RECOGNIZING GERMAN DIALECTS BY PROSODIC FEATURES ALONE

Felix Schaeffler and Robert Summers
IPSK - Department of Phonetics and Speech Communication, University of Munich, Germany

ABSTRACT

The prosodic characteristics of German dialects have seldom been the subject of research. Therefore a perceptual experiment was designed to find out whether there are prosodic cues in the speech signal strong enough to identify the origin of the speaker. Spontaneous monologues from 70 speakers were used as items. 16 subjects participated in the experiment. The monologues were taken from the Munich RVG1 corpus. A delexicalization procedure was applied to the speech data. Seven main regions of German dialects were used as categories. The aims of the study were to find out whether there are perceptable differences between the prosodic features of the dialects and whether the dialects differ in the saliency of their prosodic characteristics. The results suggest that there are prosodic characteristics for some dialect groups. Good recognition rates were found for south-west and north-west dialects. A north - south contrast seems to be a salient feature.

1. INTRODUCTION

Some languages have notable differences in the prosodic system of their dialects (e.g. Swedish). Although German has a great variety of different dialects, little is known about prosodic characteristics of these dialects. Research about German prosody normally refers to "the German language" and besides every-day-experience there are only a few reports about prosodic phenomena which are characteristic for certain dialects: One well known prosodic feature of some Rhinelandic dialects ist the so-called "correption" or "Rhineland Accentuation". [3] describes it as a sharp fall of pitch and intensity in the stressed vowel together with a shortening of the vowel (cf. also [5]).

[2] describes some more general differences between regional varieties of German. It is mentioned that southern German dialects show normally displacement of prominence peaks to the right. The accented syllable has low pitch, followed by a rise (and often a peak) on the subsequent syllables. Furthermore, in [2] three intonation patterns for falling nuclear accents in statements are mentioned:

a) a plain fall
b) a high fall from the nuclear syllable followed by a low rise
c) a delayed rise to high after the nuclear syllable, followed by a fall to mid or low

While pattern a) is a pattern of Standard German with no dialectal preference, pattern b) is found in the area of Hamburg and pattern c) is found in a large area along the Rhine valley from Switzerland to Cologne. As it is not mentioned in [2] on which amount of data these descriptions are based, their generality is hard to assess. In [3] several older investigations of prosodic characteristics of certain dialects are reported, but research has been delimited to certain areas and a small set of features.

On the other hand everyday experience shows, that at least some dialects must have special prosodic characteristics. E.g. often people report that speakers from the Palatinate "sing" or that Bavarian speakers have a "flat" intonation. Therefore it is certainly an interesting challenge to find out what prosodic differences between German dialects (or, to put the topic less restrictive, German regional variants) exist.

Phonetic investigations of prosody often bear serious problems. Descriptions of prosodic features are seldom based on spontaneous speech, and many methodological problems arise when for example trying to match measurements of F0 on perceptual entities. Therefore it was decided not to start the investigation of prosodic characteristics of German dialects with measurements of duration, pitch and intensity but to perform a perceptual experiment with delexicalized (filtered) speech items. The mentioned previous findings about prosodic differences in German dialects as well as everyday experience suggest that there are at least some prosodic differences between German dialects. To get a first impression of the possibility to recognize dialects from delexicalized speech, a pretest was performed with 4 subjects and 14 stimuli. The average recognition rate was 40%. This led to the hypothesis that it should be possible to recognize German dialects from prosodic features alone. No hypothesis was formulated concerning the saliency of prosodic features in different dialects. Rather, the results of this investigation should be used to formulate hypotheses about this matter.

Although many German people today do not speak strong dialects it is often easy to determine which region a speaker comes from, presumably mainly on the basis of sound characteristics, if syntactical cues or cues of vocabulary lack. The striking sound characteristics (beside syntactical and vocabulary differences) turn out to be a problem if interest focusses on the role of prosodic features in the perception of dialects. Segmental characteristics have to be extinguished to get a realistic impression about manner and saliency of dialectal prosody. Several delexicalization procedures (e.g. cf. [4], [6], [7]) have been described to destroy segmental information of a speech signal while preserving prosodic characteristics as well as possible. For this investigation a filtering technique was applied (see below).

2. SOME COMMENTS ON THE ITEMS

The items used for the experiment were taken from the RVG1 database, a collection of read and spontaneous speech from 500 speakers of German, including also speakers from Switzerland, Austria and northern Italy. The database was described in detail by [1]. The aim of this database was a collection of regional variants of German and not of true dialects. The difference between dialects and regional variants shall not be discussed in detail here. As we expected that the task of the performed experiment was rather difficult, a subset of speakers was used which showed a comparatively strong coloring of their language. Therefore it might be concluded that these speakers showed a speaking style somewhere between a dialect and a regional variant. Furthermore, 80% of the speakers in the RVGI corpus judged themselves as speaking a certain dialect (cmp. [1], p. 1087).

As also the speaking situation should support the use of dialectal speech, it was decided to use spontaneous monologues, where people spoke about their work, home or related issues and were to imagine

talking to somebody from their home region. The speakers in the RVGI database are classified into 9 main dialectal regions which are splitted into several subregions. The classification was mainly based on the regional background of the speaker and the speaker's parents. The 9 main regions are:

A. Low Franconian
B. West Low German
C. East Low German
(D. West Central German)
E. East Central German
F. Alemannic
G. East Franconian
(H. South Franconian)
I. Bavarian and Austrian Dialects

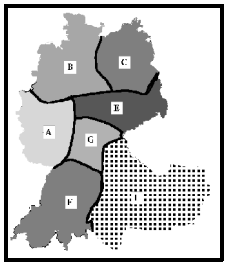

Figure 1: German dialect regions

For the experiment, region A and D were joined to one region "A" and region F and H were joined to one region "F", resulting in 7 regions. This was due to the fact that the regions A and H are rather small regions. Figure 1 shows a map of the regions used for the experiment.

3. PROCEDURE

70 speakers were selected from the RVGI corpus. The selection procedure was the following: First all speakers were excluded which had judged their own variant as "Standard German". 10 speakers (5 male, 5 female) of each region were selected from the remaining subset which showed strong dialectal coloring of their pronunciation. This judgement was performed by one of the authors by listening to the monologues.

The monologues were filtered between 70 Hz and 270 Hz with a band-pass filter. This delexicalization procedure was suggested by [4]. After several tests a filter of 6^{th} order was used for this experiment. This filter order showed the least artefacts and destroyed segmental information effectively. From each monologue a portion of 40 seconds was taken. This data was presented to 16 subjects in two sessions of thirty-five items in randomized order. The split into two sessions was necessary due to limited memory capacity on the computers used. The subjects were all from the staff of the Phonetics department at the University of Munich, but only a part of them were Phoneticians. Each subject was allowed to listen to the stimuli in whole or in parts, which was possible by marking parts of the speech signal by mouseclick. After listening the subjects were to assign each item to one of the 7 mentioned regions. The subjects were allowed to compare the stimuli and to pause as often as they wanted during the test. After judging the 70 stimuli, the subjects judged a randomly selected unfiltered subset of the stimuli to assess the subjects' ability to recognize German dialects. After the test the subjects filled in a questionnaire to get informations about the origin of the subjects, their opinion about the difficulty of the task and their strategies to solve it.

4. RESULTS

In the following, the findings about the relation of the dialectal regions and the assigned categories will be reported among other things. Therefore it is necessary to distinguish between the dialectal regions and the perceived categories. Capital letters (e.g. "A") will refer to the regions, small letters (e.g. "a") to the perceived categories.

4.1. Recognition rate

First it was tested whether the recognition rates the subjects showed were above chance recognition rates. Presumably the distribution of the recognition rate should correspond to a binomial distribution (n=70, p=0.14), if the judgements are based on guessing. The mean recognition rate of all subjects was 16 out of 70 (23%). A binomial test was performed which showed significance on the 5% level (P_{H0} (X=16) = 0.03) that the recognition rate had above chance probability.

4.2. Recognition of the different regions

To describe the relation of the perceived categories to the dialectal regions, a terminologic convention of automatic speech recognition research shall be adopted. The wrong and correct assignments can be regarded from two points of view: If one looks at a certain dialectal region, correct and incorrect assignments for this region can be counted. In this context, a correct assignment will be called a "match", an incorrect assignment will be called a "miss". On the other hand one can look at a certain perceived category and check how often this category was assigned to any dialectal region. In this context, a correct assignment will be called a "hit", an incorrect assignment will be called a "false alarm". Confusion matrices with percentual values are a good way to express the relation of the values. In the present case one has to pay attention to the fact, though, that the sum of judged items per region (160) is not equal to the sum of assignments per perceived category: the subjects had a bias towards certain categories. Table 2 shows the number of assignments for each category (in percent).

a	b	c	e	f	g	i
14,22	17,17	11,27	11,90	16,73	12,34	16,37

Table 2: percentual values of assignment for each category

Therefore two confusion matrices are shown. Table 2 is a confusion matrix which describes the relation of each dialectal region to the judgements of the subjects (in percent). Rows show matches and misses and sum up to 100%. The columns are not interpretable directly. Table 3 describes the relation of each perceived category to the dialectal regions (in percent). Columns show hits and false alarms and sum up to 100%. Again, the rows are not interpretable directly.

As table 3 shows, region "F" had the best matches (the judgement was correct in 33,1% of all cases). Almost the same rate (31%) was found for region "B". Region "I" was recognized in 25,6% of all cases. The other regions show lower recognition rates. In all of the 4 remaining regions a higher score was reached for an incorrect region. Items from Region "C" were most often judged as "b", region "G" was judged as "i", region "A" was judged as "e" or "b", region "E" was judged as "b" or "c".

	a	b	c	e	f	g	i
A	**14,4**	18,1	11,9	19,4	13,8	11,3	11,3
B	17,7	**31**,0	12,0	10,1	7,0	10,8	11,4
C	15,0	20,0	**17,5**	10,0	16,3	12,5	8,8

E	15,6	22,5	19,4	**18,1**	10,0	5,6	8,8
F	12,5	9,4	8,8	4,4	**33,1**	11,9	20,0
G	10,6	7,5	3,1	11,9	16,3	**21,9**	28,8
I	13,8	11,9	6,3	9,4	20,6	12,5	**25,6**

Table 3: Confusion matrix: Relation of regions to judgements (rows show matches and misses and sum up to 100%percent). Shaded cells denote matches.

	a	b	c	e	f	g	i
A	**14,5**	15,1	15,1	23,3	11,8	13,0	9,9
B	17,6	**25,5**	15,1	12,0	5,9	12,3	9,8
C	15,1	16,7	**22,2**	12,0	13,9	14,5	7,7
E	15,7	18,8	24,6	**21,8**	8,6	6,5	7,7
F	12,6	7,8	11,1	5,3	**28,3**	13,8	17,5
G	10,7	6,3	4,0	14,3	14,0	**25,4**	25,1
I	13,8	9,9	7,9	11,3	17,7	14,5	**22,4**

Table 4: Confusion matrix: Relation of assignments to regions (columns show hits and false alarms and sum up to 100%percent). Shaded cells denote hits.

	rank of misses and matches	rank of hits and false alarms
A (a)	e,b,a,* f,c, (i,g)	B,E,C,* A,I,F,G
B (b)	b,a,* c,i,g,e,f	B,E,C* ,A,I,F,G
C (c)	b,c,f,* a,g,e,i	E,C,(A,* B),F,I,G
E (e)	b,c,e,* a,f,i,g	A,E,G,* C,B,I,F
F (f)	f,i,* a,g,b,c,e	F,I,C,* G,A,E,B
G (g)	i,g,* f,e,a,b,c	G,(I,C),* F,A,B,E
I (i)	i,f,* a,g,b,e,c	G,I,F,* B,A, (C,E)

Table 5: Relation of regions to judgements: rank of judgements

Table 4 shows clearly that the recognition rates are influenced by the number of assignments: higher assignment rates correspond with higher recognition rates. Nevertheless the relations are still similar.

Table 5 shows the rank for each region from highest to lowest under the 2 conditions. The "*" marks the point where 50% are reached, categories or regions in brackets have equal percent values. Table 5 shows that 50% of misses and matches fell into two

	a	b	c	e	f	g	i

categories for 4 Regions (F,G,I,B). The best category here met the correct region, with the exception of region "G".

5. DISCUSSION

The data show some interesting tendencies, although the experiment should be continued with more subjects to have a better basis for interpretation and testing. All subjects reported in the questionnaire that the task was "difficult" or "very difficult". The recognition rates reflect this intuition. Although it could be proven that the mean recognition rate was not based on guessing, the recognition rates were not particulary high. It should be investigated whether the recognition rates correlate with the general ability of the subjects to recognize German dialects. The necessary data has been collected but was not analysed yet. This analysis could give insight whether some subjects had bad recognition rates because of their low general ability to recognize dialects or whether it has to be concluded that German dialects do not provide very salient prosodic differences.

Nevertheless the recognition patterns show some regularities which allow some careful conclusions. Three regions (B, F, I) were comparatively well recognized, even if it was taken into account that the respective categories were more often assigned to the items than the others. Furthermore, also the false judgements show some interesting patterns, especially for the well recognized regions.

Region "F" was the best recognized region for hits and matches. It was often confused with region "I", and seldom confused with region "E" or "B". That means that the subjects seldom assigned "e" or "b" to an item from region "F", no matter whether one refers to misses or to false alarms. Furthermore, the subjects seldom assigned "f" to an item from region "E" or "B". Again this holds for misses and false alarms. The situation is a little bit more complicated for "F" and "C". Although the subjects seldom assigned "c" to region "F", they relatively often showed false alarms and assigned "f" to "C". On the one hand this reflects the bias of the subjects in this sample to assign "f" more often than "c", on the other hand this could be a clue that prosodic similarities between Region "F" and "C" exist.

It can be stated for region B that "i", "f" (see above) and "g" was seldom assigned to it. This holds for misses as well as for false alarms. Frequently region "B" was confused with "A", "a" assignments were the second greatest for the "B" region. As a whole, most "a" judgements were false alarms and were assigned to items from region "B". "b" judgements were seldom false alarms for "F", "G" and "I", but it is striking that there were more "b" misses for "C" and "E" than correct matches and "b" misses had the second greatest value for "A". It can be suggested that "b" was some kind of "trash category", which nevertheless was still selective: It was seldom used for the southern dialectal regions "I", "F" and "G". The "trash category" hypothesis is also confirmed by the fact that "b" was the most frequently assigned category, which may be a little bit surprising as almost all of the subjects came from the southern part of Germany (especially "I") and one could suppose that they tend to choose southern regions (what is partly true as "f" and "i" were also assigned frequently, and "c" and "e" were assigned comparatively seldom).

Region "E" had relatively bad recognition rates and was more often judged as "b" or "c". This is remarkable as the Thuringian and Saxonian dialects are normally quite salient and easy to recognize. We suppose for the moment that this doesn't hold for prosodic features, which only seem to show some "northern" characteristics.

Region “G” was more frequently assigned as “i” than as “g”, if misses and matches are taken into account. The hits and false alarms show a different pattern. Region “G” got the most “g” judgements overall. From this point of view, region “G” was even better recognized than region “I”. This can lead to the assumption that the recognition rate for “G” was slightly “masked” by the great number of overall “i” judgements. The relatively clear north - south differentiation is again slightly broken up by relatively frequent “g” false alarms for region “C”. It may be that certain characteristics of the items from region “C” led some subjects to the conclusion that they deal with a southern dialect, but then also “i” judgements should show a comparable pattern, which is not the case. Region “I” and “C” were seldom confused under all circumstances. So there is no explanation for this fact so far.

Region “I” showed a similar pattern as region “F”, but was not recognized as well as this category. But it also was seldom confused with the northern categories “C” and “B”. This confirms the hypothesis that northern and southern German dialects can be told apart by prosodic features to a certain extend.

Region “A” had the worst matches. It was often confused with “B” and “E”. Besides, category “a” showed often false alarm and was assigned to wrong (especially northern) categories. This region should be investigated further to find out whether the “tendency to the north and east” intensifies.

6. CONCLUSION

It can be stated that it seems to be possible to recognize at least some German dialectal regions by their prosodic features. Maybe the clearest picture shows up for north-west German dialects and south-west German and Swiss German dialects. Speakers from this regions had relatively high recognition rates and were seldom confused. Therefore the hypothesis can be stated that these regions differ clearly in their prosodic characteristics. Also for the Austrian, Bavarian and East Franconian dialects it can be stated that they were seldom confused with northern dialectal regions. This may lead to the hypothesis that in general southern dialects differ in their prosodic patterns from northern dialects. This hypothesis is a little bit weakened by the role of East Low German, which was relatively often confused with Alemannic.

The investigation will be continued to find out whether the tendencies found so far can be confirmed. Furthermore, there are many interesting questions concerning the data which were not mentioned until now. For example, some items showed much higher recognition rates than the others. These items should be investigated with regard to their characteristics. Also it was not checked whether there are different recognition rates for male and female speakers. Maybe the most important question will be which prosodic features constitute the differences of the German dialects and how they can be described. Finally, also some special characteristics of the data should be mentioned. It is not entirely clear whether typical items could be found for all dialectal regions. For example the RVGI corpus contains relatively few speakers from the East Low German (“C”) area and most of the speakers came from Berlin and surrounding areas. This is also reflected in the speech items used for this experiment. Features like these have to be taken into account for further examination of the data.

ACKNOWLEDGEMENTS

We would like to thank our subjects for their patience while spending about one and a half hours of their time listening to and judging strange sounding signals.

REFERENCES

[1] Burger, S. and Schiel, F. 1998. RVG 1 - A Database for Regional Variants of Contemporary German. In Rubio, A. et al. (Ed.), *Proceedings of the First International Conference on Language Ressources and Evaluation 1998*, Granada, Spain.

[2] Gibbon, D. *Intonation in German.* Retrieved 1998 from the World Wide Web: **http://coral.lili.uni-bielefeld.de/~gibbon/Hirst96/german96/**

[3] Heike, G. 1983. Suprasegmentale dialektspezifische Eigenschaften. Überblick und Forschungsbericht. In Besch, W. et al. (Eds.), *Dialektologie. Ein Handbuch zur deutschen und allgemeinen Dialektforschung*. Band 1.2. Berlin: de Gruyter

[4] Lehiste, I. 1979. Perception of Sentence and Paragraph Boundaries. In Lindblom, B. and Öhman, S. (Eds.), *Frontiers of Speech Communication Research.* London: Academic Press

[5] Newton, G. 1990. Central Franconian. In Russ, C.V.J. (Ed.), *The Dialects of Modern German. A Linguistic Survey.* London: Routledge

[6] Strom, V. and Widera, C. 1996. What’s in the Pure Prosody? In Bunnel, T.H. and Idsardi, W. (Eds.) *Proceedings of the ICSLP 1996.* Philadelphia.

[7] Swerts, M. and Geluykens, R. 1994. Prosody as a Marker of Information Flow in Spoken Discourse. *Language and Speech*, 37(1), 21-43.