

PROSODIC FEATURES OF SPONTANEOUS UTTERANCE-INITIAL PHRASES IN BERNESE AND VALAIS SWISS-GERMAN

**Adrian Leemann, Department of Linguistics, University of Berne,
Switzerland adrian.leemann@isw.unibe.ch**

**Beat Siebenhaar, Department of Linguistics, University of Berne,
Switzerland beat.siebenhaar@isw.unibe.ch**

A B S T R A C T

The rationale behind the study at hand is to trace the question of whether spontaneous utterance-initial phrases are prosodically different from the rest of an utterance. In an interview setting, 16 speakers from two different Swiss-German dialect regions were recorded. The subjects, grammar school students, were asked questions about their plans after school, their free time etc. The students' answers were divided into empty pauses, filled pauses, first phrases, and the rest of the utterance. Fundamental frequency (F0) measurements indicate that there are significant changes in F0 from the filler to the first phrase and from the first phrase to the rest of the utterance in both speaker groups. Results further show the prevalence of two patterns in F0 alterations: one pattern demonstrates a steady declination in F0 from the filled pause to the first phrase and the rest of the utterance while the other pattern features a low F0 in the filled pause followed by an F0-increase in the first phrase and again ending in a declination in F0 for the rest of the utterance. It is further shown that Valais speakers present longer and more pauses than Bernese speakers do – a fact that may be important for conversation analysts. From the results it follows that if models of intonation are to be set up, first phrases and possible preceding filled and empty pauses must be given special consideration in the conceptualization of the models.

Keywords: Prosody, Conversation Analysis, Swiss-German, Turn-Taking, Intonation

INTRODUCTION

Spontaneous and prepared speech differ in several ways. Abercrombie (1966) characterizes the latter as having a standardized intonation pattern, little variation in tempo, pauses that are set according to grammatical structures and as possessing little or no disfluency. The former, in contrast, may feature the omission of syntactic elements as well as the overwhelming presence of fillers and hesitations. Abercrombie advocates the study of “genuine spoken language” (ibid.: 9) in the same way Fox Tree believes that “[t]he phenomena that are the hallmark of spontaneous talk have often been thought of as unwanted elements of speech, unfortunate by-products of speaking on the fly. However, another way of viewing these phenomena is as an integral part of the communicative enterprise” (2000: 376), a stance this study, too, affiliates itself with.

Intonation and speech rate seem to be the most researched aspects of prosody in contrastive analyses between spontaneous and prepared speech. In terms of pitch, Swerts et al. (1996) found that the F₀ tends to be higher in read-aloud tasks than in spontaneous speech; while in spontaneous speech, intonation is more varied (Syrdal 1996). Speech rate, too, is higher in conversational speech, which leads to reductions in vowels. Speech disfluencies, such as filled and empty pauses and hesitation phenomena, are, too, a typical characteristic of spontaneous speech - disfluencies per word in spontaneous English speech vary from 5-10% (Shriberg 1999). From a pragmatic perspective, fillers such as *uh* and *um* announce delays in speaking; they offer extra time for the speaker to search for the desired word or the adequate syntactic structure, which may not be accessible at that very moment (Clark and Fox Tree 2002). Shriberg (2001) further notes that disfluencies tend to occur predominantly in utterance-initial positions. The present paper explores these utterance-initial phrases on a prosodic level and within the framework of our current National Science Foundation (NSF) research project “Quantitative Approaches to Geolinguistics of Swiss-German Prosody”.

Swiss-German dialects are used by all Swiss speakers in nearly every communicative situation. It is thus not exceptional, for instance, that a member of the Swiss Government, who participates in a TV broadcast with a director of a bank, would discuss political issues in his/her own dialect. This underlines the fact that there are no significant differences in terms of the prestige between the dialects and is one of the reasons why Swiss-German dialects have been examined reasonably well on most linguistic levels. There is, however, a lack of prosodic descriptions of the dialects. This is where the current project pitches in: in recording two Alpine and two Midland dialects, we try to work out a gross geolinguistic model that is geared at revealing the main prosodic features of these dialects. In four different places (Bern, Zürich, Brig, and Chur), 20-30 subjects are recorded. The data is collected via spontaneous interviews that include questions regarding the informants' goals after graduation, what they do in their next vacation etc. The prosodically most relevant parameters, time and fundamental frequency, are then extracted and modeled. The comparison of the subjects from each location allows for a distinction to be made between region-specific and individual prosodic characteristics. The comparison between the different recording locations offers insight into the geolinguistic structure of prosody.

With respect to modeling the intonation and the timing of individual speakers and speaker groups, however, we encountered the following problem: it is a well-known fact that within a conversation setting prosody varies according to discourse structure. Brown et al. (1980) concluded that new topics, which are regularly introduced by question answer pairs, are often presented in a comparatively higher pitch. Longer pauses, too,

have been associated with shifts in topic (*ibid.*), and speech rate also varies according to discourse structure. The question thus arose of how such distinct discourse structure-related differences in prosody could be incorporated in the models resulting from our project.

From the above it follows that this study is not only phonetic in its nature but also conversation analytic. We want to scrutinize the acoustic correlates of utterance-initial phrases; phrases, which must be viewed in the larger context of the conversation, because they are articulated by speakers who are randomly and spontaneously chosen by the researcher. This conversation analytic aspect will be attended to in the first section of this paper, followed by the discussion of the phonetic component. To establish the link between phonetic research, i.e. prosody research in our case, and conversation analysis (hereafter CA), it is then shown that, despite the vast amount of literature on prosodic features of phrase-final structures, it seems that little research has been conducted on the prosodic features of spontaneous *utterance-initial* phrases and phrase-initial features.

1 CONVERSATION ANALYSIS AND PHONETICS

1.1 Conversation analytic component

For Harvey Sacks conversation analysis is the study that “seeks to describe methods persons use in doing social life [...]” (1984: 25). A core concept in CA is that of turn-taking, i.e. “[t]he talk of one party bounded by the talk of others [...], with turn-taking being the process through which the party doing the talk of the moment is changed” (Goodwin 1981: 2). Sacks et al. (1974) designed a set of rules that account for places where a next turn can be anticipated. Among other things, these instances of possible turn-takings, which they refer to as transition-relevance places, are enacted with signals, such as discourse markers as well as syntactic and semantic features in the ongoing turn. More importantly, in the context of the present study, prosodic features such as pausing, duration of segments, and intonation constitute further turn-yielding signals (Taboada 2006: 7). Selting and Couper-Kuhlen (1996: 11) welcome the fact that CA, as a socially oriented approach towards the study of language, has acknowledged the importance of prosodic features in language-in-interaction, as opposed to studying prosody from a structuralist point of view.

This overlap between CA and phonetics is also appreciated by Local (2003: 1), who adds that despite the large number of available corpora of spontaneous speech surprisingly little has been used for further, talk-in-interaction, analyses. Local poses questions of the following nature “[h]ow do speakers/listeners manipulate fine phonetic detail in producing and interpreting the moment-to-moment flow of everyday conversation?” (*ibid.*). This is, in fact, one of the aspects conversation analysts have not addressed thoroughly enough, so Selting and Couper-Kuhlen (1996).

For the transcription process, conversation analysts more often than not apply Jefferson’s transcription system (cf. Jefferson 2004 – for the German ‘Gesprächsanalytisches Transkriptionssystem’ see Selting et al. 1998), with the aim of capturing talk as it occurs in daily conversations, “[...] in all its apparent messiness [...]” (Hutchby and Wooffitt 1998: 75). Such an ambitious goal obviously meets with criticism. Kendon (in Hutchby and Wooffitt 1998: 76) points out that “[i]t is a mistake to think that there can be a truly neutral transcription system [...]. Transcriptions, thus, embody hypotheses.” In CA’s defense, Hutchby and Wooffitt (*ibid.*) state two main goals of CA: first, to describe the dynamics of turn-taking and second to elucidate the characteristics of speech delivery, including prosodic features such as stress, pauses, enunciation, intonation, pitch etc. Such transcriptions include, for example, the measurement of pauses

in tenths of seconds (*ibid.*: 81) or the most literal transcription of laughter as possible (*ibid.*: 83).

While Local and Kelly (1989b: 204) believe that pausal phenomena and audible respiratory activity are consistent in CA, they argue that tempo, pitch, loudness, vowel quality, voice quality etc. are often rendered inconsistently and arbitrarily in CA transcripts. Hutchby and Wooffitt (1998: 77) take note of such a counter argument, yet state that if a CA analyst were to pay closer attention to phonetic phenomena, transcripts would go beyond the reader's understanding thereof. Moreover, conversation analysts believe that CA has a different aim, namely "to get as much of the actual sound as possible into our transcripts, while still making them accessible to linguistically unsophisticated readers" (Sacks, Schegloff, and Jefferson 1974: 734). Such a position is, however, problematic as fine phonetic observations can shed light on details in the analysis of a conversation that are not revealed if one adheres to such methodologies.

Clark (1993), for instance, discusses the nature of answer-questions adjacency pairs. In answering a question, the respondent often delays his answer. The timing of such delays is crucial, as it makes the delay open for interpretation on the part of the questioner. Did the respondent not understand the question, not retrieve the required information? Can she/he not formulate his response? In order to find answers to such questions, exact measurements of pausal duration, measurements beyond stopwatch-timing, need to be made. A further example is rising or falling intonation in an answer to a question. Rising intonation often denotes uncertainty on the part of the respondent, in contrast to falling intonation in an answer, which does not leave open such an ambiguous interpretation. The phonetic correlate can be minimal, its interpretation, on the other hand, may not be the one intended by the speaker.

This is, in part, where this study starts off. While CA transcripts may reveal that utterance-initial phrases are indeed prosodically different from the rest of the utterance, we want to illustrate what this looks like on a more detailed level. This is achieved by means of instrumental analyses (fundamental frequency measurements and pause measurements). In other words, we want to bring to light the prosodic features of utterance-initial phrases as provided by a speaker who, heteronomously, provides answers to our questions.

1.2 Phonetic component

First descriptions of German prosody date from the late 19th century with the emergence of phonetics. Even early monographs on Swiss-German dialects (Vetsch 1910, Wipf 1910 and others) hold sections on dialectal prosody. However, early impressionistic descriptions that included statements on the geolinguistic distribution of prosodic patterns (Bremer 1893, Sievers 1912) could not be verified until the present day (Gilles 2005). Based on perception models, Isačenko and Schädlich (1964) built models for sentence intonation which questioned the prevailing syntax based models (Bierwisch 1966). These were revised by communicative oriented descriptions (Féry 1993, Selting 1995). Actual research on linguistic-based German prosody mainly has three branches: prosody as part of pragmatics (e. g. Kohler 1991) and interaction (Selting 1995), phonological representation of intonation (e. g. Grice & Baumann 2002, Gibbon 1998), and variationist research on differences of regional prosody (e. g. Gilles 2005, Peters 2004, Siebenhaar 2004, Siebenhaar et al. 2004). In addition, speech technology views prosody from a technology-based perspective.

Results from studies that describe standard German – which in some regions is still more of a construct than a real norm, especially in prosody – can only be referred to as a background to our dialect data. Kohler's (1987, 1991) perception studies on the

communicative function of pitch alignment as well as the descriptions and perceptive tests of the interplay between different phonetic aspects in determining phrase boundaries in spontaneous speech (B. Peters et al. 2005), are based on a German standard. However, it is not mentioned that this is the standard German spoken in Northern Germany. Correspondingly, Atterer and Ladd (2004) have shown that even in laboratory read speech there are prosodic differences in standard German that can be reduced to regional aspects. They have shown that speakers from southern Germany generally show later peaks than speakers from northern Germany do. Consequently, regional intonation has become a focus of research in the last decade as seen in publications by Gilles, Peters, Selting, and Auer. They describe intonation patterns that are specific to a region or that have a different communicative function, in one region as opposed to another. Their work is grounded in the description of the contours of final nucleus syllables in their functional distinction of termination and continuation. The comparison shows a geolinguistic difference between southern and northern regions with preferences for different patterns, but the phonetic distinctions are not as apparent as they are on the segmental level (Gilles 2005).

Siebenhaar and co-workers (Siebenhaar 2004, Siebenhaar et al. 2004, Häsler et al. 2005) made first modern attempts at a description of Swiss-German dialectal prosody. The prosody of interviews of three speakers from two different dialect regions is analyzed in such a way as to subsequently formulate models for a dialectal speech synthesis system. The aim was that the models should build a methodological basis with which to compare dialectal prosody. These analyses, indeed, indicate clear differences between the speakers. Unfortunately, the results could only partially be traced to dialectal reasons, as the sample was too small. However, it can be said that timing seems to be more stable than intonation. The results correspond to the findings of Keller (1994, based on data by Caelen-Haumont 1991) who demonstrated that the durational domain is subject to more rigid constraints than the F0 domain. He showed that duration correlates more between speakers within a given syllable than this is the case with F0. The timing models achieve a correlation with real data that is nearly as high as that achieved with read speech. While in spontaneous speech timing is quite predictable and specific to every linguistic variety, intonation seems to be more variable and dependent on situational and/or functional factors.

The mentioned descriptions of German prosody focus on either timing, peak alignment of accents, or on the intonation contours of final nucleus syllables. Utterance-initial phrases, however, have not been researched extensively.

1.3 Prosodic features of utterance-initial phrases

With respect to turn-medial and turn-final phrases, Stephens and Beattie (1986) found that subjects who are presented with an audio recording were able to discriminate between turn-final and turn-medial utterances in the case of disagreements in conversation, thus highlighting the role of prosodic features in the regulation of turn-taking. In the analysis of an interview with Margaret Thatcher, Beattie et al. (1982) suggest that non-verbal turn-yielding signals include a low drop in pitch and loudness. Maclay and Osgood (1959: 20) propose that pausal phenomena, too, serve to identify the end of phrases and sentences. In comparison to research on turn-medial/turn-final utterances, turn-initial phrases have not been investigated as thoroughly.

Much of the research in the context of prosodic features of utterance-initial phrases has evolved around analyses of fillers in terms of their intonation and duration (cf. Shriberg 1999, Swerts 1998), and not around the acoustic analysis of *entire* turn-initial phrases as such. Analogously, there is abundant literature on the pragmatic (cf. e.g. Clark

1993, Corley 2003, Clark and Fox Tree 2002) and social psychological (cf. e.g. Cook and Lalljee 1973, Siegman and Pope 1965) contextualization of such discourse markers. In describing the acoustic correlates of disfluent speech, Shriberg (1999) mentions that the lengthening of syllables before the actual point of interruption is a typical feature of everyday, disfluent speech. Despite the duration modification in lengthened syllables, however, the fundamental frequency remains largely unaffected. Further she shows how vowels of filled pauses, most often acoustically similar to schwa, are articulated significantly longer than where the same vowel occurs in fluent contexts. With respect to the intonation of filled pauses, Shriberg finds a low F0 and a linear or a slightly gradual fall in pitch. Swerts (1998), alternatively, examines a possible correlation between filled pauses and discourse structure. He shows that pause fillers are more likely to occur in initial-phrases, if preceded by major discourse boundaries. Additionally, he concludes that the fillers in initial-phrases are segmentally as well as suprasegmentally different from those in phrase-medial or phrase-final positions.

2 METHODS

Given the assumption that utterance-initial phrases are prosodically different from utterance-medial or utterance-final phrases, it needs to be clarified how such a hypothesis can be tested. The data used in this study was retrieved from a corpus of spontaneous speech that was collected within the National Science Foundation (NSF) research project “Quantitative Approaches to Geolinguistics of Swiss-German Prosody” at the University of Berne. 25 subjects, all of whom attended grammar school and were aged between 18-22 at the point of the documentation, were recorded in Brig, which represents the Western alpine variety of Swiss-German. 25 subjects were recorded in Berne, a city that stands for the Western midland dialects of the country. From this pool of recordings, 16 were used for analyses in this paper: eight Valais (four female, four male) and eight Bernese (four female, four male) recordings. Given the premise that all Swiss speakers are able to understand each others’ dialects, as the entire spectrum of dialects is present in the media, the interviewer spoke in his local dialect. Accommodation phenomena are not expected on the part of either the interviewer or the interviewees, except perhaps in the case of interviewees who speak the Valais dialect (Schnidrig 1986).

With the aim of extracting as much spontaneous language as possible, the subjects were asked to answer a number of questions as part of a spontaneous interview. The interview consisted of questions such as “What do you think you will do once you have graduated?”, “What do you do in your spare time?”, “What does your next vacation look like?” etc. This form of interview is considered the most suitable method with which to collect naturally occurring language, since it sheds broader light on non-marked language use with a stranger. The interview constitutes roughly half of each of the 20-minute recording sessions. As said by Selting (1995: 243ff), such questions present non-restrictive, open questions, along with a renewed focusing (WH-questions or verb-first questions). The subject’s answer comprises the beginning of a narrative contribution to the conversation.

The conversation between the researcher and the informant was manually labeled with PRAAT (Boersma and Weenink 2006). What was of interest to us were the answers of the interviewees to questions of the interviewers. If appropriate, the first label that pertained to the subject is a filled pause (FP). Secondly, the first phrase (PHRASE) was marked off from the rest of the utterance (REST). This analysis was carried out on a perceptual basis: changes in intonation, pauses, and final lengthening were regarded as indicators for phrase boundaries. Possibly occurring non-filled pauses (#) between the

question and the PHRASE were labeled as well. Various measuring points were elicited. The most relevant ones are summarized below:

- A possible # preceding and/or following an FP.
- The type of FP: *aa*, *ää*, *mm*, *ja* (= *yes*), *ja+* (= "*yes*" plus other particle), *also* ("well"), "vowel" (= all other vowels), "repetition" (= repetition of a word of the question), *nei* (= "no"), "conjunction", and *ts*. It has to be stressed that 'real' *ja*, *nei* (yes/no) answers do not count as FP but as PHRASE.
- The sequence of #FP#, PHRASE, and REST (i.e. which of these components was present in the answer to the interviewer's question).
- Type of literal question (yes/no,¹ open).
- Fundamental frequency (F0) mean for each FP, PHRASE, REST occurrence.
- differences in F0 (in %) between FP/PHRASE and PHRASE/REST.
- FP duration in ms for each occurrence.
- Duration of # before FP in ms.

In the following sections, the findings of our study are described and discussed. Our concerns are fourfold: firstly, we explore whether or not utterance-initial phrases, as provided by the speakers in the sample at hand, are prosodically different from the rest of the utterance. Secondly, if there are differences, we question whether they can be attributed to the specific dialect regions. Thirdly, if they cannot be attributed to the speakers' dialectal backgrounds, we address what other factors they could be attributed to. Lastly, we want to attend to some of the repercussions of these findings for our current project and, from a broader perspective, on prosody research *per se*.

3 DATA ANALYSES

The following analyses focus on some of the relevant distinctions that could be extracted from our data. The first section deals with the use of pauses while the second addresses the fundamental frequency.

3.1 Use of pauses

First insights revealed a different use of filled and empty pauses between the two speaker groups.

3.1.1 FPs and type of question

The first issue that needs to be addressed is whether or not the type of question - open question or literal yes/no question - has an impact on the type of FP. The result of a contingency test is not significant. Thus, each type of FP is equally possible in the replies to both types of questions for speakers of both dialect groups.

3.1.2 FPs and #s

The total amount of answers the students provided comprises 353, out of which the speakers delivered 251 turn-initiations with an FP, i.e. in 71.1% of the cases they provided some sort of hesitation marker in their answers. In 50% of these FPs there is a preceding #, in 27% the FP is preceded and followed by a #, in 17% the FP stood alone, and in 5% the FP follows the question directly without a #, but it is followed by a #. If we look at the type of FP that was used by the students, we get the following distribution, *Figure 1*:

¹ These yes/no questions were also intended to be open questions, yet it turned out that they were formulated as "Can you tell me what you want to do after school?" which literally is a yes/no question.

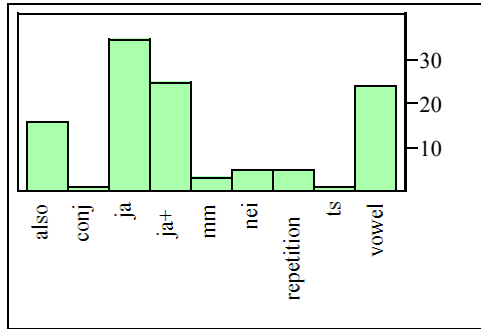


Figure 1: Frequency distribution of FP types.²

The distinction between FPs with no semantic meaning and FPs with minimal semantic meaning does not show any significant differences. With a frequency of nearly 40%, *ja* is by far the most recurrent type of FP. Therefore, this positive feedback (which is different from the *ja*, denoting an answer to a yes/no-question) indicates a generally positive attitude on the part of the interviewee towards the interviewer.

Further tests showed that there are differences between our Valais and Bernese speakers in terms of the types of FPs they use. Our Valais speakers use more *also*, and *ja+* than Bernese speakers do, while the Bernese fill their pauses with *ja* and *mm* more frequently. It also has to be considered whether there are differences in the use of FPs in their varying sequences, i.e. # following FPs, # preceding FPs, and the duration of # before FPs between Valais and Bernese speakers. A contingency analysis of the sequences of FPs (FP#, FP, #FP#, and #FP) between Bernese and Valais speakers suggests that they make use of or omit # differently, cf. Table 1:

	#FP	#FP#	FP	FP#
BE	82 52.9%	30 19.4%	36 23.2%	7 4.5%
WS	53 46.1%	43 37.4%	12 10.4%	7 6.1%

Table 1: Contingency table of FPs by Bernese and Valais speakers.

The contingency table indicates that #FP and FP alone are used more often by the Bernese speakers with nearly 53% and 23% respectively - while the Valais speakers show a large number of #FP#, namely 38%. Overall, the Valais speakers use significantly more #s to frame the FP than Bernese speakers do. In fact, these differences are significant for # before FP and after FP.³

A further test is used to address another significant issue, namely whether the #s before FPs provided by Valais speakers are quantitatively different from those of the Bernese speakers. An ANOVA indicates that this is indeed the case; it turns out that the #s before the FPs of the Valais speakers (777 ms) are significantly longer than those of the Bernese speakers (601 ms).⁴ When looking at the duration of the FPs, we find a mean of 480 ms. The Valais speakers' FPs tend to be longer than those of the Bernese speakers, but the difference is just under significant.

² *ja+* stands for the filler *ja* and a subsequent sequence of sounds, such as /ja: *m:/ or /ja: d*s:/ which corresponds to *yes, um* and *yes, that*.

³ # before FP: Fisher's Exact Test, 2-Tail, p=0.04; # after FP: Fisher's Exact Test, 2-Tail, p=0.0009.

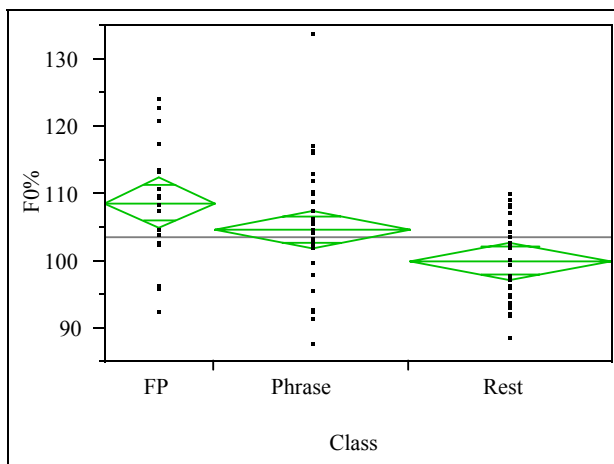
⁴ F = 5.9917, p = 0.0152, DF = 1.

3.1.3 Summary and discussion

If we recapitulate these findings, we get the following differences between Bernese and Valais speakers: the type of FP used and the duration thereof for the two groups is different. Both groups use or omit # differently. The Valais speakers have more # after and before FPs, and the Valais speakers' #s before the FPs tend to be longer. The reasons for these differences remain unclear. Valais speakers may have had more production difficulty than Bernese speakers, possibly they were less comfortable with the discussed topics, or they may have been less honest in answering the questions as opposed to the Bernese speakers. These are some of the aspects that need to be taken into consideration for the interpretation of filled and empty pauses, so Fox Tree (2002). Pope and Siegman (1965) assume that interview questions that are low in specificity may correlate with caution and hesitation markers in the interviewee's speech. However, frequency counts show that both groups are asked a nearly identical number of open or yes/no questions. Therefore, it may be that the interviewers' questions in the Valais recordings were perceived as having a low degree in specificity. While such explanations need to be viewed as possible causes, it seems more likely, however, that the results can in fact be ascribed to the different dialect regions. This different use of pauses can be viewed as evidence for regionally different communicative behavior, which should be taken into account by conversation analysts.

3.2 Fundamental frequencies

At this stage, we need to contemplate the fundamental frequency of FPs in the larger context of the FP-PHRASE-REST sequence. The following calculations were made by using the F0% values, where the basis for every subject is its mean F0 of the REST. This is beneficial in that subjects with different fundamental frequencies can directly be compared. The results indicate that there are two distinct patterns of F0 distribution from FP to PHRASE to REST. On the one hand, there is a gradual declination in F0 as the speaker progresses from the FP to the PHRASE to the REST. An example of this pattern is the F0% distribution of subject BE01m given in *Figure 2*.⁵



*Figure 2: Gradual declination of F0% from FP to PHRASE to REST.*⁶

Out of 16 speakers, five seem to adhere to this pattern of F0 declination. In stark contrast, however, all the other speakers pursue a different pattern of F0 modification in the course

⁵ BE01m stands for Bern, subject Nr. 1, male; WS stands for Wallis, i.e. Valais.

⁶ What looks like a diamond in these figures indicates the confidence interval (95%), while the dots show the dispersion, and the diamond-central line the median.

of the FP to the PHRASE to the REST: low F0 in FP to high F0 in PHRASE to low F0 in REST, which may take a shape as depicted in *Figure 3* (subject WS25f):

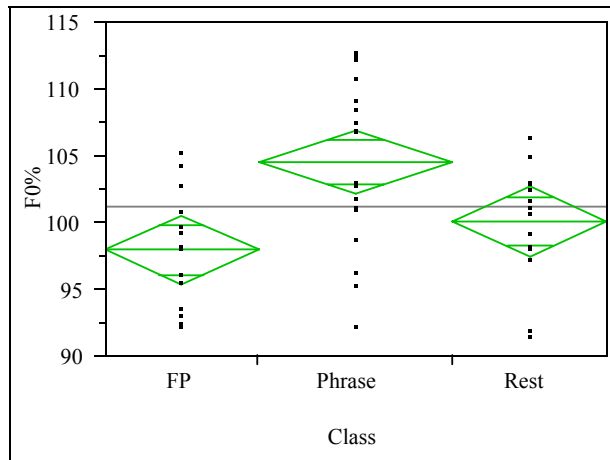


Figure 3: Low – high – low F0%-pattern from FP to PHRASE to REST.

The fact that all observed speakers can be placed in either of these groups is quite extraordinary. Both patterns show a declination in F0 from PHRASE to REST; yet, what causes an FP to be higher or lower in F0 from the following first PHRASE? The following section addresses factors that may affect these F0 patterns, including regional variation, type of FP, sex, and the type of question.

3.2.1 Regional variation

In order to find out whether regional variation may be the cause for these patterns, the relative overall means (%) of FP-PHRASE-REST of the two groups were compared. Whether there are significant differences in F0 between FP-PHRASE-REST in the Bernese group and in the Valais group was tested with a oneway ANOVA.

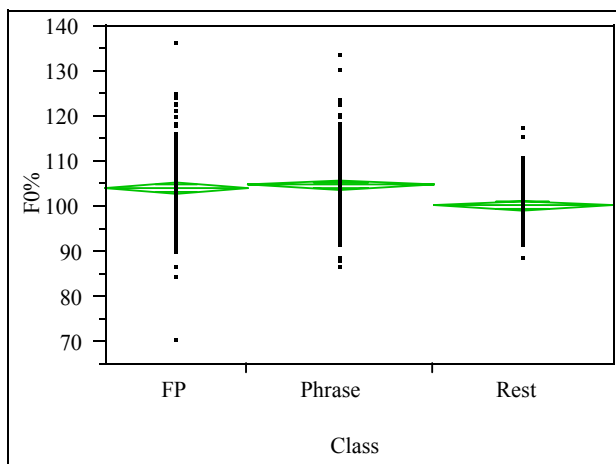


Figure 4: Oneway Analysis of F0% by FP, PHRASE, REST, Bernese speakers.

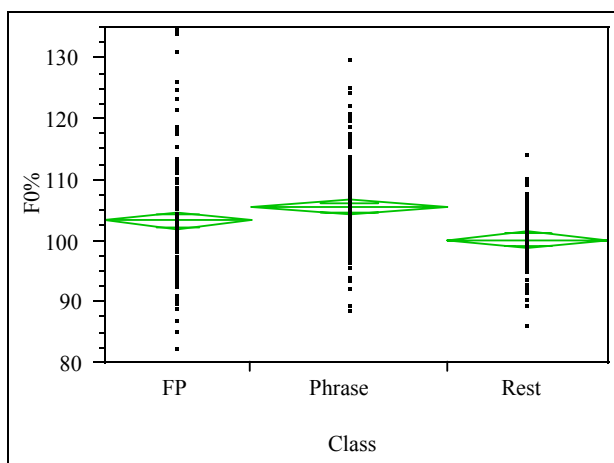


Figure 5: Oneway Analysis of F0% by FP, PHRASE, REST, Valais speakers.

While for the Bernese speakers (Figure 4) the PHRASE and the FP demonstrate significant differences from the REST, the FP is not significantly different from the PHRASE.⁷ The t test confirms that the differences between the PHRASE and the REST, the FP and the REST, as well as between the PHRASE and the FP are significant within the Valais group of speakers (Figure 5).⁸ These results allow for the conclusion that in the sample at hand, utterance-initial phrases are indeed different in terms of their F0 as opposed to the following phrases. More importantly, however, the results show that whether or not a speaker follows the low-high-low or the declination pattern is not contingent upon the speaker variety.

3.2.2 Type of FP

A contingency table (Table 2) should reveal a possible connection between the type of FP and the rise or fall of F0 from FP to PHRASE. As Bernese and Valais subjects show the same tendencies of F0 patterning, both groups were examined together, which allows the running of a Chi-Square-Test with all cells having expected counts above 5. This Chi-Square test results in a significant difference between the cells, cf. Table 2.⁹

Count Col %	<i>also</i>	<i>ja</i>	<i>ja+</i>	<i>mm</i>	“vowel”	
fall	3 12.0%	47 51.7%	11 37.9%	11 61.1%	13 25.0%	85
rise	22 88.0%	44 48.4%	18 62.1%	7 38.9%	39 75.0%	130
	25	91	29	18	52	215

Table 2: Contingency table of rise or fall from FP to PHRASE by FP type (FP types with a frequency lower than 10 were excluded).

Results indicate that for every type of FP, both F0 movements from FP to the following PHRASE (rise and fall) are possible. For *mm*, however, a falling movement is more likely, while for a “vowel”, *ja+*, and *also*, a rising movement is more probable. An assumed interrelation between fall and rise on the one hand and open and literal yes/no questions (cf. 3.2.4) on the other does not show significant differences, except for *ja+*. A literal yes/no question that is answered with a *ja+* can equally be realized with a higher or

⁷ T-Test PHRASE-REST: $p = 2.06e-9$, T-Test FP-REST: $p = 0.0000014$, T-Test PHRASE-FP: $p = 0.419$.

⁸ T-Test PHRASE-REST: $p = 3.22e-8$, T-Test FP-REST: $p = 0.0021$, T-Test PHRASE-FP: $p = 0.0236$.

⁹ Chi-Square: $p = 0.0002$

lower fundamental frequency than the following PHRASE, while the answer to an open question is normally (10/11) realized with an FP lower in F0 than the following PHRASE. At this point, a finer distinction of the FPs' communicative functions could provide a more precise picture.

3.2.3 Sex

To find out whether the speakers' sex has an impact on the F0 of the FP, an ANOVA is run of the F0% factored by the speakers' sex. The results show that the Bernese speakers do not show significant differences between male and female speakers.¹⁰ The Valais speakers, conversely, demonstrate differences in the F0 between the two sexes that are below a threshold level of .05,¹¹ see *Figure 6*:

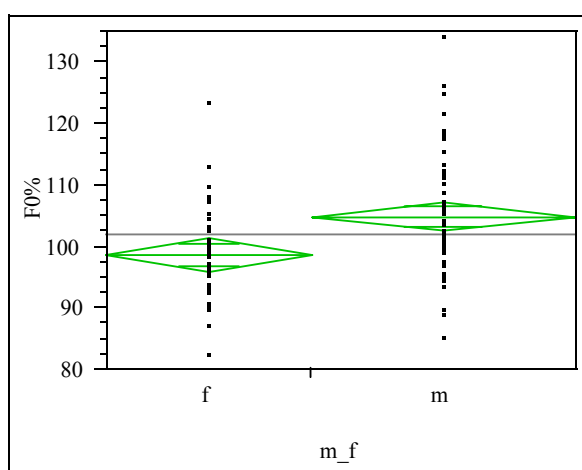


Figure 6: Oneway Analysis of F0% by female/male Valais speakers.

The diagram suggests that the Valais women's fundamental frequency of the FPs is below the fundamental frequency of the REST, while that of the men is somewhat higher. Yet, this difference between Valais men and women does not depend on the fact that only the men keep to both patterns while only the women draw on the low-high-low pattern. The men do not show any F0% differences that are contingent upon the patterns they adhere to. Thus, in the case of these eight subjects, it is the difference between the men and the women that is significant.

3.2.4 Type of question

It was of further interest to explore whether the type of question has an influence on the F0 of the FP and on the first PHRASE. First, tests showed that no positive correlation exists between the type of question (literal yes/no, open) and the F0 of the FP. Second, there is a clear correspondence between the type of question and the F0 of the first PHRASE. For both varieties, the mean F0 value for answers to open questions was significantly (3 %) higher than in answers to yes/no questions.¹² Furthermore, it turned out that Valais speakers showed a distinct pattern of behavior in relation to the succession of FP and first PHRASE in answers to open questions as opposed to yes/no questions (cf. *Table 3*).

¹⁰ $F=3.28$; $p=0.0723$, $DF=1$.

¹¹ $F=12.06$; $p=0.0008$, $DF=1$.

¹² $F=13.77$; $p=0.0002$; $DF=1$.

Count Col %	yes/no	open	
fall	26 44.8%	13 25.0%	39
rise	32 55.2%	39 75.0%	71
	58	52	110

Table 3: Contingency table of rise or fall from FP to PHRASE by question type for Valais speaker.

While the F0 could equally rise or fall from the FP to the first PHRASE for answers to yes/no questions, the answers to open questions showed a significantly higher share of rising movements.¹³ The Bernese subjects demonstrate the same tendency, yet without significant differences.

Additionally, a possible connection between the question types and the F0 pattern (low-high-low or declination pattern) was tested. It turned out that answers to yes/no questions have a similar rise or fall from FP to PHRASE for both groups. Differences can be detected in answers to open questions (cf. Table 4), however, where speakers with a declination pattern do not show a preference for rise or fall, while speakers with a low-high-low pattern prefer a rise. This means that the speakers with a declination pattern answer yes/no and open questions similarly, while the subjects that adhere to the low-high-low pattern prefer a fall in F0 from FP to PHRASE when answering yes/no questions.

Count Col %	Declination	Low-high-low	
fall	25 53.2%	21 24.4%	46
rise	22 46.8%	65 75.6%	87
	47	86	133

Table 4: Contingency table of rise or fall from FP to PHRASE by F0 Pattern for open questions.

3.2.5 Summary and discussion

The above results can be summarized as follows: we encounter two patterns of F0 from FP-PHRASE-REST. On the one hand, there is a gradual declination pattern in F0, while on the other hand there is a low-high-low pattern in F0. We also found that this pattern could not be attributed to either Valais dialect or Bernese dialect. Yet it can be said that utterance-initial phrases are significantly higher than the remaining phrases. Concerning the type of FP, results show that every type of FP could either be higher or lower than the first PHRASE but that a simple “vowel”, *ja+* or *also*, were predominantly followed by a rise to the PHRASE, while *mm* was followed by a downstep to the PHRASE. Speakers of both dialect groups show the same behavior. Answers to open questions are normally realized with a rise from FP to the first PHRASE, while answers to yes/no questions are equally realized with a rise or a fall. This distinction is significant for Valais speakers. While Bernese speaker show the same conduct, it is not statistically significant.

¹³ Fisher’s Exact Test, 2-Tail, p=0.0453.

It seems that the rise or fall from FP to PHRASE depends on the type of question, yet only in correlation with the speakers' dialect background. However, the type of F0 pattern he or she adheres to generally seems to play an equally important role. With respect to the first discovery, one can, of course, only conjecture. It could be that the Valais speakers' pitch range, which is wider than other Swiss-German (including the Bernese) dialects, is one of the causes for the significant differences of the Valais speakers vis-à-vis the Bernese. The second finding may be an issue for conversation analysts in that the drawing of conclusions based on the observation of fundamental frequency changes in utterance-initial positions is exacerbated. The F0 pattern and the interpretation thereof may be due to individual differences or differences related to the speaker's specific dialect background.

4. CONCLUSION

The present study on prosodic aspects of utterance-initial phrases and utterance-initial filled pauses has touched upon an area of research that, in the context of Swiss-German and German in general, has largely been unprecedented. First, it was shown how the speakers use filled and empty pauses differently. The key hypothesis was whether utterance-initial phrases show differences in prosody from the rest of the utterance. This hypothesis can be verified for the present sample. The two speaker groups show significant differences in F0 between the FP, the PHRASE, and the REST. A low-high-low pattern as well as gradual declination pattern in F0 was detected. These patterns cannot be ascribed to regional variation nor to the speakers' sex; yet the number of pauses and their duration do indicate regional differences.

The repercussions of these findings for our current NSF project are considerable. If an intonation and/or timing model of each of the dialects is established, these utterance-initial phrases must either be excluded from the model, so as not to skew its explanatory power, or they must be accounted for in detail and included within the models in a sensible way. From a general prosodic perspective, these findings underline the fact that models of prosody should take into consideration the context of an utterance within the conversation. The models ought to be adapted accordingly. From a CA perspective, these results show that it is sensible to have precise prosodic information not only on the phonological but also on the phonetic level, as minor differences on the suprasegmental level may alter the meaning of an utterance. Further, the interpretation of phonetic information must be viewed with respect to regional differences, as speakers embedded in one dialectal context can make use of phonetic distinctions differently than speakers of another dialect area, with a different communicative behavior, do. In addition, it would be interesting to examine individual choices of phonetic patterns.

It is evident that this piece of research only scratches the surface. A thorough analysis of our data from a conversation analytic point of view would allow for the assignment of different communicative functions to various filled pauses, which would in turn provide further explanatory power to different phonetic patterns. Also, the different functions of the questions and answers within the communicative context could be considered. Despite this need for further research within our project, the paper at hand shows a possible scenario for further work on the interface between phonetics and conversation analysis.

REFERENCES

- ABERCROMBIE, David. (1966). *Studies in phonetics and linguistics*. London: Oxford University Press.
- ATTERER, Michaela and Robert D. Ladd. (2004). On the phonetics and phonology of 'segmental anchoring' of F0: evidence from German. *Journal of Phonetics* 32, 177-197.
- AUER, Peter. (2001). 'Hoch ansetzende' Intonationskonturen in der Hamburger Regionalvarietät. In Schmidt, Jürgen Erich (ed.), *Neue Wege der Intonationsforschung*. Hildesheim: Olms (Germanistische Linguistik 157-158), 126-165.
- BEATTIE, Geoffrey, Anne Cutler and Mark Pearson. (1982). Why is Mrs. Thatcher interrupted so often? *Nature* 300.23, 744-747.
- BIERWISCH, Manfred. (1966). Regeln für die Intonation deutscher Sätze. Untersuchungen über Akzent und Intonation im Deutschen. In *Studia Grammatica VII*. Akademie-Verlag: Berlin, 99-199.
- BOERSMA, Paul and David Weenink. (2006). Praat: doing phonetics by computer. <http://www.fon.hum.uva.nl/praat>, (6.7.2006).
- BREMER, Otto. (1893). *Deutsche Phonetik*. Leipzig.
- BROWN, G. et al. (1980). *Questions of Intonation*. Baltimore: University Park Press.
- CAELEN-HAUMONT, Geneviève. (1991): Stratégies des locuteurs et consignes de lecture d'un texte: Analyse des interactions entre modèles syntaxiques, sémantiques, pragmatique et paramètres prosodique. Unpublished PhD Thesis: Université d'Aix-en-Provence.
- CLARK, Herbert H. (1993). On the course of answering questions. *Journal of Memory and Language* 32, 25-38.
- CLARK, Herbert H. and Jean E. Fox Tree. (2002). Using *uh* and *um* in spontaneous speaking. *Cognition* 84, 73-111.
- FÉRY, C. (1993). *German Intonational Patterns*. Tübingen: Niemeyer.
- FORD, Cecilia and Elizabeth Couper-Kuhlen. (2004). *Sound patterns in interaction*. Amsterdam : John Benjamin.
- FOX TREE, Jean E. (2000). Coordinating spontaneous talk. In Wheelon, Linda (ed.), *Aspects of language production*. Hove, East Sussex: Psychology Press.
- FOX TREE, Jean E. (2002). Interpreting Pauses and *Ums* at Turn Exchanges. *Discourse Processes* 34.1, 37-55.
- GIBBON, Dafydd. (1998). German intonation. In Hirst, Daniel and Albert di Cristo (eds.), *Intonation Systems*. Cambridge: Cambridge University Press, 78-95.
- GILLES, Peter. (2005). *Regionale Prosodie im Deutschen*. Berlin: de Gruyter.
- GOODWIN, Charles. (1981). *Conversational Organization: Interaction between Speakers and Hearers*. New York: Academic Press.
- HÄSLER Katrin, Ingrid Hove and Beat Siebenhaar. (2005). Die Prosodie des Schweizerdeutschen – Erkenntnisse aus der sprachsynthetischen Modellierung von Dialekten. *Linguistik online* 24 (3/2005), 187-224.
- HUTCHBY, Ian and Robin Wooffitt. (1998). *Conversation analysis*. Cambridge: Polity.
- GRICE, M. and S. Baumann. (2002). Deutsche Intonation und GToBI. *Linguistische Berichte* 191, 267-298.
- ISAČENKO, A. and H.-J. Schädlich. (1964): *Untersuchungen über die deutsche Satzintonation*. Berlin: Akademie-Verlag.
- JEFFERSON, Gail. (2004). Glossary of transcript symbols with an introduction. In Lerner, Gene (ed.), *Conversation Analysis*. Amsterdam: John Benjamins Publishing, 13-31.
- KELLER, Eric. (1994). Fundamentals of Phonetic Science. In Keller, Eric (ed.), *Fundamentals of Speech Synthesis and Speech Recognition*. John Wiley: Chichester, 5-21.
- KELLY, John and John Local. (1989). On the use of general phonetic techniques in handling conversational material. In Derek, Roger and Peter Bull (eds.), *Conversation: An Interdisciplinary Perspective*. Multilingual Matters: Clevedon, 197-212.
- KOHLER, Klaus. (1987). Categorical pitch perception. In Viks, Ü. (ed.), *Proceedings ICPhS 11*. Tallinn: Academy of Sciences of the Estonian SSR, 331-333.
- KOHLER, Klaus. (1991). *Studies in German intonation*. Kiel: Institut für Phonetik und digitale Sprachverarbeitung, 25.
- LALLJEE, Mansur and Mark Cook. (1973). Uncertainty in first encounters. *Journal of Personality and Social Psychology* 26.1, 137-141.
- LOCAL, John. (1992). Continuing and restarting. In Auer, Peter and Aldo di Luzio (eds.), *The contextualization of Language*. Amsterdam: John Benjamins, 273-296.

- LOCAL, John. (1996). Conversational phonetics: some aspects of news receipts in everyday talk. In Couper-Kuhlen, E. and M. Selting (eds.), *Prosody in Conversation: Ethnomethodological Studies*. Cambridge: Cambridge University Press, 175–230.
- LOCAL, John. (2003). Phonetics and talk-in-interaction. In *Proceedings of the 15th International Congress of Phonetic Sciences*. Barcelona, Spain, 115–118.
<http://www-users.york.ac.uk/~lang4/icphs.pdf>, (24.7.2006).
- LOCAL, John and John Kelly. (1989). *Doing Phonology*. Manchester: Manchester University Press.
- MACLAY, Howard and Charles E. Oswood. (1959). Hesitation phenomena in spontaneous English speech. *Word* 15, 19–44.
- PETERS, Benno et al. (2005). Melodische Satzakkentmuster in prosodischen Phrasen deutscher Spontansprache. Statistische Verteilung und sprachliche Funktion. In Kohler, Klaus J. et al. (eds.), *Prosodic Structures in German Spontaneous Speech*. Kiel (AIPUK 35a), 7–54.
http://www.ipds.uni-kiel.de/kjk/pub_exx/aipuk35a/aipuk35a_2.pdf, (7.7.2006).
- PETERS, Jörg. (2004). *Intonatorische Variation im Deutschen. Studien zu ausgewählten Regionalsprachen*. Habilitationsschrift, Universität Potsdam.
http://www.let.ru.nl/gep/jp/Habil/Peters_Habil.pdf, (7.7.2006).
- SACKS, Harvey. (1984). Notes on Methodology. In Atkinson, Maxwell and John Heritage (eds.), *Structures of social action: studies in conversation analysis*. Cambridge: Cambridge University Press, 21–27.
- SACKS, H., E.A. Schegloff and G. Jefferson. (1974). A Simplest Systematics for the Organization of Turn Taking for Conversation. *Language* 50, 696–735.
- SCHMID, Karl. (1915). Die Mundart des Amtes Entlebuch im Kanton Luzern (Beiträge zur Schweizerdeutschen Grammatik 7). Frauenfeld: Huber.
- SCHNIDRIG, Kurt. (1986). *Das Dusseln: ein Subsidiärdialekt im Deutschwallis*. Europäische Hochschulschriften (Reihe 1, Deutsche Sprache und Literatur 938). Peter Lang: Bern.
- SELTING, Margret. (1995). *Prosodie im Gespräch*. Tübingen: Max Niemeyer.
- SELTING, Margret (2003). Treppenkonturen im Dresdenerischen. *Zeitschrift für germanistische Linguistik* 31, 1–43.
- SELTING, Margret and Elizabeth Couper-Kuhlen. (1996). Towards an interactional perspective on prosody and a prosodic perspective on interaction. In Couper-Kuhlen, Elizabeth and Margaret Selting (eds.), *Prosody and Conversation: Interactional Studies*. Cambridge: Cambridge University Press, 11–56.
- SELTING, Margret, et al. (1998). Gesprächsanalytisches Transkriptionssystem (GAT). *Linguistische Berichte* 173, 91–122.
- SIEBENHAAR, Beat. (2004). Berner und Zürcher Prosodie. Ansätze zu einem Vergleich. In Glaser, Elvira et al. (eds.), *Alemannisch im Sprachvergleich. Beiträge zur 14. Arbeitstagung für alemannische Dialektologie in Männedorf (Zürich) vom 16. - 18.9.2002* (ZDL-Beiheft 129). Stuttgart: Franz Steiner, 419–437.
- SIEBENHAAR, Beat et al. (2004). Prosody of Bernese and Zurich German. What the development of a dialectal speech synthesis system tells us about it. In Gilles, Peter and Jörg Peters (eds.), *Regional Variation in Intonation* (Linguistische Arbeiten 492). Tübingen: Niemeyer, 219–238.
- SIEGMAN, Aron Wolf and Benjamin Pope. (1965). Effects of question specificity and anxiety-producing messages on verbal fluency in the initial interview. *Journal of Personality and Social Psychology* 2.4, 522–530.
- SHRIBERG, Elizabeth. (1999). Phonetic consequences of speech disfluency. In *Proceedings of the International Congress of Phonetic Sciences* 1, 619–622.
- SHRIBERG, Elizabeth. (2001). To “errrr” is human: ecology and acoustics of speech disfluencies. *Journal of the International Phonetic Association* 31.1, 153–169.
- SIEVERS, Eduard. (1912). *Rhythmisch-melodische Studien*. Heidelberg : Carl Winter.
- STEVENS, Jane and Geoffrey Beattie. (1986). On judging the ends of speaker turns in conversation. *Journal of Language and Social Psychology* 5.2, 119–134.
- SWERTS, Marc et al. (1996). F0 Declination in Read-Aloud And Spontaneous Speech. *Proceedings of the International Conference on Spoken Language Processing* 3, 1501–1504.
- SWERTS, Marc. (1998). Filled pauses as markers of discourse structure. *Journal of Pragmatics* 30, 485–496.
- SYRDAL, A. K. (1996). Acoustic Variability in Spontaneous Conversational Speech of American English Talkers. In *Proceedings of the ICSLP 96*, 438–441.
- TABOADA, Maite. (2006, to appear). Spontaneous and non-spontaneous turn-taking. *Pragmatics* 16.2.
- VETSCH, Jakob. (1910). Die Laute der Appenzeller Mundarten (Beiträge zur Schweizerdeutschen Grammatik 1). Frauenfeld: Huber.
- WIPF, Elisa. (1910). Die Mundart von Visperterminen im Wallis (Beiträge zur Schweizerdeutschen Grammatik 2) Frauenfeld: Huber.