

Vorabdruck von: Siebenhaar, Beat und Adrian Leemann (2012): "Methodological reflections on the phonetic–phonological continuum, illustrated on the prosody of Swiss German dialects". In: Wälchli, Bernhard, Andrea Ender und Adrian Leemann (Hg.): *Methods in Contemporary Linguistics*. Berlin: De Gruyter: 21–44. (= Trends in Linguistics 247)

Methodological reflections on the phonetic– phonological continuum, illustrated on the prosody of Swiss German dialects

Beat Siebenhaar and Adrian Leemann

1. Introduction

Since Trubetzkoy (1939) we discriminate between phonetics and phonology, where phonology categorically interprets language-specific continuous acoustic signals and thereby conceptually separates between a component of meaning and the stream of speech, which both are correlated in a second step. Today, the allegedly obvious separation is being questioned on a number of levels. This softening of what used to be formerly rigid boundaries between phonetics and phonology is particularly prevalent in a description of prosody (cf. Byrd and Choi 2010: 32).

In the context of intonation research, this uneasy connection between phonetics and phonology is hinted at in Bolinger (1972). He notes that the phonetic representation of intonation, for instance, cannot simply be determined by considering grammatical, phonological, aspects of sentences, as illustrated in the infamous “Accent is predictable (if you’re a mindreader)” *Language* article. What Bolinger is referring to is the then becoming dominant school of thought of metrical phonology, where prominence is understood as an abstract feature that can be derived from the metrical strength of syllables (Lieberman and Prince 1977). This framework was adopted by Pierrehumbert (1980) who formulates an autosegmental-metrical approach towards intonation, where key syllables in utterances are described as discrete tones. This system has been formalized in the ToBI transcription system. The underlying assumption is that the temporal coordination of fundamental frequency and phonetic segments is highly rule-governed, where the highs and lows of the fundamental frequency (f_0) contour predictably line up with metrically strong syllables (Pierrehumbert 1980). Yet, there are studies that insinuate otherwise. Kochanski et al. (2005) as well as Silipo and Greenberg (2000) re-address the role of stress, i.e. metrically strong syllables, in predicting f_0 by analyzing a corpus of spontaneous speech in British and American English. The studies conclude that metrically strong

syllables are exceptionally marked with loudness, duration, and distinct spectral tilt – not necessarily f_0 movements.

Over the past three decades, temporal aspects, too, aroused the curiosity of linguistic research. With the greater part of actual research we analyze durations of segments within the acoustic signal; an alternative access - articulatory phonology (Browman and Goldstein 1992) – observes gestures of the articulatory tract. While the distinction of long and short (and over-long, where they exist) vowels and consonants was discovered long before linguistics as subject proper was established and is thus, reflected in the orthography of languages that feature the phonological quantity distinction, the effects of the gradual lengthening and shortening of sounds only became evident with acoustic measurements based on visualisation of speech and especially with the technological representation of speech in speech synthesis and speech recognition systems. As is the case with intonation, the marking of stress, accents, and phrase boundaries is particularly interesting. The appreciation of these concepts is to a large extent dependent on the phonological system of the language in question, which is assumed to be categorical, while the phonetics of an utterance are conceived of as being gradient. The argumentation is similar for intonation and timing: Continuous changes of fundamental frequency are – in the actually most respected theory – categorized into high and low tones, which are tied to accented syllables and phrase boundaries, followed by an unspecified interpolation that subsequently applies. The same holds for timing, where gradient changes of segment durations are categorized as short and long (and where they exist over-long) sounds, and also applies to accents and phrase boundaries. The other way around, the categorical phrase boundaries and accents are represented in continuous duration changes. The relation between these gradient changes in f_0 and duration and the underlying phonological categories is still unclear. Yet, to this day, it is not entirely straightforward, how these phonological categories are represented in prosody. Both, phonetic and phonological research converge in the typological discussion of rhythm of languages (Ramus et al. 1999, Low et al. 2000).

These considerations suggest that there is more to describing and understanding f_0 and temporal patterns than considering categorical, metrical, i.e. phonological, aspects of sentences. By means of examples of the Bernese “Quantitative Approaches to Geolinguistics of Swiss German Prosody” corpus, we illustrate the problematic interplay between phonetics and phonology in the context of prosody. After overviewing key concepts of prosody and a short description of the data, we will show that creating a corpus

of spontaneous speech already brings with it many decisions located at the boundaries of phonetics and phonology. In the second part, which addresses temporal aspects of prosody, the phonological classification of long and short vowels as well as the phonetic correlate of phrase boundaries are put into question. In a third part, evidence is presented which underlines the detachment of stress from f_0 movements. Thereby, the central phonological and ultimately methodological assumption that underlying stress patterns predict f_0 movements is put into question. A phonetic intonation model, which allows one to bypass this assumption of f_0 prediction, the Fujisaki, or Command-Response model (Fujisaki and Hirose 1982) model, is presented and its application on the current set of data is illustrated.

2. Key concepts

Before jumping into the relevant topics at hand, key concepts of intonation research, prominence, stress, and the modeling of intonation are touched upon so as to lay the theoretical groundwork for the subsequent presentation of Swiss German intonational and temporal data and the discussion thereof.

2.1. Prominence

Prominence on the word level is frequently denotes word accent or lexical accent. The acoustic correlates of prominence are intricate and seem to be language-dependent, and most importantly, it is sensible to differentiate between production and perception: In prominence production, the most critical indicator for varieties of English, for instance, is duration, followed by intensity and, least importantly f_0 . In prominence perception, however, f_0 occupies a more critical role (see Kochanski et al. 2005). Not all languages mark prominence concurrently with the above-mentioned parameters in prominence production. French, for example, shows reduced correlation of these parameters. Vaissière (1983: 66) even claims that

it is possible that specific interrelations between the three suprasegmental features (f_0 , duration, and intensity) [...] are the most salient characteristics differentiating between languages, dialects and individual ways of speaking. If this is true, most of the existing descriptions of prosodic systems [...] are incomplete, since they describe only one parameter at the time. (Vaissière 1983: 66)

As will be shown below, it seems that particularly the Alpine dialects under scrutiny exhibit a somewhat different suprasegmental code as opposed to Midland dialects.

2.2. Stress

Stress is a highly intangible prosodic feature (Lehiste 1970: 106). Stress and accent are often used interchangeably, which adds to the terminological confusion. Stress is governed by the lexicon of a language and is marked by prominence. Syllables that carry stress are perceived as more salient. Stress is assigned according to strong and weak syllables, a notion that grew out of metrical phonology (see Liberman and Prince 1977). In this framework, prominence is understood as an abstract feature, which derives from the metrical strength of syllables, consequently, the interconnectedness between stress and prominence. However, prominence is not necessarily lexical stress but it can also be associated with boundary marking.

2.3. Modeling prosody

Intonation models can generally be categorized into more concrete or more abstract approaches (cf. Cutler and Ladd 1983: 2ff.). The former category is frequently referred to as phonetic models, the latter as phonological models of intonation. The two approaches differ vastly with regard to the degree of abstractness postulated of the prosodic representation.

The abstract take towards intonation analyzes the prosodic structure and its relation to phonology and other aspects of grammar so as to generate an inventory of abstract categories, eventually creating a formalization of intonational function and form. By the formulation of rules, the phonological, symbolic approach transposes the abstract phonological description of intonation contours into its concrete phonetic form. Basically, *f0* contours are understood as the addition of atomistic local events: phrase accents and pitch on the one hand, and boundary tones on the other (cf. Pierrehumbert 1980). Most importantly for the present paper, much of the work in intonational phonology implicitly presupposes that prominence is first and foremost a function of *f0*. Ladd (2006: 48–49), for example, states that

A pitch accent may be defined as a local feature of a pitch contour – usually, but not invariably a *pitch change*, and often involving a local minimum or maximum – which signals that the syllable with which it is associated is *prominent* in the utterance. [...] If a word is prominent in a sentence, this

prominence is realized as a pitch accent on the “stressed” syllable of the word (Ladd 2006: 48–49).

On corpora of different varieties of English, Kochanski et al. (2005), Silipo and Greenberg (2000), demonstrate that many prominent syllables do display high pitch, yet, many non-prominent syllables follow the same pattern. They conclude that “prominence and pitch movements should be treated as largely independent and equally important variables” (Kochanski et al. 2005: 1052).

In the phonetic approach, claims are made about the concrete, close-to-the-signal phonetic form of intonation. Intonation is understood as the addition of multiple components, consisting of baselines, globally declining phrase components, and local word accents (cf. Öhman 1968, Fujisaki and Hirose 1982). It is the realization of intonation that represents the primary scientific goal. *f0* contours can be modeled blindly, i.e. without, in a first step, taking into account whether *f0* contours are anchored with stressed syllables or not. In a second step, *f0* excursions can be associated with the segmental level. This procedure allows one to deduce the effect of metrical stress on actual *f0* movements.

Timing, on the other hand, has received less attention in prosodic research, except for the quantity opposition on the segmental level, because it is not as functionally loaded as *f0* and it is often regarded as a corollary of *f0*. Therefore, the modeling of segment duration is normally rule-based and more often than not explored in the context of data-driven statistical models for speech-synthesis-systems. In these models, duration changes are usually derived from phonological components such as stress, accent, phrase boundaries, as well as the surrounding segments and the position of the segments within larger entities (foot, word, phrase). Moreover, speaking style, focus and speech rate, which are out of the scope of phonology, are integrated into these models (cf. Klatt 1976, Siebenhaar et al. 2001, van Santen 1998). In many instances, these temporal aspects are directly linked to intonation; yet, as mentioned above, prominence can be marked without *f0*-changes. Nevertheless, there are no genuine linguistic models for timing that function independent of intonation. In this sense, analyzing temporal aspects of spontaneous speech is by itself a methodological approach on prosody that goes beyond the actual intonation only analyses. Moreover, respecting time as a linguistic phenomenon – articulating a linguistic unit is intrinsically temporal – opens a view on linguistics, which are not only based on a graphic symbolization of language.

3. Data

The goal of the empirical study was to find prosodic differences between four Swiss German dialects, where the term dialect is used in the German sense of a geographically defined variety. It is only since the end of the twentieth century, that the focus of research in prosody moves from standard languages to regional and dialectal variation. That shift towards dialectal speech implies a revision of the empirical basis from laboratory, word or phrase list data, to data that is based on spontaneous, natural speech (cf. Bucheli Berger, Glaser, and Seiler, present volume, for a syntactic description of natural speech Swiss German dialects, and Schmid, present volume, for a rhythmic description of Italian dialects). The focus of our analyses lies on an acoustic description, i.e. on a phonetic analysis, of these four dialects. Results of these analyses are published in Leemann (2012), Leemann and Siebenhaar (2007, 2008a, 2008b, 2010). In the present contribution, the center of interest lies not on the data analysis per se; instead, the data are used to illustrate the practical and theoretical problems at the interface between phonetics and phonology.

The data consists of approximately two hours of spontaneous speech. Forty subjects aged twenty from four different dialect regions of German-speaking Switzerland were interviewed. All four dialects belong to the Alemannic dialect family. Speakers (5 females and 5 males per dialect) from two Alpine varieties, Valais (VS) and Grisons (GR), and two Midland dialects, Bern (BE) and Zurich (ZH) were recorded in spontaneous interviews. Approximately three minutes per speaker were manually labeled on a segmental and syllabic level and analyzed for temporal aspects. *f0* contours were explored using the Fujisaki intonation model.

4. Phonetics and phonology in data preparation

In the first steps of data preparation, it becomes obvious that phonetics and phonology can hardly be separated and are co-dependent. This intimate link between phonetics and phonology affects the decision-making process of an empirical study that aims to explore prosodic differences between dialects – this aspect shall be discussed in this section.

For the analysis of the prosodic aspects, segments of the interviews had to be isolated and labeled. This labeling itself requires decisions on behalf of the labelers which are guided by phonetic and phonological considera-

tions. Even the prosodic level, which in fact represents the dependent variable to be investigated, influences the decisions. To begin with, the basic segments that are to be analyzed had to be decided on. Most linguistic analyses on prosody focusing on intonation choose the syllable as basic unit. However, whether onset, nucleus and coda are equally affected by stress or speech rate changes seems to be language dependent (Barry et al. 2007). Moreover, while the nucleus is more or less unambiguously defined in phonetics as the most sonorant or most articulatorily open gesture between two less sonorant or more closer parts, the definition of the syllable in phonology seems to be an issue of much more controversy. In German, for instance, consonant clusters and schwa deletion characterize the discussion if there are syllabic consonants or if consonants have to be described as extrasyllabic. This is especially relevant if one considers the south German schwa deletion in prefixes (*Gschpängscht* < *Gespenst* ‘ghost’). Considering this background, we opt for a segmentation level narrower than the syllable. This level is closer to the phone/phoneme as the basic prosodic unit. The syllable is a derived category based on the sonority hierarchy. For the analysis of *fθ*, however, the syllable was chosen as the appropriate unit. The syllable represents the structural anchor point for abstract prosodic features, such as tone or stress, for example.

The segmentation follows a top-down approach, from utterance to phrase and phone, and bottom-up from phones to utterances. The practically justified combination of the two approaches allows for a distinction in ambiguous cases. However, in spontaneous speech, only the definition of “utterance” is not problematic itself, while the definitions of the other units are questionable. The utterance is a speech unit that is pragmatically separated by the question of the interviewer on the left and by the end of the sound chain on the right, the latter of which is generally given by the speaker himself/herself, as the interviewers usually did not intervene. The segmentation of the utterance into phrases poses a greater problem, as the definition of “phrase” can be grounded in grammatical, semantic, pragmatic and prosodic features. With many discontinuities and hesitations, spontaneous speech often disregards syntactical shapeliness (cf. Bucheli Berger, Glaser, and Seiler, present volume), so that pragmatic (conversational) and semantic aspects of sense units are attributed greater significance. As prosody represents the focus of the current study, prosodic features attributed to phrase boundaries such as pitch changes, final lengthening, pauses and changes of voice quality (Cruttenden 1997) should ideally not affect the decision. However, given the interrelation of the afore-mentioned aspects,

the decision as to where to place the phrase boundary is more often than not opaque. None of the mentioned criteria are separately unambiguous, but the interplay between them provides an inter-individually comprehensible decision on where to set a phrase boundary: In the recording of interactive spontaneous speech, perception is where all aspects meet (cf. Gilles 2005: 42–45). Thus, the decision as to where a phrase boundary is labeled is ultimately a pragmatic decision of the investigator based on perception, whether a sense unit was terminated, whether a grammatical unit was terminated, whether the interviewer intervened and so forth. To some extent, the decisions were cross-checked with the project members.

The labeling of the segments is tedious as well. To begin with, it is difficult to say if the labeling is a phonetic or a phonological procedure: In order to define the duration of a sound, the sound must be brought in relation with an independent dimension. This dimension can be the canonical phonological representation. The systematic reductions of spontaneous speech, however, strongly obscure a canonical representation. Let us exemplify this with a word that is often used on different levels of reduction. The full form of *eigentlich* ‘actually’ in Bern is [ˈɛɪ̯gəlɪx], with a variant closer to the standard German [ˈɛɪ̯gətɪx]. *Eigentlich* is often used as a discourse particle, which is subject to, sometimes quite rigorous, reductions. In this use, the first step of reduction is the loss of the accent [ɛɪ̯gəlɪx]. Centralization of the unstressed [ɪ] follows: [ɛɪ̯gələx]. In a next step, the central schwa is syncopated: [ɛɪ̯gləx]; the [ɪ] disappears [ɛɪ̯gəx], the schwa of the last syllable is syncopated: [ɛɪ̯gx], the complex coda is reduced to a simple fricative [ɛɪ̯x], the diphthong is monophthongized [ɛx] and finally reduced to a schwa [əx]. These reductions are critical complications for the transcription process but even more so, they exacerbate a systematic segmentation of the signal. While the transcription suggests a stepwise reduction, the acoustic signal shows a gradual reduction of the duration and quality of the individual sounds due to the gestural reduction. Thus, labeling, which is based on acoustic features, has to set clear-cut boundaries in the continuum, where one sound can be shadowed by another. It must be emphasized at this point that a highly precise labeling is crucial for the temporal analysis. The calculation of the mean duration of a sound class relies entirely on the labeling thereof. Let us go back to *eigentlich*, which, too, illustrates this problem. In order to calculate the mean duration of schwa, the question poses itself as to which form, which schwas, need to be included. Is it only the schwa of the medial syllable in the full form [ˈɛɪ̯gəlɪx / ˈɛɪ̯gətɪx] or also the schwa resulting from the reduction of the unstressed [ɪ] in the form [ɛɪ̯gələx] or

can it also be the schwa of the fully reduced form [əx]? In our project, we decided to take into account all forms; hence all the mentioned reduction forms are accepted as forms of the lexicon. However, the reductions are marked, so that in a second analysis, one could return to the original forms and consider the reductions separately.

As the definition of phonemes is to a great extent based on word phonology, only the schwas in the full form are regarded as schwa-phonemes. The other schwas result from regular phonological processes and therefore do not represent phonemes proper. From a prosodic point of view, all the schwas in the systematic reduction forms can be regarded as representations of the other (full) phonemes from which the concrete realizations can be measured.

The discussion of the problems in data preparation, exemplified with defining phrase boundaries and transcription, shows that the boundary between decisions based on phonetics and decisions based phonology cannot be drawn strictly. The border proves to be rather a continuum where methodological reflections in the perspective of the goal of a specific project have a great impact on the concrete decisions. Furthermore, it is confirmed once again that, on the one hand, the phonetic continuum can hardly be transferred to a phonological classification and, on the other, that a classification of data of spontaneous speech is hardly possible on a purely phonetic ground.

5. Temporal aspects

5.1. Duration of schwa

One of the central questions that follows from the previous example is whether the phonetic realizations of the schwa phoneme in the narrower sense and the schwas in the broader sense – including reduced variants – behave differently in the timing domain. It turns out that for three of the four dialects, schwas resulting from vowel reductions are shorter than schwas representing phonological schwas in the former, narrower sense. For the VS dialect, however, schwa phonemes and schwas resulting from reductions are of the same duration. From this we conclude that the four dialects under scrutiny exhibit different strategies of reduction, one that keeps vowel duration more constant (VS) while the others show more variable vowel duration (BE, ZH, GR).

5.2. Stress and focus

As is the case for standard German, Swiss German dialects show lexical stress too. In our data, stress and narrow focus is marked. Narrow focus was marked according to aspects of givenness of information, contrasting information, as well as emphasizing information during the course of the interview. With only few exceptions, there are only stressed syllables that are focused. For all dialects, vowels in focused syllables are significantly longer than those in non-focused but word-stressed syllables, which are again significantly longer than those in unstressed syllables and schwas. The same can be said of consonants; yet, these differences are not always significant. In the context of timing, pragmatic focus and the phonological stressed–unstressed dimension are well reflected in phonetic duration differences. As it has been shown for the schwa reductions, the difference between focused, stressed and unstressed segments is more distinct in BE, GR, ZH than in VS.

5.3. Quantity and phrase final lengthening

The German phoneme system distinguishes short and long vowels. Yet, except for /a~a:/ and /ε~ε:/ the quantity contrast entails an opposition of tenseness. Therefore, some grammars (e.g. Duden 2005: 26) abandon the opposition in quantity in favor of an opposition in tenseness (/a~ɑ/ and /ε~æ/). In contrast to standard German, most Swiss German dialects (and all four dialects discussed here) demonstrate a quantitative distinction of long and short vowels while vowel quality remains the same (*gi:gələ* : *gigələ* ‘to play the violin : to giggle’, *be:t* : *bet* ‘flower bed : bed’). Swiss German dialects also have a quantity distinction of obstruents (*vaðə* : *vətə* ‘calf : cotton wool’; cf. Fleischer and Schmid 2006 for the Zurich dialect). Willi (1996) has shown that the opposition between fortis and lenis plosives in Zurich German is not a distinction achieved through voicing but through consonantal duration. This distinction in duration is also substantiated for the Thurgovian dialect by Kraehenmann (2003), who conceives of the long and short obstruents terminologically as singletons and geminates. From this we assume, that the distinction of long and short segments should be preserved in all phrase positions. Short segments may not be lengthened, or long segments shortened, so much as to cause perceptual ambiguity at the segmental level. Looking at the long and short vowels in our data, the claim that long and short vowels are always distinct cannot be maintained fully.

Phrase-final lengthening affects vowel duration to such an extent, that short vowels in phrase-final syllables exhibit the same length as long vowels in phrase-medial syllables. Figure 1 shows the typical distinction of long and short vowels in phrase-medial and phrase-final syllables, here in the VS data, which are representative of all dialect groups except for the non-accented vowels in GR. The first and penultimate syllables are not taken into account because they show an intermediate duration. Figure 1 indicates that short vowels are on average shorter than long vowels. Yet, the short vowels of the ultimate syllables (u) of a phrase show the same length as the long vowels in phrase-medial syllables (m). From this we conclude that final lengthening, a prosodic feature of phrasing, affects vowel quantity in such a way that the segmental, phonological distinction is no longer maintained over the different position. When we compare the left and the right figure, it is apparent that short vowels are lengthened if they do bear lexical stress. Long vowels are much less affected by stress; stressed and unstressed long vowels are not significantly different. Yet in all four dialects, phrase-final lengthening affects unstressed syllables more than stressed syllables, as they are as long or even longer than the stressed syllables in the same position. Moreover, the figure indicates a high variation that for stressed syllables 8.7% of all mid phrasal short vowels are longer than the mean duration of the mid phrasal long vowels, and more than 4.1% of the long mid phrase vowels are shorter than the mean of the mid phrasal short vowels. The numbers remain at the same level if one only considers long and short a, which eliminates inter-vowel distinctions.

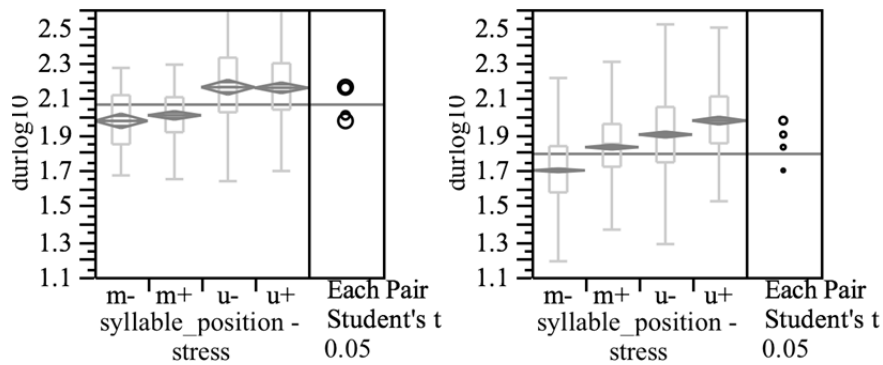


Figure 1. Box plot of duration and confidence intervals of long vowels (left) and short vowels (right) in phrase-medial (m) and ultimate (u) position, stressed (+) and unstressed (-). The overlapping circles on the right of the

figure show that the difference of stressed and unstressed long vowels is not significant.

5.4. Degree and extent of phrase final lengthening

Phrase-final lengthening has been documented in many studies covering numerous languages but the degree and extent of lengthening varies between languages (cf. Fletcher 2010: 540). Our project shows that the degree and extent of phrase-final lengthening even varies within the selected Alemannic dialect group. Figure 2 shows the duration of schwas that are phonologically represented as such in different positions in the phrase for the ZH and the VS dialect. If we compare only schwas, there is no interference from different vowel qualities, quantities, and stress and we can analyze the 'pure' influence of the position of a syllable in a phrase on the duration of the vowel. The figures disclose that phrase-final lengthening is much more distinct in ZH than in VS. On the one hand, the lengthening is more prominent in phrase-final syllables, on the other hand, its effect on penultimate syllables is clearly evident. Moreover, phrase boundaries are also marked with a phrase-initial lengthening in ZH, while this is not the case for VS. The two other dialects (BE and GR) behave along the lines of the ZH dialect.

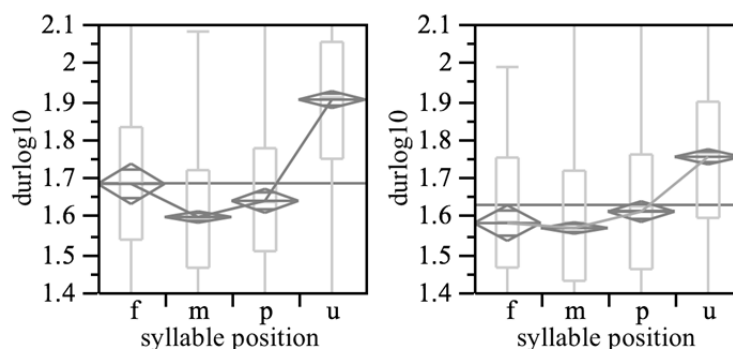


Figure 2. Box plot of duration and confidence intervals of phonological schwas in first (f), medial (m), penultimate (p) and ultimate (u) syllables of phrases in the ZH dialect (left) and the VS dialect (right).

Despite the different characteristics of phrase-final lengthening (and the additional phrase-initial lengthening), the connection between temporal changes and perceptual structuring of utterances is clearly visible for both

dialects. A phonological interpretation of the phonetic continuum seems thus appropriate. It should be noted, though, that the statistical dispersion in each position is very high, which points to the fact that the duration of an individual sound cannot unambiguously be interpreted and connected to a certain position of the syllable in the phrase. Even in the ZH variant, where we encounter a very distinct phrase final-lengthening, 8% of all schwas in mid-phrase position are longer than the mean of the schwas in phrase-final position; in return, 8 % of all schwas in phrase-final position are shorter than the mean of the mid-phrase schwas. For VS, this value even amounts to 20%.

6. Intonation

6.1. Methods

The methodological framework chosen to analyze f_0 contours in the present contribution is somewhat unorthodox. We do not follow the dominant autosegmental phonology methodology (Goldsmith 1976, Liberman and Prince 1977) and the derived transcription system therefrom, i.e. ToBI (Pierrehumbert 1980). Given the distinct dialectal diversity of German-speaking Switzerland, it is considered appropriate to apply a model that has the ability to detect phonetic details with great specificity. These objective measures can then serve as the basis for phonological interpretations. Further methodological concerns as to the reasons for opting for a phonetic intonation model will be illustrated in the discussion. Intonation contours are therefore explored using the Fujisaki, i.e. the Command-Response model.

The Command-Response Model 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 output signals of the two mechanisms are added onto the smallest asymptotic value (F_b) of the f_0 contour to be generated. For analysis purposes, the model decomposes the f_0 contour into a set of components from which timing and frequency information can be estimated. The PC can be applied for a description of the global declination tendency of f_0 . The AC is understood as a device for marking segments more f_0 -prominent on the local level. f_0 contours in our data were analyzed by means of Mixdorff's FujiParaEditor (2008). The f_0 behavior in each of the afore-mentioned variables was analyzed using parametric and

non-parametric statistical tests against the background of detecting dialect-specific as well as cross-dialectal differences. Dialect-specific multiple linear regression models were generated, which allow for a distillation of the relative contribution of independent variables towards explaining f_0 variability in a given parameter in a specific dialect.

In the subsequent presentation of the results, a particular focus will be placed on how the variable stress does – or does not – affect f_0 behavior. This variable deserves particular attention since, as mentioned earlier, the methodological framework of intonational phonology implicitly assumes that f_0 modulations occur on or in the vicinity of stressed syllables.

6.2. Distribution of stressed syllables in accent commands

Most ACs contain only one syllable with lexical stress. 15% of all ACs incorporate two or more stressed syllables. Interestingly, however, more than a third of all accents do not contain any stressed syllables at all. This finding is congruent with the insights put forth by Kochanski et al. (2005) and Silipo and Greenberg (2000). A great number of unstressed syllables in their corpus of spontaneous speech are marked with distinct f_0 movements. This finding corroborates the meaningfulness of treating f_0 and stress as separate variables.

Secondly, this result may further serve as evidence of what is frequently found in the literature on both Swiss German (see Hegetschweiler 1978: 24) as well as Swiss High German intonation (Ulbrich 2005: 320): Swiss German default accents often demonstrate a low f_0 in an otherwise stressed syllable, and a high f_0 in subsequent, otherwise unstressed syllables. This incongruence between stress and pitch has been observed particularly for the Alpine varieties. In the ToBI framework, such accents can be labeled as L*+ H (cf. Fitzpatrick-Cole 1999).

6.3. Amplitude of stressed syllables in accent commands

Overall, we find the highest amplitudes in ACs that contain 1 or more stressed syllables. If the AC does not contain any stressed syllables, it is generally lower in amplitude. This finding underlines the phenomenon that, in the stream of speech, metrical stress can cause higher f_0 excursions, and is congruent with the vast amount of literature on acoustic correlates of stress in German (see for example Isačenko and Schädlich 1966). If we take into consideration the findings put forth at 6.2, we can conclude that even

though f_0 excursions may be caused by stressed syllables, this need not necessarily be the case. What seems to be happening, however, is that f_0 excursions that are caused by stressed syllables are higher in amplitude than f_0 excursions for ACs without stress. In other words, metrical stress does not have to be accompanied by local f_0 movements (accent commands), but if it is, stress seems to cause distinctly higher AC amplitudes.

All dialects exhibit roughly the same proportions of ACs with 0 stressed syllables, yet, we find that the differences in amplitude between ACs with 0 stress and ACs with 1 or more stressed syllables are more distinct for the Midland varieties than the Alpine varieties. We find a striking North-South divide with the Alpine varieties showing similar amplitudes for all AC types, regardless of whether the AC contains no or several stressed syllables. This ties in with Wipf's (1910: 22) observation on VS Swiss German that unstressed syllables can also carry higher tones, as well as Meinherz' (1920: 38ff.) remark that weak syllables in the Grison dialect often carry higher pitch accent than highly dynamic ones. In comparison, we observe a distinct difference between no stress ACs (low amplitude ACs) and ACs with one or more stressed syllables (much higher amplitudes) for the Midland varieties. Put differently, the contribution of metrical stress to AC amplitude seems to occupy a more critical role in the Midland varieties, particularly in the BE variety.

6.4. Effect of stress in multiple linear regression models

The most striking differences between the Alpine and Midland groups are found in the relative weight of the *linguistic* predictors in the *AC amplitude* models, including the predictor *stress*. Figure 3 shows the multiple linear regression (MLR) calculated on each dialects' speakers' AC amplitudes.

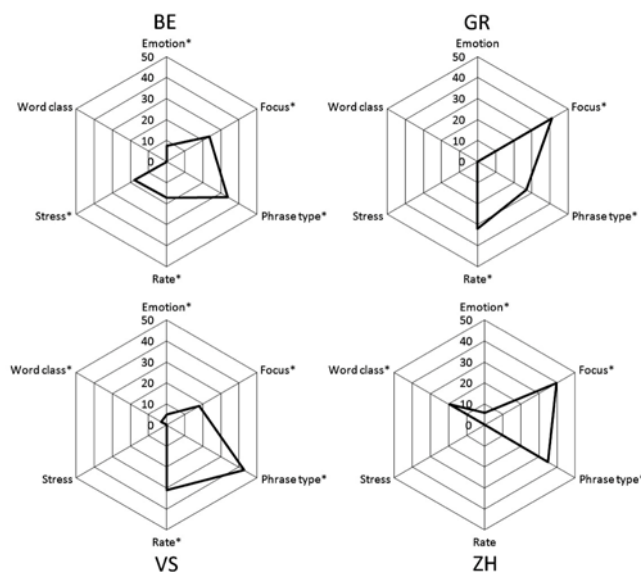


Figure 3. MLR of AC amplitudes for all four dialects (From: Leemann 2012. Reprinted with kind permission from John Benjamins Publishing Company, Amsterdam/Philadelphia).

The variables taken into consideration in this MLR are emotion (5 levels – neutral, bored, angry, happy, sad), focus (2 levels – no focus / focus), phrase type (3 levels – continuing, terminating, question), articulation rate (speaker specific in syllables / second), stress (2 levels – stress / no stress), and word class (2 levels – lexical / grammatical). The MLRs in Figure 1 points to the fact that stress, as a linguistic predictor, bears little power in f_0 movement prediction in all dialects, except for the BE dialect (adjusted $R^2 = .13$; $F(14, 2537) = 29$, $p < .0001$). In the ZH dialect, *stress* proves to be a highly significant predictor in bivariate tests; in the generated models, however, stress just fell short of reaching significant levels.

An explanation as to the GR speakers' low sensitivity to lexical stress may be lie in the GR speakers' contact with Romansh and Italian, two Romance languages also spoken in the canton of GR. Italian shows penultimate and antepenultimate stress and exhibits right-headed rhythmic groups frequently featuring low-high f_0 movements (see Hirst and Di Cristo 1998: 24, Rossi 1998: 220). Romansh, too, exhibits lexical accents in word-final or penultimate position (see Cavigelli 1969). Since in most Germanic languages, feet are left-headed, while Italian and Romansh are right-headed, one may speculate that the Grisons dialect can be regarded as a mix-version

of these two stress systems. Note, also, that Grisons varieties frequently feature the archaic feature of non-reduced word-final syllables, which may too, add to distinct *f0* modulations in unstressed syllables. One may conclude from this is that if the Grisons, over centuries, alternatively incorporated both rhythmic group patterns, it could be hypothesized that stress will eventually lose its importance, since stress is no longer perceived as discrete. Therefore, we hypothesize that the generally devalued variable stress in the Grisons dialect is likely to have little effect on the variance of *f0* contours.

As for the VS speakers' low sensitivity towards stress, illustrated in Figure 1, the same arguments as put forth for the GR's low sensitivity towards stress may apply. French, with which the VS speakers are in contact in the West, is a language in which the prominence markers *loudness*, *duration*, and *fundamental frequency* are correlated only little. These prominence marking parameters are set according to the first and the last syllable of the word: the first syllable normally shows a rise in *f0*, while the word-final syllable may exhibit a variety of prominence contrasts, frequently, however, a rise in *f0* (see Welby 2006). The exposition of the Valais dialect to the prominence systems of French may over centuries have led to an interesting mix. This language contact may have contributed to complex and somewhat unpredictable *f0* variability that Wipf (1910) alludes to. In addition, Valais varieties, too, commonly feature the archaic feature of non-reduced word-final syllables. These may too contribute to distinct *f0* modulations in unstressed syllables.

We can conjure alternate interpretations concerning the distinct difference between Alpine and Midland dialect behavior. Exploring language and migration history may provide one way of tapping into these differences. Given the mountainous terrain, Alpine varieties may have served as linguistic refuges over the past centuries and - in that sense - may represent what Johanna Nichols (1993) refers to as residual zones. Here, the highest Alemannic varieties were preserved, retaining what are now described as archaic features. On a segmental level, these differences can be reconstructed in part (Wiesinger 1983: 829, Hotzenköcherle 1984). However, a historical reconstruction of prosodic – particularly intonational – features is an impossible endeavor given the apparent lack of audio data from past centuries.

7. Discussion

First of all, during the prosodic analysis of spontaneous speech one faces many challenges that cannot be solved on phonetic or phonological grounds alone, because phonetics and phonology are closely interrelated. The transcription as well as the segmentation processes themselves do not allow for an analysis of purely phonological entities – since we are given only a purely phonetic realization in the signal of which a phonological representation has to be abstracted. This basic phonetic realization contains reduction of sounds, coarticulation, allegro forms, language change and linguistic variation. It is these phenomena which do not allow for a uniform phonological representation of words, of sounds, and of phrases. Phonetic considerations, perception, semantics and syntax intervene when it comes to defining the basic units of the analysis. Even prosody itself cannot be excluded in defining phrase boundaries, for example, and if we do include prosodic cues in our definition of phrase boundaries, it is not clear if there is a phonological or a phonetic view on it. The dichotomous view on phonetics and on prosody is fuzzy, to say the least. Decisions in data preparation are therefore methodologically highly relevant and, accordingly, must be stated very clearly.

Evidence from a large corpus of Swiss German dialectal speech underlines the detachment of stress from phonetic data – which is particularly true in the context of spontaneous speech. In the temporal domain the phonological distinction of stressed and unstressed syllables is at least partially reflected in phonetic duration, albeit with a great variance, so that a direct link of stress and duration cannot be made, especially because the position in the phrase – beside others not mentioned here (cf. van Santen 1998) – affects segment duration and interferes with stress. However, phonologically short segments are lengthened by stress while phonologically long vowels show little or no effect of stress on duration.

Concerning intonation, the problematic notion of stress is more obvious. Especially for *f₀*, stress and / or prominence seem to be concepts of degree rather than binarisms. In the autosegmental framework, however, it is not the aim to capture continuous *f₀*-movements that signal prominence. To begin with, it is not clear whether ToBI (Silverman et al. 1992, Grice and Baumann 2002 for German), the formal notation system that grew out of the autosegmental framework, is intended to provide phonetic transcriptions of intonation, phonological transcription, or possibly neither of the two (Grabe 1998). In addition, Taylor (2000: 1709) critically indicates that

“there has been no evidence to show that there are strict boundaries between intonational units which signal abrupt changes in meaning”. He continues to say that if intonational sound *SA* gives rise to meaning *MA* and sound *SB* gives rise to meaning *MB*, then a sound half-way between *SA* and *SB* can certainly give rise to a meaning somewhere between *MA* and *MB* (*ibid.*). Along these lines Fox (2000) adds:

[T]he continuous phonetic scale is reflected in a parallel continuous scale of meaning. It is therefore difficult to identify on the basis of the criterion of distinctiveness of meaning a restricted number of phonologically distinct entities which underlie the very large number of occurring manifestations (Fox 2000: 275).

In other words, *f0* movements seem to be a matter of degree and continuity, not a binary, high-low, concept. Methodologically, then, the use of a quantitative model, which allows one to model every *f0* movement there is, seems more optimal. For the temporal aspect, the phonological claim is the same, and here, the traditional phonological distinctions are by and large found in the data. However, the duration of a particular sound is very variable, so that also in timing an unambiguous attribution of a duration pattern to a stress value or to a specific syllable position within the phrase is not possible.

Furthermore, opting for a quantitative account of prosodic features of Swiss German constitutes a significant contrast to a majority of intonation studies working in abstract and symbolic frameworks. Here, the first methodological step consists of analyzing and parametrizing the *f0* contour. Only in a second step we establish the linguistic analysis of these mathematical parameters and their relation to the individual segments. This provides innovative insight into dialectal *f0* contours that is not conceivable with symbolic, syntactic, or functional conversational analytical analyses. Hence, the findings in the current study can complement, specify, and support existing findings on *f0* patterns and on statements on temporal aspects of Swiss German. In addition, even minor differences in *f0* realizations and in durational relations, albeit on a subphonemic level, may in the end prove to be perceptually relevant for a cross-dialectal comparison – as it has been attested for the segmental level (cf. Haas 1978). The different temporal and intonational patterns in marking phrase boundaries will most probably not be of phonological difference, nevertheless, they show different models of marking the same function. These differences characterize each dialect with a specific sound that is perceived and stereotypically attributed (Zimmermann 1998).

References

- Barry, William, Bistra Andreeva, and Ingmar Steiner
2007 The phonetic exponency of phrasal accentuation in French and German. In *Interspeech 2007*. Antwerpen: 1010–1013.
- Bolinger, Dwight
1972 Accent is predictable (if you're a mind-reader). *Language* 48: 633–44.
- Browman, Catherine P., and Louis Goldstein
1992 Articulatory Phonology: An Overview. *Phonetica* 49: 155–180.
- Byrd, Dani, and Susie Choi
2010 At the juncture of prosody, phonology, and phonetics – The interaction of phrasal and syllable structure in shaping the timing of consonant gestures. In *Papers in Laboratory Phonology 10*, Cécile Fougéron et al. (eds.), 31–60. Berlin: Mouton de Gruyter.
- Cavigelli, Pieder
1969 *Die Germanisierung von Bonaduz in geschichtlicher und sprachlicher Schau*. Frauenfeld: Huber.
- Cruttenden, Alan
1997 *Intonation*. Cambridge: Cambridge University Press.
- Cutler, Anne, and Robert D. Ladd
1983 *Prosody: models and measurements*. Berlin: Springer.
- DUDEN
2005 *4. Grammatik. 7th Edition*. Mannheim, Leipzig, Wien, Zürich: Dudenverlag.
- Fitzpatrick-Cole, Jennifer
1999 The alpine intonation of Bern Swiss German. In *Proceedings of the XIVth International Congress of the Phonetic Sciences (ICPhS)*, San Francisco: 941–944.
- Fleischer, Jürg, and Stephan Schmid
2006 Zurich German. *Journal of the International Phonetic Association* 36 (2): 243–253.
- Fletcher, Janet
2010 The Prosody of Speech: Timing and Rhythm. In *The Handbook of Phonetic Sciences, Second Edition*, William J. Hardcastle, John Laver, and Fiona E. Gibbon (eds.), 524–602. Oxford: Blackwell.
- Fox, Anthony
2000 *Prosodic Features and Prosodic Structure. The Phonology of Suprasegmentals*. Oxford: Oxford University Press.
- Fujisaki, Hiroya, and Keikichi Hirose
1982 Modeling the dynamic characteristics of voice fundamental frequency with applications to analysis and synthesis of intonation. In *Pre-*

- prints of the Working Group on Intonation. 13th International Congress of Linguistics, Tokyo: 57–70.*
- Gilles, Peter
2005 *Regionale Prosodie im Deutschen. Variabilität der Intonation von Abschluss und Weiterweisung.* Berlin/New York: de Gruyter.
- Goldsmith, John
1976 *Autosegmental and Metrical Phonology.* New York: Garland.
- Grabe, Esther
1998 *Comparative Intonational Phonology: English and German.* PhD Thesis. Wageningen: Ponsen en Looien.
- Grice, Martine, and Stefan Baumann
2002 *Deutsche Intonation und GToBI. Linguistische Berichte* 191: 267–298.
- Haas, Walter
1978 *Sprachwandel und Sprachgeographie. Untersuchungen zur Struktur der Dialektverschiedenheit am Beispiele der schweizerdeutschen Vokalsysteme.* Wiesbaden: Steiner.
- Hegetschweiler, Peter
1978 *Comparing Native English Intonation and English Intonation of Swiss Germans.* Unpublished MA thesis, University of Zürich.
- Hotzenköcherle, Rudolf
1984 *Die Sprachlandschaften der Schweiz.* Aarau, Frankfurt, Salzburg: Sauerländer.
- Hirst, Daniel J., and Albert Di Cristo
1998 *A survey of intonation systems.* In *Intonation Systems: A Survey of Twenty Languages.* Daniel J. Hirst, and Albert Di Cristo (eds.), 1–44. Cambridge: Cambridge University Press.
- Isačenko, Alexander V., and Hans-Joachim Schädlich
1970 *Untersuchungen über die deutsche Satzintonation.* In *Untersuchungen über Akzent und Intonation im Deutschen,* Manfred Bierwisch (ed.), 7–67. Berlin: Akademie Verlag.
- Klatt, Dennis H.
1976 *Linguistic uses of semental duration in English: Acoustic and perceptual evidence.* *Journal of the Acoustical Society of America* 59 (5): 1208–1221.
- Kochanski, Greg, Esther Grabe, John Coleman, and Burton Rosner
2005 *Loudness predicts prominence: Fundamental frequency lends little.* *Journal of the Acoustical Society of America* 118 (2): 1038–1054.
- Kraehenmann, Astrid
2003 *Quantity and Prosodic Asymmetries in Alemannic. Synchronic and Diachronic Perspectives.* Berlin/New York: Mouton de Gruyter.

- Ladd, D. Robert
 2006 *Intonational Phonology*, 2nd ed. Cambridge: Cambridge University Press.
- Leemann, Adrian
 2012 *Swiss German Intonation Patterns*. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Leemann, Adrian, and Beat Siebenhaar
 2007 Intonational and Temporal Features of Swiss German. In *Proceedings of the XVIth International Congress of the Phonetic Sciences (ICPhS)*, Saarbrücken: 957–960.
- Leemann, Adrian, and Beat Siebenhaar
 2008a Swiss Alpine and Midland Intonation. In *Proceedings of Speech Prosody 2008*, Campinas, Brazil: 289–292.
- Leemann, Adrian, and Beat Siebenhaar
 2008b Perception of Dialectal Prosody. In *Proceedings of Interspeech 2008*, Brisbane, Australia: 524–527.
- Leemann, Adrian, and Beat Siebenhaar
 2010 Statistical Modeling of *F0* and Timing of Swiss German Dialects. In *Proceedings of Speech Prosody 2010*, Chicago, 1–4. 9. Mar. 2012 (<http://speechprosody2010.illinois.edu/papers/100180.pdf>)
- Lehiste, Ilse
 1970 *Suprasegmentals*. Cambridge, MA: MIT Press.
- Liberman, Mark, and Alan Prince
 1977 On stress and linguistic rhythm. *Linguistic inquiry* 8: 249–336.
- Low, Ling E. E., Esther Grabe, and Francis Nolan
 2000 Quantitative characterizations of speech rhythm: Syllable-timing in Singapore English. *Language and Speech* 43 (4): 377–401.
- Meinherz, Paul
 1920 *Die Mundart der Bündner Herrschaft*. Frauenfeld: Huber.
- Mixdorff, Hansjörg
 2008 *Fujisaki Parameter Extraction Environment*. 9. Mar. 2012 (<http://public.beuth-hochschule.de/~mixdorff/thesis/fujisaki.html>)
- Nichols, Johanna
 1992 *Linguistic diversity in space and time*. Chicago: The University of Chicago Press,
- Öhman, Sven
 1967 Word and sentence intonation: a quantitative model. *STL-Quarterly Progress Status Report* 2–3: 20–54.
- Pierrehumbert, Janet
 1980 The Phonology and Phonetics of English Intonation. Unpublished PhD Thesis, MIT.

- Ramus, Franck, Marina Nespor, and Jacques Mehler
1999 Correlates of linguistic rhythm in the speech signal. *Cognition* 72: 1–28.
- Rossi, Mario
1998 Intonation in Italian. In *Intonation Systems: A Survey of Twenty Languages*, Daniel J. Hirst, and Albert Di Cristo (eds.), 219–239. Cambridge: Cambridge University Press.
- Siebenhaar, Beat, Brigitte Zellner Keller, and Eric Keller
2001 Phonetic and Timing Considerations in a Swiss High German TTS System. In: *Improvements in Speech Synthesis*. Keller, Eric, Gérard Bailly, Alex Monaghan, Jacques Terken, and Mark Huckvale (eds.), 165–175. Chichester: Wiley.
- Silipo, Rosaria, and Steven Greenberg
2000 Prosodic stress revisited: Reassessing the role of fundamental frequency. *Proceedings of the NIST Speech Transcription Workshop*, College Park, MD.
- Silverman, Kim, Mary Beckman, John Pitrelli, Mori Ostendorf, Colin Wightman, Patti Price, Janet Pierrehumbert, and Julia Hirschberg
1992 TOBI: A Standard for Labeling English Prosody. *Proceedings of the 1992 International Conference on Spoken Language Processing*, vol. 2: 867–870.
- Taylor, Paul
2000 Analysis and Synthesis of Intonation using the Tilt Model. *Journal of the Acoustical Society of America* 107 (3): 1697–714.
- Trubetzkoy, Nikolaj
1939 *Grundzüge der Phonologie*. Prague: Cercle linguistique de Prague.
- Ulbrich, Christiane
2005 *Phonetische Untersuchungen zur Prosodie der Standardvarietäten des Deutschen in der Bundesrepublik Deutschland, in der Schweiz und in Österreich*. Frankfurt am Main: Peter Lang.
- Vaissière, Jacqueline
1983 Language-Independent Prosodic Features. In *Prosody: Models and Measurement*, Anne Cutler and D. Robert Ladd (eds.), 3–66. New York: Springer.
- van Santen, Jan
1998 Timing. In *Multilingual Text-to-Speech Synthesis: The Bell Labs Approach*, Richard Sproat (ed.), 115–139. Dordrecht, Boston, London: Kluwer Academic Publishers.
- Welby, Pauline
2006 French intonational structure: Evidence from tonal alignment. *Journal of Phonetics* 34: 343–371.

Wiesinger, Peter

- 1983 Die Einteilung der deutschen Dialekte. In *Dialektologie. Ein Handbuch zur deutschen und allgemeinen Dialektforschung. Zweiter Halbband*, Werner Besch et al. (eds.), 807–900. Berlin/New York: de Gruyter.

Willi, Urs

- 1996 *Die segmentale Dauer als phonetischer Parameter von 'fortis' und 'lenis' bei Plosiven im Zürichdeutschen. Eine akustische und perzeptorische Untersuchung*. Stuttgart: F. Steiner.

Wipf, Elisa

- 1910 *Die Mundart von Visperterminen im Valais*. Frauenfeld: Huber.

Zimmermann, Gerhard

- 1998 Die 'singende' Sprechmelodie im Deutschen. Der metaphorische Gebrauch des Verbuns 'singen' vor dem Hintergrund sprachwissenschaftlicher Befunde. *Zeitschrift für Germanistische Linguistik* 26: 1–16.