

Pitch Range is not Pitch Range

Christiane Ulbrich

School of Communication
University of Ulster
c.ulbrich@ulster.ac.uk

Abstract

This paper presents a phonetic analysis of pitch range as perceived and measured on utterance and syllable level. A previous analysis of read speech showed that German speakers produced a larger pitch-range on utterance level, whereas Swiss German speakers produced a larger pitch-range on syllable level. This analysis was based on the production of broadcasters reading news messages and a fairytale, both stylistically very restricted and largely standardized. Therefore, in the present study semi- and spontaneous utterances are analyzed to provide evidence that these findings are cross-linguistic rather than discourse-specific. The evidence was provided by auditory annotation and acoustic measurements.

1. Introduction and Background

Prosodic features are not only observed using different analytical frameworks [e.g. 11, 13]. Cross-linguistic studies have shown regional variation in prosody [2, 5, 6, 8]. This paper does not deal with what is usually termed a 'dialect'. Instead, it focuses on differences between two standard varieties (henceforth SVs) of German spoken in Germany and Switzerland. Even though, there is still a dispute about the existence of three rather than one German SVs, most authors agree that there are considerable differences between the two SVs [14, 4]. Regardless this ongoing debate, recent publications have provided evidence that there are significant differences on all linguistic levels between the three German SVs [1, 16]. A first comparison of prosodic features in Austrian, German and Swiss showed significant difference on the prosodic level [16]; including differences in pausing, pitch accent realisation and their tonal alignment, f_0 -declination, and pitch range. However, in the previous study the corpus consisted of read speech only, recorded from news broadcasters to allow for a direct comparability. The speakers were professional speakers reading a fairytale and eleven news messages. Due to their profession these are well trained speakers reading a highly stylised text. The present paper seeks confirmation of the findings in the realisation of pitch-range and f_0 -interval on utterance and syllable level in additional discourse types.

Prosodic features have been shown to have large inter- and intra-speaker variation [10; 15] and to be dependent on the speaking style [12]. Therefore recordings of semi- and spontaneous speech were collected from the same speakers who contributed to the read speech corpus to keep the speech material directly comparable.

Significant differences on the prosodic level were found in a previous auditory and acoustic analysis of three German SVs. The results of perception tests and of the auditory annotation led to the hypothesis that Germans generally speak more monotonously compared to the Swiss speakers. However, acoustic measurements showed that the absolute f_0 -interval in German utterances was larger compared to SG speakers. The local

f_0 -interval measured in the area pitch accents was found to be larger in SG than in GG. Comparing the previous results to the findings of the analysis of pitch range and f_0 -interval in semi- and spontaneous speech is intended to clarify whether the observed cross-linguistic differences are triggered by specific instructions and/or stylistic features of read speech and are therefore rather text-specific or if they are truly cross-linguistic.

2. Methods

2.1. Speech Material

The data of the read speech are taken from a larger corpus investigating the prosody of German, Austrian and Swiss news readers in public broadcasting agencies and have been collected in 2002 [16]. For the present comparison of three different discourse types six of the originally recorded speakers – three male Swiss (DRS II) and three male German speakers (ARD) – have been recorded in 2004. The semi-spontaneous speech data were collected from a retold version of a recipe. All speakers had to use the same keywords; nouns and predicates used in the original recipe. This was intended to control sentence length and permit lexical comparability of the occurrence of pitch accents. The spontaneous speech data were selected during an interview. They contained the answers to the question: "If you were free to choose, where would you like to live?". This question was chosen to control for emotions which has been shown to affect the prosodic level of spoken language [9]. Only declarative sentences were selected and matched under the following conditions:

- number of syllables
 - position of pitch accents
 - phonological syllable structure of accented syllables and phonetic environment
 - phonetic content of accented syllables and their environment
- From each corpus part (read (R), semi- (SE) and spontaneous (SP)) seven sentences were selected per speaker (N6). The acoustic analysis was carried out using PRAAT [4]. Measurements of f_0 on utterance and syllable level were taken in the periodic part of the vowel for each syllable. Two control listeners phonetically transcribed the speech material on the segmental and the prosodic level. The phonetic prosodic transcription was adapted from the phonetic tier of the IViE system which allows for a syllable-by-syllable annotation [7].

2.2. Previous Results

The hypotheses for the present study were derived from perception tests and an ensuing auditory and acoustic analysis of the read speech corpus collected for the comparison of prosodic characteristics in the three German SVs spoken in Switzerland, Austria and Germany. During the perception tests with differently experienced participants German speakers

were found to produce a smaller pitch range and, generally, to produce read speech more monotonously compared to Swiss (and Austrian) speakers. The acoustic measurements showed that the German speakers actually produce a larger f_0 -interval if the utterance level and an absolute f_0 -interval are considered. However, if the realization of pitch accents is considered, f_0 -measurements revealed that Swiss speakers produced a larger f_0 -interval in the realization of local pitch targets.

The f_0 -interval resulting from acoustic measurements of an absolute f_0 -maximum and an absolute f_0 -minimum was found to be larger in GG speaker's utterances compared to the Swiss speakers. Figure 1 gives an illustration of the f_0 -measurements.

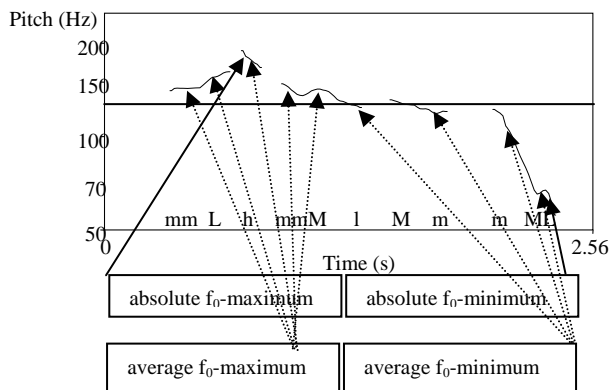


Figure 1: Illustration of measurements of an absolute and averaged f_0 -maximum and f_0 -minimum on utterance level including phonetic prosodic transcription

However, a closer look at the measurements revealed that the f_0 -maximum and f_0 -minimum in GG speaker's utterances were largely determined by the position of the pitch accent. The absolute f_0 -maximum was measured on the first accented syllable within the utterances whereas the absolute f_0 -minimum was found at the end of the declarative utterances. SG speakers on the other hand realized absolute f_0 -maxima and f_0 -minima not exclusively in those utterance positions. They were found to be produced also within the declarative sentences in the realization of pitch accents as well as in IP (Intonational Phrase) boundary tones. Therefore, an additional f_0 -interval was compared for the different groups of speakers, calculated from f_0 -measurements per syllable. F_0 -measurements above a speaker-specific mean value (f_0 -median) were averaged to the $f_{0\text{mean-maximum}}$. F_0 -measurements below a speaker-specific mean value were averaged to the $f_{0\text{mean-minimum}}$.

Considering the area of pitch accents (syllable level) SG speakers were found to produce a larger $f_{0\text{mean}}$ -interval on either the accented syllable itself (upwards or downwards glide) or a f_0 -movement in the syllable preceding the accented syllable took place towards the opposite direction as the tonal target to be produced. This means, there is a low or falling f_0 before a high tonal target and a high or rising f_0 before a low tonal target. Figure 2 gives a schematic illustration of the measurements on the syllable level for a falling or low target (LH) and a rising or high target (HL) indicated by capital letters in the phonetic prosodic transcription. It also illustrates the cross-linguistic difference previously found.

In the read speech corpus the difference between the absolute f_0 -interval and the averaged $f_{0\text{mean}}$ -interval was significantly larger in GG speakers utterances (6,9st) compared to SG speakers (2st) [16]. Considering the speaking style of the read

speech corpus the occurrence of the absolute f_0 -maxima and f_0 -minima could be due to the highly standardized discourse type. The findings left the question open if these prosodic characteristics are text-specific in nature rather than cross-linguistic. Therefore, the analysis was carried out on the semi- and spontaneous speech data using the exact same methods as in the read speech corpus.

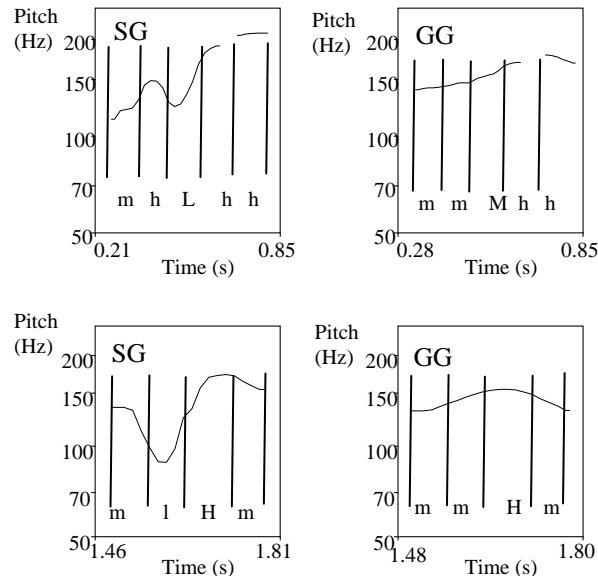


Figure 2: Illustration of f_0 -measurements on syllable level including phonetic prosodic transcription

In all three corpora seven sentences were selected:

- 17 syllables per sentence
- Phonetic context of the accented syllable had to contain two unaccented preceding syllables
- Phonological syllables structure (CV-structure) was matched (in the read and the semi-spontaneous corpus it was possible to keep the phonetic content directly comparable, not so in the spontaneous data)

The comparison on the syllable level was carried out for high (H) and low (L) pitch targets separately. The duration of phonetic segments was normalized.

The following table shows the f_0 -measurements for the analysis of the f_0 -interval on utterance level statistically analyzed using SPSS ANOVA in a repeated measurements design taken for each speaker (Total/S) and for the total of all speakers (Total/all).

Table 1: f_0 -measurements for the utterance level

corpus	N of sentences	N of syllables/ sentence	f_0 -max f_0 -min/ sentence	f_0 -int $f_{0\text{mean}}$ -int
Total/S	21	51	21/21	21/21
Total/all	126	306	126/126	126/126

Since high and low targets were not produced on the same syllables by all speakers a selection of five high and five low targets for each part of the corpus took place. Table 2 shows the number of pitch targets analyzed in the present study.

Table 2: pitch targets for the syllable level

corpus	high pitch targets	low pitch targets
Total/S	15	15
Total /all	90	90

F_0 -measurements on the syllable level were taken within the two syllables preceding the pitch target and on the pitch target. Using these two measurements two local f_0 -intervals were calculated: pre- f_0 -1 and pre- f_0 -2. This is illustrated in Figure 3.

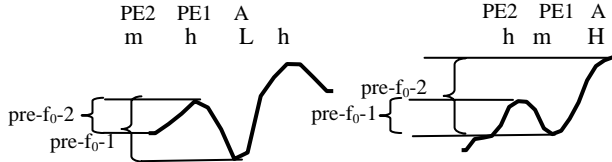


Figure 3: Stylized pitch contours, phonetic prosodic transcription and illustration of pre- f_0 -intervals 1 and 2.

2.3. Hypotheses

The present investigation of read speech led to the following hypotheses for the three speaking styles (R, SE, SP):

- 1) Absolute f_0 -interval of GG speakers is larger than the absolute f_0 -interval of SG speakers
- 2) The difference between absolute f_0 -interval and averaged $f_{0\text{mean}}$ -interval is larger in the realization of GG compared to SG speakers
- 3) For both high and low pitch targets pre- f_0 -1 is larger than pre- f_0 -2 in realizations produced by SG speakers compared to GG speakers

3. Results

The findings of the present study will be presented separately for the utterance and the pitch accent level for the three speech corpora (R, SE, SP). The results are presented using semitones (st).

3.1. Absolute and averaged f_0 -interval

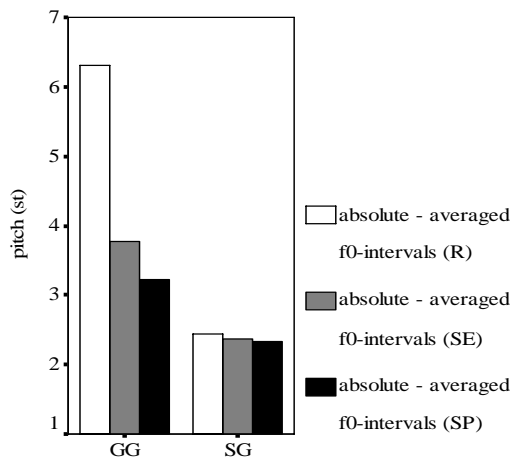


Figure 4: f_0 -difference between f_0 -interval and $f_{0\text{mean}}$ -interval for GG and SG speakers in read, semi- and spontaneous speech

Significant differences between the GG and SG speakers were found in the comparison of the f_0 -distance between absolute and averaged f_0 -interval (f_0 -int; $f_{0\text{mean}}$ -interval: $F_{(1,4)}=15,736$; $p<0,05$). Also, the results for the speaking style showed significant differences $F_{(2,8)}=6,147$; $p<0,05$). Nonetheless, there is an interaction between the origin of the speaker and the speaking style ($F_{(2,8)}=5,454$; $p<0,05$), due to the fact that absolute and averaged f_0 -interval differ significantly only in the read speech corpus and not in semi- and spontaneous speech. These findings lead to the conclusion that the larger difference between absolute and averaged f_0 -interval in read speech produced by GG speakers is rather a text-specific than a cross-linguistic prosodic feature.

3.2. F_0 -interval of pitch targets

The results of the acoustic analysis on syllable level will be detailed for each of the two syllables preceding either a high or a low pitch target.

3.2.1. Pre- f_0 -2 and pre- f_0 -1 of high pitch targets

The statistical analysis of the pre- f_0 -2-interval resulting from f_0 -measurements in the accented syllable (A; see figure 3) and the syllable not immediately preceding the pitch target (PE2) was not found to differ significantly between the GG and SG speakers for high pitch targets. However, significant differences were found regarding the discourse-type ($F_{(2,8)}=13,959$; $p<0,05$). F_0 in R was significantly lower compared to SE and SP. The results for both pre-pitch-target f_0 -intervals before high pitch targets are illustrated in figure 5.

The picture changes when pre- f_0 -1 is considered resulting from f_0 -measurements in the accented syllable (A; see figure 3) and the syllable immediately preceding the pitch target (PE1) Significant differences were found under both conditions; discourse type ($F_{(2,8)}=125,47$, $p<0,0001$) and origin of speaker ($F_{(1,4)}=24,734$; $p<0,05$). The pre- f_0 -1-interval was found to be significantly smaller in GG than in SG realizations. With respect to the discourse type the pre- f_0 -1-interval in R was smallest and largest in SP.

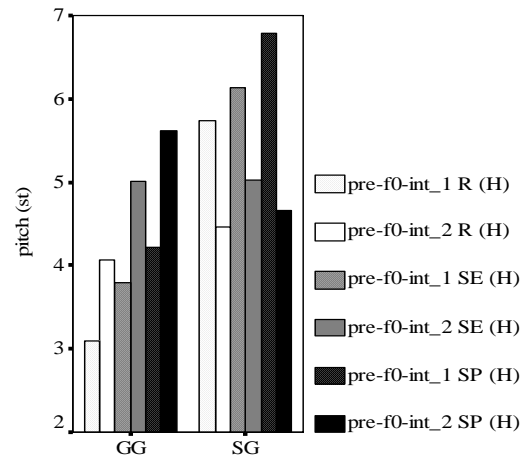


Figure 5: cross-linguistic comparison of two pre-accent f_0 -intervals in SG and GG high pitch accent realizations

3.2.2. Pre- f_0 -2 and Pre- f_0 -1 of low pitch targets

The statistical analysis of the pre- f_0 -2-interval resulting from f_0 -measurements in the accented syllable (A; see figure 3) and

the syllable not immediately preceding the pitch target (PE2) was not found to differ significantly between the GG and SG speakers for low pitch targets. Only depending on origin of the speaker significance was approached ($F_{(1,4)}=5,81$; $p<0,074$). The pre- f_0 -2-interval was larger in GG realizations compared to the SG speakers. The results for both pre-pitch-target f_0 -intervals before low pitch targets are illustrated in figure 6.

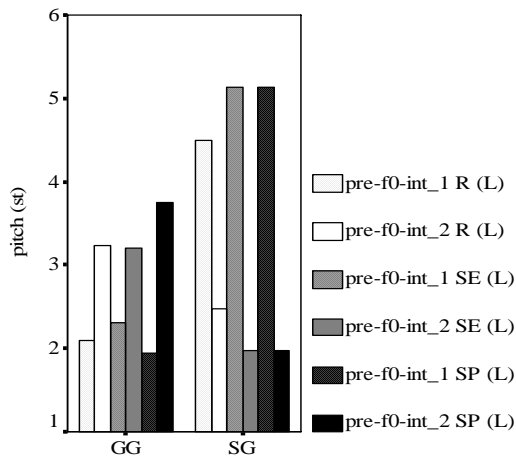


Figure 6: cross-linguistic comparison of two pre-accent f_0 -intervals in SG and GG low pitch accent realizations

Also in pre- f_0 -1 resulting from f_0 -measurements in the accented syllable (A; see figure 3) and the syllable immediately preceding the pitch target (PE1) before low pitch targets, pre- f_0 -1 was significantly smaller in GG realizations than in realizations produced by SG speakers ($F_{(1,4)}=29,247$; $p<0,05$). No significant differences for discourse type were found.

4. Conclusion

The present study shows cross-linguistic prosodic features of speakers of the German and the Swiss German SV. The analysis was carried out to provide evidence for differences in pitch range on utterance and syllable level in different discourse types, since a previous analysis was exclusively based on read speech. The statistic analysis of absolute and averaged f_0 -intervals showed that the previous findings provide rather evidence for discourse-specific prosodic features. Only in read speech the distance between absolute and averaged f_0 -interval is larger for GG speakers compared to SG speakers. Since the feature has been observed in a larger corpus of read speech containing different text types this leads to the conclusion that the realization of an f_0 -maximum on the first accented syllable and the realization of an f_0 -minimum during the final fall is either a prosodic feature of German read speech or due to broadcast specific reading style. The latter conclusion is more likely since speech training and instructions is highly specific in the public sector. The cross-linguistic comparison nonetheless provides evidence that an absolute f_0 -interval is larger in GG utterances compared to SG speakers. That confirms the previous analysis that pitch variation in SG utterances larger and more frequent than in GG utterances.

Comparing the results for two local pre-accent f_0 -intervals within three different discourse types showed that previous results for read speech on the syllable level show cross-linguistic difference rather than discourse-specific prosodic features. The data showed that before a tonal target is realized

by GG speakers, f_0 moves towards that specific target, whereas SG speakers realize a f_0 movement opposite the target to be produced. The observation that pre- f_0 -2 does not differ significantly between the two groups of speakers – while pre- f_0 -1 does – shows that the larger local f_0 -interval of SG speakers preceding the pitch target is restricted to a minimal movement of only the pre-accented unstressed syllable. This is best described in a phonetic prosodic annotation. The results suggest that the auditory impression of a smaller pitch range in GG compared to SG might rather be due to less and smaller f_0 -movements on the syllable level. This however remains a hypothesis which needs to be verified by perception tests.

5. References

- [1] Ammon, U., 2004. Variantenwörterbuch des Deutschen. Berlin: De Gruyter.
- [2] Auer, P.; Gilles, P.; Peters, J.; Selting, M., 2000. Intonation deutscher Regionalsprachen – Vorstellung eines Forschungsprojekts. In: *Akten des 1. Kongresses Internationalen Gesellschaft für deutsche Dialektologie*. Stellmacher et al. (Eds.), Wiesbaden: *Interaction and Linguistic Structures*. Freiburg: Steiner ZDL Beihefte).
- [3] Boersma, P.; Weenink, D., PRAAT <http://www.fon.let.uva.nl/praat/>.
- [4] Clyne, M. G., 1995. *The German Language in a changing Europe*. Cambridge University Press.
- [5] Grabe, E.; Nolan, F. and Farrar, K., 1998. IViE – a Comparative transcription system for Intonational variation in English. *ICSLP 5*, Sydney, Australia.
- [6] Grabe, E.; Post, B.; and Nolan, F., to appear. Modelling intonational Variation in English. The IViE System. In: *Proceedings of Prosody 2000*, 2-5 October, Krakow,
- [7] Grabe, E. IViE-Labeling Guide <http://www.phon.ox.ac.uk/~esther/ivyweb/guide.html>.
- [8] Gilles, P. 2005. Regionale Prosodie im Deutschen. In: *Linguistik – Impulse & Tendenzen*. In: Günther, S. et al. (Eds.): *Linguistic – Impulse & Tendenzen*.
- [9] Kehrein, R. 2002. Prosodie und Emotionen. In: *Reihe Germanistische Linguistik 231*.
- [10] Ladd, D. R. (1996) *Intonational Phonology*. CUP.
- [11] Ladd, D.R.; Terken, J. 1995. Modeling inter- and intraspeaker pitch range variation. In: *Proc. International Congress of Phonetic Science*, Stockholm.
- [12] Mixdorff, H., Pfitzinger, H.R., Grauwinkel, K. 2005: Towards Objective Measures for Comparing Speaking Styles. *Proc. of the 10th Int. Conf. on Speech and Computer*, 131-134.
- [13] Möbius, B., Pätzold, M., Hess, W. (1993): "Analysis and synthesis of German F0 contours by means of Fujisaki's model", *Speech Communication*, vol. 13, 53-61.
- [14] Schmidt, J.E. 2005. Die deutsche Standardsprache: Eine Varietät – drei Oralisierungsnormen. In: Ludwig M. Eichinger / Werner Kallmeyer (Eds.): *Standardvariation – Wie viel Variation verträgt die deutsche Sprache?* Berlin/New York, 278–305.
- [15] Shriberg, E.; Ladd, D.R., Terken, J.; Stolcke, A. 1996. Modeling pitch range variation within and across speakers: Predicting F0 target when "speaking up". *ICSLP*, 1-4, Philadelphia, PA.
- [16] Ulbrich, C. (2005 in. pr.): *Phonetische Untersuchungen zur Prosodie der Standardvarietäten des Deutschen*. In: *Hallesche Schriften zur Sprechwissenschaft und Phonetik 17*.