Identification of Regional Varieties by Intonational Cues. 
An Experimental Study on Hamburg and Berlin German*

Jörg Peters¹, Peter Gilles², Peter Auer², Margret Selting³

¹University of Nijmegen, NL
²University of Freiburg, FRG
³University of Potsdam, FRG

Abstract

Two experiments examined the commonly held belief that regional varieties of German can be identified by intonational features alone. In both experiments, listeners were presented with regional intonational contours of German. In the first experiment, listeners judged contours of Hamburg urban vernacular compared with contours of Northern Standard German. In the second experiment, listeners judged contours of Berlin urban vernacular compared with contours of both Northern Standard German and Low Alemannic German. The performance of listeners was found to vary with their linguistic experience. Listeners who were familiar both with the local variety and with some nonlocal variety by personal contact performed better than listeners who were familiar with the local variety only. Moreover, also listeners not familiar with Hamburg German and Berlin German, respectively, were found to perform the identification test with some success. This led to the conclusion that overall success rates do not only depend on true recognition of local contours but may additionally be enhanced by using some kind of elimination strategy. A second factor that affected performance was the choice of speaker for generating the carrier utterances. In the first experiment, all carrier utterances were produced by a speaker of Northern Standard German. In the second experiment, two sets of carrier utterances were used. The first set was obtained from a speaker of Northern Standard German and the second set from a speaker of Berlin urban vernacular. As expected, Berlin contours were better identified when presented with an utterance that was produced by a speaker of Berlin urban vernacular. However, no uniform effect was found for the different contours that were examined.

* Acknowledgments: The research reported here was supported by the German Research Foundation (DFG) through grant no. Se 699/3-1 and Au 72/12-1. We are grateful to Frank Kügler and Tobias Streck for collecting part of the experimental data.

Address for correspondence: Jörg Peters, University of Nijmegen, Centre for Language Studies, Postbus 9103, 6500 HD Nijmegen, The Netherlands; e-mail: <j.peters@let.kun.nl>.
1 Introduction

It is a common belief that intonation plays a prominent role in the identification of languages and dialects. Adolf Socin, a 19th century linguist, even stated that speakers of different German varieties, like the Palatinate or Bavarian dialect, could be distinguished by intonational features alone (Socin, 1888, p. 485).

Later research has used a variety of techniques to assess the role of intonation for language identification. Bush (1967) used low-pass filtered speech generated from samples of American English, British English, and Indian English. Atkinson (1968) processed speech samples of English and Spanish by generating a pulse train that retained the frequency and amplitude of the original speech signal. Ohala and Gilbert (1981) 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 (1983) used the output of a laryngograph, that is, a signal closely related to the original glottal waveform, that was obtained from speech samples of English and French. In a recent study on the identification of regional varieties of German, Schaeffler and Summers (1999) 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 (1981, p. 130) 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 more generally. In the experiments mentioned above, listeners relied on low-level information. The stimuli presented involved pitch information and limited information on amplitude and timing but no higher level information, for example information on the linking between pitch and segmental or syntactical structure. In this respect, the studies mentioned above are in line with the aims of research on automatic language identification (cf. Caseiro & Trancoso, 1998; Leavers, Wiehler, & Burley, 1996; and Thymé-Gobbel & Hutchins, 1996).

From the linguist’s point of view, language identification by processing of low-level information seems to be less interesting. Even if we come to know that a pair of languages or dialects may be distinguished by listening to isolated pitch information, we learn little about differences of their intonational systems. Without information about the relationship between observed pitch patterns and segmental structure, listeners are not even able to identify basic intonational units. The question arises whether experiments of this sort have anything to offer to those who seek to assess the role of intonation proper in language identification, that is, how pitch relates to segmentally based linguistic structure.

An experiment using an alternative method was reported by Romano (1997). Romano tested the hypothesis that some of the prosodic differences between regional varieties of Italian persist in spoken standard Italian. To verify this hypothesis speakers from six villages of the Salento (South East of Italy) were asked to read sentences in standard Italian. According to Romano, the resulting utterances were characterized by prosodic variation but by an almost total absence of segmental cues. 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 in judgments of \textit{wh}-questions, exclamations, and emphatic patterns. This was especially true of listeners judging utterances of their own variety. In judging utterances from non-native varieties recognition rates were rather low indicating a random distribution of answers. The likewise moderate success of this experiment may be due to the fact that listeners were asked to identify speakers who came from a rather restricted area of Southern Italy. This study, however, is remarkable in that it includes higher-level information in the task of language identification and thus considers the role of intonation proper rather than the role of isolated pitch information.

The study reported here is intended to contribute to this line of research in several respects. First, it provides evidence for the identification of a different set of linguistic varieties, that is, Hamburg urban vernacular (HUV) and Berlin urban vernacular (BUV). Secondly, non-intonational features of the carrier sentences were controlled by use of pitch resynthesis (PSOLA). This procedure enabled us to combine a single utterance with different pitch contours instead of using more or less exact repetitions of utterances in natural speech. Third, we examined the linguistic background of the subjects in the identification task as an additional factor, which may affect their performance. Ohala and Gilbert (1981), Maidment (1983), and Romano (1997) all note the importance of linguistic experience for the outcome of their experiments. In particular, familiarity with all varieties in question was found to enhance success rates. Finally, we asked whether the choice of the carrier voice may affect success rates. Romano (1997) presented carrier sentences that were produced by local speakers using a standard Italian pronunciation to minimize the influence of non-intonational features. In the experiment on BUV contours, we examined possible effects of non-intonational features by presenting listeners with carrier sentences that were obtained both from a speaker of Northern Standard German (NSG) and a speaker of BUV. In the next two sections both experiments are reported. A final section provides a general discussion.

2 Experiment 1

In Experiment 1, we examined the identification of single contours found in HUV. Specifically, we wanted to test two hypotheses: (1) Listeners are able to identify local pitch contours of HUV compared with those of NSG. (2) The linguistic background of listeners affects their performance in the identification task. Natives of Hamburg perform better than non-native residents of Hamburg.

2.1 Method

Materials. Three contours that were found to be frequently used in Hamburg and surrounding areas served as examples for creating a set of target sentences. We obtained these contours from previous analyses of spontaneous conversations between natives of Hamburg (cf. the work cited below). For comparison, we chose contours of NSG that were judged by the authors to be functionally equivalent, that is, they were judged to have comparable functions when used in natural conversations in NSG. Figure 1 illustrates the HUV and NSG contours on the left and right panel, respectively. Figure 2
gives F0 traces of corresponding utterances used in the experiment. For contour I, two variants are represented, which were tested separately (Contours Ia and Ib). For Contour III, two realizational variants are displayed, which were not differentiated in the experiment. All representations are labeled along the guidelines of GToBI (Grice & Baumann, in press).

**Figure 1**


<table>
<thead>
<tr>
<th>Contour</th>
<th>HUV</th>
<th>NSG</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ia</td>
<td><img src="image" alt="HUV Contour Ia" /></td>
<td><img src="image" alt="NSG Contour Ia" /></td>
</tr>
<tr>
<td>Ib</td>
<td><img src="image" alt="HUV Contour Ib" /></td>
<td><img src="image" alt="NSG Contour Ib" /></td>
</tr>
<tr>
<td>II</td>
<td><img src="image" alt="HUV Contour II" /></td>
<td><img src="image" alt="NSG Contour II" /></td>
</tr>
<tr>
<td>III</td>
<td><img src="image" alt="HUV Contour III" /></td>
<td><img src="image" alt="NSG Contour III" /></td>
</tr>
</tbody>
</table>
Figure 2
F0 traces of HUV contours and their NSG counterparts (one example per contour)
Contour I of HUV has been described in detail by Gilles (2001). In terms of the British school of intonational analysis, Contour I involves a nuclear tone that consists of a fall-rise. In GToBI, Contour I can be characterized by $H^* L - H^\%$ with $H^*$ being linked to the nuclear syllable. The contour indicates continuation of the present discourse topic.

The regional character of Contour I of HUV seems to be due to the tune-text association rather than to the tonal structure itself. HUV tends to realize the commonly found $H^* L - H^\%$ sequence without truncation even if only little voiced material follows the nuclear syllable. Contour Ia represents the most salient variant with $H^*$ being linked to the last syllable of the intonational phrase (see Fig. 1, first row, left panel). In this case, additional lengthening of the stressed syllable allows the speaker to realize the falling-rising movement on a single syllable. Contour Ib of HUV represents a less salient variant with $H^*$ being associated with the penultimate syllable of the intonational phrase (see Fig. 1, second row, left panel). In this case, the falling-rising movement of F0 is compressed on the last two syllables.

The NSG counterparts of Contour Ia and Ib were described by Ladd (1996, p. 134). In NSG, the sequence $H^* L - H^\%$ is often found to be replaced by the sequence $L^* H - %$ when little segmental material is available (see Fig. 1, first and second row, right panels). In this way, NSG speakers manage to realize the final rise without compressing the nuclear pitch contour on one or two syllables like in HUV.

Contour II, described by Auer (2001), is characterized by an extra-high falling pitch movement at the beginning of the utterance (see Fig. 1, third row, left panel). In terms of a British style analysis, it can be characterized as consisting of an extra-high prehead, a low onset syllable, a rising head, and a falling nuclear tone. In GToBI, Contour II can be represented by $%^H L^* (H^*) H^\% L - %$, with $'(H^*)'$ indicating one or more optional H-tones intervening between the onset and the nuclear syllable. According to Auer (2001), Contour II is used to formulate the upshot or conclusion of more complex turns. As the NSG counterpart of $%^H L^* (H^*) H^\% L - %$ we chose $H^* H^* (H^*) L - %$ (see Fig. 1, third row, right panel). Like Contour II of HUV, this contour is found at the end of complex turns in NSG.

Contour III of HUV represents another high-falling contour that is frequently used (see Fig. 1, upper and lower half of row 4, left panel). The contour comes in two variants. Both variants of the contour start with an extra-high tone on the first accented syllable. If the utterance contains one or more prenuclear accented syllables the extra-high tone occurs on the first of these syllables. In a British style analysis this syllable is the “onset” (the first syllable of the head). If the utterance lacks prenuclear accents the extra-high tone occurs on the nuclear syllable. In GToBI, the variants of Contour III can be characterized by $^H^* L - H^\%$ (with $^H^*$ on the nuclear syllable) and $^H^* L^* H - H^\%$ (with $^H^*$ on the onset syllable), respectively. According to Auer (2001, p. 142ff), Contour III reflects the general tendency of Hamburg speakers to start with a high pitch. The occurrence of Contour III, however, seems to be syntactically restricted. In contrast to Contour II of HUV, Contour III was exclusively found in V1-sentences (yes/no questions and declaratives with an empty subject position). As counterparts in NSG we chose contours that differ only by starting with a high instead of an extra-high tone on the first accented syllable (see Fig. 1, upper and lower half of row 4, right panel).
To create a first set of target sentences we selected sentences from the above-mentioned conversations. Written versions of these sentences were read by a 25 year-old female speaker of NSG (see Appendix A). We generated two variants of each sentence by combining it with a local and a nonlocal contour. The local and nonlocal variants were obtained from the set of contours of HUV and NSG represented in Figure 1. For generating the variants we used pitch resynthesis (PSOLA). The PSOLA technique allowed us to manipulate F0 values without changing spectral characteristics of the utterance. Thus, we were able to construct two variants of a single utterance differing by F0 characteristics only.

We recorded three carrier sentences for each trial (identification of HUV-Contours Ia, Ib, II, and III when compared with their NSG counterparts). Generating two variants of each sentence, we obtained a total number of 24 stimuli (4 trials × 3 examples × 2 variants). For statistical analysis, the scores obtained for each variant (HUV contour vs. NSG contour) were averaged over the three examples per trial and subject.

Procedure. We carried out two tests, a discrimination task and an identification task. The discrimination task was intended to make sure that subjects were able to discriminate between the contours at all. A tape was prepared that presented each pair of sentences with local and nonlocal contours twice in random order. The listeners were asked to decide for each pair of sentences whether the sentences were same or different. The test session was preceded by a short training session in which the listeners were told whether their judgments were true or false.

The identification task was intended to reveal whether listeners were able to recognize which pitch contour pertained to HUV when presented both with an utterance bearing a HUV contour and an utterance bearing a NSG contour. We presented each pair of utterances with a transcript of the context as found in the original recording of the conversation. The transcript contained a foregoing context and, in some cases, a following context as well, comprising 2–11 intonational phrases. We decided to present both variants pairwise to avoid giving the extensive context information twice. In that case, the judgment of the second variant could have been affected due to a better familiarity of the listener with the context information that would have been given a second time.

Subjects were asked to begin each trial by reading the transcript. Then, both variants of the sentence were presented pairwise, with one repetition. After listening to the stimuli, subjects stopped the tape and compared both variants with respect to their local character, using a seven-point rating scale. The scale ranged from “sounds by no means like Hamburg German” to “sounds very much like Hamburg German.” Subjects were allowed to rate both variants alike but not to omit single ratings.

Two tapes were prepared presenting the stimuli in different order. Subjects were randomly selected for listening to one of the tapes. Two pairs of sentences were used for a training session, in which listeners could acquaint themselves with the task, but no feedback was given. The recordings of the carrier sentences were made on digital audio tape (DAT). The resynthesized versions of the sentences were played back from a tape recorder.

1 Pitch resynthesis was carried out with the help of the analysis program PRAAT, developed at the University of Amsterdam (Paul Boersma & David Weenink).
Subjects. Thirty-four listeners took part in the experiment. On the basis of a questionnaire, which included detailed questions on linguistic background and places of residence through lifetime, the subjects were divided over two groups of 17 listeners each. The listeners of the first group were classified as natives of Hamburg (= Natives). They were born in Hamburg and had been living there all of their lives (except for holiday trips and the like). The listeners of the second group were classified as non-native residents of Hamburg. They were staying in Hamburg when participating in the experiment and had been living there for more than two years but were born in other regions of Germany (= Non-Native Residents). Subjects who were born in Hamburg but left this place afterwards for longer periods were not considered. The listeners of both groups were about equally divided over both sexes with ages ranging between 19 and 39. Most of them were drawn from the student population of the University of Hamburg and were paid for their services.

2.2 Results

Discrimination task. Listeners discriminated more than 90% of the pairs of target sentences. No significant difference was found between the performance of Natives and Non-Native Residents. Most errors were made in discriminating between utterances bearing Contour III and their counterparts in NSG. Nevertheless, we included all scores that we had obtained in the identification task into the analysis.

Identification task. Four separate analyses of variance (repeated measures) were performed on the scores that were obtained in each trial. Independent variables were Listener Group (Native vs. Non-Native Resident of Hamburg) and Contour (HU V contour vs. NSG contour). The results are presented in Figure 3.

Significant main effects for Contour were found for all conditions. No main effect was found for Listener Group. Significant Contour × Listener Group interactions were found for all contours except for Contour Ia. Table 1 summarizes the results of these statistical tests.

Inspection of the diagrams in Figure 3 suggests that the interaction between Contour and Listener Group for Contour Ib, Contour II, and Contour III is due to the different effects of linguistic background on the rating behavior. Non-Native Residents appear to rate the Hamburg contour higher than the NSG contour under each condition.

2 In an informal presentation of the results of Experiment 1 by Gilles, Peters, Auer, and Selting (2001), data from 41 subjects were included in the analysis. For balancing the groups of subjects, data of seven randomly selected subjects were excluded from analysis. One subject, formerly classified as native, was reclassified as resident after further inspection of the linguistic background.

3 In Gilles et al. (2001), unreliable scores were excluded from analysis. Both analyses yielded roughly the same results.

4 Note that rating scales were reversed for easier comparison with the data presented in Experiment 2. Originally, lower scores, rather than higher scores, served to indicate “closeness” to HUV.
whereas *Natives* fail to distinguish between both contours except for Contour Ia. These observations were confirmed by a post hoc test (paired *t*-test) on CONTOUR for each group of listeners. The differences between the ratings of the Hamburg contour and the NSG contour by *Non-Native Residents* were significant under all conditions. The differences between the corresponding ratings by *Natives* were significant for Contour Ia but not for Contour Ib, II, and III. Table 2 summarizes the results of these tests.


**TABLE 1**

Results of analyses of variance (repeated measures). *F*’s are given if *p* < .05

<table>
<thead>
<tr>
<th>Listener group</th>
<th>Contour</th>
<th>Listener group × contour</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><em>F</em>(1,32)</td>
<td><em>F</em>(1,32)</td>
</tr>
<tr>
<td>Contour Ia</td>
<td>n.s.</td>
<td>41.32</td>
</tr>
<tr>
<td>Contour Ib</td>
<td>n.s.</td>
<td>14.07</td>
</tr>
<tr>
<td>Contour II</td>
<td>n.s.</td>
<td>36.22</td>
</tr>
<tr>
<td>Contour III</td>
<td>n.s.</td>
<td>17.91</td>
</tr>
</tbody>
</table>

**TABLE 2**

Results of post hoc tests (paired *t*-tests, 2-tailed) on groups of listeners by levels of CONTOUR

<table>
<thead>
<tr>
<th>Listener group</th>
<th>Contour of target sentence</th>
<th><em>p</em> &gt; .05</th>
</tr>
</thead>
<tbody>
<tr>
<td>Native</td>
<td>Hamburg contour versus NSG contour</td>
<td>Contour Ia</td>
</tr>
<tr>
<td>Resident</td>
<td>Hamburg contour versus NSG contour</td>
<td>Contour Ia, Ib, II, III</td>
</tr>
</tbody>
</table>

### 2.3 Discussion

The main effect of CONTOUR confirms the hypothesis that listeners are able to identify HUV contours when presented both with HUV contours and contours of NSG (Hypothesis 1). The interaction found between CONTOUR and LISTENER GROUP for Contour Ib, II, and III confirms the hypothesis that the linguistic background of the listeners affects their performance (Hypothesis 2). Therefore, the main effect of CONTOUR must not be interpreted as a global effect. However, the direction of the interaction was not as we had expected. Non-Native Residents performed better in the identification task than Natives did and not vice versa. The post hoc tests revealed that only Non-Native Residents identified all four HUV contours. Natives failed to identify HUV contours except for Contour Ia.

This result is in need of explanation, as at first sight it may seem plausible that listeners familiar with HUV from birth would perform better in distinguishing between local and nonlocal variants of their mother tongue. On the other hand, Non-Native Residents of Hamburg may be better equipped to identify local forms of HUV compared with variants of NSG as they had been exposed both to HUV and to other local varieties of German. When moving to Hamburg, they may have had the opportunity to develop a more or less conscious sensitivity for differences between their own vernacular and the one spoken in Hamburg. Thus, the results of the experiment are fully consistent with the finding of previous studies mentioned in the Introduction that familiarity with all varieties that were included in the task was found to enhance success rates.

However, there is an alternative explanation. Non-Native Residents could have performed better than Natives not by true recognition but simply by excluding the NSG contours with which they might have been more familiar than the Natives of Hamburg.
Preliminary evidence for this conclusion comes from an additional experiment that we carried out with subjects who neither were born in Hamburg nor were residents of Hamburg ( = Non-Residents). In this Experiment 33 volunteers participated in a slightly modified version of the identification task via the internet. The results may be summarized as follows: Non-Hamburgers were about as successful in identifying the local variants of Contour I as Non-Native Residents. In identifying the local variants of Contour II, they performed worse than Non-Native Residents but they performed still better than Natives. These results suggest that some kind of elimination strategy was used to rate HUV contours higher than the NSG contours rather than a special knowledge of Hamburg intonation.

In Experiment 2, we carried out a similar identification task on Berlin intonation but changed the test design such that it would become easier to determine which strategy was actually used by the listeners.

3 Experiment 2

A second experiment was designed to find out whether contours of BUV could be identified as well. Experiment 2 differs from Experiment 1 in two respects:

First, in order to prevent listeners from identifying local contours simply by excluding NSG contours as possible candidates they were asked to compare three rather than two variants of a given contour. In addition to a variant of BUV and of NSG, we presented a functionally equivalent variant of each contour typical of Low Alemannic German (LAG). Secondly, we examined the influence of voice quality and other factors that did not vary with the experimental condition. For this purpose, we generated the test items from utterances that were produced both by a speaker of NSG and by a speaker of BUV.

Experiment 2 served to test three hypotheses: (1) Listeners are able to identify contours of BUV when compared with contours of NSG and LAG. (2) The linguistic background of the subjects affects their performance in the identification task. Listeners who have been exposed to large amounts of both BUV and some non-BUV variety perform better than listeners who do not meet this requirement. (3) BUV contours superimposed on utterances by a BUV speaker are better identified than those superimposed on utterances by a NSG speaker.

3.1 Method

Materials. Three contours that were found to be frequently used in Berlin and surrounding areas served as examples for creating a set of target sentences. As in Experiment 1, we used 10 spontaneous conversations between natives as a data source. Figure 4 gives stylized illustrations for each contour. Figure 5 gives F0 traces of corresponding utterances used in the experiment.

---

5 The test is available at <http://fips.igl.uni-freiburg.de/~peter/webtest>. See Gilles et al. (2001) for details.
Contour I of BUV was described by Peters (2001). The nuclear syllable bears a high accent that is normally realized with a preceding valley in the first half of the syllable (see Fig. 4, first row, left panel). In terms of a British style analysis, it involves a rise-fall followed by a low pitch. In GToBI, Contour I may be characterized by L* + H and a terminal fall (L* + H L–%). Whereas contours with a L* + H accent and a terminal fall seem to be common in many Upper German dialects (cf. Gibbon, 1998; Fitzpatrick-Cole, 1999), they rarely occur in Northern Germany. In BUV, this contour seems to be restricted to the expression of contrastive meanings (see Peters 2001 for different subtypes of this contour). As counterparts in NSG and LAG we chose H* + L L–% and L* + H H + L(∗) L–%, respectively (see Fig. 4, first row, middle and right panel). H* + L L–% is supposed to bear a contrastive meaning in NSG, even if evidence from systematic analyses is still missing. The LAG contour, L* + H H + L(∗) L–%, like all LAG contours used in the experiment, was attested in recordings made of a 25 year-old speaker from Offenburg, placed in the Low Alemannic area. This contour is assumed to contain both a bitonal accent in nuclear position and a bitonal phrase accent that is secondarily associated with the last stressed syllable.6

---

6 See Grice, D’Imperio, Savino, and Avesani (in press) and Peters (2001a) for earlier proposals of bitonal phrase accents. The use of a parenthesized star (*) to indicate the tone that is associated with a TBU has been adopted from Grice and Baumann (in press).
Figure 5
F0 traces of HUV contours and their NSG and LAG counterparts (one example per contour)
Contour II was described by Selting (2000). The contour has a high nuclear accent that is realized with a delayed peak, that is, the peak only occurs on the first unstressed syllable after the nuclear syllable (see Fig. 4, second row, left panel). Moreover, the nuclear accent is realized with a rising pitch throughout the nuclear syllable rather than with a valley prior to the rising movement like in Contour I. In GToBI, utterances with Contour II can be characterized by $H^*<L-%$ (with $'H<$ indicating the delayed peak). Contour II is used to suggest a more restricted utterance interpretation than the default contour for declarative utterances, which is $H^*L-%$ in BUV. According to Selting (2000), the contour contextualizes subjective evaluations or attitudes, which suggest that the violation of a social or cultural norm is being “taken lightly.” As counterparts in NSG and LAG we chose $H^*L-%$ and $L^*+H^{(*)}+L L-%$, respectively (see Fig. 4, row 2, middle and right panel). $H^*L-%$ is commonly assumed to be the default contour for declarative utterances in NSG. We selected this contour as we currently do not know of any other contour of NSG that may be used in a situation in which Berlin speakers would use $H^*<L-%$. The contour $L^*+H^{(*)}+L L-%$ (right panel) was attested by the LAG speaker under the same conditions as the BUV contour.

Contour III was described by Selting (2000a). This contour consists of a rise on the nuclear syllable followed by a high pitch plateau (see Fig. 4, third row, left panel). In GToBI, it can be characterized by $H^*H-%$. A second variant noted by Selting (2000a) shows a delayed rising movement in the nuclear syllable and may be characterized by $L^*+H H-%$. The most salient form of Contour III shows no declination trend after the nuclear syllable. In the recognition test, we considered the first variant ($H^*H-%$) only. This variant tends to be used consecutively in long series with up to 10 or even more items. Whereas high plateaus may be found in other varieties of German as well (cf. Grabe 1998), Berlin speakers stand out by using this contour more frequently and in longer series. The $H^*H-%$ variant is used in “open lists” and for presenting a series of actions or events in biographical story telling as recurrent, routine, and therefore expectable. In general, the contour projects more to come. All items of a list can bear Contour III, including the last one. In the present experiment, complete lists were presented consisting of three or four items, respectively. As counterparts in NSG and LAG we chose series of $H^*L-H%$ and $L^*+H L-%$ contours, respectively (see Fig. 4, third row, middle and right panel). In $H^*L-H%$, the phrase accent is secondarily associated with the first postnuclear stressed syllable and the final rise starts only in the last syllable. In $L^*+H L-%$, which is attested by the Offenburg speaker, the L– phrase accent is associated with the final phrase boundary only. Accordingly, F0 reaches the baseline only at the end of the utterance.

Taking original utterances from recorded conversations between Berlin speakers as a model, we created two sets of target utterances. A first set was obtained by recording the carrier utterances from a 30-year-old male speaker who was a native of Berlin (= BUV Carrier Voice). A second set was obtained by recording the carrier utterances from a 37-year-old male speaker of NSG (= NSG Carrier Voice). The BUV utterances differed from the NSG utterances by segmental features, rhythm, and voice quality. Segmentally, BUV utterances differed from NSG utterances mainly by spirantization of /g/, replacement of the personal pronoun ich by ick, and monophthongization of diphthongs in particular lexemes of BUV. These differences are illustrated in Appendix B, where literary transcriptions of both NSG and BUV carrier utterances are given.
Additional differences between both sets of carrier utterances concern rhythmic properties, overall rate of speech (with NSG spoken at a slower rate), and voice quality.

Using pitch resynthesis (PSOLA), we generated three variants from each carrier utterance. The contour of the first variant was obtained from the set of BUV contours represented in Figure 4. The contours of the second and third variant were obtained by choosing functionally equivalent contours of NSG and LAG, respectively.

From each speaker, we recorded three carrier sentences for each trial (identification of Contour I, Contour II, and Contour III). Generating three variants of each sentence, we obtained a total number of 54 stimuli (2 speakers × 3 trials × 3 examples × 3 variants). For statistical analysis, scores were averaged over the three examples as described for Experiment 1.

Procedure. An identification task was designed along the lines of Experiment 1. The stimuli were retrieved from digitally stored audio files via a graphical user interface. Subjects listened to the target utterances as often as they desired. Like in Experiment 1, all sentences were presented with a foregoi ng context, and some sentences also with a following context. Listeners were asked to acquaint themselves with the context both by reading a transcript and by listening to an accompanying audio file. The audio file was extracted from the original tape recording of the conversation. After listening to all variants of a given sentence and to the context file, subjects were asked to rate each variant with respect to the local character of its contour using a seven-point scale ranging from “speech melody sounds by no means like Berlin German” to “speech melody sounds very much like Berlin German.”

The experiment was carried out in two sessions. In a first session, subjects listened to stimuli obtained from the speaker of BUV. In a second session, which took place on another day, subjects listened to stimuli obtained from the speaker of NSG. The order of the examples was varied between all subjects and sessions.

Subjects. Forty-five listeners took part in the experiment. Like in Experiment 1, the subjects were selected on the basis of a questionnaire. According to their linguistic background, they were divided over three groups of 15 listeners each. The listeners of the first group were natives of Berlin. They were born and had been living in Berlin or surrounding areas (= Natives). The listeners of the second group had been exposed to large amounts of both BUV and some other varieties of German by personal contact. Being non-natives of Berlin most of them were residents of Berlin or had been living in the surrounding area for a longer time. For the sake of uniform terminology they will all be classified as residents (= Non-native Residents). The listeners of the third group were living outside of the Berlin area and had hardly been exposed to BUV by personal contact (= Non-Residents). The listeners of each group were about equally divided over both sexes with ages ranging between 19 and 44. Most participants were drawn from the student population of the Universities of Potsdam and Freiburg and were paid for their services.

Results. Three separate analyses of variance (repeated measures) were performed on the scores that were obtained in each trial. Dependent variables were Listener Group (Native vs. Non-Native Resident vs. Non-Resident), Contour (BUV Contour vs. NSG...
Contour vs. LAG Contour), and carrier voice (BUV speaker vs. NSG speaker). The results are presented in Figure 6.

For all contours, we found a main effect of contour and carrier voice but no main effect of listener group. For Contour I and Contour II, we found also an interaction both between contour and listener group and between contour and carrier voice. In the case of Contour I and Contour II, the LAG contour was rated lowest, whereas in the case of Contour III the NSG contour achieved the lowest scores. Table 3 summarizes the results of the statistical tests.

**Table 3**
Results of analyses of variance (repeated measures). F’s are given if p < .05

<table>
<thead>
<tr>
<th>LISTENER GROUP</th>
<th>CONTOUR</th>
<th>CARRIER VOICE</th>
<th>LISTENER GROUP × CONTOUR</th>
<th>CONTOUR × CARRIER VOICE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>F(2,42)</td>
<td>F(2,84)</td>
<td>F(1,42)</td>
<td>F(4,84)</td>
</tr>
<tr>
<td>Contour I</td>
<td>n.s.</td>
<td>106,38</td>
<td>7,13</td>
<td>8,81</td>
</tr>
<tr>
<td>Contour II</td>
<td>n.s.</td>
<td>116,56</td>
<td>27,22</td>
<td>3,32</td>
</tr>
<tr>
<td>Contour III</td>
<td>n.s.</td>
<td>59,09</td>
<td>17,42</td>
<td>n.s.</td>
</tr>
</tbody>
</table>

Post hoc tests were carried out to identify the sources of interaction. A test on the rating behavior of each listener group revealed for Contour I that the interaction between contour and listener group might have two sources (see also Fig. 6, left panels). First, Natives and Non-Native Residents rated the NSG contour differently. Non-Native Residents rated the NSG contour lower than Natives, whereas no significant difference was found between their rating of the BUV contour and of the LAG contour. Secondly, both Natives and Non-Native Residents rated the BUV contour and the LAG contour differently from Non-Residents. Non-Residents rated the BUV contour lower than both Natives and Non-Native Residents. Furthermore, they rated the NSG contour lower than Natives. On the other hand, they rated the LAG contour higher than both Natives and Non-Native Residents. The interaction between contour and listener group in the case of Contour II may mainly result from differences in the rating behavior of Non-Residents versus Natives and Non-Native Residents. Non-Residents rated the BUV contour lower than Non-Native Residents and the NSG contour lower than Natives, whereas no difference was found in the evaluation of the LAG contour. Table 4 summarizes the results of these tests.
Figure 6
Scores for contours by Listener Group (left panels) and for contours by Carrier Voice (right panels). Higher scores indicate perceived closeness to BUV.
### TABLE 4

Results of post hoc tests (Scheffé tests) on LISTENER GROUP by levels of CONTOUR

<table>
<thead>
<tr>
<th>Contour of target sentence</th>
<th>p &lt; .05</th>
</tr>
</thead>
<tbody>
<tr>
<td>Contour I</td>
<td></td>
</tr>
<tr>
<td>BUV contour</td>
<td>Native versus Non-Residents</td>
</tr>
<tr>
<td>NSG contour</td>
<td>Native versus Non-native Residents</td>
</tr>
<tr>
<td>LAG contour</td>
<td>Native versus Non-Residents</td>
</tr>
<tr>
<td>Contour II</td>
<td></td>
</tr>
<tr>
<td>BUV contour</td>
<td>Non-native Residents versus Non-Residents</td>
</tr>
<tr>
<td>NSG contour</td>
<td>Native versus Non-Residents</td>
</tr>
<tr>
<td>LAG contour</td>
<td>none</td>
</tr>
<tr>
<td>Contour III</td>
<td></td>
</tr>
<tr>
<td>BUV contour</td>
<td>none</td>
</tr>
<tr>
<td>NSG contour</td>
<td>none</td>
</tr>
<tr>
<td>LAG contour</td>
<td>none</td>
</tr>
</tbody>
</table>

A post hoc test on differences in the evaluation of sentences obtained from the BUV speaker versus the NSG speaker revealed the following sources of interaction (see also Fig. 6, right panels). The interaction between CONTOUR and CARRIER VOICE for Contour I and Contour II may be due to the fact that choice of speaker affected the rating of local and nonlocal contours differently. In the case of Contour I, the choice of the Berlin speaker increased mean scores on the BUV contour but had no effect on the rating of the NSG contour or of the LAG contour. In the case of Contour II, the choice of the Berlin speaker increased the mean scores on both the BUV contour and the NSG contour but had no effect on the rating of the LAG contour. A similar tendency was found for Contour III (with reversed order in the ranking of the NSG contour and the LAG contour). In this case, however, the interaction did not reach statistical significance. Table 5 summarizes the results of these tests.

### 3.2 Discussion

The main effect of CONTOUR supports the general hypothesis that listeners are able to identify local contours of BUV when presented with BUV contours, NSG contours, and LAG contours (Hypothesis 1). Moreover, the interaction between CONTOUR and LISTENER GROUP shows that the linguistic background of listeners affects their performance in the identification task (Hypothesis 2).

The question arises, whether the higher scores for BUV contours are due to true recognition of BUV contours as part of the Berlin intonational system or to some kind of elimination strategy as suggested in the discussion of the results of Experiment 1. On the one hand, listeners who have been exposed to large amounts of BUV (Natives and Non-Native Residents) performed better than listeners who did not meet this requirement. This is true for Contour I and Contour II, at least. Thus, familiarity with BUV
seems to enhance success rates. On the other hand, in case of true recognition of BUV contours we would expect Natives and Non-Native Residents to rate the NSG and LAG contours about equally low. At least for the NSG scores on Contour I this is not true. In this case, familiarity with both BUV and some non-BUV variety results in better performance as expected from Experiment I (see also Hypothesis 2 for Experiment II). Non-Native Residents rated the NSG variant lower with respect to the BUV variant than Natives and thus showed better discrimination between BUV and NSG contours.

This result may be taken as evidence that the identification of the BUV contour (by rating it as closest to BUV) cannot fully be attributed to recognition of a well-known regional contour alone. Similar doubts may arise with respect to the strategy chosen by the Non-Residents. As they lack specific knowledge of the intonational system of BUV there is no basis for true recognition of BUV contours and therefore the BUV variants should not receive higher scores than other variants. In Contour II, however, both the BUV and NSG contour were scored significantly higher than the LAG contour. As many listeners who were classified as Non-Residents actually had been exposed to large amounts of LAG in their lifetime (most of them were recruited from Freiburg University) this result may most easily be explained by the assumption that these listeners chose an elimination strategy in excluding those contours that they were most familiar with.\footnote{This explanation does not fit Contour III where the LAG contour was scored higher than the NSG contour. This may be due to the fact that the LAG contour is more similar to the BUV contour than the NSG contour (see Fig. 4, row 3).}

The main effect of carrier voice found for Contours I–III confirms the hypothesis that the choice of a BUV carrier voice does improve overall identification rates. On average, the difference between the Berlin and the Non-Berlin scores increases when the carrier utterance shows additional features of BUV (Hypothesis 3). However, the interaction between contour and carrier voice found for Contour I and Contour II demonstrates that the carrier voice does not uniformly affect rating behavior. It appears
to improve performance by increasing scores for the BUV contours rather than by lowering scores for the non-BUV contours. This is true for the LAG contours, at least. The right panels of Figure 6 show for Contours I–III that the scores for the LAG contours remain largely unaffected by the choice of carrier voice (see also Table 5). In contrast, the results for the NSG contours are more difficult to integrate in the whole picture. In the case of Contour I, the choice of the BUV carrier voice has no effect on the scores but in the case of Contour II and Contour III the BUV carrier voice raises the scores like the scores of the BUV contours.

4 General Discussion

The research reported in this paper was motivated by the widespread belief that regional varieties of German can be identified by intonational cues alone. The experiments reported here differ from most previous research in presenting listeners with utterances bearing different intonational contours rather than with isolated pitch information. By this method, we attempted to assess dialect identification by intonation proper, that is, pitch patterns that are systematically related to segmentally based linguistic structures, rather than by its isolated acoustical (or auditory) correlates. Experiments carried out by Maidment (1983), Ohala and Gilbert (1981), Schaeffler and Summers (1999) and others have demonstrated that pure pitch information, enhanced by limited information on amplitude and timing, allows listeners to identify languages or dialects with moderate success. However, these experiments did not examine the role of intonational features taken as linguistic features of utterances in the identification process. Rather, they examined the role of global auditory features of the converted speech signal that possibly derive from differences in prosodic structure. This method can only provide indirect evidence for the significance of intonational cues for language or dialect identification.

Both experiments confirmed the hypothesis that listeners are able to identify local contours of Hamburg or Berlin when compared with contours of other varieties. Both experiments also demonstrate that the rating behavior of listeners depends on their linguistic background. Listeners who were familiar both with the local variety of German and some nonlocal variety performed better than listeners who were familiar with the local variety alone. The listeners who were classified as Non-Native Residents in both experiments may be assigned to the first group, whereas the listeners classified as Natives may be assigned to the second group.

Listeners who had not been exposed by personal contact to HUV and BUV, respectively, performed differently in both experiments. Non-Residents of Hamburg performed about equally with Residents of Hamburg in Experiment 1, whereas Non-Residents of Berlin performed worse than Non-Native Residents of Berlin in Experiment 2. This difference may be due to the fact that listeners in Experiment 2, but not in Experiment 1, were prevented from identifying the local variant simply by excluding the NSG variant, with which they were familiar, as a possible candidate. We may conclude that nativeness to the local contour alone may not be an advantage in identifying local pitch

---

8 Paralinguistic intonational features may also contribute to language or speaker identification but were left out of consideration in the present study.
contours. Rather, familiarity with both the local and some nonlocal variety appears to be crucial.

This result may be surprising as it is known from research on L2 learning that nonnatives are often insensitive to distinctive features of another language even if they have been exposed to large amounts of speech of this language. According to Labov, Karen, and Miller (1991) and Otake and Cutler (1999), this also applies to speakers of different dialects. In the present case, however, listeners were asked to choose between different varieties rather than to identify phonological categories in a single variety. If asked to tell a local variety apart from some standard variety by intonation, familiarity with a second local variety may be more important than nativeness to the local variety as it allows one to identify the local variety by using a simple elimination strategy.

Other findings point in the same direction. Ohala and Gilbert (1981) report better results in identifying samples of English, Japanese, and Cantonese for bilinguals than for monolinguals. Similarly, Maidment (1983), who presented native speakers of English with laryngographical signals obtained from English and French speech, reports better results for those who were also familiar with French (at least if no feedback was given in the identification task). In both cases, experience with all languages from which the test signals were generated improved performance.

The natives of our experiments resemble the monolinguals of these experiments. Nativeness, if not combined with familiarity with other varieties, may contribute to the identification of local contours by true recognition but it does not allow further enhancement of success rates in the identification task by a simple elimination strategy. Exposure to NSG or other varieties through the broadcast media may improve performance but, as is evident from both experiments, it does not distort the differences.

An additional finding of Experiment 2 was that listeners performed better when they listened to the Berlin carrier voice than when they listened to the NSG carrier voice. However, the choice of carrier voice had no uniform effect on all the contours tested. The choice of a local carrier voice raised scores on the BUV contours but it did not affect scores on the LAG contours. These facts may let us conclude that the perception of intonational features will not necessarily be overridden by the perception of segmental features, voice quality, or other prosodic characteristics. However, the scores obtained for the NSG contours presented with both carrier voices do not fit well this line of argument. In two of three cases the choice of the BUV carrier voice raised the scores on the NSG contours as well. A possible explanation could be that the contours presented as NSG Contour II and NSG Contour III were found to be used in BUV as well, even if in slightly different conversational contexts.

Comparing the scores on local contours and NSG contours of both experiments, one may get the impression that the Non-Native Residents in Experiment 1 performed somewhat better than the Non-Native Residents in Experiment 2. Possibly, the inclusion of LAG contours may have pushed up the scores for the NSG contours such that they come closer to those for the BUV contours. On the other hand, the

---

9 Trilinguals scarcely performed better than monolinguals (53.2% vs. 50.3%). However, only four trilinguals took part in the experiment.
contours of HUV may be more salient than the contours of BUV. This argument is supported by findings of Schaeffler and Summers (1999). The authors report an average identification rate of 31.0% for varieties spoken in North-West Germany (including HUV) but an average identification rate of just 17.5% for varieties spoken in North-East German (including BUV).

Finally, listeners were found to differ in the identification of the various local contours. In Experiment 1, natives of Hamburg identified Contour Ia but not Contour Ib, Contour II, and Contour III. In Experiment 2, a different patterning was found for Contour III than for Contour I and Contour II. The difference found in Experiment 1 may be largely due to differences in the saliency of the contours tested. Romano (1997), who reports best recognition rates for wh-questions and emphatic patterns, comes to a similar conclusion for his data. The results for Contour III in Experiment 2 may be explained by the presence of similar contours in other varieties of German that may have leveled out the differences between the three listener groups. In conclusion, any attempt to assess the role of intonation in the identification of dialects or languages should take into account that contours may be more or less salient.

Received: April 17, 2001; revised manuscript received: April 03, 2002; accepted: July 17, 2002

References


Appendix

A. Carrier sentences of Experiment 1:

Contour Ia:
- a. *Aber heute is = es ja so:: …*
  ‘But today things are like…’
- b. *Die kriegen 1800 Ma:rk …*
  ‘They get 1800 marks…’
- c. *Und es ist eben schw:e:r …*
  ‘Well, it’s difficult…’

Contour Ib:
- a. *Wir sind sieben Paare…*
  ‘We are seven couples…’
- b. *Aber man kann natürlich auch nich erwarten …*
  ‘But, of course, you cannot expect…’
- c. *Dann ist die Produktion gesteigert …*
  ‘Then, the production increases…’

Contour II:
- a. *Und dann ist der Sonntag gelaufen.*
  ‘And then Sunday is gone.’
- b. *Es fällt mir nichts ein.*
  ‘Nothing occurs to me.’
- c. *Was will der mit soviel Pflegeversicherungsgeld.*
  ‘Why does he need so much money from nursing care insurance?’

Contour III:
- a. *Weißt du’ n Rat?*
  ‘Do you know a way out?’
- b. *Finden sie denn, dass es jetzt schlechter geworden ist?*
  ‘Do you really think it has been getting worse now?’
- c. *Finden sie?*
  ‘Do you think so?’

B. Carrier sentences of Experiment 2:

Contour I:
  ‘I must already come tomorrow’
- b. *NSG: Da gibts méhrere Sachen.*
  ‘There is more than one thing left to say.’
- c. *NSG: Da bin ich dort alleíne hingefahrn.*
  ‘Then, I went there alone (by car)’

*Language and Speech*
Contour II:  

a. NSG: *Mach ich das.*  
BUV: *Mach ick dette.*  
‘I did it.’

b. NSG: *Jetzt weiß ich ja nicht, ob der mich angesprochen hat.*  
BUV: *Jetzt wees ick ja nich, ob der mich anjesprochen hat.*  
‘I didn’t know whether he addressed me.’

c. NSG: *Weil ich ja Musik im Ohr hatte.*  
BUV: *Weil ick ja Musik im Ohr hatte.*  
‘As I listened to music.’

Contour III:  

a. NSG and BUV: *Mal schwimmen / mal baden / mal so …*  
‘Sometimes we went swimming / sometimes we went to the pool / some times we did this or that …’

b. NSG and BUV: *’N paar Japaner repariern / ’n bisschen da machen / ’n bisschen mit den Kollegen / Kollejen rumquatschen …*  
‘Repairing some Japanese cars/ doing this or that/ talking a little with colleagues …’

c. NSG: *Wir haben mit denen gequatscht / wir wurden über die Straße gebracht / oder der hat mal die Straße abgesperrt / oder sonst was…*  
BUV: *Wir haben mit denen jequatscht / wir wurden über die Straße jebracht / oder der hat mal die Straße abjesperrt / oder sonst wat …*  
‘We talked with them / they helped us to cross the street / or he blocked off the street / or he did something else …’