Swiss German Intonation Patterns

Studies in Language Variation

The series aims to include empirical studies of linguistic variation as well as its description, explanation and interpretation in structural, social and cognitive terms. The series will cover any relevant subdiscipline: sociolinguistics, contact linguistics, dialectology, historical linguistics, anthropology/anthropological linguistics. The emphasis will be on linguistic aspects and on the interaction between linguistic and extralinguistic aspects — not on extralinguistic aspects (including language ideology, policy etc.) as such.

For an overview of all books published in this series, please see *http://benjamins.com/catalog/silv*

Volume 10

Swiss German Intonation Patterns
by Adrian Leemann

Swiss German Intonation Patterns

Adrian Leemann
University of Zurich

John Benjamins Publishing Company
Amsterdam / Philadelphia

The paper used in this publication meets the minimum requirements of the American National Standard for Information Sciences – Permanence of Paper for Printed Library Materials, ANSI Z39.48-1984.

Library of Congress Cataloging-in-Publication Data

Leemann, Adrian.
Swiss German intonation patterns / Adrian Leemann.
p. cm. (Studies in Language Variation, ISSN 1872-9592 ; v. 10)
Includes bibliographical references and index.
1. German language--Dialects--Switzerland. 2. German language--Intonation.
3. Switzerland--Languages. I. Title.
PF5132.L44 2012
437'.9494--dc23 2012012903
ISBN 978 90 272 3490 2 (Hb ; alk. paper)
ISBN 978 90 272 7384 0 (Eb)

John Benjamins Publishing Co. · P.O. Box 36224 · 1020 ME Amsterdam · The Netherlands
John Benjamins North America · P.O. Box 27519 · Philadelphia PA 19118-0519 · USA

Table of contents

Abbreviations used

AC	accent command
AM	autosegmental-metrical
ASCII	American Standard Code for Information Interchange
BE	Bern Swiss German
f0	fundamental frequency
Fb	baseline value of fundamental frequency
GR	Grisons Swiss German
IP	intonation phrase
IPO	Instituut voor Perceptie Onderzoek
LFC	Low Frequency Contributions
LPC	Linear Predictive Coding
MLR	Multiple Linear Regression
PC	phrase command
RFC	rise/fall/connection
SDS	Sprachatlas der Deutschen Schweiz (Linguistic Atlas of German-speaking Switzerland)
SNF	Swiss National Science Foundation
T_0	phrase command onset
T_1	accent command onset
T1dist_rise	temporal distance between accent command onset and segment onset
T_2	accent command offset
T2dist_rise	temporal distance between accent command offset and segment offset
T2dist_fall	temporal distance between accent command offset and segment offset
ToBI	Tone and Break Indices
TTS	text-to-speech
VIF	Variance Inflation Factor
VS	Valais Swiss German
ZH	Zurich Swiss German
α:	natural angular frequency of the phrase control mechanism
β:	natural angular frequency of the accent control mechanism
γ:	relative ceiling level of the accent component

SAMPA reference

Throughout this study, *Speech Assessment Methods Phonetic Alphabet* (SAMPA) transcription is used. The applied 7-bit ASCII Swiss German SAMPA was developed at the *Laboratoire d'Analyse Informatique de la Parole* (LAIP), Lausanne (Siebenhaar, Zellner, Keller 2002). It largely resembles the SAMPA originally developed by Wells (1997). The key differences are: Swiss German SAMPA /&/ for Wells' SAMPA /{/, /*/ for /@/, /G/ for /N/, and /'/ for /"/. Lexical *stress* is conceived of as a feature of vowels. _(ptk) stands for the occlusion phase preceding fortis plosives, V(bdg) stands for the occlusion phase preceding lenis plosives, #& represents a filled pause, #h a respiration pause, and # a silent pause.

CHAPTER 1

Introduction

Intonation, or *voice fundamental frequency*, is a key feature of speech found in all languages around the globe (Hirst & Di Christo 1998: 1). Mainly, intonation is a tool for the structuring of information in spoken language, much like punctuation and formatting in a written text. Its scope, however, is much wider: intonation is the ubiquitous, underlying score of speech. After all, it is no coincidence that intonation is commonly referred to as "the melody of speech". It is a channel of information that exceeds the realm of phonology, grammar or semantics: it is a means to express emotion, humor, and attitude, and hint at characteristics of the speaker. This is exactly why Stalder (1819: 7–8), in his early observation on the intonation differences of Swiss German dialects, alluded to "the stiffness and seriousness of the Bernese, – the hasty and quick of the Entlibucher, – the sluggishness in the articulation of the upper Freiämter, – the singing of the shepherds in the high mountains of Uri, Bern, Appenzell, and the Valais" (translation by AL). Speech melody seems to be an indicator of much more than information structuring and literally sets the "tone" in a conversation. Because it seems so deeply rooted in human behavior and is thus such a natural and integral part of speech, its existence in daily life is very rarely perceived and truly acknowledged as such. How many foreign language classes actually take the time to teach you the speech melody of the language you are learning? How often in our daily conversations do we actually take note of the intonational features of language?

The present study aims to do exactly that: to take note of the *fundamental frequency [f0] contours of spontaneous speech*. To be more precise, this study examines the *f0* contours of the Bern, Grisons, Valais, and Zurich Swiss German dialects. As of today, no systematic account of Swiss German dialectal intonation exists, not to mention a systematic study of *f0* behavior that is based on spontaneous speech. Hotzenköcherle (1962: 240) therefore rightly notes in the introductory volume of the *Linguistic Atlas of German-speaking Switzerland* (*Sprachatlas der Deutschen Schweiz*, 1962–2003) that the study of suprasegmental features of Swiss German dialects is considered "an important and alluring task for future monographic research".

The study of intonation is, however, a difficult endeavor because intonation is truly multifunctional. As Vaissière (2004: 256) puts it: "[o]f all dimensions of speech, intonation is clearly the most difficult to study". Intonation can vary according to a number of dimensions, the most central of which are social factors, such as *sex* and *age* (for example, see Bolinger 1989: 9ff.), language-dependent factors, such as *tones* or *pitch accents* (see Cruttenden 1986: 1ff.), linguistic and paralinguistic functions like *phrase* and *lexical accents*, *prominence*, *information structure*, *focus*, *contrast*, or *conversational setting* (Baumann 2006a; Selting 1995), as well as *emotion* and *attitude* (see Murray & Arnott 1993; Kehrein 2002). In short, intonation is conditioned by a large number of variables on the linguistic, the paralinguistic, and the non-linguistic level of speech which is why, according to Bolinger (1989), intonation constitutes the "greasy part of speech". This, of course, poses major difficulties for both a definition as well as an empirical investigation of intonation. This is why, until today, no universally accepted definition of intonation exists. There is neither consensus as to the object of research and the aim of intonational studies, nor agreement as to how intonation should be represented (Vaissière 2004: 256).

This study sets out to provide an adequate description of the *f0* contours of spontaneous speech of the Bern, Grisons, Valais, and Zurich dialect and aims to fill the gap on Swiss German dialectal intonation. The interdisciplinary character of this study, incorporating acoustics, dialectology, as well as sociophonetics, hopefully contributes to a better understanding of the internal mechanisms of *f0* in spontaneous speech and sheds light on the weight of the different variables that interact in order to shape the *f0* contours of speech.

Within the framework of a research project at the University of Bern, speech data of 40 subjects from four different regions of German-speaking Switzerland were retrieved.[1] These regions include the Canton of Bern, Grisons, Valais, and Zurich. A vast majority of studies on intonation resort to the study of laboratory speech. In this study, however, the study of spontaneous material was preferred for two reasons: spontaneous speech data allows for the idiosyncratic dialectal features to permeate uninhibitedly (see Gilles 2005). Secondly, and equally important, it permits an investigation of *f0* variability in the context of linguistic, as well as paralinguistic and non-linguistic functions of intonation, since some of these components of intonation only surface in natural speech and informal settings (unless they are artificially induced or portrayed, of course).

1. SNF-Project 100011-116271/1: "Quantitative Ansätze zu einer Sprachgeographie der schweizerdeutschen Prosodie". Department of Linguistics, University of Bern, 2005–2008.

The theoretical framework chosen to analyze and represent *f0* contours is an analysis-by-synthesis (Bell et al. 1961) procedure using the Command-Response model (Fujisaki & Hirose 1982), developed in Tokyo (Japan) and originally created for the Japanese language. This bridges yet another gap in the research literature on intonation: this is the first large-scale study that applies the Command-Response approach on a corpus of spontaneous, dialectal speech. The variables in this study are of linguistic, paralinguistic, as well as non-linguistic nature. *f0* variation in the variables was tested via an analysis of two global intonation parameters as well as three local parameters. Since this study aims to provide a fertile basis for future intonation research on Bern, Grisons, Valais, and Zurich Swiss German, as many linguistic, paralinguistic, and non-linguistic variables were incorporated as the scope of the study would allow. Hence, in the light of the numerous findings as well as the mostly inexistent material for comparison, the interpretations given in this study are by all means hypothetical and require further support. After all, the main goal of this study is to provide a first, thorough description of the *f0* behavior of four Swiss German dialects, not to formulate fully-fledged, linguistic interpretations. Rather, they are meant to spur new, exciting research geared solely towards the understanding of specific variables.

This study is structured into a total of 14 chapters. The second chapter will provide an introduction of the key concepts of intonation. After that, we will turn to a discussion of different ways of representing intonation contours in the section on intonation models. Chapter 4 will then present the model chosen for this study: the Command-Response model. Subsequently, we will review the rather scarce account of existing literature on Swiss German intonation before we turn to the statistical analysis. The statistics section is headed by Chapter 6, which specifies the methods used in this study. Chapter 7 illustrates the application of the Command-Response model and the modeling constraints applied. Note that the modeling approach applied in this study has a strong explorative character, since the analysis method is not formally defined in the papers presented by Fujisaki and co-workers, particularly not with regard to spontaneous speech. Chapters 8 to 11 give the results of the statistical analyses, starting with overall results and then progressing according to type of variable: linguistic variables, paralinguistic variables, and non-linguistic variables. The global and local *f0* behavior in each of the variables is analyzed in bivariate tests using parametric and non-parametric statistical tests against the background of detecting dialect-specific patterns as well as cross-dialectal differences. Each of these chapters features an immediate discussion of the findings. In Chapter 12, I will present dialect-specific, multiple linear regression models and logistic regressions for the investigated model parameters. The regressions allow for a distillation of the relative contribution of independent variables towards explaining *f0* variability in a given parameter value. At the heart of this study are the *dialect profiles*, presented in Chapter 13, in which the major findings are revisited, discussed, and placed into a broader, dialectological and sociolinguistic context. Finally, we proceed to concluding remarks and a brief outlook.

CHAPTER 2

Intonation

The *fundamental frequency* [hereafter *f0*] of spoken language, is commonly referred to as the *intonation* of language. The term *intonation*, which is derived from Greek *tonos* (tension), denotes the rise and fall of voice pitch over entire phrases and sentences. It is generally considered to belong to the category of *suprasegmentals* which, as the term implies, include those properties of an utterance that lay beyond a single segment. Rather, they affect a string of segments, ranging from one syllable to entire sentences. The term suprasegmentals is often used synonymously with the term *prosody*, derived from Greek *prosodia*, which Allen (1973: 3) describes as "a 'tune' to which speech is intoned, […] the melodic accent which characterized each full word in Greek". Apart from intonation, prosody further includes *loudness*, *quantity*, *speech rate*, *rhythm*, *voice quality*, *phrasing*, and *pausing* (see Möbius 1993a: 9). These prosodic parameters are inter-dependent: duration and intensity, for example, can affect the perception of pitch (see Lieberman 1980; Niebuhr 2007).

Every language exhibits intonational features. In fact, it was shown that prosodic features are among the first linguistic features children acquire, and – as observed in aphasic patients – they also seem to be the last feature lost (Hirst & Di Cristo 1998: 2). From an evolutionary perspective, it is theorized that the underlying principle of pitch is to convey a physical impression of the sound source, i.e. the size of the vocalizer (see Vaissière 2004: 252). Because of the sexual dimorphism in the vocal anatomy of humans, the *f0* of males is generally lower than the *f0* of females. If threatened, males tend to make use of these low-pitched sounds in order to protect their family and to intimidate their enemies. Cross-culturally and cross-linguistically, low pitch therefore is associated with dominance, physical largeness, and potential threat, while high pitch conveys submissiveness, small physique, and harmlessness. This phenomenon is known as the frequency code (see O'Hala 1983).

The extraordinary characteristic about intonation is that it carries *meaning*. A falling intonation, for example, more often than not denotes finality, whereas a rising intonation signals incompleteness (Fox 2000: 269ff.). This is not the case, however, for other suprasegmental features. For example, *stress* and *tone* or, on the segmental level, *phonemes* and *syllables*, are not inherently meaningful. Furthermore, intonation is *gradient* and not discrete, meaning that the shape of a fall or rise can be described accurately and numerically. This gradience also translates into gradience in meaning (Fox 2000).

Generally, one can distinguish between three types of languages with respect to how intonation modulations are exploited: *tone languages*, *pitch accent languages*, and *intonation languages*. In tone languages, tone is a characteristic of the lexicon: the change of a tone in an otherwise unaltered segmental environment can trigger a change in meaning. For tone languages, *f0* is thus primarily used to signal lexical contrasts (Vaissière 2004: 242). In pitch accent languages, on the other hand, tonal patterns of the word represent the most basic constituent for the shape of *f0* movements (see Vaissière 2004: 242). In intonation languages, the speech melody is a feature of phrases and sentences, which, in languages without lexicalized tones, is manifested in terms of a rather intricate intonational system. Intonation and tone are not mutually exclusive, yet tone languages make less rigorous use of intonation patterns than intonation languages (Cruttenden 1986: 9ff.). The majority of intonation languages are characterized by a default, low, falling pitch and a contrasting, raised pitch. Most frequently, the latter is used in questions, whereas the former contour indicates statements (Hirst & Di Cristo 1998: 1). There are, however, many varieties that show reverse tendencies, such as the rising intonation of Belfast English declaratives or the falling interrogatives of Bengali (Gussenhoven 2004: 54). So evidently, there are language-inherent differences in intonation in the world's languages, which means that intonational features may not always translate into the same linguistic functions.

Intonation research is a highly inter-disciplinary field of research, involving disciplines such as linguistics (phonetics in particular), speech pathology, foreign language education, and speech technology. Over the past decades, the study of *intonation* has continually gained impetus and the number of conferences and workshops in this area of research is growing steadily. In foreign language education, for example, intonation and prosody research finds direct application in classrooms: students can be given acoustic as well as visual feedback on their prosodic performance (Siepmann 2001: 13ff.). Speech technology represents another multifaceted branch of intonation research, which includes speech analysis, synthesis, and speech recognition (Mixdorff 1998: 9). Speech synthesis research, for example, is geared towards improving the naturalness of synthetic speech. Due to this interdisciplinary character of intonation research, the terminology used in this field has become exceedingly hybrid. Möbius (1993: 7) aptly points out that intonation research is marked by a Babylonian confusion of tongues. Hence, in order to minimize possible confusion, the following sections are geared at introducing and defining the terminology and concepts used throughout this study.

The organization of the subsequent sections mirrors the structural hierarchy of intonation. Since this study is based on a corpus of spontaneous speech, special attention is paid to the intonation of spontaneous speech. After a definition of the broad concept of intonation and its underlying physiology, the focus will shift to the topmost overall structure, the intonation phrase. The subsequent concepts include the declination of the intonation phrase, pitch reset, and – on the level of global intonation – the phenomenon of pitch range. On the local intonational level, the often-debated terms prominence, accent, and stress will be introduced and discussed. Accent is then further subcategorized into word accent and sentence accent. Having arrived at the lowest intonational level, we will take a closer look at microprosodic intonational variation. After this introduction to the key concepts of intonation, we will touch upon the relevant informational, paralinguistic, and non-linguistic functions of intonation.

Before delving into the matter of defining intonation, it is crucial to draw attention to a couple of important issues. Because the present study is mainly concerned with the *phonetic* modeling of Swiss German dialects, the following section is weighed accordingly. In other words, the terminology introduced corresponds to the terminology used in intonational phonetics as opposed to intonational phonology. Furthermore, the next section neither sets out to cover the vast array of definitions presented in previous literature on intonation nor to contribute to the unification of terminologies.[1] Finally, it should be borne in mind that the definitions used and discussed in the present study largely stem from the German and English research tradition and are thus mainly based on the German and English language.

2.1 Defining intonation

Baumann (2006a: 4) points out that, in most literature, intonation is either understood in a broad or a narrow sense.[2] In a narrow sense, intonation is conceptualized as the continuous contour of speech melody, with *f0* as the acoustic correlate, which designates the quasi-periodic number of cycles per second of the

1. For a detailed discussion of intonation-related terminology, see Inozouka 2003.

2. Beckman (1986) provides a definition of intonation that is situated halfway between these two extreme points of view. She advocates a pluriparametric view of intonation by underlining the critical role of intensity and durational features in intonational representation.

speech signal, measured in Hertz (Botinis et al. 2001: 264; Gilles 2005: 3). In contrast, intonation in a broad sense is equated with prosody, including pitch range, phrasing, stress, accentuation, rhythm, and tempo. The view adopted in the present study corresponds to the narrow view, with *f0* being the primary acoustic parameter of intonation. The author believes that a systematic link between intonation and the grammatical system can only be detected by means of such a restricted definition (Günther 1999: 62). It is this notion of intonation, realized with all its manifold linguistic functions on both the word and sentence level, which constitutes the core of this study.

The physiological mechanisms indispensable to the production of speech sounds are the supralaryngeal vocal tract, the larynx, and the subglottal system. Intonation is generated for the most part by the laryngeal structure (i.e. length of the vocal cords and muscular tension) and by subglottal pressure. The larynx, a feature found in all terrestrial animals, was originally devised to protect the lungs of the primal lungfish from water intrusion but, as evolution progressed, has adapted for phonation (Lieberman & Blumstein 1988: 11). In the words of Lieberman and Blumstein (1988), speech sound is generated as follows:

> The primary role of the larynx in the production of speech is to convert a relatively steady flow of air out from the lungs into a series of almost periodic, i.e. "quasi-period," puffs of air. The larynx does this by rapidly closing and opening the airway by moving the vocal cords together or pulling them apart.
>
> (Lieberman & Blumstein 1988: 4)

Vocal cord frequency thus refers to the number of complete cycles in a given period of time, usually within a second. The required time to complete one full cycle is called a period and is indicated as 1 Hertz (Hz). The faster the vocal fold vibration, the higher the periods per second, i.e. the higher the produced sound (Gussenhoven 2004: 2). The male *f0* usually varies between 60 Hz and 240 Hz, while the female *f0* generally falls somewhere between 180 Hz and 400 Hz (Cruttenden 1986: 4). If the frequency is lower than 40 Hz, the vocal fold vibration is perceived as a series of separate events. If it exceeds 40 Hz, however, it is considered continuous.

Fundamental frequency is only found in voiced sounds. Interestingly enough, the human ear perceives continuous speech as a nonstop *f0* pattern, albeit about one quarter of the sounds in English, for example, is voiceless (Cruttenden 1986: 4). In the same vein, human perception allows us to perceive a male *f0* on the telephone, despite the fact that male voices are normally below 300 Hz and telephones usually only transmit frequencies between 300 Hz and 3400 Hz. The correct *f0* is calculated via periodicity analysis (see Kohler 1977).

The perceptive equivalent of *f0* is referred to as *pitch* (see Inozouka 2003: 53; Baumann 2006a: 4). Lehiste (1970) points out that the lowest pitch still perceivable to humans is at approximately 16 Hz, the highest at 20,000 Hz. In the present study, *f0* is understood as the actual vibration of the vocal folds, whereas pitch is used in the sense of the *f0* as perceived by the human ear.

f0 variation results from the modification of the length and tension of the vocal cords brought about by cartilage movement and intrinsic muscles activity (Möbius 1993: 72). Figure 2.1 shows the cartilage and the intrinsic muscles of the human larynx (adopted from Lieberman & Blumstein 1988: 98).

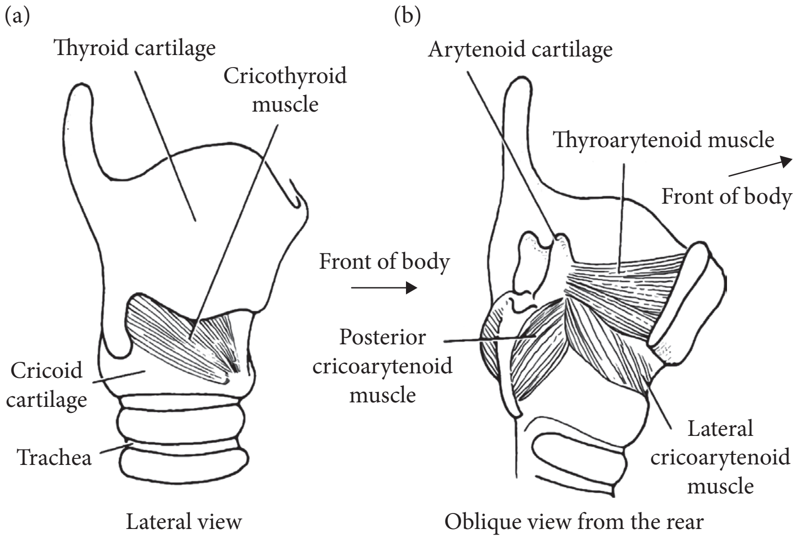

Figure 2.1. The human larynx (adopted from Lieberman & Blumstein 1988: 98)

The major constituent of the larynx is the ring-shaped cricoid cartilage, attached to the top part of the trachea. The thyroid cartilage, identifiable from the outside as the male Adam's apple sits on top of the cricoid cartilage. The male thyroid cartilage grows larger during puberty and causes an elongation of the vocal folds. This does not only result in the development of the Adam's apple but also generates an overall lower *f0*. The vocal cords – a combination of ligaments, muscles, and tissue – are located inside the thyroid cartilage, attached to the two arytenoid cartilages, the thyroid and cricoid cartilage (Lieberman & Blumstein 1988: 100). Via electromyography (EMG), a technology for measuring electrical muscular activity, Collier (1975) discovered that the cricothyroid muscle is mostly responsible for the direction, the range, and the speed of *f0* changes. Muscular contraction brings about a higher *f0*, while relaxation causes a lowering of *f0*.

Further, he shows that a decrease of subglottal air pressure results in a gradually falling baseline. Thus, whenever the cricothyroid muscle is inactive, *f0* is governed by subglottal air pressure (Collier 1975: 254). Atkinson (1978) further provides evidence that laryngeal tension has a strong impact on higher leveled *f0* movements, whereas pulmonary pressure dominates *f0* at lower values. More specifically, the cricothyroid muscle, supported by the lateral cricoarytenoid muscles (see Figure 2.1), controls the highest *f0* values, whereas mid-range *f0* values are governed by the sternohyoid muscle (an extrinsic muscle to the larynx). What both Collier (1975) and Atkinson (1978) were able to demonstrate is that intonation production is the product of a joint effort of several different components of an intricate, underlying physiological mechanism (Möbius 1993: 74).

2.2 Intonation phrase

The most significant production unit of intonation is the *intonation phrase* [hereafter referred to as IP]. This term was introduced originally by Pierrehumbert in 1980 and is often used synonymously with the term *phrase*. Each IP contains at least one accent that serves as the tonal anchor point for the intonation contour (Gilles 2005: 6). An IP may consist of one syllable only or may span syntactic phrases, clauses, or entire sentences. A larger utterance may therefore contain more than one IP (Botinis et al. 2001). In general, IPs in spontaneous conversation are shorter than IPs in prepared speech, yet the duration of the IPs depends on the idiosyncratic style of the speaker. Some speakers realize comparatively shorter IPs, others produce longer IPs (Féry 1988: 43). Shorter IPs are usually accompanied by shorter pre- and post-boundary pauses, while longer IPs are typically associated with longer pre- and post-boundary pauses (Krivokapic 2007). The boundaries of IPs usually feature clearly perceivable, either high or low boundary tones in phrase-final position (see Pierrehumbert 1980). Phrase-initial tones are possible as well but are not necessarily required (Féry 1988: 45). Unfortunately, however, IP boundary marking is not always quite as straightforward since it involves a number of factors apart from intonation as well.

IPs are delimited by a bundle of different prosodic features which can either occur together or independently.[3] This cluster of prosodic features entails *f0 movements* at the beginning and end of an IP, *pausing*, *phrase-final syllable*

3. The recognition of prosodic boundaries is one of the major concerns in automatic speech processing research (see Ostendorf 2000).

lengthening, *decrease of articulation rate*, and *pitch resets* (see Cruttenden 1986). Further, Peters et al. (2005) note that laryngealization, glottalization, and change of intensity also contribute to IP boundary marking. What is still unclear is the weight of these individual features, however. For example, Vaissière (1983: 57) points out that "[t]he physiological basis of the relations among pause, breathing, declination and resetting is difficult to establish, since speakers may pause without breathing, or reset the baseline without pausing." As a result, the prosodic parameters responsible for IP demarcation are discussed and weighed differently in the literature.

Further cues for IP boundary marking can also be derived from syntactic structure. This is especially true for prepared speech and, to a lesser extent, also for spontaneous speech. Taylor (1994: 16) notes that it is unclear whether syntax determines prosody via complex mapping or whether syntax constitutes one pillar of IP placement. It seems quite likely, however, that prosodic structure is affected by metrical factors. Accordingly, prosodic boundaries can be placed in the middle of syntactic constituents (Taylor 1994). The correlation of IP boundaries and syntactic boundaries therefore is not causal but casual (Botinis et al. 2001: 269). Because syntax and IP often do not overlie, the IP is considered by some scholars to be associated with sense groups, i.e. words grouped by meaning (see for example O'Connor and Arnott 1961: 3).

Lieberman (1980: 192) emphasizes yet another factor of IP demarcation, namely the importance of breathing. Breath groups are used by the speakers to divide a train of words into sentences, where speakers usually draw breaths at the end of conceptual units, such as sentences and clauses (ibid.; Vaissière 1983: 54; for Swiss German see Hove 2004). Typically, breathing and pauses trigger a reset in pitch (Vaissière 1983: 57). Hence, Cruttenden (1986) and Peters et al. (2005) consider pauses (including respiratory pauses) as a reliable cue for IP boundary marking. Additionally, pauses frequently seem to accompany new conversation topics and often take place after the first word of an IP. Cruttenden (1986: 38) attributes such pauses, which provide the speaker with additional time to plan the rest of the sentence, to typical "performance errors". He adds, however, that pausing is highly idiosyncratic, and that pauses cannot always be taken as cues for IP boundaries. Duez (1982) demonstrates that pauses in political speech are more aligned with the underlying grammatical structures, whereas in spontaneous speech, grammatical structure and pause placement must not necessarily coincide. Since repetitions, incomplete sentences, and false starts are a feature of unmonitored speech, IP boundaries in spontaneous speech are often vague (Cruttenden 1986: 36). A case in point is the incongruence between syntactic structure and IP placement. In German, for example, IP boundaries can but do not have to coincide with larger syntactic boundaries (see Bierwisch 1966).

2.3 Declination and pitch reset

Declination refers to the gradual lowering in *f0* during the continuation of an utterance. In the words of Lieberman (1980: 195) "[t]he falling fundamental frequency contour, which is structured by the vegetative aspects of respiration, is the universal language signal signifying the end of an 'ordinary' sentence". The easiest way for a person to make a sound during expiration will involuntarily result in this kind of *f0* declination (ibid.). Hence, physiologically, declination is grounded in the decreasing transglottal air pressure in the course of an utterance. Some scholars therefore argue that declination is not controlled by the speaker and should be interpreted as a result of physiological speech production mechanisms (see Lieberman 1967; Vaissière 1983). Conversely, Pierrehumbert (1980) assumes that the speaker has intentional control over declination, and that declination should be considered a means of vocal expression. Not denying the principle of phonetic declination, she argues that downdrift also bears a phonological function, which she calls *downstep*. This so-called downstep constitutes a grammaticalized version of declination.

As is illustrated in Figure 2.2, the declination tendency is typically framed by a topline, the realization of *f0* peaks, and a baseline, which serves as a reference point for the valleys of an intonation contour (see 't Hart et al. 1990).

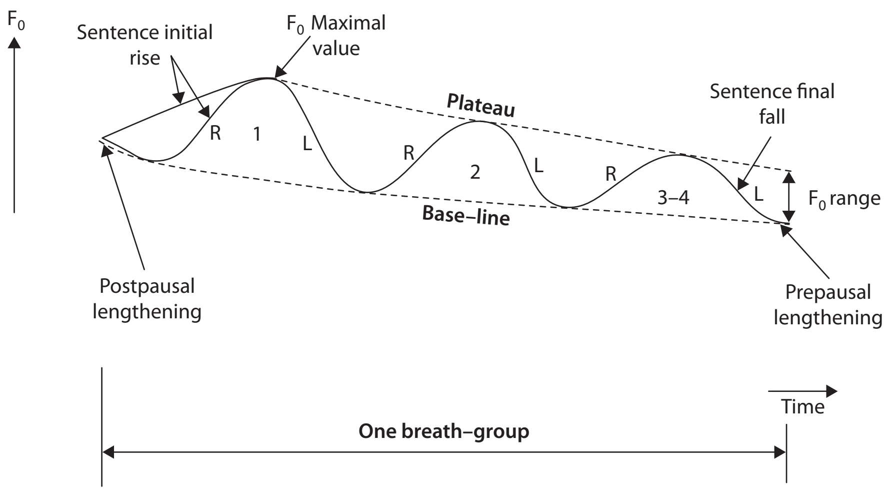

Figure 2.2. Properties of an intonational phrase as observed in several languages (adopted from Vaissière 1983: 55). The dotted lines represent top- and baseline of the IP

Declination applies to both the top- as well as the baseline of the intonation contour. Baseline declination is apparent in that the fundamental frequency of unaccented syllables is higher for IP initial position than for phrase-final position. The declination of the topline, on the other hand, denotes the decline of absolute *f0* for pitch prominent syllables in the course of an IP. On a higher level, declination also operates across intonation-groups or whole paragraphs (Cruttenden 1986: 126ff.). The degree of declination normally depends on the length of the utterance, with declinations being steeper for shorter utterances and less steep for longer utterances (see Vaissière 1983; see Adriaens 1991: 60). It is crucial to note, however, that declination does not occur in all intonational phrases. Declination effects are high for laboratory speech this is not the case for spontaneous speech in which the declination tendency seems to be less obvious or sometimes not present at all (Vaissière 1983: 57).

Declination is closely linked with a *reset in pitch*, i.e. the readjustment of the declination line. Since declination causes the *f0* to reach a fairly low point, the *f0* needs to be reset. Generally, this happens after a pause but the resetting of the baseline can also occur in the absence of pauses (see Fujisaki et al. 1979). Therefore, the relationship between pitch reset and pausing is still unclear (Vaissière 1983: 57). According to Pierrehumbert (1979), speakers of English frequently reset their pitch in such a way that it coincides with a syntactic boundary. In this sense, resetting is used as a boundary marker and the degree of the reset is indicative of the strength and importance of the syntactic boundary (Vaissière 1983: 57). Pitch resets thus suggest special prominence, which is why they carry a direct communicative function (Fox 2000: 309).

As has already been mentioned, pitch resets in spontaneous speech need not coincide with larger syntactic boundaries. This loose link between syntax and pitch reset is exemplified in the following Figure 2.3.[4]

4. The figure illustrates a contour that was generated with the Command-Response model. This type of figure occurs commonly in the present study. The top panel illustrates the speech waveform, the +++ show the extracted *f0*. The bold line displays the model *f0*, while the dotted line stands for the phrase component. At the bottom of the figure, the accent commands are depicted in rectangular shapes. Syllable boundaries are marked with the vertical dotted lines. In all of these figures, the contour is displayed in the log F domain (see Mixdorff 1998: 8). For more information about the Fujisaki model see Chapter 4.

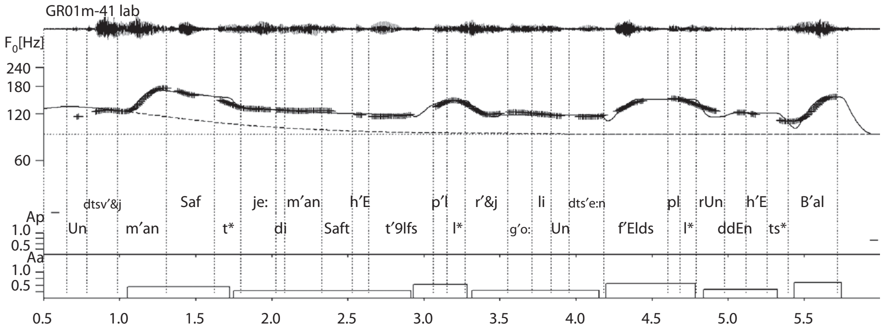

Figure 2.3. Example of a Grison speaker's IP – containing several syntactic phrases

In German, the speech sequence in the above Figure denotes the following (syntactic phrases are marked off with square brackets): [*und zwei Mannschaften*] [*jede Mannschaft hat elf Spieler*] [*einen Goalie und zehn Feldspieler*] [*und dann hat's einen Ball*] (Engl.: [*and two teams*] [*each team has eleven players*] [*one goalkeeper and ten field players*] [*and then there's a ball*]). One would expect that at the introduction of the new topic [and then there is a ball], the speaker would start a new IP since the larger syntactic unit, the main clause [and two teams] followed by two subordinate clauses [each team has eleven players] [one goalkeeper and ten field players], is completed. Instead, the *f0* continues to decline even into the following phrase and the new topic [and then there is a ball] and thus stretches over the entire utterance. A further problem arises regarding the distinction between pitch resets and word accents. For the syllable f'EldS, for instance, one could assume that there is a minor pitch reset. On the other hand, the rise in *f0* could also be interpreted as the pitch movement of the word accent for this particular syllable.[5]

2.4 Stress and accent

Werner (2000: 3) does not exaggerate by saying that the definitions for *accent* and *stress* (or the German counterparts *Akzent* and *Betonung*) are hotly fought over in intonation research. For the most part, theoretical debates on accent refer to the acoustic correlates of accent, yet they also refer to its phonological significance (2000: 114). In the words of Ladd (1996: 286), "the terminology in the general area

5. As illustrated by the modeling of this phrase in Figure 2.3, the pitch movement of the syllable *f'EldS* was interpreted as the pitch movement of the word accent.

of 'accent' is really a mess". In order to shed some light on these terminological issues, the following section will introduce some of the expressions frequently used in the intonation literature or related to accent in a broader sense. These terms include prominence, stress and accent, accent group, and phrase accent. Naturally, it is difficult to tackle these issues as separate concepts since the terms and definitions overlap. Prominence and stress, for example, to some degree overlie conceptually: a stressed syllable is a prominent syllable, and a syllable can be made prominent by marking it with stress. Nevertheless, it makes sense to try to uphold narrow definitions, as this is the only way in which subsequent findings can be discussed in a precise manner.

2.4.1 Prominence

Prominence is a means of marking syllables and placing emphasis. A prominent syllable stands out of a given context, implying increased importance (Cruttenden 1986: 7). Mayer (1997: 15) distinguishes between two types of prominence: *lexical prominence* and *phrasal prominence*. If prominence occurs on the word level, it is normally referred to as word accent or lexical accent. In reference to the discussion of declination, it should be noted that the later a prominent syllable occurs in an IP, the lower its peak *f0* value (Taylor 1994: 17). This of course is due to the declination tendency of the IP.

The acoustic correlates of prominence are complex and language-dependent. More importantly, it is imperative to distinguish between *prominence production* and *prominence perception*. For prominence production, the most critical indicator for varieties of English is duration, followed by intensity and, least importantly, *f0*. For prominence perception, on the other hand, *f0* was shown to occupy a much more critical role (see Kochanski et al. 2005; Hirst 1983).[6] Since the aim of this study is to model the produced intonation contours of speakers of Swiss German dialects, the following discussion will mostly focus on prominence production.

Prominent syllables are commonly assumed to be louder, longer, and higher in *f0*.[7] Since both *f0* as well as intensity are produced by increasing and decreasing pulmonary effort, subglottal pressure, and vocal fold tension, those two parameters

6. However, Bannert (1983) discovered that *f0* may also be perceived as higher in constituents with higher intensity – thus, underlining the role of intensity in intonation perception.

7. It should be borne in mind, however, that prominence can also be achieved by a reversal of these parameters. Isačenko and Schädlich (1970: 24ff.), for example, have shown that a lowering of the *f0* on specific segments – if the surrounding segments show high pitch – has the same effect as increasing the pitch in a low *f0* segmental environment.

prone to correlate positively in most instances (Möbius 1993a:15). However, the role of intensity for prominence marking is not yet clear and its linguistic function in continuous speech is difficult to determine. Intensity is also the least studied of the three parameters (Vaissière 1983: 63). Not all languages mark prominence concurrently with all three above-mentioned parameters. Rather, a number of languages have ascertained their own suprasegmental code (ibid.). Dutch and English have been reported to feature a concurrence of all three parameters in lexically stressed syllables (Vaissière 1983: 62–63). In these languages, the prosodic parameters for marking prominence are timed according to the lexically strong syllables. Hence, they are referred to as stress-timed languages. The temporal interval between stressed syllables is uniform, i.e. these languages show a tendency towards isochrony based on stress (Cruttenden 1986: 25). An example language with reduced correlation of the three prosodic parameters for prominence marking is French. In the case of French, a syllable-timed language, the parameters for prominence marking are bound to the first and last syllable of the word, or, the rhythmic group: on the first syllable, we typically find a minor rise in *f0*, while the word-final syllable may exhibit a variety of prominence contrasts (Vaissière 1983: 63).[8] These examples show that the relative weight of each of the three prosodic factors for prominence marking is language-specific and that it is the interrelations of these features which, according to Vaissière (1983), are the most salient characteristics of a language variety:

> It is possible that the specific interrelations between the three suprasegmental features (*f0*, 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)

This finding provides fertile grounds for future research on Swiss German dialects. Earlier literature on the Valais dialect, for example, suggests that intensity and pitch do not coincide frequently in prominence marking (see Wipf 1910). Instead, other compensatory means for marking prominence seem to be applied. The representation of *f0* is the only prosodic parameter that will be under scrutiny in this study. In future research, however, the study of intensity and its (possible) correlation with *f0* might prove to be of interest.

8. The question of whether or not isochrony in fact holds in reality is another issue, since it is claimed that for syllable-timed languages, for example, acoustic correlates of stress-to-stress intervals cannot be verified decisively. This issue is, however, not pursued further in the present paper (see Auer & Uhmann 1988; Ramus 1999; Low et al. 2000).

2.4.2 Stress

Stress is one of the most elusive prosodic features (Lehiste 1970: 106). Adding to the terminological confusion, stress and accent are often used interchangeably in the literature. Stress, often used synonymously with lexical accent or lexical stress, is determined by the lexicon of a language and is marked by prosodic prominence. In a given linguistic context, the syllables that carry stress are perceived as more salient. Commonly, one distinguishes between word stress as an abstract feature, as proposed by Trager and Smith (1951), Chomsky and Halle (1968), and re-interpreted by Liberman and Prince (1977), and the actual phonetic realization of stress (see Günther 1999: 48ff.).

Trager and Smith (1951) as well as Chomsky and Halle (1968) suggest that stress is primarily a property of single vowels and that prominence placement is governed by stress rules. Based on this concept, Liberman and Prince (1977) elaborate this approach and show that stress is not an absolute feature of vowels but is, in fact, relative. Stress is assigned according to strong and weak syllables, which are present on both the lexical and the phrasal level. This approach falls under the general taxonomy of metrical phonology. In this theoretical framework, prominence is understood as an abstract feature, deriving from the metrical strength of syllables in any given utterance (Niebuhr 2007: 6). Spoken in isolation, each word carries a lexical accent which can be predicted by means of metrical rules.

Generally, disyllabic words (e.g. *carrot*) consist of a relatively stronger syllable (*s*) and a weaker syllable (*w*). In words with more than two syllables (e.g. *Pamela*), the relative strength of the syllables is assigned by metrical theory. In words with primary and secondary stress (e.g. *sensibility*), the main stress of the word is placed on the syllable that is governed by *s* nodes only (Hayes 1985: 1ff.).[9]

The correspondence between abstract, prosodic characteristics and acoustic features is a very intricate matter (see Hirst & Di Cristo 1998: 5ff.). With respect to stress, it is the role of intensity, in particular, which is a source of ambiguity. In the case of German, the acoustic correlates of stress normally entail an increased *f0*, as well as higher intensity and longer duration (see Mixdorff 1998). Isačenko and Schädlich (1970), too, advocate that intensity changes are only a secondary means for placing stress, the most central factor being changes in *f0*. Usually, stress falls on the first syllable of the stem of a native word, which in German generally constitutes the penultimate syllable of the word. In the English language, however, Trager and Smith (1951) point out the significance of intensity as a means

9. Psycholinguistic studies have shown that such prominence patterns are represented in the mental lexicon of the speaker (see Levelt 1989).

of marking stress. This claim is also supported by Beckman (1986) and Silipo and Greenberg (1999) for American English and, recently, by Kochanski et al. (2005) for British English. In the latter study, intensity and duration appear to carry more information on stress than *f0*. Ivic and Lehiste (1963), on the other hand, conclude that the reliable cue for stress marking in Serbo-Croatian is not intensity but, in fact, duration (ibid. 1963, quoted in Lehiste 1970: 134). In other words, acoustic correlates of stress seem to be highly language-dependent.

2.4.3 Accent

Cruttenden (1986: 16) equates *stress* with any kind of prominence, whereas the term *accent* is only used if prominence is achieved through *f0* movements. Word-stress syllables, so Cruttenden (1986: 17), are marked as *stressed* in the dictionary and are potential carriers of *accents*, i.e. *f0* movements – given certain exceptions (see Féry 1988; Günther 1999). In theory, any syllable can be accented (e.g. in cases of contrastive accents where lexically weak syllables can be accented) but, normally, it is the metrically strong syllables which carry accent. Mixdorff (1998) argues along similar lines and reports that, in German, accented syllables feature a distinct *f0* movement, higher intensity, and longer duration. If *f0* movements are not present, he refers to that syllable as stressed but "de-accented" (ibid. 1998: 19).

Given the above discussions, there seems to be little evidence to suggest that stress indeed constitutes a prerequisite of accent. For this reason, accent, in the present study, is defined independently of stress. *Accent*, in the current study, refers to a realized *f0* movement, regardless of the presence of lexically strong (i.e. stressed) or weak (i.e. unstressed) elements. This *f0* movement can concern only one, or several segments, which by being "accented" are understood as being made *f0* prominent. *Stress*, on the other hand, refers to the abstract word stress, which is understood as being governed by stress rules (see Trager & Smith 1951; Chomsky & Halle 1968). *Stress* can be signaled by any, or a combination, of the three prosodic parameters intensity, duration, and *f0*. More often than not, *stress* and *accent* do, but need not, correlate, since both *f0* as well as intensity are produced by an increase and decrease of pulmonary effort, subglottal pressure, and vocal fold tension. Moreover, I agree with Vaissière (1983), who proposes that each of the three prosodic parameters for marking prominence (intensity, duration, and *f0*) should be regarded independently as their relative weight seems to be language-specific.

2.5 Pitch range

Pitch range, often used synonymously with pitch span, refers to the variation in pitch height between *f0* maxima and *f0* minima in speech and has to be defined separately from pitch level (Ladd 1996). Pitch level denotes the baseline, i.e. local *f0* minima, which can be raised or lowered from utterance to utterance. Pitch range carries paralinguistic functions: a higher pitch range is often associated with emotions such as joy, anger, fear, or surprise (see Murray & Arnott 1993). Furthermore, it also exhibits a communicative function, in that it can mark the beginnings and endings of new topics. A high pitch range, for example, introduces a new topic, whilst a low pitch span signals the end of a topic (Cruttenden 1986: 129). Pitch range variation may apply only to parts of an IP (Gussenhoven 2004: 76). Thus, the boundaries between pitch range and local accents often are not clear-cut, which makes it difficult to judge whether a high accent, for example, is due to local *f0* prominence or a momentary shift in pitch range (Taylor 1994: 20). Mennen Schaeffler, and Docherty (2008) explain that pitch range is particularly difficult to quantify. Most commonly, long-term distributional measures are used for pitch span measurement (2008: 527).

2.6 Functions of intonation

The information carried by speech can be categorized into three major groups. Following Fujisaki's (2004) distinction, these groups are:

Linguistic information: "the symbolic information that is represented by a set of discrete symbols and rules for their combination" (Fujisaki 2004: 1).

Paralinguistic information: "the information that is not inferable from the written counterpart but is deliberately added by the speaker to modify or supplement linguistic information" (Fujisaki 2004: 2).

Non-linguistic information: "these factors are not directly related to the linguistic and paralinguistic contents of the utterances and cannot generally be controlled by the speaker" (ibid.).

Intonation operates on all of these three levels of information. On the *linguistic* level, intonation – or, to be more precise, stress and its possible acoustic correlate *f0* – can serve as a means of marking metrically strong syllables. Further, *f0* is indicative of word class demarcation: grammatical words frequently feature lower *f0* than lexical words (see Mixdorff 1998; Möbius 1993a, for example). On a *paralinguistic* level, intonation is a tool for conversational structuring in that it can highlight certain constituents that carry important information. The term *non-linguistic* refers to features over which the speaker has no direct control. These include physical factors such as sex, age, anatomical idiosyncrasies, or speech habits, such as the rate of articulation. The non-linguistic function of intonation further refers to the ability of intonation to express, for instance, emotion.

It should be pointed out that the above categorization is idealistic, since there is no consensus regarding the association of intonational features with specific