

Perception of Dialectal Prosody

Adrian Leemann¹, Beat Siebenhaar²

¹ Department of Linguistics, University of Berne, Switzerland

² Department of Linguistics, University of Berne, Switzerland, and Institut für Germanistik,
Universität Leipzig, Germany

adrian.leemann@isw.unibe.ch, siebenhaar@uni-leipzig.de

Abstract

Previous studies on the perception of language prosody and dialectal prosody have shown that languages and regional dialects can be identified by prosodic cues alone. This pilot study tests this for 4 Swiss German dialects. 70 participants were presented with filtered speech material devoid of segmental cues. The filter was applied for frequencies between 250 Hz-7000 Hz. Despite this filtering, 3 of 4 dialects were recognized by the participants. Identification rates were considerably higher for dialects which present distinct prosodic features, in this case relatively slow speech rate in one instance and high pitch range in the other.

Index Terms: dialectal prosody, speech perception, dialect identification

1. Introduction

During the communicative process we not only communicate contents but also a great deal about ourselves. In the case of Swiss German, for example, speakers of the dialects automatically anchor their geographic origin [1]. The basis on which dialectal speakers of Swiss German determine where their dialect-speaking interlocutor is from are sound characteristics, syntactic cues, and cues in the realm of the lexicon.

The question we are asking ourselves in this paper is whether or not prosody alone, as part of the sound cues, is a significant feature by which the regional origin of Swiss German speakers can be identified. In order to test whether this is the case it first needed to be confirmed that prosodic differences between Swiss German dialects in fact exist – a topic which, in Swiss Linguistics, has not been addressed thoroughly, thus a gap which our project is trying to fill in part. In our Swiss National Science Foundation (SNSF) project (*Quantitative Approaches to a Geolinguistics of Swiss German Prosody* 2005-2008) we have been able to show that significant prosodic differences between Swiss German dialects exist. To this day, we have partially analyzed 3 of 4 target dialects (Bern (BE), Zurich (ZH), Valais (VS), and the Grisons (GR) – GR is yet to be analyzed) – while BE and ZH both represent Midland varieties, BE in the West - ZH in the East, VS and GR stand for Alpine varieties, again, divided into a Western and into an Eastern variety.

By the application of the Fujisaki intonation model [2], a number of intonational characteristics of the investigated 3 dialects were established. For the VS speakers it has become clear that they show the highest pitch range on the local and global level. The BE dialect distinguishes itself with late pitch onsets and late pitch offsets, i.e. the BE show later pitch rises and later pitch falls, with regard to syllable starting point and syllable end point. The ZH speakers, in contrast, show early pitch onsets and early pitch offsets; pitch onsets which begin before the actual syllable start [3]. However, these

characteristics in pitch onset and pitch offset times are only phonetic and not phonological in nature (see [4] and [5], for example).

On the timing level, certain characteristics of the BE, ZH, and VS dialects have also been detected. ZH speakers show a comparatively high articulation rate (5.9 syllables/sec versus VS 5.8 syllables/sec, and BE 5.1 syllables/sec.). Furthermore, the BE generally show longer vowels when compared to VS and VS speakers present a shortened vowel when the following segment is less sonorant. The BE dialect shows extensive initial and final syllable lengthening – a tendency which is not this distinct in the VS dialect [6].

The results of this project have shown that there are significant prosodic differences between the investigated dialects. It needs to be tested whether these prosodic differences are also perceptible. There are several studies geared at finding out more about the significance of prosody in dialect and language identification; the most relevant ones for the present study are discussed below. In the context of Swiss German these studies do not exist, however.

1.1. Recognition of languages and dialects by prosodic cues alone

One of the most prominent studies is that of O'Hala et al. [7], who investigated whether languages could be identified by prosodic cues alone. They examined Chinese, English, and Japanese. For the experiment they used a delexicalized speech signal which did not contain any segmental structure, yet the speech contained information about intonation, timing, and amplitude. This reduced speech signal was then played to participants. The identification rate was 56.4% (chance level guessing would have been 33.3%). The study showed that, in the context of different languages, prosodic cues alone allow for identification.

Whether this was also possible on the regional level was tested by Gilles et al. [8]. The source material used was a standard German text from a diatopically unmarked speaker. By means of speech resynthesis, 2 variants were generated which differed only in terms of intonation. 1 variant contained typical Hamburg intonation features; the other variant remained unmarked in its intonation. The participants were asked to judge whether what they hear is close to a Hamburg intonation or not. The authors concluded that regional specific intonation contours allow a geographical localization, even if segmental-phonetic features are absent.

Another study which tackled this issue is that by Schaeffler and Summers [9]. They played a delexicalized signal with speech data of 7 different German dialect regions to 16 participants. Despite the fact that the recognition rates for most dialects were not particularly high, the authors infer that there seems to be a North-South contrast in prosodic systems, as especially reliable rates were found for the South-West and North-West dialect regions.

1.2. Perception of Swiss German dialects

As mentioned earlier, no such study has yet been conducted in the context of Swiss German. Nevertheless, Ris [10] impressionistically points out that Swiss German dialects are commonly perceived as very different from each other by the Swiss population. He mentions that a number of dialects are perceived as particularly marked, among them also Bern, Zurich, Valais, and Eastern Switzerland (which includes the Grisons (GR)). The BE variety, so Ris, is perceived as “slow”, “homely” and “snug”, among other attributes. ZH is thought of as “fast”, “neutral”, “modern”, and “adaptive”. VS as “unintelligible”, “lovely”, and “indigenous”. GR is perceived as “clear”, “unappealing”, and “talkative”.

Because there has not been any study that attends to the issue of perceptual suprasegmental differences in the context of Swiss German dialects, we are testing this with the present pilot study. In contrast to the mentioned perception studies we do not compare languages or dialect groups in the present study; rather, we test the differentiation of dialects within one dialect group.

2. Method

The experiment consisted of two parts. In the first part, the participants, 23 University of Berne Students and 84 University of Zurich students (N=107), were presented with one dialect sample for each of the 4 dialects in question, i.e. 4 stimuli of a duration of 7 seconds each. The uneven distribution of participants is due to the varying availability of the students. Other than all the participants being native speakers of a Swiss German dialect, the participants' linguistic background was not controlled for. Here, the samples were retrieved from 4 male speakers considered as typical from our SNSF project database, which consists of spontaneous speech data from 20 speakers per dialect. The aim of this first short experiment was to test whether the participants can identify the four dialects in an unfiltered version.

For the main part of the experiment, 2 speakers from each dialect region were again selected from the corpus. Each speaker provided 2 speech samples which were again 7 seconds in length each, i.e. 4*4 stimuli = 16 stimuli. The speakers were chosen because they are in the core of the model for the dialect in consideration (see [6]) and, moreover, the authors perceived them as “prototypical” male speakers of the dialects in question. The speech files for both parts of the experiment were terminating phrases or complete sentences. They were recorded with Edirol R-9 and Marantz PMD 671 recorders.

In order to test the perceptive effect of prosody only, the speech signal needs to be devoid of segmental information, which is why the speech signals for the second part of the experiment were delexicalized.

2.1. Delexicalization procedure

In a first step, the fundamental frequencies of all speakers were normalized to 140 Hz in Praat [11], while the pitch floor value was set to 50 Hz and the pitch ceiling value to 300 Hz. This normalization procedure was conducted in order to prevent varying fundamental frequencies to have an affect on the participants' perception of the speech samples. The data was subsequently filtered with a stop Hann band filter at 250 Hz-7000 Hz, followed by a smoothing at 500 Hz. In order to normalize the files' amplitudes the peaks were scaled at 0.99. A resynthesis method (cf. [8]) was not considered because the segmental differences would clearly bias the results, as shown

by the recognition rates of the unmodified signal that are at over 85%. Moreover, a regionally unmarked Swiss German, as we find it in standard German, does not exist.

2.2. Experiment design

In addition to the above-mentioned 20 stimuli (4 unfiltered dialect samples, 16 filtered dialect samples), 2 stimuli which we had previously modified were incorporated in the second experiment in a matched guise fashion. Here, 1 phrase from a BE speaker was modified in its timing and intonational features to that of a VS speaker; in the other phrase, the procedure was inversed – one VS speaker's prosody was modified to that of a BE speaker. The parameters according to which this modification in prosody was undertaken can be found in [3] and [6]. For the intonational modification Mixdorff's *FujiParaEditor* was used [12].

The stimuli in both parts of the experiment were presented in randomized order and were played to the participants only once. After listening to the stimulus, the participants were given 5 seconds to indicate on a questionnaire whether what they heard was articulated by a VS, BE, ZH, or GR speaker. Furthermore, they were asked to specify the certainty of their judgment, i.e. “perhaps” or “probably”. Thus, the chance level of both experiments is $\frac{1}{4}$, i.e. 25%.

3. Results

The results from the first experiment show that the unfiltered, lexicalized version of the ZH phrase was recognized with a rate of 91%, the VS phrase with 89%, the GR phrase with 88%, and the BE phrase with 85%. Clearly, dialect identification is well above chance level with a 7 second original sound file.

The second experiment consists of recognizing the filtered sound files. For the analyses, only participants who provided judgments to all stimuli were taken into consideration. The number of Zurich participants with missing values is much larger than the number of Berne participants with missing values. This is due to the lower control in the larger lecture hall where the Zurich experiment was conducted, while in Berne, the test was performed individually or in small groups. 70 out of a 107 participants provided judgments to all stimuli, 22 BE participants and 48 ZH participants.

The identification rate of the four dialects lies at 32% (i.e. 7% above chance level ($p < 0.0001$)), Figure 1 shows the distribution of the correct judgments by dialect.

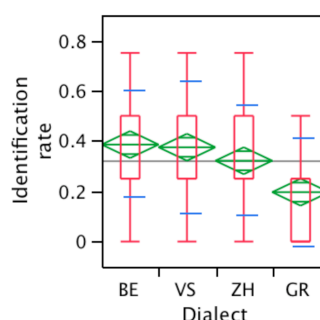


Figure 1: Identification rate of the 4 dialects.

In the following subsections, all 4 dialects and their recognition rates are presented. When relevant, the differences between how the BE subject group and how the ZH group identified the dialect will be shown. Finally, the results of the modified files will be attended to.

3.1. Recognition of GR dialect

The GR dialect was identified particularly poorly with an identification rate of only 19% - less than chance level. When we look at the distributions of the 4 GR speech samples we find that in only one case the mode lies on the GR dialect. An interesting picture emerges if we consider how the BE and ZH participants identified the GR dialect speakers. Figure 2 shows the correct dialect identification rate on the y axis and the 2 subject groups on the x axis. Diamonds mark the upper and lower confidence intervals of 2.5% each. The differences between the BE and ZH participants are significant ($p=0.0127$). This entails that the ZH group identifies the GR dialect significantly better than the BE group does, however, it does so very poorly (cf. also section 3.2.).

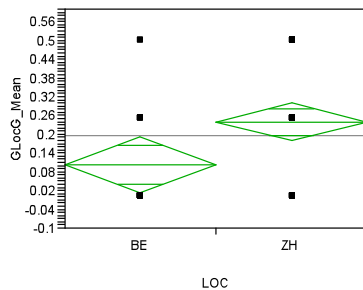


Figure 2: BE and ZH participants' identification rate of GR dialect.

3.2. Recognition of VS dialect

The VS speech items are perceived and judged differently. While in one example the participants showed a clear preference for the VS dialect (41%), the other examples show lower identification rates. The VS speakers are generally identified with a rate of 37.5%, 12.5% above chance level ($p<0.0001$). Again, we find an interesting result when considering how the BE and ZH participants identified the VS dialect, Figure 3 shows the differences in identification of the VS samples between BE and ZH participants. It clearly shows that the BE participants found it easier to identify the VS dialect, with an identification rate of 47%. The ZH speakers, in contrast, identified the VS speakers at 33% ($p=0.048$).

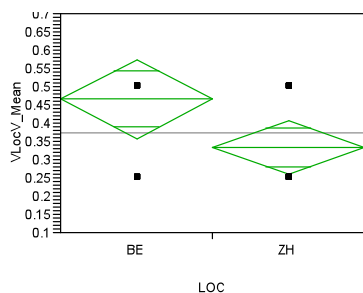


Figure 3: BE and ZH participants' identification rate of VS dialect.

The fact that the ZH participants identified the GR samples significantly better and the BE participants recognized the VS speakers much more readily is almost certainly due to a more extensive contact between VS speakers and BE speakers in Berne, and ZH speakers and GR speakers in and around Zurich. Statistics from the 2000 census substantiate this claim: Of all daily commuters to Zurich, 0.73 % come from GR, while there are only 0.15 %

from VS. The relation is reversed in Berne, where 1.06 % of all daily commuters come from VS and only 0.2 % from GR [13]. This correlation of contact and recognition corresponds to Clopper and Pisoni' results [14] for non modified signals.

3.3. Recognition of BE dialect

The BE examples show the most judgments for the correct dialect in 3 out of 4 speech samples. The BE speakers were identified with a rate of 39%, i.e. 14% above chance level ($p<0.0001$).

Interestingly, the BE participants recognize their own dialect samples a little bit less (35%) than the ZH subject group recognizes the BE dialect samples (40%).

3.4. Recognition of ZH dialect

The ZH samples are again assessed differently. One example was identified with a rate of 37% the other three tokens did not achieve such high ratings. In one instance, we even find a ZH sample file identified as a VS dialect sample. For the ZH dialect samples there is an identification rate of 32%, 7% above chance level ($p=0.0038$). We find a similar tendency in ZH dialect recognition as we did in the recognition of the BE dialect, in that the group whose own dialect is being judged performs more poorly at detecting the dialect than the other group: here, the BE participants show an identification rate of 34%, while the ZH group only show 31% of correct identification; these differences are not significant, however.

Another interesting finding in the ZH recognition data is that in all ZH samples the BE option is chosen the least – a result which reveals that the participants rate BE and ZH as being clearly different from each other. These results straightforwardly show that a conception of BE prosody and what it should sound like exists - a ZH example is surely not a BE example.

3.5. Recognition of prosodically modified samples

As pointed out earlier, we modified 1 BE speaker file to that of a VS prosody speaker and 1 VS speaker file to that of a BE prosody speaker. Both the files were randomly incorporated within the 16 stimuli of the second part of the experiment, resulting in a total of 18 stimuli in the second part of the experiment. The results of both ZH and BE participants are summarized in Table 1 below. Percentages indicate how the modified speakers' dialects were judged.

	BE with VS prosody	VS with BE prosody
Recognized as ZH	36%	27%
Recognized as BE	21%	24%
Recognized as GR	16%	16%
Recognized as VS	27%	33%

Table 1: Judgments of modified speech samples.

Figures in Table 1 show that for the BE speaker with VS prosody most participants indicated to hear a ZH speaker (36%), followed by a VS speaker (27%). For the VS speaker with BE prosody, 33% judged the speaker as a VS speaker and 27% as ZH speaker.

The identification of the 2 modified files cannot be examined in greater detail in the present pilot study, as only one example per modified dialect sample exists. The distributions of the given judgments to all the stimuli were tested for their difference to the probability distribution with chi square. In the case of the judgments for VS, ZH, and BE,

the null hypothesis was refuted $\frac{3}{4}$ of the time. In terms of the GR dialect, only in one instance was H0 refuted; here the judgment was wrong to begin with. The modified files both accept the null hypothesis, i.e. the modeling was not able to elicit a definite preference for judgments towards any dialect.

3.6. Uncertainty in judging the speech samples

The participants were asked to indicate the certainty of their judgments. The main experiment was perceived as a very difficult task. This was on the one hand communicated to the authors after the experiment; on the other hand, this becomes clear when we look at the certainty of their judgments. 1 indicating certainty, 2 indicating uncertainty, the participants believed the VS dialect to be the easiest to detect with a 1.83 mean overall rating, while the values for the others ranged between 1.86 (GR) and 1.90 (ZH). These uncertain judgments are also due to the experiment condition: 5 seconds per judgment is rather short and, in addition, the subjects were able to listen to each stimulus only once. However, even these restricted conditions allow for a distinction of the dialects.

4. Discussion

The confirmed difficulty the participants experienced when taking part in this experiment go hand in hand with the findings of this study. Despite the fact that 3 out of 4 dialects were recognized above chance level it seems difficult to recognize the speakers' dialects only on a prosodic level. We did find remarkable results in that the Southern varieties, VS and GR, are recognized better by the BE participants in the VS case, and by the ZH participants in the GR case. As previously pointed out, it can be assumed that this is due to the dialect speakers being in contact with each other [13], as the majority of VS students study in BE and a great number of GR students are ZH orientated for their tertiary education.

Furthermore, a tendency was detected in that the BE participants performed more poorly in recognizing their own dialect, and the ZH participants acted analogously in the recognition of their own tongue. Most likely, this has to do with the fact that one generally recognizes dialects other than one's own dialect more accurately – as in one's own perception, the "other" variety is normally more marked.

If we revisit the identification rates once again, we find that the rates for the BE dialect are the highest. One may speculate that the BE speakers' slow articulation rate and other typical BE timing characteristics possibly give away their geographic origin. As for the second most recognized dialect, the VS dialect, it may be their distinct intonational features which led the participants to attribute the correct geographic origin. For the rather low identification rates of the ZH dialect and particularly of the GR dialect it is more difficult to come up with plausible speculations. As Ris [8] points out, the ZH variety is perceived as a more "neutral" dialect – and thus more difficult to identify by just prosodic cues alone, possibly. The reasons why the GR dialect is generally identified so poorly may be due to the participants' unfamiliarity of the dialect.

The modified dialect samples did not provide clear results and are in need of further elaboration. With only 2 modified files, we were not able to establish a definite attribution to one variety.

5. Conclusions

This pilot study has shown that Swiss German dialects, which all belong to the same dialect group, can be recognized by prosodic features alone. This finding is in need of

qualification, however. The identification rates are not particularly high. Nevertheless, a number of interesting findings have emerged. On the one hand, proximity between dialect speakers does help in the identification of dialects. On the other hand, results show the tendency that it is more difficult to detect one's own dialect than it is to detect other dialects.

This pilot study has brought to light a number of issues that may be refined in a large-scale follow up experiment. Possibly, longer dialect sample sound files could be selected. O'Hala et al. [7] have shown that identification rates increase when the participants are exposed to longer sound files. Furthermore, female dialect speakers along with male speakers could be selected for the experiment, possibly results would be different when female speakers' dialects are judged. Finally, the applied filter may be adjusted and optimized.

6. Acknowledgements

We thank Stephan Schmid for conducting the experiment with University of Zurich students, Manuela Guntern for lending a hand in the data preparation process, and Heidy Suter for providing insights towards the discussion section.

7. References

- [1] Werlen, I., "Mundarten und Identitäten", in *Dialekt in der (Deutsch)Schweiz – Zwischen lokaler Identität und nationaler Kohäsion*. Lenzburg: Forum Helveticum Schriftenreihe, 15, 26-32, 2005
- [2] Fujisaki, H., "A note on the physiological and physical basis for the phrase and accent commands in the voice fundamental frequency contour", in O. Fujimura [Ed], *Vocal physiology: voice production, mechanisms, and functions*, New York: Raven, 347-355, 1988.
- [3] Leemann, A. and Siebenhaar B., "Swiss Alpine and Midland Intonation", in *Proceedings of Speech Prosody 2008*, to appear 2008.
- [4] Fitzpatrick-Cole, J. "The alpine intonation of Bern Swiss German", in *Proceedings of ICPhS XIV San Francisco*, 941–944, 1999.
- [5] Schmid, S. and Fleischer J., "Zurich German", *Journal of the International Phonetic Association*, 36/2, 243–253, 2006.
- [6] Leemann, A. and Siebenhaar B., "Intonational and Temporal Features of Swiss German", in *Proceedings of the ICPhS XVI Saarbrücken*, 957-960, 2007.
- [7] O'Hala, J. J. and Gilbert, J. B., "Listeners' ability to identify languages by their prosody", in P. Léon [Ed] *Problèmes de Prosodie II: Expérimentations, modèles et fonctions*. Studia Phonetica, 18, 123-131, 1981.
- [8] Gilles, P., Peters, J., Auer, P., and Selting, M., "Perzeptuelle Identifikation regional markierter Tonhöhenverläufe. Eine Pilotstudie zum Hamburgischen", *Zeitschrift für Dialektologie und Linguistik*, 68/2, 155-172, 2001.
- [9] Schaeffler, F. and Summers R., "Recognizing German dialects by prosodic features alone", in *Proceedings of the ICPhS XIV San Francisco*, 2311-2314, 1999.
- [10] Ris, R., "Innerethik der deutschen Schweiz", in P. Hugger [Ed], *Handbuch der schweizerischen Volkskultur*, Band II, Zürich: Offizin, 749–766, 1992.
- [11] Praat – Boersma, P. Weenik D., 2008.
Online: <http://www.fon.hum.uva.nl/praat/> accessed on 14 April 2008.
- [12] Mixdorff, H. Fujisaki Parameter Extraction Environment, 2008.
Online: <http://www.tfz-berlin.de/~mixdorff/thesis/fujisaki.html> accessed on 14 April 2008.
- [13] BFS, Neuchâtel, Eidgenössische Volkszählung. Schätzungen: Sektion räumliche Analysen, 2003. Online: <http://www.media-stat.admin.ch/stat/pendler> accessed on 14.4.2008.
- [14] Clopper, C. G. and Pisoni D. B., "Effects of region of origin and geographic mobility on perceptual dialect categorization", in *Language Variation and Change*, 18, 193–221, 2006.