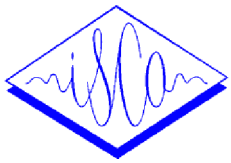
ISCA Archive
http://www.isca-speech.org/archive



Speech Prosody 2008
Campinas, Brazil
May 6-9, 2008

Swiss Alpine and Midland Intonation

Adrian Leemann and B. Siebenhaar

Department of Linguistics
University of Berne, Switzerland
{adrian.leemann; beat.siebenhaar}@isw.unibe.ch

Abstract

The present paper discusses the intonational features of 3 Swiss German dialects: Valais Swiss German (WS), representing the Alpine variety, and Zurich (ZH) and Berne (BE) Swiss German, representing the Midland dialects. By application of the Fujisaki intonation model, 24 speakers of the mentioned dialects are investigated according to their global as well as local intonational features. The BE and WS show similar intonational features in 4 out of 9 observed variables. Yet the WS group seems to distinguish itself with a high pitch range; the BE, on the other hand, is especially characterized with a late pitch onset in local accents. The ZH speakers, in contrast, stand out in that they show distinct intonational features which are largely unparalleled in the other 2 dialects. The paper ends with an outlook of a practical extension to the Fujisaki model, the integration of a slow-rise component – a component introduced by Mixdorff in writing, yet never implemented into his Fujisaki parameter extraction program.

10.21437/SpeechProsody.2008-64

1. Introduction

Over the past decades, there have been numerous studies that addressed Swiss German dialects. The bulk of this research is contained in the *Linguistic Atlas of German-speaking Switzerland* (Sprachatlas der deutschen Schweiz) [10] as well as in the *Swiss Idiotikon* [9]. These works specifically focus on the sound structure, morphological structure, and on the lexicon of Swiss German dialects. In terms of prosody, however, only little research has been conducted.

Early dialect monographs, collected in the *Contributions to Swiss German Grammar*, impressionistically describe prosodic features of a number of Swiss dialects. In the last century, Wipf [11], for example, portrays WS Swiss German as a "singing", "bouncy" dialect, which puts its word accents "onto the most insignificant of syllables". Apart from such sketchy accounts of the melodic features of Swiss German dialects, there are only a handful of academic pursuits which aimed at investigating Swiss German prosody in a more scientific manner.

Fleischer and Schmid [3] specifically examined the vowel and consonantal inventory of ZH Swiss German. In passing, they noted that not much research was conducted with respect to ZH Swiss German intonation. Nevertheless, they added that ZH Swiss German seems to display a large overall pitch range and a default pitch accent that consists of a low-rising contour. This finding goes hand in hand with that of Fitzpatrick-Cole [2]: She concluded that another Swiss German dialect, BE Swiss German, too, shows a default pitch accent L*+H, as opposed to the default Standard German pitch accent H*+L.

Because of this evident lack of research into Swiss German suprasegmentals, we are currently working in a *Swiss National Science Foundation* Project, entitled "Quantitative Approaches to a Geolinguistics of Swiss German Prosody". The project started in November 2005 and will reach its end in November 2008. The research design of the present project rests upon the methodologies used in the predecessor project, which, by means of synthetic modeling, aimed at a preliminary suprasegmental description of two dialects with the speech of 3 Swiss German speakers [6].

2. Method

The present study is based on a corpus of spontaneous speech, retrieved through spontaneous interviews. 3 locations were chosen: Berne and Zurich, both Midland dialects; yet, Zurich stands for an Eastern, Berne for a Western dialect. Additionally, Brig, located in the Southwestern Valais region of Switzerland, serves as the basis for our Alpine data in the present paper. At a later stage in our project, we will also analyze data from the second Alpine dialect which we recorded: Chur, which is located in the Southeastern part of Switzerland. A total of 80 Gymnasium students were recorded. The intonational data of 24 speakers (10 BE, 10 WS, and 4 ZH speakers) will be presented in this paper.

The students were asked to answer questions such as "What do you intend to do after the Baccalaureate?". We used R-9 and Marantz PMD 671 recorders, subsequently transcribed, segmentally labeled in Praat [1], and annotated, i.e. enriching the data with factors which are known to have an effect on intonational representation. Pitch values were extracted, followed by a smoothening of the pitch contours in Praat [1]. The Fujisaki model was applied for the intonation analyses. At this stage of our research, 8 female and 16 male speakers were analyzed. On average, we investigated 3 minutes and 2 seconds (excluding pauses) of spontaneous speech per speaker. Or, put differently, 990 syllables (including filled pauses) per speaker.

2.1. Fujisaki parametrization

The Fujisaki model [4] views intonation as the superposition of two components, a phrase component (Ap) and an accent component (Aa). The phrase component models the temporally stretched intonation movements, while the accent component represents local accents, often word accents, which can, however, stretch over a number of syllables. Put into formula, the Fujisaki model is governed by the following equations (adopted from Mixdorff [7]):

$$\ln F_0(t) = \ln Fb + \sum_{i=1}^{I} Ap_i Gp(t-T_{0i}) + \sum_{j=1}^{J} Aa_j [Ga(t-T_{1j}) - Ga(t-T_{2j})]$$

$$Gp(t) = \begin{cases} \alpha^2 t \exp(-\alpha t), & \text{for } t \geq 0, \\ 0, & \text{for } t < 0. \end{cases} \quad (1)$$

$$Ga(t) = \begin{cases} \min\left[1 - (1+\beta t)\exp(-\beta t), \gamma\right], & \text{for } t \geq 0, \\ 0, & \text{for } t < 0. \end{cases}$$

The phrase control mechanism, *Gp(t)*, is the impulse response mechanism. In our approach, the natural angular frequency α is speaker-specific, taking on a value between 0.2 and 3.0. In *Ga(t)*, the step response of the accent control mechanism, β, its natural angular frequency, was chosen as a constant throughout the speakers, with β = 20. *Fb* is the minimum value of a speaker's fundamental frequency.

The parameters for the phrase commands, *Ap* (phase command amplitude) and *T0* (impulse of the phrase command), as well the parameters for the accent commands, *Aa* (accent command amplitude) and *T1* (starting point of the accent command) and *T2* (end point of the accent command), are often referred to as Fujisaki parameters. They are extracted with a tool provided by Mixdorff [8]. Subsequently, the parameters are tied to syllables, where each command is interpreted linguistically. Most often, Aps are tied to pitch resets, i.e. phrase initiations, while Aas mark local pitch accents.

The reason why we opt for the Fujisaki model is twofold, the primary reason being that the model allows for the generation of virtually any given F0 contour through a mathematical formulation – which in turn permits quantification of intonational events. Secondly, only a handful of parameters are needed to adequately generate a given F0 contour.

3. Results

Here, the most relevant results of the statistical analyses will be presented, starting with the Ap and Aa amplitude values, followed by a discussion of how Aas are distributed. Finally, inter-group differences between the temporal distances of T1 and the beginning of a syllable, as well as between T2 and the end of a syllable will be presented.

3.1. Aa and Ap amplitudes

3.1.1. Ap Amplitude

Along with Aa amplitude, Ap amplitude contributes to the overall pitch range. The 3 groups are significantly different from each other (F < 0.0001), see Figure 2. The ZH speakers show the least distinct Ap amplitudes, in contrast to the WS speakers.

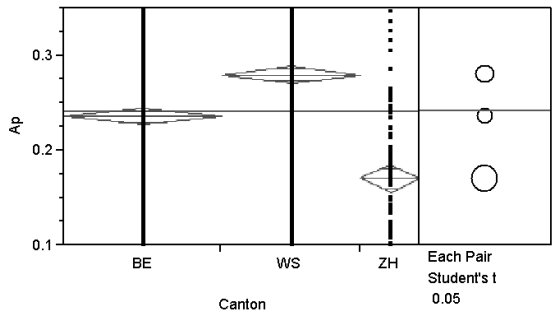


Figure 2: *Phase command amplitude according to speaker group.*

3.1.2. Aa Amplitude

Aas mark accented syllables, i.e. pitch accents. The extent to which, i.e. their amplitudes, and how they are distributed in the 3 groups can be seen in Figure 3. WS shows the most distinct Aa amplitudes (p WS/BE < 0.001, p WS/ZH < 0.001), while BE and ZH are not significantly different from each other (illustrated by the student's t test on the right: in case of an overlap of the circles, the difference between the groups is not significant).

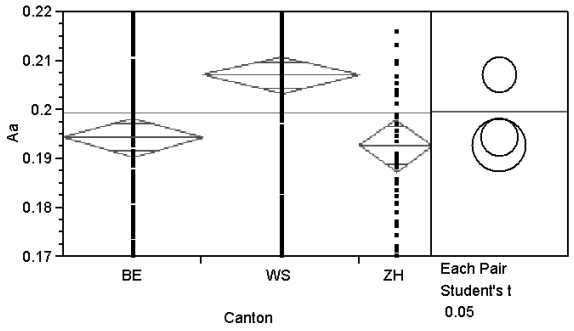


Figure 3: *Accent command amplitude according to speaker group.*

3.2. Distribution of Aas

3.2.1. Stress

In order to find out whether lexical stress of a word may have an effect on the distribution of Aas, lexical stress, as existent in the speaker's regional dialect's lexicon, was included as a variable for analysis. Table 1 summarizes the findings of these analyses.

	Aa WS	Aa BE	Aa ZH
Stressed	52%	53%	49%
Unstressed	39%	41%	41%
Schwa	9%	6%	10%

Table 1: *Accent command distribution according to stressed, unstressed, or schwa syllables.*

The major differences between the two groups can be found in their distribution of Aas on stressed and on schwa syllables. It is particularly the BE group which stands out with only 6% of all its Aas on schwa syllables (ChiSq BE/WS <.0001; ChiSq BE/ZH <.0.001).

3.2.2. Word class

A further variable that was tested for influence is the word class onto which Aas may fall. It was distinguished between lexical and grammatical words. Table 2 summarizes the results.

	Aa WS	Aa BE	Aa ZH
Lexical	72%	68%	69%
Grammatical	28%	32%	31%

Table 2: *Aa distribution according to lexical or grammatical category.*

Table 2 shows how the ZH and BE group behave almost identically with regard to the ratio of placing Aas on grammatical and lexical words. The only significant difference is that the WS group places more Aas on lexical words (ChiSq WS/BE < 0.0001; ChiSq WS/ZH = 0.0009).

3.2.3. Phrase position

The position within a phrase in which Aas occur was further considered. The phrases were split up into first syllables (f), medial syllables (m), penultimate syllables (p), and ultimate (u) syllables. In addition, the accent command placement was subdivided into Aas which fall onto stressed and unstressed syllables.

Results demonstrate that the ZH group behaves somewhat differently than the BE and WS groups; Table 3 illustrates this distribution. The ZH group shows unusually many Aas on stressed ultimate syllables (ChiSq ZH/WS <0.0001; ChiSq ZH/BE =0.0007).

	Aa WS	Aa BE	Aa ZH
f	63%	59%	55%
m	57%	55%	57%
p	62%	61%	68%
u	46%	48%	66%

Table 3: *Aas placed on stressed syllables, according to phrase position.*

The same calculations were made with Aas that fall onto lexically unstressed syllables, also according to syllable position within the phrase, see Table 4.

	Aa WS	Aa BE	Aa ZH
f	35%	41%	28%
m	28%	24%	30%
p	38%	34%	42%
u	29%	30%	55%

Table 4: *Aas placed on unstressed syllables, according to phrase position.*

ZH places extraordinarily many Aas on unstressed penultimate (ChiSq ZH/BE = 0.0428) and particularly ultimate (ChiSq ZH/WS <.0001; ChiSq ZH/BE <.0001) syllables. The ratio of Aas placed on stressed schwa syllables parallels this distribution between the 3 dialects.

3.3 Distance values

In the following section, we will look at the temporal distance (in seconds) between the pitch onset of Aas (T1) and the starting point of a syllable (xmin), as well as the temporal distance between the pitch offset of Aas (T2) and the end point of a syllable (xmax).

3.3.1. Temporal distance: T1 and xmin

This value informs us about the extent with which the onset for the Aa occurs before or after the starting point of the actual syllable. Figure 5 illustrates these results.

The BE begin their Aa onset 18 milliseconds after the actual starting point of the syllable. The WS place their onset for the Aa shortly before the starting point of the syllable while the ZH show an Aa onset much before the actual starting point, namely 113 milliseconds before the actual syllable (all p values <.0001).

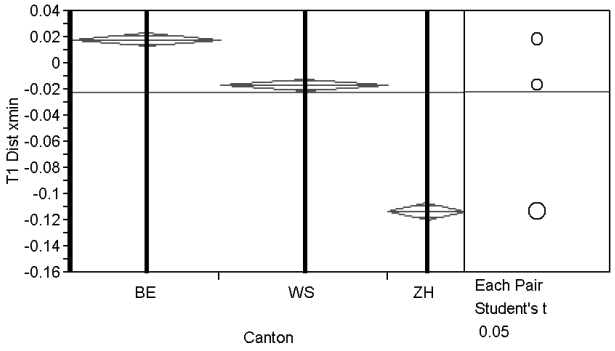


Figure 5: *T1-xmin values.*

3.3.2. Temporal distance: T2 and syllable end point

This value provides insights into the extent with which the Aa occurs before or after the end of the actual syllable. When calculating with temporal distance values between T2 and xmax, we need to further integrate the variables stress and the types of phrases within which the Aas occur, in order to obtain an accurate picture as to how these values are distributed according to their functions.

Figure 6 shows the temporal distance values between the 3 dialects, sorted according to whether or not the word carrying the Aa is stressed, unstressed, or whether a schwa syllable is stressed.

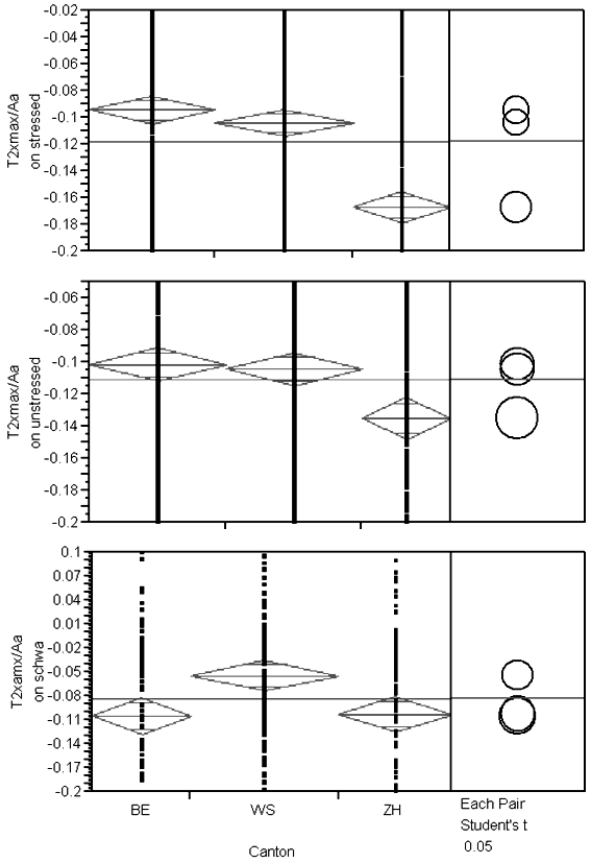


Figure 6: *Temporal distance values between T2 and xmax, according to lexical stress*

All 3 groups show T2s before the end of the syllable, regardless of whether or not the Aa falls onto a stressed, unstressed, or a schwa syllable. The BE and WS group show similar figures in their T2/xmax values in Aas placed on stressed and unstressed syllables; yet the two groups diverge in the last category, T2/xmax values when the Aa falls onto a schwa syllable (WS/BE p=0.0011, WS/ZH p = 0.0012), with the WS showing an average T2 closer to the actual syllable end compared to ZH and BE. The ZH group, however, shows T2 points especially in stressed and unstressed syllables carrying an Aa, which occur earlier than in the BE and WS group (all p values <.0004) is especially so if the Aa falls onto a stressed syllable (p <.0001).

For the reason that phrase types particularly depend on the intonational structure of nuclear accents, analyses were run, which show the temporal distance values between T2 and xmax of nuclear accents, divided into continuing (C) and terminating (T) phrases. BE and WS both show that in T phrases, nuclear accents tend to end earlier as opposed to in C phrases. The ZH group does not seem to make this distinction. Unfortunately, p values turned out slightly under the set alpha value of .05 in these lasts tests.

4. Discussion

The BE and WS group exhibit similar behavior regarding 4 out of a total of 9 investigated factors. This is the case for the ratio of Aas put on stressed, unstressed, and schwa syllables, for the similar distribution of Aas that fall onto unstressed syllables in the context of their position with the phrase, and in both of the investigated T2/syllable end position temporal distance values. No other two groups show this many similarities. The ZH dialect does not seem to fit in with the other two dialects. With these results, we are getting a step closer to explaining the factors which govern intonation in the investigated dialects.

For WS Swiss German, it seems as though it is particularly the high pitch range that seems to occupy a crucial role. None of the other two dialects shows such high values in its Ap and Aa amplitudes. The WS place more Aas on lexical words than ZH or BE; in addition they show a T2 which is comparatively close to the xmax when the Aa falls on a schwa syllable.

Results further seem to show that BE Swiss German is particularly characterized by their T1s in Aas occurring slightly after the actual xmin, as opposed to WS and especially ZH which show T1s before syllable onset. Possibly, the BE's slow articulation rate, 5.1 syllables/sec (as opposed to ZH's 5.9 syllables/sec, and WS' 5.8 syllables/sec), may interact with intonational phenomena – an issue which our project is yet to tackle.

However, it seems that, out of all 3 dialects, it is the ZH dialect that exhibits the most extraordinary features. It clearly shows the least distinct Ap amplitude (which somewhat stands in contradiction to Fleischer and Schmid's [3] observation, that ZH Swiss German shows a large overall pitch range) it puts most Aas on schwa syllables, it places a great number of Aas on ultimate stressed syllables, and especially many on unstressed syllables in ultimate phrase position. Finally, it was shown that ZH starts and ends its Aas earlier than WS and BE.

The WS and BE groups, both representing Western Switzerland varieties, stand in contrast to ZH, an Eastern variety, in nearly half of the investigated factors governing intonation. Once data from the Chur variety, a Southeast Alpine variety, are analyzed, a geolinguistic structuring of Swiss German intonation will become even clearer.

5. Outlook

In a further step, the present data will be labelled according to emotional categories, so as to check and describe the extent with which emotions correlate with the Fujisaki parameters. Furthermore, we are in the process of complementing Mixdorff's *Fujisaki Parameter Extraction Environment* [8] with a third command to the Fujisaki model – a command which Mixdorff himself established in his PhD thesis [7], yet never included within the editor.

In order to model question-final rises, for example, the Fujisaki model either enables the placing of a sequence of Aas, which increase in amplitude towards the end of the phrase, to model the slow rise, or one can set a number of phrase commands in close proximity (see Fujisaki [5]). Either of these options is difficult to interpret linguistically, as Aps generally constitute a pitch reset at the beginning of a phrase and Aas are used to model local accents independent of their position within a phrase.

With the aim of preventing the modelling of intonational contours in such a way, Mixdorff's slow-rise component offers the freedom to model rising, temporally-stretched, intonational events, such as can be found in questions, or high-rising terminals. We are now in the process off integrating the slow-rise command where applicable, in order to model slow-rises more fittingly.

6. References

[1] Boersma, P.; Weenink, D., Praat. http://www.fon.hum.uva.nl/praat, last visited 22.11.2007.

[2] Fitzpatrick-Cole, J., 1999. The alpine intonation of Bern Swiss German. *Proceedings of the ICPhS*, San Francisco, 941-944.

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

[4] Fujisaki, H.; Hirose, K., 1984. Analysis of voice fundamental frequency contours for declarative sentences of Japanese. *Journal of the Acoustical Society of Japan* (E), 5(4), 233-241.

[5] Fujisaki, H.; et al. 1998. Analysis and Interpretation of fundamental frequency contours of British English in terms of a command-response model. In *ICSLP*, paper 0830.

[6] Häsler, K.; Hove, I.; Siebenhaar, B., 2005. Die Prosodie des Schweizerdeutschen – Erkenntnisse aus der sprachsynthetischen Modellierung von Dialekten. *Linguistik Online* 24 3/2005: 187-224.

[7] Mixdorff, H., 1998. *Intonation Patterns of German*. Unpublished PhD Thesis. TU Dresden. http://www.tfh-berlin.de/~mixdorff/thesis/index.html, last visited 22.11.2007.

[8] Mixdorff, H. Fujisaki Parameter Extraction Environment. http://www.tfh-berlin.de/~mixdorff/thesis/fujisaki.html, last visited 22.11.2007.

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

[10] *Sprachatlas der deutschen Schweiz*, 1962-1997. I-VI Bern, VII-VIII Basel: Francke.

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