Identifying regional varieties by pitch information: A comparison of two approaches

Jörg Peters[a], Peter Gilles[b], Peter Auer[b], and Margret Selting[c]
[a]University of Nijmegen, The Netherlands
[b]University of Freiburg, Germany
[c]University of Potsdam, Germany
E-mail: j.peters@let.kun.nl, peter.gilles@germanistik.uni-freiburg.de,
peter.auer@germanistik.uni-freiburg.de, selting@rz.uni-potsdam.de

ABSTRACT

It is a commonly held belief that languages and dialects can be identified by pitch information alone. In most previous experiments, subjects were presented with pitch information as well as with limited information on amplitude and timing but not with higher-level, i.e. linguistic, information as represented by intonation contours. The question arises as to whether higher-level information may significantly enhance success rates. To evaluate both approaches, two experiments were carried out. In the first experiment, listeners were presented with isolated pitch information extracted from recordings of four varieties of German. In the second experiment, listeners were presented with intonation contours of the same varieties, which were superimposed on neutral carrier utterances. The results suggest that success rates may indeed be enhanced by using intonation contours. Moreover, the linguistic background of the listeners was found to affect performance in both tests.

1 INTRODUCTION

Several experiments have been carried out to substantiate the belief that languages or dialects can be identified by pitch information alone. In most experiments, subjects were presented with isolated pitch information and limited information on amplitude and timing. Bush [1] used low-pass filtered speech generated from samples of American English, British English, and Indian English. Atkinson [2] processed speech samples of English and Spanish by generating a pulse train that retained the frequency and amplitude of the speech signal. Ohala and Gilbert [3] processed speech samples of English, Japanese, and Cantonese by generating a triangular pulse train that retained the fundamental frequency, amplitude, and timing characteristics of the speech signal. Maidment [4] used the output of a laryngograph, i.e. a signal closely related to the original glottal waveform that was obtained from speech samples of English and French. Recently, Schaeffler and Summers [5] used low-pass filtered speech again.

Although all investigators found better-than-chance levels of language identification, the success rates reported were not remarkably high. Ohala and Gilbert [3] note that the moderate success rate of their experiment may be due to the method of converting the speech signal into a pulse train. This method may destroy crucial prosodic information like syllable or word boundaries. A similar argument may apply to the other studies mentioned, as in all experiments listeners were presented with lower-level, i.e. non-linguistic, information only.

An experiment that did include higher-level information was reported by Romano [6]. Romano tested the hypothesis that some of the prosodic differences between regional varieties of Italian persist in spoken standard Italian. Speakers from six villages of the Salento (South East of Italy) were asked to read sentences in standard Italian. According to Romano, their utterances showed significant prosodic variation but only few segmental cues indicating the dialectal background of the speakers. Listeners from the same six villages were asked to allocate the speaker of each utterance to one of the six areas of the Salento. Best recognition rates were found when listeners rated utterances of their own variety. In rating utterances from non-native varieties, recognition rates were rather low, indicating a random distribution of answers.

This approach has been refined in two experiments reported by Gilles et al. [7] and by Peters et al. [8]. In the first experiment, listeners were presented with utterances of a speaker of Standard German. Using pitch resynthesis (PSOLA), one half of these utterances were superimposed with intonation contours typical of Hamburg German. The other half retained the corresponding non-regional contours of Standard German [7, 8]. In a second experiment, regional contours of Berlin German were tested. In this case, contours from a third variety, Low Alemannic German, were included to prevent listeners from using some kind of elimination strategy, i.e. identifying the Berlin contours simply on the basis of their not being Standard German [8]. Both experiments confirmed the hypothesis that listeners are able to identify contours of Hamburg or Berlin German when compared with contours of other varieties. Both experiments also demonstrated that the rating behaviour of listeners is affected by their linguistic background. Listeners who were familiar with both the local variety of German and some non-local variety performed better than listeners who were only familiar with the local variety.

ISBN 1-876346-48-5

Despite the reasonable success rates of these experiments, which included higher-level pitch information, the question arises whether single utterances that bear regional intonation contours but lack the respective segmental cues may actually be better recognized than isolated pitch information obtained from longer stretches of speech. To evaluate both approaches, two experiments were carried out using stimuli from the same set of varieties but differing by the kind of information presented.

2 EXPERIMENT 1

Introduction. Listeners were presented with isolated pitch information extracted from recordings of four German varieties: the urban vernaculars of Dresden (DD), Duisburg (DU), Mannheim (MA), and Freiburg (FR). DD belongs to East Middle German, DU to Northern West Middle German, MA to Southern West Middle German, and FR to Western Upper German. Two hypotheses were tested: (1) Listeners are able to recognize utterances from DD, DU, MA, and FR by lower-level pitch information alone. (2) The linguistic background of the listeners affects their performance in the identification task. Listeners perform better in identifying stimuli created from their native variety than in identifying stimuli from non-native varieties.

Materials. We extracted the F_0 signal from speech samples, which were taken from spontaneous conversations between speakers of DD, DU, MA, and FR, respectively. With the help of the analysis program PRAAT (©1992-2002 P. Boersma & D. Weenink), we superimposed the F_0 signal on a schwa-like sound with a cut-off frequency at 5 kHz. As a result, we obtained humming sounds, which retained the pure pitch information of the original speech signal.

Procedure. Four speech samples were selected for each variety obtaining a total number of 16 stimuli. Each sample was about 15 seconds long and was randomly selected from monological passages of the original recordings. Listeners retrieved the stimuli from digitally stored audio files via a graphical user interface (cf. http://fips.igl.uni-freiburg.de/peter/experiment/). To minimize order effects, two different user interfaces were prepared, each presenting the stimuli in a quasi-random order. The listeners were randomly selected for using one of the two interfaces. They were asked to assign each stimulus to one of the four varieties and were allowed to listen to the stimuli as often as they desired.

Subjects. There were 51 listeners participating in the experiment, 21 being native speakers of DD and 30 being native speakers of FR. The listeners of both groups were divided about equally between both sexes with ages ranging between 19 and 34. Most of them were drawn from the student populations of the Universities of Dresden and Freiburg.

Results. Dresden listeners identified 29.7% of the stimuli, Freiburg listeners 27.6%. Only for the Dresden listeners was the overall identification rate found to be significantly above chance level, which corresponded to a recognition rate of 25% (p = 0.033, N = 310; p = 0.108, N = 479; Binominal test, one-sided, α = 5%).

A different picture emerged when we examined the recognition rates for each variety separately. Dresden listeners succeeded in recognizing the stimuli from Dresden but failed in all other conditions (DD: p = 0.005, N = 72; DU: p = 0.474, N = 81; MA: p = 0.084, N = 77; FR: p = 0.349, N = 80; Binominal test, one-sided). On the other hand, Freiburg listeners succeeded in recognizing the stimuli from Freiburg but failed in all other conditions (DD: p = 0.123, N = 120; DU: p = 0.356, N = 119; MA: p = 0.542, N = 120; FR: p = 0.001, N = 120) (see Figure 1).

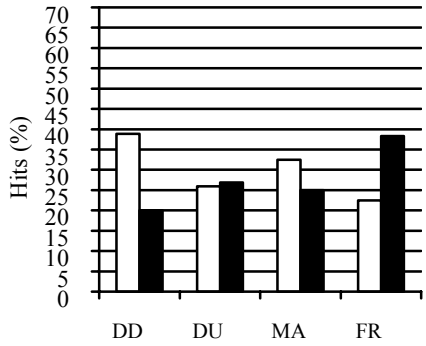

Figure 1: Recognition of humming sounds by listeners from Dresden (white bars) and Freiburg (black bars).

In summary, both listener groups performed reasonably well only in rating stimuli from their native variety (Hypothesis 1). Accordingly, both groups of listeners differed clearly in the ratings of the Dresden and Freiburg stimuli (Hypothesis 2).

3 EXPERIMENT 2

Introduction. In this experiment, we examined the identification rates when listeners were presented with higher-level pitch information. The following two hypotheses were tested: (1) Listeners are able to recognize utterances that bear regional intonation contours from DD, DU, MA, and FR but provide no segmental cues to the speaker's linguistic background. (2) The linguistic background of the listeners affects their performance in the identification task. Listeners perform better in recognizing contours of their native variety than contours of non-native varieties.

Materials. Test utterances were recorded from a 38-year-old male speaker of Standard German and digitized at 22,050 Hz. Using pitch resynthesis (PSOLA), we superimposed these utterances with intonation contours, each of which was found to be typical of DD, DU, MA, or FR, respectively. In a British style analysis, these contours may roughly be characterized as in Table 1. All contours, except for SCANSION, correspond to nuclear tones of the British tradition.

ISBN 1-876346-48-5

	Contour 1	Contour 2
DD	Double Plateau	Scooped Fall
DU	Scansion	Fall-Rise-Fall
MA	Early Fall	Early Fall with Rise
FR	Rise-Level-Slump 1	Rise-Level-Slump 2

Table 1. Contours presented in experiment 2.

The DOUBLE PLATEAU of DD consists of a high nuclear accent, a plateau that stretches until the last foot, a further rise, and a second plateau (cf. [9]). The SCOOPED FALL consists of a fall that reaches the baseline only late in the utterance and may be represented by a concave curve (cf. [10], [11]). SCANSION refers to the tendency of Duisburg speakers to place high pitch accents on the stressed syllable of nearly every foot of the utterance, sometimes even on postnuclear syllables. The FALL-RISE-FALL consists of a high nuclear accent, which is followed by a falling movement, a second high accent on the last stressed syllable, and a final fall. The EARLY FALL consists of a nuclear high accent, whose falling movement starts early in the accented syllable and rapidly reaches the baseline. The EARLY FALL WITH RISE shows an additional rise to mid level at the final boundary of the intonational phrase. The RISE-LEVEL-SLUMP (a term coined by Cruttenden [12]) consists of a nuclear rising accent, a high plateau, and a fall, which starts on the last stressed syllable. The two versions are actually two variants of the same contour differing only by the position of the nuclear syllable (ultimate vs. penultimate). A more detailed description of these contours will be included in a later presentation.

Procedure. For each intonation contour, we created 3 sample utterances, obtaining a total number of 24 stimuli (8 contour variants x 3 samples). As in Experiment 1, listeners retrieved the stimuli from digitally stored audio files via a graphical user interface. Again, two different user interfaces were prepared, each presenting the stimuli in a quasi-random order. The listeners were randomly selected for using one of the two interfaces. They were asked to assign each stimulus to one of the four varieties and were allowed to listen to the stimuli as often as they desired.

Subjects. There were 61 listeners participating in the experiment, 30 being native speakers of DD and 31 being native speakers of FR. The listeners of both groups were divided about equally between both sexes with ages ranging between 19 and 34 (except for one subject who was 51). All speakers who took part in the first experiment took also part in the second experiment.

Results. In both listener groups, overall identification rates were found to be significantly higher than chance level (25%). Dresden listeners identified 37.5% of the stimuli ($p < 0.001$, $N = 720$), Freiburg listeners 49.7% ($p < 0.001$, $N = 744$; Binominal test, one-sided). The difference between the overall identification rates of both listener groups reached statistical significance ($\chi^2 = 22.247$, $p < 0.001$, $N = 1464$, two-sided).

Next, we examined the recognition rates for each variety separately. Both the Dresden listeners and the Freiburg listeners showed recognition rates above chance level for the contours of each variety (Dresden listeners: DD: $p < 0.001$, $N = 180$; DU: $p < 0.001$, $N = 180$; MA: $p = 0.035$, $N = 180$; FR: $p < 0.001$, $N = 180$; Freiburg listeners: DD: $p = 0.031$, $N = 186$; DU: $p < 0.001$, $N = 186$; MA: $p < 0.001$, $N = 186$; FR: $p < 0.001$, $N = 186$; Binominal test, one-sided) (see Figure 2).

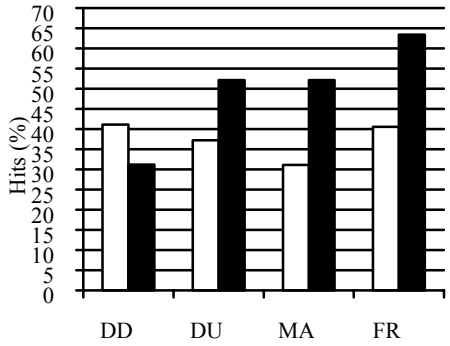

Figure 2. Recognition of intonation contours by listeners from Dresden (white bars) and Freiburg (black bars).

In addition, we found that Dresden and Freiburg listeners performed differently in rating the contours of each variety. The Dresden listeners performed better than the Freiburg listeners in recognizing the Dresden contours ($\chi^2 = 3.910$, $p = 0.048$, $N = 366$, two-sided). The Freiburg listeners performed better than the Dresden listeners in recognizing the contours of all other varieties (DU: $\chi^2 = 8.243$, $p = 0.004$, $N = 366$; MA: $\chi^2 = 16.644$, $p < 0.001$, $N = 366$; FR: $\chi^2 = 19.200$, $p < 0.001$, $N = 366$).

In summary, both listener groups recognized the contours of all varieties reasonably well (Hypothesis 1). However, Dresden listeners performed better in identifying the contours from Dresden whereas Freiburg listeners performed better in identifying the contours from Freiburg, Mannheim, and Duisburg. Thus, the linguistic background does seem to have affected the rating behaviour (Hypothesis 2).

4 COMPARISON OF BOTH EXPERIMENTS

Finally, we compared the success rates in both experiments. Both listener groups performed better in Experiment 2 than in Experiment 1 (Dresden listeners: $\chi^2 = 5.818$, $p = 0.016$, $N = 1030$; Freiburg listeners: $\chi^2 = 59.207$, $p < 0.001$, $N = 1223$; 2-tailed). The difference in performance, however, shows up more clearly in the Freiburg listeners than in the Dresden listeners (see Figure 3). The main reason for this difference may be the different performance in rating the contours of the non-native varieties DU and MA in Experiment 2 (see § 3).

ISBN 1-876346-48-5

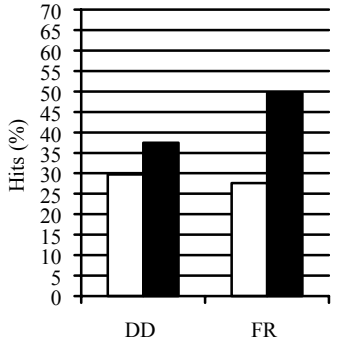

Figure 3: Recognition rates of Dresden and Freiburg listeners in Experiment 1 (white bars) and Experiment 2 (black bars).

5 CONCLUSIONS

The results of this study suggest that overall recognition rates may indeed be enhanced by using higher-level pitch information as represented by intonation contours. The choice of higher vs. lower level information, however, did not uniformly enhance recognition rates. Experiment 1 showed that in rating stimuli from a native variety pure pitch information still provides reasonable success rates. The Dresden listeners even recognized stimuli of their native variety about equally well under both experimental conditions (cf. Figures 1 and 2, leftmost white bars). Likewise, the Freiburg listeners recognized stimuli of their native variety well under both conditions but their recognition rates differed from the recognition rates of the Dresden listeners in two respects. First, they showed better results in recognizing stimuli of their native variety if presented with contours instead of pure pitch information. Second, they performed better than the Dresden listeners in the recognition of Duisburg and Mannheim stimuli when presented with intonation contours but failed to do so when presented with pure pitch information. A possible explanation might be that the Freiburg listeners were better acquainted with the varieties of Duisburg and Mannheim than Dresden listeners, due to recent political history or different cultural affiliation. According to the findings reported in [8], this could also explain why the Freiburg listeners performed better than the Dresden listeners in recognizing their own variety. In [8] it was shown that listeners who were familiar with both the local variety and some non-local variety performed better than listeners who were familiar with the local variety only (cf. § 1). Interestingly, this factor did not have any effect when listeners were only presented with lower-level information.

ACKNOWLEDGEMENTS

The research reported here is part of a project on intonational variation in German urban vernaculars at the Universities of Potsdam and Freiburg, supported by the *German Research Foundation* (DFG) under project No. SE 699/3-4 and AU 72/12-3.

REFERENCES

[1] C. Bush, "Some acoustic parameters of speech and their relationships to the perception of dialect differences," *TESOL Quarterly, 1*, pp. 20-30. 1967.

[2] K. Atkinson, "Language identification from non-segmental cues," *Working papers in Phonetics (UCLA), 10*, pp. 85-89, 1968.

[3] J. J. Ohala and J. B. Gilbert, "Listeners' ability to identify languages by their prosody," in *Problèmes de Prosodie, II: Expérimentations, modèles et fonctions*, P. R. Léon and M. Rossi, Eds., pp. 123-131. Ottawa: Didier, 1981.

[4] J. A. Maidment, "Language recognition and prosody: Further evidence," *Speech, Hearing and Language: Work in Progress U. C. L. No. 1*, pp. 133-141, 1983.

[5] F. Schaeffler and R. Summers, "Recognizing German dialects by prosodic features alone," in *Proceedings of the XIVth International Congress of Phonetic Sciences. San Francisco, August 1-7*, pp. 2311-2314, 1999.

[6] A. Romano, "Persistance of prosodic features between dialectal and standard Italian utterances in six sub-varieties of a region of southern Italy (Salento)," in *Eurospeech '97 Proceedings. ESCA 5th European Conference on Speech Communication and Technology, Rhodos, 1997*, G. Kokkinakis, N. Fakotakis, and E. Dermatas, Eds., pp. 175-178, 1997.

[7] P. Gilles, J. Peters, P. Auer, and M. Selting, "Perzeptuelle Identifikation regional markierter Tonhöhenverläufe: Ergebnisse einer Pilotstudie zum Berlinischen und Hamburgischen," *Zeitschrift für Dialektologie und Linguistik, 68*, pp. 155-172, 2001.

[8] J. Peters, P. Gilles, P. Auer, and M. Selting, „Identification of regional varieties by intonational cues. An experimental study on Hamburg and Berlin German". *Language and Speech,45(2),* pp. 115-139, 2002.

[9] M. Selting, "Dresdner Intonation: Treppenkonturen", *InLiSt No. 28*, August 2002.

[10] M. Selting, "Dresdner Intonation: Fallbögen, *InLiSt No. 29*, August 2002.

[11] I. Gericke, "Die Intonation der Leipziger Umgangssprache," *Zeitschrift für Phonetik, Sprachwissenschaft und Kommunikationswissenschaft, 16 (4),* pp. 337-369, 1963.

[12] A. Cruttenden, "Mancunian intonation and intonational representation", *Phonetica, 58*, pp. 53-80, 2001.