.

Comparing the durations of the nuclei of mid-phrase stressed syllables, the two dialects differ significantly (fig. 1). Only mid-phrase syllables were taken into accout to avoid influence of final lengthening. Not only are the BE vowels longer in general; moreover, there is a difference in the relative duration of the nuclei. In stressed syllables (+), long vowels and diphthongs do not differ significantly in BE, while in WS long vowels are longer than diphthongs. The difference of accented long and short vowels is significant for every speaker.

Figure 1: Duration of nuclei in mid-phrase syllables by type and stress. Diph: Diphthong, LV: long vowel, SV: short vowel, +: stressed, -: unstressed (Diamonds mark mean and 95%-confidence interval).

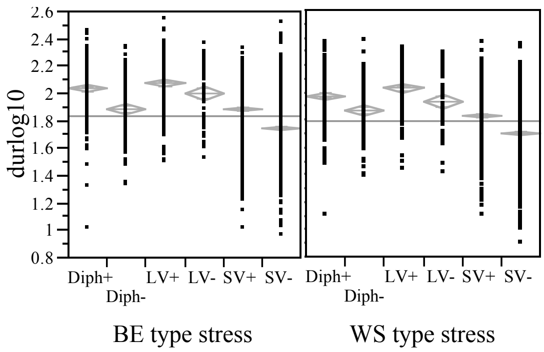
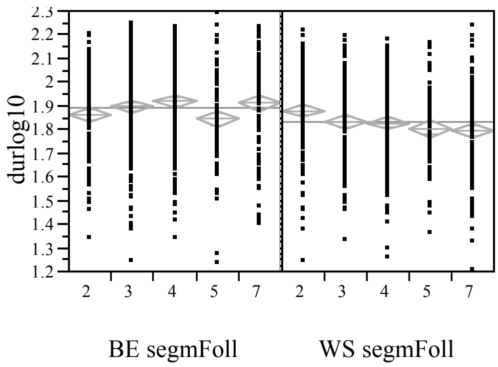


Figure 2: Duration of short vowels in mid-phrase stressed syllables by type of following consonant. 2: liquid 3: nasal 4: fricative 5: lenis plosive closure 7: fortis plosive closure (Diamonds mark mean and 95%-confidence interval).

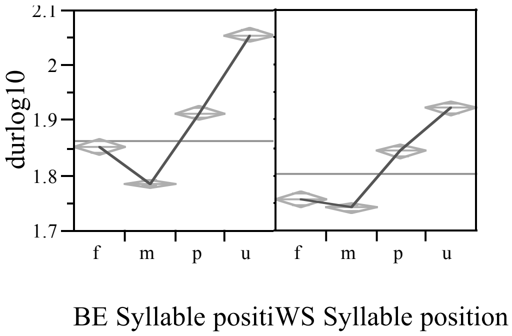


In read speech, influence of the neighboring segments on duration is well described [cf. 7, 10, 12]. In spontaneous speech research is rare. In our corpus, only the effect of the following segment shows a relevant correlation. Because of the influence of stress, phonological length, and final lengthening, only stressed short vowels in mid-phrase syllables are considered. Figure 2 suggests that the vowel is shortened in the WS dialect with a falling sonority of the following segment. This trend is not visible in the BE data. Except for the plosives (type 5 & 7) there is even a lengthening before a more sonorant consonant.

Figure 3 represents the duration of the nuclei dependent on the position of the syllable in the phrase. Four positions are distinguished: first, middle, penultimate and ultimate syllable within the phrase. Final lengthening is apparent and initial lengthening is obvious for the BE dialect, while it is hardly noticable for the WS dialect.

Generally, differences between the positions are more distinct in the BE than in the WS dialect. Compared to nuclei in middle syllables, the penultimate nuclei are lengthened by about 30% in both dialects. However, the BE speakers have more marked final lengthening in the last syllable of a phrase, which is lengthened by 92% compared to only 60% by the WS speakers.

Figure 3: Duration of nuclei by position in the phrase. f: first m: medial p: penultimate u: ultimate (Diamonds mark mean and 95%-confidence interval).



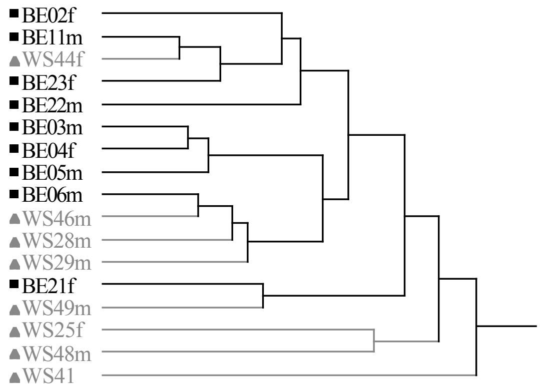
The slower speech rate of the BE speakers is not only due to longer mean duration of the vowels but also to a more distinct final lengthening.

The following parameters were used to model the temporal variation of the speakers by means of a generalized linear model: nucleus type, sonority of consonants, sonority of preceding and following segment, stress, grammatical status of the word containing the segment, assimilated plosives, position of the segment in the syllable, and position of the syllable in the phrase. The BE speakers were modelled more accurately (r=0.68) than WS group (r=0.65). Models for the individual speakers were used to structure the types of speakers.

Figure 4 shows a cluster analysis. At first glance it becomes clear that WS speaker show less coherence among each other than BE speakers do. From left to right the WS speaker converge to common clusters at a later stage; from right to left, they are first to be discarded. BE speakers more often form clusters of more than two members.

However, BE and WS are not always clearly distinct: in three clusters, members of the two dialects are mixed. Hence, BE21f and WS49m are speakers with more disfluencies and final lengthening. WS44f, which behaves similarly to three BE subjects, speaks coherently. BE06m strongly accentuates stressed syllables. Thus, regarding the timing, a group of BE speakers shows common features; and we have other, mainly WS speakers, that are further away from this centre. The center speakers can be seen as typical representatives of the BE dialect and can be described for prosodic features. The situation for the WS speakers is different as they do not form a distinct common temporal behaviour.

Figure 4: Dendrogram of a cluster analysis (Ward method).



4. RESULTS – INTONATION

In this section we will first address the Aas realized by the two speaker groups. In a second step, we will discuss the role of the phrase component in terms of controlling the intonation as well as address a link between Aa and Ap analyses. Finally, these results are commented on in the preliminary conclusion.

4.1. Accent commands

WS speakers are assumed to show more accent commands than BE speakers (see Wipf [11]). This is indeed the case when considering two ways of measurement: Aas/sec. and syllables/Aa: WS speakers realize 2.08 Aas/sec., BE speakers only 1.78. Also, the WS speakers show 2.8 syllables/Aa, the BE speakers 2.97 on average. Even though this is only a minor difference in Aa production, the perception that the WS speakers realize more Aas is emphasized by their higher articulatory speed. The question now remains as to where this higher share of Aas in the WS speaker group stems from.

One reason for a different perception of the groups could be that they have a different distribution of accents on grammatical and lexical words. Yet a contingency test shows that both, BE and WS, put more accents on lexical, as opposed to on grammatical words. The ratio for the WS speakers is fairly different than expected: of all the accent commands, 23% are placed on grammatical words, while that number reaches 35% with the BE. The differences between the two groups are significant (chi square = 12.714, p= 0.0053).

A further reason for the higher number of Aas in the WS speaker group may be that they put more Aas on unstressed and schwa syllables. Table 1 summarizes the findings of this test:

Table 1: Stressed, unstressed, and schwa syllables that are given an Aa by WS and BE group.

	Aa WS	Aa BE
Stressed	54%	53%
Unstressed	39%	41%
Schwa	7%	6%

Both groups show similar distributions. The differences are small yet significant (chi square=12.758, p=0.0258). The main difference between the two groups lies in the number of Aas put on schwa syllables, where the WS group demonstrates 1/6 more Aas than the BE.

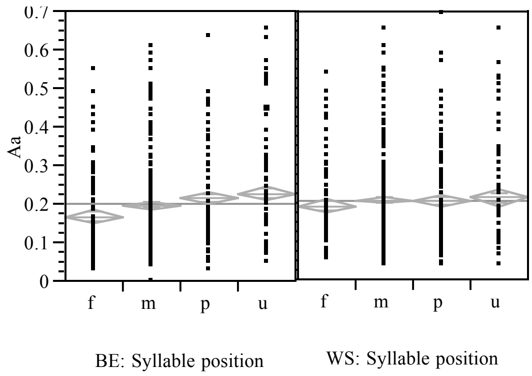
4.2. Phrase commands

If the amplitude of the Ap is high, the phrase component may constitute the major control mechanism for intonation phrasing – if not, phrase structuring seems to be controlled by accent commands instead of global accents. An ANOVA showed that BE21f as well as BE11m show virtually identical amplitudes; the same goes for WS49m and WS44f, which also form their own group. While WS46m and BE04f are significantly different from each other and from the rest of the speakers. When looking at the amplitudes in general, the amplitudes are not distinctive – in contrast to standard German for example [5]. The only speaker with a relatively high phrase command is WS46m.

4.3. Interaction between Ap and Aa

To support the above claim that the phrase component is not the main cue for phrasing in the two dialects, we tested if there is a correlation between the existence of Aas and the syllable position within the phrase. Both speaker groups exhibit the highest number of Aas in first syllable position. The BE put Aas on 47% of all first syllables – the WS on 45% of all the first syllables. When looking at the amplitude of the Aa and how it correlates with the syllable position of the phrase we get the following picture: The BE speakers show Aas with higher amplitudes the further back in the phrase the Aa occurs. This difference is significant ($F=8.7$, $p<.0001$ cf. fig. 5). The WS speakers show the same tendency, however, the differences are not significant ($F=1.11$, $p= 0.3425$).

Figure 5: Aa amplitude by position in the phrase. f: first m: medial p: penultimate u: ultimate (Diamonds mark mean and 95%-confidence interval).



It is noteworthy that the type of the break (terminal or continuing) does not show a significant effect on the value of the Aa in either dialect.

5. CONCLUSION

The timing results show similar tendencies for both dialects concerning the relative mean duration of segments or final lengthening. However, there are also general differences. One of these key differences is higher speech rate in WS. Also to be mentioned: the different relation of accented and non-accented nuclei, the more distinct final lengthening in BE or the different influences of neighboring segments. The distinct prosodic perception of the two dialects is therefore based not on major differences, except for speech rate, but rather on a slightly different parameter setting in many factors. Overall, speakers can be arranged by their timing information, however, the clustering only restrictedly reflects dialectal distinctions.

On the intonation level, WS speakers produce slightly more Aas. For one, lexical words are carrying more Aas in the WS group, which in turn suggests that lexical words are marked more rigorously. Secondly, the group demonstrates slightly more Aas on schwa syllables as opposed to the BE group. With regard to Ap, it appears that, due to its low amplitude values, phrase structuring must be predominantly governed by local Aas and not global phrase commands in the sample at hand. Lastly, the BE show a significant increase in Aa amplitude towards the end of the phrase. Thus, it can be assumed that the penultimate and ultimate syllables are crucial for the determination of the phrase – the WS reveal similar tendencies, yet not in a statistically significant way.

This last aspect corresponds to the regionally different phrase marking in timing, other direct correlations can be found. As the intonation data are not yet present for all speakers, structuring as with the timing data could not yet be performed. However, perceptual differences in intonation hint towards an investigation of microprosodic variation. Nevertheless, the analyses performed up to now shed a new light on differences in dialectal prosody that are not phonologically but phonetically motivated, which are worth pursuing.

6. REFERENCES

[1] Boersma, P., Weenik, D. Praat. http://www.fon.hum.uva.nl/praat/ visited 5-Mar-07

[2] Fitzpatrick-Cole, J. 1999. The alpine intonation of Bern Swiss German. *Proc. 14th ICPhS* S. Francisco, 941–944.

[3] Fleischer, J., Schmid S. 2006. Zurich German. *Journal of the International Phonetic Association* 36/2, 243–253.

[4] Fujisaki, H. 1988. A note on the physiological and physical basis for the phrase and accent components in the voice fundamental frequency contour. In: Fujimura, O. (ed.), *Vocal physiology: voice production, mechanisms, and functions*. New York: Raven, 347-355.

[5] Mixdorff, H. (1998). *Intonation Patterns of German*. Unpublished PhD Theis. TU Dresden. http://www.tfh-berlin.de/~mixdorff/thesis/index.html, visited 5-Mar-07.

[6] Mixdorff, H. Fujisaki Parameter Extraction Environment. http://www.tfh-berlin.de/~mixdorff/thesis/fujisaki.html visited 5-Mar-07

[7] Riedi, M. 1998. *Controlling Segmental Duration in Speech Synthesis Systems*. Zürich: vdf Hochschulverlag.

[8] *Schweizerisches Idiotikon. Wörterbuch der schweizerdeutschen Sprache*, 1881 ff. Frauenfeld: Huber.

[9] *SDS = Sprachatlas der deutschen Schweiz*, 1962–1997. I–VI Bern, VII–VIII Basel: Francke.

[10] Siebenhaar, B. et al. 2001. Phonetic and Timing Considerations in a Swiss High German TTS System. In: Keller, E. et al. (eds), *Improvements in Speech Synthesis*. Chichester: Wiley, 165–175.

[11] Wipf, E. 1910. *Die Mundart von Visperterminen im Wallis*. Frauenfeld: Huber

[12] Zellner Keller, B. 2002. Revisiting the Status of Speech Rhythm. *Proc. Speech Prosody 2002* Aix-en-Provence, 727–730.