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

Speech Prosody 2010
Chicago, IL, USA
May 10-14, 2010

# Statistical Modeling of $F_0$ and Timing of Swiss German Dialects

*Adrian Leemann*[1], *Beat Siebenhaar*[2]

[1]Linguistics Department, University of Berne, Switzerland
[2]German Department, University of Leipzig, Germany
adrian.leemann@isw.unibe.ch, siebenhaar@uni-leipzig.de

## Abstract

This study examines the timing and fundamental frequency behavior of 40 Swiss German speakers from four different dialect regions. Prosodic behavior of each dialect is analyzed using statistical tests against the backdrop of detecting dialect-specific patterns as well as cross-dialectal differences. We find that variation in $F_0$ as well as in timing is vast across the dialects, with one of the dialects, the Southwestern Alpine variety, exhibiting particularly salient $F_0$ and timing features. The results confirm earlier impressionistic observations on this unexplored topic of Swiss German prosody.

**Index Terms**: Swiss German, Prosody, Timing, Intonation, Dialectology, Command-Response Model

10.21437/SpeechProsody.2010-35

## 1. Introduction

Despite Switzerland's long tradition in dialectology, prosody has largely been neglected. Most of the early remarks on prosodic characteristics of Swiss German dialects feature descriptions such as the "singing of the shepherds in the high mountains of [the] Valais" [1], the "slightly cradling" melody of the Grisons dialect [2], the "snug", and "slow" idiom of the Bernese [3], and the "fast" and "neutral" dialect of Zurich [4]. Recent typological descriptions of Bernese and Zurich prosody show a general difference to standard German. [5] claims that the default pitch accent in the Berne dialect consists of a low-rising contour (L*+H) compared to the Standard German falling accent (H*+L). Similar contours have been observed in other Southern German dialects [6, 7, 8].

Research on the temporal features of Swiss German, too, represents a research desideratum. Temporal aspects on the phrase level have been investigated by analyzing the position and duration of pauses in different speaking styles [9, 10, 11]; yet, temporal characteristics of regional varieties have mainly been ignored. Initial attempts on standard German are presented in [12] and on Swiss German dialects in [13]. However, individual and stylistic variation is vast, which renders generalizations about the influence of regional varieties difficult. Phrase-final lengthening on the regional level has been addressed by [14], and our own studies indicate that final lengthening is regionally different. Furthermore, the phonemic distribution of segment duration is important within the temporal domain. Within a language, different dialects can feature different oppositions. While most German dialects only differ with respect to vowel quantity, Alemannic and Bavarian dialects also distinguish consonant quantity [15, 16]. Moreover, the Standard and Northern German opposition between voiced and voiceless plosives is expressed as a temporal distinction between long and short consonants in the Southern dialects [17]. This distinction is even observed in word-initial position [18]. Non-phonological intrinsic segment duration has mainly been studied in the context of speech synthesis systems of standard languages, regional or dialectal differences have not yet been analyzed. Given this obvious lack of research on fundamental frequency [hereafter $F_0$] and timing characteristics of regional varieties of Swiss German, this study sets out to fill this gap.

## 2. Methods

In the framework of a Swiss National Science Foundation (SNSF) project (Quantitative Approaches to Geolinguistics of Swiss German Prosody 2005-2008), 40 subjects from four different dialect regions of German-speaking Switzerland were interviewed. Speakers (5 females / 5 males per dialect) from 2 Alpine varieties, Valais (VS) and Grisons (GR), and 2 Midland dialects, Bern (BE) and Zurich (ZH) were recorded in spontaneous interviews. The recorded data were transcribed, segmented, and annotated with linguistic variables (stress (2 df), word class (2 df), segmental features (see 3.2 Timing)), paralinguistic variables (phrase type [continuing/ terminating/ question/ discontinuity, i.e. 3 df], focus [narrow/broad, i.e. 1 df], prosodic paragraphing (numeric) [19]), and extralinguistic variables (emotion (4df), articulation rate (2 df), and sex (1 df)) of interest. Approximately 3 minutes of spontaneous speech per speaker (123,000 segments total) were analyzed. $F_0$ contours were then explored using the Fujisaki Intonation Model [20]. Methods underlying $F_0$-analysis are presented at 2.1, methods used for timing analyses are given at 2.2.

### 2.1. Fundamental frequency

The Command-Response Model [20] is hierarchically structured and formulated as a linear model. As input signals, the model receives phrase commands (*PCs*) in the form of impulse functions and accent commands (*ACs*) in the form of rectangular functions. The input signals are processed by the phrase and accent control mechanisms. The output signals of the two mechanisms are added onto the smallest asymptotic value ($F_b$) of the $F_0$ contour that is to be generated. For analysis purposes, the model decomposes the $F_0$ contour into a set of components from which timing and frequency information of the $F_0$ contours can be estimated. $F_b$ is assumed to be a speaker-specific parameter, α, the natural angular frequency of the phrase control mechanism, is set at 2/sec, β, the natural angular frequency of the accent control mechanism, at 20/sec in the present study. The phrase component can be applied for a description of the global declination tendency of $F_0$, i.e. the higher the *PC*, the steeper the declination. The unit in which declination is observed in the present approach is the prosodic phrase. The accent component is understood as a device for marking segments more $F_0$–prominent on the local level, i.e. the higher the AC, the more $F_0$–prominent the given syllables.

$F_0$ contours were parametrized by means of Mixdorff's *FujiParaEditor* [21]. $F_0$ behavior in each of the linguistic, paralinguistic, and non-linguistic variables was analyzed using parametric and non-parametric statistical tests against the background of detecting dialect-specific patterns as well as cross-dialectal differences. Dialect-specific multiple linear regression models (MLR) for all parameters were then

developed, which allows for a distillation of the relative contribution of independent variables towards explaining $F_0$ variability in a given parameter value in a specific dialect.

### 2.2. Timing

As a means of comparing timing features on the phrasal level, syllables were counted. In order to obtain segment-specific results, however, segment duration data is primarily used for the analyses. Statistical procedures are performed with base 10 logarithmic data that follows a near normal distribution.

## 3. Results

### 3.1. Fundamental frequency

#### *3.1.1. Berne*

The signature properties of the BE dialect involve *long AC durations* (=$T_2$-$T_1$, i.e. end point minus starting point of the local accent) as well as *late* $T_1$s relative to segment onset. Late rises, in particular, seem to be *the* prototypical feature of the BE variety. Comparatively long *AC durations* can almost certainly be attributed to the generally slower articulation rate of BE speakers [22], which corroborates the stereotype of the BE speakers as "slow" speakers [3]. It is indeed plausible that the combination of late rises, long AC durations, and slow articulation rate would evoke the notion of tranquility, "snugness" or "homeliness" alluded to in previous accounts.

#### *3.1.2. Grisons*

The GR dialect exhibits a great number of exceptional features in *PC magnitude*, i.e. the global declination parameter. Prototypical intonation phrases (IPs) start with high $F_0$, which is sustained throughout the phrase, and end with a still relatively high $F_0$. In sequence, such IPs may be conceived to have this "cradling" melody [2] alludes to. Also, a number of MLRs in the GR variety exhibit only a few significant predictors, suggesting that the $F_0$ contour is fairly robust to linguistic, paralinguistic, and non-linguistic influences. Figure 1 shows the MLR that was performed on the GR *AC amplitude* values with a radar chart. Each radius represents a variable fed into the MLR. The length of the radius is proportional to the percental magnitude of the respective variable, i.e. the longer the radius, the greater the variable's effect in the model.

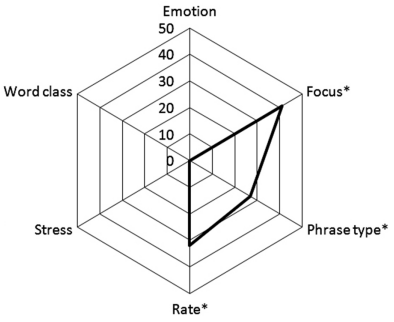


Figure 1: Radar chart of the MLR on the GR speakers' *AC amplitude* values.

Asterisked variables in Figure 1 imply statistically significant effects in the MLR ($R^2$ =.12, $F$(9, 2797) = 44, $p$ < .0001). Only the variables focus, phrase type, and articulation rate show significant effects, i.e. local $F_0$ is modulated particularly to emphasize focused constituents, phrase type, as well as to distinguish between speakers of different articulation rates. If we look at the amount of $F_0$ variation explained in the GR model (12%), we find that the variability in $F_0$ behavior can be explained comparatively well with the given predictors; hence, the dialect is relatively easy to predict.

#### *3.1.3. Valais*

The VS dialect stands out with a great number of exceptional features regarding *AC amplitude*. In bivariate tests, its *AC amplitudes* reflect distinctly different responses in many factors. More importantly, in the linear regression models, it shows the greatest array of variables for explaining variation. In other words, VS $F_0$ behavior is highly sensitive to external variables, particularly if paralinguistic and non-linguistic in nature – this is exemplified in the MLR on the VS speakers' AC amplitude values, as shown in Figure 2.

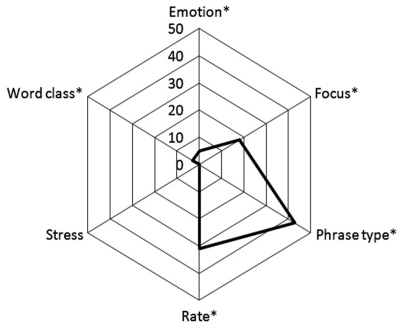


Figure 2: Radar chart of the MLR on the VS speakers' *AC amplitude* values.

Figure 2 exemplifies the VS speakers' high sensitivity of local accents towards word class, emotion, focus, phrase type, and articulation rate ($R^2$ =.09, $F$(14, 2827) = 21, $p$ < .0001). In addition to exhibiting a high number of predictors, the relative weight of the predictors often differs from the corresponding weight in other dialects. In spontaneous speech, this peculiar dialect-internal structuring of VS intonation results in what may be commonly perceived as an exotic and rather impalpable melody. Possibly, this is what characterizes the VS variety as a 'singing' variety [1]. If we look at the amount of $F_0$ variation explained in the VS model (9%), it becomes evident that the VS' intonational structure is comparatively complex and literally less *predictable*.

#### *3.1.4. Zurich*

The ZH dialect stands out with a number of exceptional features in *PC durations*, which hint at generally longer PC durations. Yet, if compared to the idiosyncratic $F_0$ behavior of the other dialects, ZH Swiss German does not feature any truly flashy properties. This may be why ZH German is known as a rather *neutral* dialect [4]. The distinctive *PC durations* could be a result of the ZH speakers fast articulation rate – another stereotype [4] empirically confirmed in this study. Faster articulation rate is shown to result in a reduction of phrase boundaries, hence the more distinct overall phrase length.

### 3.2. Timing

On the temporal level, the dialects differ with respect to a number of observed variables. Selective results are presented below. To begin with, the length of phrases (in syllables) were compared between the dialects. Speakers of the Western dialects (BE and VS) exhibit shorter phrases than speakers of the Eastern dialects, while ZH speakers show the longest phrases (ANOVA $F$(3, 8049) = 38; $p$ < .001). It is only the BE

and VS varieties which do not feature significant inter-level differences (t-tests); all other differences are significant ($p < .001$). Phrase length influences overall segment duration, where shorter phrases contain more final syllables which, in any case, exhibit final lengthening.

Secondly, the number and duration of pauses, which are, too, relevant for the discriminative attribution of speech rate of the dialects [3, 4], differ in quantity and distribution between the dialects. In all dialects, terminal phrases are normally followed by a pause. The pausing behavior after continuing phrases, on the other hand, is different between the dialects. The VS speakers by far show the least number of pauses after continuing phrases ($X^2$ (3, 6141) = 85, $p < 0.001$). In addition, VS speakers realize the shortest pauses after continuing phrases ($F(3,4793) = 29$; $p < .001$).

Thirdly, the durations of all syllable nuclei according to their position in the phrase are compared – see Figure 3. If, across the dialects, the first (f), penultimate (p), and ultimate (u) positions are compared to the mean durations of medial syllables (m), we find distinct lengthening in all three positions. Thus, we not only find phrase-final but also phrase-initial lengthening. Yet, there are remarkable differences between the dialects: VS speakers make the least distinct differences between the four positions, as shown in the comparably flat duration increment from one syllable position to another in Figure 3. On the contrary, ZH speakers, with their similarly short middle und penultimate nuclei, indicate much more pronounced initial and particularly phrase-final lengthening in ultimate syllable position. BE and GR speakers overall feature longer vowel duration. In this group, too, the BE speakers mark phrase-final boundaries more distinctly.

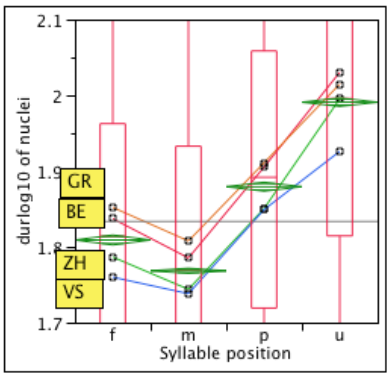


Figure 3: ANOVA of vowel duration by position of syllable in phrase (match code: dialect region).

The presented differences in timing between the dialects represent only a fraction of the established results. Further tests confirmed a distinctly different temporal ordering of the segments between the dialects. Due to the high number of observed variables, general tendencies are difficult to deduce. Therefore, general linear models (GLMs) were created to model segment duration in the four dialects. The following variables were applied for the prediction of segment duration: Intrinsic duration of sound classes, position of the syllable in the phrase, focus and stress, position of the segments in the syllable, type of the following and preceding segment, and assimilation of double plosives. Table 1 shows the predictive power of the GLMs modeled separately for each dialect.

| BE | ZH | VS | GR |
|---|---|---|---|
| 47.9 % | 46.5 % | 44.1 % | 53.1% |

Table 1: Predictive power of the GLMs.

The table suggests that, overall, the models explain nearly half of the dialect-specific variation in segment duration. In comparison, models for reading style of a single speaker explain more than 75% of the total variation. The relatively poor results in our models may thus be attributed to the spontaneous nature of the analyzed speech, which obviously exhibits significantly more variation than read speech. Moreover, model-intrinsic variation is increased because our models include 10 speakers each. 10 speaker-specific models were created, which show an improvement in the models' explanatory power by 3-5%. Since, however, we aim for a dialect comparison, only the aforementioned aggregate models are used for further calculations and interpretations. The highest percentages in Table 1 (GR and BE) indicate that the factors mentioned above provide a better prediction than for the other models (ZH and VS), as manifested in their lower percentages. High values stand for a small variability while small values signify distinct variation. From this, we conclude that the VS speakers vary the most in segmental duration, while the GR speakers by far demonstrate the most uniform timing.

Let us zoom in a little further into the GLMs. An assessment of the Type III sums of squares allows us to filter out each factors' predictive power within the model. Table 2 lists the values of the single factors in each model.

| | BE | ZH | VS | GR |
|---|---|---|---|---|
| Intrinsic duration (12 df) | 57.8% | 46.4% | 57.9% | 56.2% |
| Position of the syllable in the phrase (4 df) | 16.3% | 21.7% | 15.7% | 13.5% |
| Focus and stress (4 df) | 9.7% | 13.1% | 12.6% | 13.4% |
| Type of the following segment (12 df ) | 8.7% | 9.1% | 8.7% | 8.1% |
| Type of the preceding segment (12 df) | 4.4% | 6.0% | 2.5% | 5.2% |
| Position of the segments in the syllable (3 df) | 2.5% | 2.6% | 1.2% | 1.7% |
| Assimilation of double plosives (4 df) | 0.6% | 1.1% | 1.4% | 1.8% |

Table 2: Predictive power of each factor.

In all dialects, intrinsic duration is the dominant factor, i.e. explains most variation. Yet, the remaining factors appear to be structured in a dialect-specific way: For the ZH dialect, the position of the syllable in the phrase is more important than for the other dialects. The same holds for the position of the segment in the syllable and the influence of the preceding segment. For Bern, focus and stress bear little predictive power.

## 4. Discussion

In the course of the $F_0$-analyses, distinct geolinguistic features of the Alpine/Midland and Eastern/Western dialects emerged. The most striking difference between the Alpine and Midland groups is found in the relative weight of the *linguistic* predictor *stress* in the *AC amplitude* models. Its comparatively weak importance in the Alpine dialects (see Figures 1 and 2) may be a result of these Germanic dialects' proximity to right-headed Romance languages such as French (syllable-timed), Italian, and Romansh. The clash of two systems with different rhythmical patterns and different prosodic means for prominence marking may have neutralized the importance of *stress* and its link to $F_0$ movements, resulting in insignificant effects of *stress* in the Alpine dialects' *AC amplitude* models. As for the findings derived from the statistical models, a

dialect-overarching assessment of the coefficients of determination of the created models implies that the fractions of the variances explained in the dependent variables are generally low. The main reason behind these low values may be found in the spontaneous nature of the data. Spontaneous speech is highly multi-layered and variable. It is characterized by a constant interplay of variables that were categorized into a *linguistic* layer, a *paralinguistic* layer, and a *non-linguistic* layer. It is these characteristics and the multi-level structure that make it exceedingly difficult to predict and/or explain spontaneous speech $F_0$ patterns.

Timing analyses showed that phrase structuring differs between the varieties. Particularly noteworthy is the VS speakers' tendency of producing phrases, which, in addition to being distinctly short, often do not feature a post-phrasal pause. Also, if there is a pause, it is significantly shorter than in the other dialects. Moreover, VS phrase boundaries are less marked in the temporal domain than in the other dialects. This low-key marking of phrase boundaries in the temporal domain parallels low PC magnitudes in the $F_0$ domain. ZH speakers posit an example of clearly distinct behavior. Similar to the VS speakers, they are fast speakers yet their tempo is more likely to be due to longer phrases. If, however, they realize phrase boundaries, they are clearly marked with pauses and a significant lengthening before and after the boundary. This behavior is comparable to the other Midland dialect (BE), which, however, is generally spoken much more slowly.

Statistical modeling has confirmed that intrinsic duration is the main factor behind temporal control. However, the weight of the other factors in the GLMs appears to be dialect-specific: Phrase marking as well as the position of the consonants in the syllable bears more weight in the Midland dialects. In the models for the Alpine dialects, on the other hand, assimilation of plosives is more important. The low weight of the factor focus/accent in BE is to be mentioned.

## 5. Conclusions

The descriptions of the prosodic properties of Swiss German dialects iterated at the beginning of this section undoubtedly sound very unscientific and impressionistic. Hence, it is all the more astonishing that these descriptions can essentially be corroborated via ANOVAs and GLM in the temporal domain and via $F_0$ analysis based on Command-Response Modeling [20], bivariate statistical analyses, and MLRs in the $F_0$ domain. The established methods for an $F_0$ modeling of spontaneous speech based on the Command-Response Model [20] as well as the quantitative statistical methods applied in this thesis seem to be well suited for such an investigation.

The analysis of timing allows for a differentiated view on general remarks that have been made in reference to the dialects' rate. ZH and VS fall under the cliché of being "fast"-spoken dialects [4, 22], a feature which is confirmed by the results of this study. Yet, the present study has also shown that both varieties adhere to a different style of speaking fast. In contrast, BE is perceived as slow, while GR is not. Based on our production data, results of the present study suggest that both dialects are similarly slow but show differences in phrase boundary marking. Perhaps it is the $F_0$ behavior that covers up the given similarities in the temporal domain.

The obtained results on $F_0$ and timing features of the four dialects overlap only in part, which points to a relative independence of timing and intonation. Interestingly, both analyses suggest that VS prosody is particularly difficult to predict. Speculations as to the underlying causes are threefold: VS speakers' indeed do vary more on all of the examined levels. Secondly, VS speakers may exhibit distinct dialect-internal variation (which is in fact confirmed by an informal check, at least on the temporal level). Thirdly, other factors not included in the models may explain further variation. All of these constitute aspects worthy of pursuing in future research.

## 6. References


[1] Stalder, F. J., "Die Landessprachen der Schweiz oder Schweizerische Dialektologie", Aarau: Sauerländer, 1819.

[2] Meinherz, P., „Die Mundart der Bündner Herrschaft", Frauenfeld: Huber, 1920.

[3] Schwarzenbach, R., "Die Stellung der Mundart in der deutschsprachigen Schweiz. Studien zum Sprachgebrauch der Gegenwart", Frauenfeld: Huber, 1969.

[4] Ris, R., "Innerethik der deutschen Schweiz", in: P. Hugger [Ed], Handbuch der schweizerischen Volkskultur, vol. II., Zürich: Offizin, 749-766, 1992.

[5] Fitzpatrick-Cole, J., "The alpine intonation of Bern Swiss German", Proceedings of the XIVth International Congress of the Phonetic Sciences (ICPhS), San Francisco, 941-944, 1999.

[6] Barker, G. S., "Intonation patterns in Tyrolean German: An autosegmental-metrical analysis", unpublished PhD Thesis, University of California, Berkeley, 2002.

[7] Kügler, F., "The intonational phonology of Swabian and Upper Saxon", Tübingen: Niemeyer, 2007.

[8] Gilles, P., "Regionale Prosodie im Deutschen: Variabilität in der Intonation von Abschluss und Weiterweisung", Berlin: de Gruyter, 2005.

[9] Gustafson-Čapková, S. and Megyesi, B., "A Comparative Study of Pauses in Dialogues and Read Speech", in Proceedings of Eurospeech, 931–935, 2001.

[10] Peters, B, Kohler, K. J., and Wesener, T., "Phonetische Merkmale prosodischer Phrasierung in deutscher Spontansprache", in: K. Kohler et al. [Eds], Prosodic Structures in German Spontaneous Speech, 143–185, 2005.

[11] Zellner, B., "Pauses and the temporal structure of speech", in: E. Keller [Ed], Fundamentals of Speech Synthesis and Speech Recognition, Chichester: Wiley, 41–62, 1994.

[12] Monaghan, A., "An Auditory Analysis of the Prosody of Fast and Slow Speech Styles in English, Dutch and German", in: E. Keller et al. [Eds]. Improvements in Speech Synthesis, Chinchester: John Wiley & Sons, 204–217, 2001.

[13] Hove, I., "Pausen in spontan gesprochenem Schweizerdeutsch", Deutsche Sprache 32: 97–116, 2004.

[14] Streck, T. (online), "Finale Dehnung im Deutschen", Freiburg, 2004, (http://www.freidok.uni-freiburg.de/freidok/volltexte/2004/1491/pdf/streck_finale_dehnung_Wiss_Arbeit.pdf).

[15] Hinderling, R. [Ed], "Bayerischer Sprachatlas", Heidelberg: Winter, 1996 - …

[16] Hotzenköcherle, R., "Sprachatlas der deutschen Schweiz", Founded by Heinrich Baumgartner and Rudolf Hotzenköcherle, R. Hotzenköcherle [Ed], in cooperation with Konrad Lobeck, Robert Schläpfer and Rudolf Trüb and with the assistance of Paul Zinsli, Basel: Francke, 1962-1997.

[17] Willi, U., "Die segmentale Dauer als phonetischer Parameter von "fortis" und "lenis" bei Plosiven im Zürichdeutschen. Eine akustische und perzeptorische Untersuchung", Stuttgart: Steiner, 1996.

[18] Kraehenmann, A. "Quantity and Prosodic Asymmetries in Alemannic. Synchronic and Diachronic Perspectives", Phonology & Phonetics 5, Berlin/New York: Mouton de Gruyter, 2003.

[19] Tseng, C. et al., "Fluent speech prosody: Framework and modeling", Speech Communication 46, 284-309, 2005.

[20] Fujisaki, H., "The Command-Response Model for F0 Contour Generation and Its Implications in Phonetics and Phonology", Proceedings of PCC2008, Beijing, 2008.

[21] Mixdorff, H., "Fujisaki Parameter Extraction Environment", Online: http://www.tfh-berlin.de/~mixdorff/thesis/fujisaki.html

[22] Leemann, A and Siebenhaar B., "Intonational and Temporal Features of Swiss German", Proceeding of ICPHS XVI, Saarbrücken, Germany, 957-60, 2007.