WHAT WE (DON’T) PERCEIVE AS FOREIGN ACCENT

This study deals with the perception of foreign accented speech by native Austrian listeners. What kinds of deviations from the language norm are allowed in order to still sound native-like? The results of three experiments show that a slight Polish accent can easily be differentiated from regionally colored Austrian German speech. This fact holds in global accent ratings (Experiment 1), as well as in error identification (Experiment 2) and goodness judgments of single sounds (Experiment 3). Aside from sociological consequences which will not be considered in this paper, an understanding of what is perceived as deviant from language norms will help to understand variation within the native language and can be applied e.g. to improve L2 pronunciation training.

FACTORS AFFECTING THE PERCEPTION OF FOREIGN ACCENT

Learning a second language (L2) people usually retain a foreign accent. Their language deviates from target norms along several parameters. The most traitorous aspect thereby is pronunciation. Extensive research revealed three major factors that affect the ultimate attainment in sounding like a native speaker: age of learning (first assumed as a 'critical period' by Lenneberg 1967), length of residence in the L2 environment (Flege 1988; Flege/Fletcher 1992; Flege et al. 1999) and continued use of the native language (L1) (Flege et al. 1997; Guion et al. 2000).

Less popular is the question about what exactly is perceived as foreign accent. The focus of this study will be on pronunciation. The term "foreign accent" is used referring only to the phonetic-phonological level of language. Its characteristics are usually divided into suprasegmental vs. segmental accent. Segmental accent can be further split into the mispronunciation of segment sequences (e.g. consonant clusters) and deviant realizations of single elements. The latter are either subsegmental or causing a category change.

Comparing research on the contribution of segmental deviations to global foreign accent, different methods and analyses have to be considered. For example in the study reported below, vowel length as well as (missing) vowel reduction will be taken as segmental characteristics, although they of course influence speech rhythm, which is not segmental in

---

Although the author is aware of the occasional distinction of the terms "foreign" vs. "second" language (L2) and "acquire" vs. "learn", in this paper they will be used interchangeably. German is the "second" language of all accent speakers in the experiments. They learned German inside and outside classroom settings.
nature. Calculating the relation of deviant segments to global accent ratings, variation occurs about what is counted as "error". Whereas Munro/Derwing (2001) take whole sound substitutions as their basis of analysis, many authors (e.g. Cunningham-Andersson/Engstrand 1989; Anderson-Hsieh et al. 1992; Magen 1998) count subsegmental deviations. Cunningham-Andersson/Engstrand (1989) try to weight different kinds of errors. Riney/Takagi (1999) confirm that the improvement of some segments raises the global accent ratings whereas a change of others does not have this effect.

Taking an attempt to generalize over different studies, it is not clear which factors contribute the most to perceived accent strength. Beside different methods of analysis, factors affecting the results are the different speaker populations, the languages under investigation as well as the experimental setup: the length of presented stimuli, the size of the rating scale and the ratio of accent vs. control speakers (cf. Flege/Fletcher 1992).

Still, it is a fact that even linguistically naive listeners are able to reliably rate accent strength and comprehensibility of non-native speech (Anderson-Hsieh/Koehler 1988; Cunningham-Andersson/Engstrand 1989; Munro 1995; Derwing et al. 1998; but see Thompson 1991 and Cunningham-Andersson 1997). Several studies (e.g. Magen 1998) find that speech manipulations within the range of a few milliseconds improve the impression of global accent. Flege (1984) shows that foreign accent can be detected in a signal consisting of only the burst of a /t/.

The aim of this study is to present examples from the perception of Polish accent in Austrian German (cf. Reinisch 2005). German is a good testing ground for research on foreign accent because it is subject to great regional variation. The listeners are confronted with native and non-native deviations from a "language norm". The term of reference here will be "Austrian standard". Unfortunately, this is a rather vague concept as no such norm exists prescriptively. Moosmüller (1991, 1995) provides an empirically based definition:

"If we take for the formation of a standard variety the acceptance of that variety as overregional and the speakers as persons with prestige (= members of the higher social classes), then varieties of the middle and upper social classes from the cities of Vienna and Salzburg (= Middle Bavarian) are accepted as Austrian standard."


This description will be adapted here, because the majority of participants in the experiments indicated a great uncertainty about the existence of an Austrian standard. Some defined it in terms of vocabulary. Only one listener mentioned news speakers on Austrian television as models.

Almberg/Husby (2000) emphasize that not only the listeners' native language, but also their native dialects are of major importance for the perception of foreign accent. Listeners speaking non-target dialects judge different factors to cause the accent than speakers of the target varieties. An artificial improvement of L2 Norwegian intonation patterns towards one
diatetal variety disturbed listeners of the non-target variant. They preferred sentences in which vowel length had been manipulated. For German listeners a similar effect is expected.

A STUDY ON POLES LEARNING GERMAN

Three experiments tested Austrian listeners' ability to identify a weak Polish accent and to differentiate it from slightly regionally colored speech in their native language. Experiment 1 consisted of global accent ratings. Experiment 2 investigated if the listeners were able to name single segments that contribute to foreign accent. Finally, Experiment 3 checked the reliability of Experiment 2 by asking to rate the accent of given segments presented in the whole phrase context.

Before discussing the experiments, it is necessary to take a closer look at the languages under investigation. What kinds of mispronunciations can be expected? An early model of L2 acquisition, the Contrastive Analysis (Lado 1957), predicts difficulties with the production of sounds that are very dissimilar from the speaker's native sound inventory. Sounds that closely match the L1, should not pose a problem. However, newer L2 acquisition models point in the opposite direction. The ones taken into account here are the Native Language Magnet Theory (NLM) (Kuhl 1993; Kuhl/Iverson 1995), the Perceptual Assimilation Model (PAM) (Best 1994, 1995) and the Speech Learning Model (SLM) (Flege 1992, 1995). The first two are not specially designed to explain L2 speech but to model phonetic acquisition in general. Whereas NLM and PAM focus on speech perception, SLM concerns mainly production. Although the models differ in theoretical background, their predictions about the relative difficulty to pronounce a sound in a certain foreign language are similar. The most important factor is the division of the sound space in the two languages, the relative position of category boundaries and their "best exemplars". The closer an L2 sound matches a typical realization of an L1 sound, the lower the chance that the learner perceives the difference. These sounds are the hardest to acquire without an accent. Learners simply use the sounds they know. On the other hand, they should be able to identify and acquire new sounds that fall between native categories or have their best exemplars far from the L1 target. The language descriptions and the analysis of Experiment 2 will refer to these predictions.

SOUND SYSTEMS

Table 1 depicts the consonant inventories of German and Polish. Shaded fields indicate sounds that appear only in Polish, bold framed fields contain sounds exclusive to German. Unfortunately, reality is not as straightforward as it may seem from the picture. Some notational conventions blur the differences crucial for the language learner, others overempha-
size a difference that is not at all that important. These are the most interesting cases in terms of L2 acquisition models. They are the focus of this chapter. I will begin with the not so obvious language characteristics namely voicing contrasts. Austrian plosives are voiceless unaspirated, produced with only a lenis-fortis distinction. Polish contrasts voiceless unaspirated with truly voiced plosives. From the perceptual point of view both Austrian categories are close to the Polish voiceless sounds. This should result in a collapse of the two phonemes into the Polish voiceless category. Nevertheless, two other factors will override perception here: the wish to keep a contrast between the German categories, and orthography.

| Consonants of German and Polish: bold framed fields = sound existing only in German; shaded fields = sounds existing only in Polish; ( ) = sounds without clear phonemic status, (v) not clearly produced with friction noise |
|---|---|---|---|---|---|---|---|---|---|
| Plosive | bilabial | labiodental | alveolar | post-alveolar | retroflex | palato-alveolar | palatal | velar | glottal |
| Plosive | p | b | t | d | k | g | (?) |
| Frikative | f | s | z | ç | x | h |
| Affrikate | pf | tʃ | dʒ | ʃ | ʒ | ç | z | (ç) |
| Nasal | m | n | n | n | n | n | (n) |
| Lateral | l | l | l | l | l | l | l |
| Trill | r | r | r | r | r | r | r |
| Glide | w | v | w | v | w | v | w |

An overemphasis of a difference between the languages in Table 1 results from the representation of the shibilants and alveolar affricates. What is labeled "post-alveolar" for German and "retroflex" for Polish is in fact quite similar. Talking only about Polish it does not matter if the /ʂ/ is called/ produced retroflex or post-alveolar as long as it is different from /ç/ (and so for the voiced counterparts and affricates). In comparison to German the labels do play a role because the German /ʃ/ is produced somewhat more to the front and more palatal than the Polish equivalent. According to the L2 acquisition models German /ʃ/ should be completely assimilated to Polish /ʂ/. The question is, whether listeners will realize this subcategorical difference.

[ç] is assigned exclusively to German. As it is analyzed as allophone of /x/ and also spelled <ch>, Polish learners most likely will match it to their velar category /x/ <ch>. It

---

2 For details and problems on establishing the phoneme inventories of the two languages, especially Polish, see Reinisch (2005) and references mentioned there.

3 The language of reference here is Austrian German. Many of the described phenomena appear in different German varieties, but not all of them correspond to German standard. In the latter case I write "Austrian", otherwise, "German".
might get the status of a non-ideal exemplar of the category which should facilitate good production. In addition, Polish words like *Chiny* [tʃɪnɨ]/ [tʃiːnɨ] 'China' are produced with a sound very similar to [ç]. The ultimate attainment of German [x] and [ç] will most probably depend on the awareness of the speakers that they have the same difference in their own language. One negative effect must be taken into account here: the phonological context relevant for the right allophonic choice. As all Polish assimilations are regressive in nature, the German progressive influence of front vowels on /x/ might go unnoticed. Austrian listeners on the other hand might count incorrectly produced [ç] as a regional feature typical for western Austria.

Another interesting problem for listeners is the perception of /r/. In Polish this sound is almost always produced as an alveolar trill. In German velar as well as alveolar variants are common, but learners are usually taught to produce [ɤ]. In many but not all German varieties /r/ is vocalized postvocally. The reactions of listeners will be shown in Experiment 2.

Greater differences between German and Polish exist in the vowel systems. German obviously has more vowel phonemes than Polish (14 vs. 6). They are depicted in Table 2. Shared sounds are shaded, the only vowel exclusive to Polish is bold framed.

<table>
<thead>
<tr>
<th>i:</th>
<th>y:</th>
<th>ü</th>
<th>u:</th>
</tr>
</thead>
<tbody>
<tr>
<td>i</td>
<td>y</td>
<td>u</td>
<td></td>
</tr>
<tr>
<td>e'</td>
<td>ø'</td>
<td>o'</td>
<td></td>
</tr>
<tr>
<td>ε</td>
<td>œ</td>
<td>o</td>
<td>c</td>
</tr>
<tr>
<td>a:</td>
<td>a</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2: Vowels of German and Polish: shaded fields = shared vowels; bold framed = vowel existing only in Polish

Thinking of the six Polish categories and their corresponding best exemplars as potential attractors for the 14 German categories, a complete assimilation of /i: e a ø u:/ to /i e a ø u/ can be expected. The only vowel that exists in Polish, but not in German is /i/. Its closest counterparts are probably /i/ and /i/, although the first one might also get attracted by Polish /i/. Front rounded vowels should be classified as "new" sounds and therefore be learned well. They are not close to any Polish vowel. Unclear is the status of German /e:/ and /o:/.

One general characteristic of Polish vowels is the lack of phonemic length contrasts. L2 acquisition models do not make any predictions about the role of vowel length. One could expect that Polish learners simply produce all vowels short like in their LI. They do not possess a distinctive feature specifying it.
Speakers

At the time of recording all 5 Polish native speakers (3 female, 2 male) were between 19 and 29 years old and had been living in Austria (Vienna) for at least 2 months. Those two (P3, P5), who came to Austria only a short time ago, had former intensive training in Warsaw at the Austrian Institute or during summer schools in Vienna. The others learned German only in Austria. P4 and P1 had been living in Vienna for 2 years, P2 – the oldest – for 14 years. All speakers were late learners of German. They hardly had any contact with the language before 15. Most importantly, all speakers stressed their desire to speak Austrian German, because they like it better than the usually taught northern variants. The self-reported use of German on a daily basis ranged from 40% to 95%, the use of Polish from 5% to 60%. The Polish speaker whose speech material was used to demonstrate the experimental tasks matched the profile of the others. According to the author his accent was somewhat stronger than the accents of the speakers in the experiments. They were all judged to speak very well.

Two native Austrian speakers (1 female, 1 male) served as controls. A1 was born and spent most of his life in Vienna. A2 is from Tyrol, although at the time of recording she had been living in Vienna for 8 years. Both spoke standard Austrian German with slight regional features. A female student from Upper Austria provided the native German sentences for the task demonstration.

Language material

The speakers described a drawn black-and-white picture of a room. The names of the objects in the picture covered the whole range of German phonemes. Recordings were carried out in a quiet room using a SONY DCD-D100 DAT walkman and AKG C420L headset microphone. Three phrases per speaker were selected and arranged in two versions, each containing the same set of phrases in a different pseudo-random order. The speech material was chosen according to fluency and grammatical correctness to exclude undesired influences of these factors on accent ratings. Therefore, unavoidably, the phrases differed in content and length. They were 3.4s to 7.2s long and contained 6 to 21 words. The volume of each phrase was manually adjusted to the "loudest" speaker using WaveLab® 3.0. Interphrase intervals were set at 4s. No other manipulations have been carried out. The speech material was the same in all three experiments. Experiments 2 and 3 included the first ordering only. Before each experiment the relevant task was demonstrated using phrases from one male Polish and one female Austrian speaker who were not included in the experiments. The listeners were told that these can be taken as the extremes of the accent scale. The phrases together with the results from Experiment 1 and 2 are listed in the appendix.
LISTENERS

16 native speakers of Austrian German (10 female, 6 male), mainly from eastern Austria, with little or no experience with Polish accent, participated as listeners in three separate sessions spread over a period of three weeks. Engaging the same listeners in all three experiments eliminated variability due to different judges. The time span of about a week between the sessions should avoid adaptation to the speech samples as well as minimize tiredness effects. The listeners were tested individually or in small groups in a quiet room. The speech material was presented via integrated loudspeakers of a PC Notebook using Windows MediaPlayer® in Experiment 1 and Praat (Boersma/Weernik 2002) in the other two experiments. Although some background noise and noise caused by the computer could not be avoided, the listeners judged the sentences to be highly comprehensible. Until the end of the last session the listeners were not informed about the native language of the speakers they judged.

EXPERIMENT 1: GLOBAL ACCENT RATINGS

Experiment 1 consisted of global accent ratings of the phrases described above. Each phrase had to be rated twice, once in each of the two blocks. Within the blocks each phrase appeared only once. 4 seconds after the end of a phrase the next one followed automatically. The proper experiment was preceded by 5 trial phrases after which questions were answered. Before the session, the listeners had the possibility to see the picture the speakers described. Familiarity with the content should increase intelligibility and allow for concentration on the accent. The listeners were told that they would hear parts / phrases of picture descriptions. It was stressed that they should rate accent strength only. Grammaticality, fluency as well as dialectal features in the signals should be ignored. Ratings had to be made on prepared sheets by circling a number from one to nine with 1 meaning "native-like pronunciation" and 9 "strong foreign accent". Only the endpoints of the trial items were semantically labeled. The listeners were encouraged to use the whole range of the scale. The experimental session lasted about 7 minutes.

The aim of Experiment 1 was to show that native Austrian listeners are able to distinguish non-native phrases with a slight Polish accent from Austrian controls, even if the native pronunciation contains some minor regional features. Both groups were close to what could be called Austrian standard, but the deviations went in different directions. Besides, the question was how detailed the listeners would differentiate the accented phrases.

---

4 Actually, one female listener was born and grew up in Switzerland, one in Southern Tyrol (Italy), but both of them have native competence in Austrian German.
All of them were completely intelligible and grammatically correct (this was confirmed in open questions posed after the session). Hesitation elements occurred in native and non-native speech. The challenge for accent differentiation was the fact that phrases of only seven speakers had to be spread over a 9-point scale. Given this possibility, the listeners were expected to differentiate between the single phrases and not only between the speakers.

RESULTS AND DISCUSSION (1)

Median, mean and mean standard deviations of the ratings were calculated for speakers per block and per phrase. Judgments in block 1 and 2 were nearly the same, although a tendency for somewhat harsher mean ratings in block 2 could be noticed. As can be seen in Table 3 the listeners easily differentiated native and accented speech. Natives were almost always identified as such. Only two listeners did not accept A2 – the speaker from Tyrol – as native. This indicates that regional variation does influence the ratings to some extent, at least if the variety is unknown to the listener as it was the case here. The two listeners reported no contact to speakers of Austrian regional varieties other than Viennese.

<table>
<thead>
<tr>
<th></th>
<th>P1 (f)</th>
<th>P2 (m)</th>
<th>P3 (f)</th>
<th>P4 (f)</th>
<th>P5 (m)</th>
<th>A1 (m)</th>
<th>A2 (f)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Median</td>
<td>5</td>
<td>3</td>
<td>5</td>
<td>2</td>
<td>7</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Mean</td>
<td>4.81</td>
<td>3.71</td>
<td>5.29</td>
<td>2.6</td>
<td>6.46</td>
<td>1</td>
<td>1.3</td>
</tr>
<tr>
<td>SD</td>
<td>1.78</td>
<td>1.78</td>
<td>1.65</td>
<td>1.88</td>
<td>1.98</td>
<td>0.14</td>
<td>0.85</td>
</tr>
</tbody>
</table>

Table 3: Global Accent Ratings; ratings collapsed over listeners, blocks and phrases; SD = standard deviation; f = female; m = male

A somewhat more complicated picture emerges from the ratings of the accent speakers. At first glance, none of them reached native-like ratings. They seem to be evenly spread over the whole scale. However, two factors have to be discussed here. First, Table 3 shows only results per speaker and not per phrase. The cause is the lack of differentiation between phrases except for P1, whose phrases got median ratings of 4, 5 and 6, respectively (compare Table 4 in the Appendix). The assumption about a better differentiation on a larger scale was not corroborated. Although instructed otherwise, the listeners reported that they even tried to judge the same speaker similarly each time they thought to recognize his/ her voice. Some also noted that they had difficulties using the whole scale as all speakers were "very good". Therefore recommendations of a 9 to 11 point scale in studies like Flege et al. (1999) or Guion et al. (2000) have to be taken with some caution. They are not optimal for small speaker populations with little accent variation.

Second, the huge standard deviation must not be neglected. Most probably it arose due to the listeners' inability to use the whole rating scale. Taking into account standard deviations, the ordering of speakers according to their accent is not so straightforward anymore.
The only difference that remains is between speaker P4, the best one, and speaker P5, the worst one. However, an additional comparison of ratings per listener did show a tendency to assign ratings that match the ordering of the mean. Subtracting the standard deviation off P4' scores, she does not differ anymore from the native speakers. Some listeners actually assigned her the best rating. The case of P4 will be discussed in more detail with regard to the other experiments.

**EXPERIMENT 2: IDENTIFICATION OF ACCENTED SEGMENTS**

About one week after the first session the same listeners participated in Experiment 2. Its aim was to reveal which elements contribute the most to foreign accent, i.e. the mispronunciation of which sounds and what kinds of deviations are perceived most negatively. The experimental design follows Koster/Koet (1993). Participants listened to the same speech material as in Experiment 1. This time all the phrases were written down on the answering sheet (in standard German orthography). The listeners had to mark these sounds, which they thought they contribute most to the foreign accent of the speaker. In addition they should annotate what the mistake is like. Examples were given of the kind "this 'a' is too short, or this 'a' sounds like an 'o'". The experiment was preceded by the same trial phrases as in the first experiment. Each phrase was played as often as the listeners wished, 4 times on average.

**RESULTS AND DISCUSSION (2)**

The listeners reacted to a whole range of speech characteristics they thought to be deviant from their language norm. A summary of the marked sounds is given in the Appendix. All sounds that were judged to cause foreign accent are underlined. The following discussion includes only examples of sounds that were marked by at least 4 listeners (25%). Hardly any sound was labeled accented by more than half of the listeners. They were not expected to transcribe the whole phrases, but mark the most striking pronunciation errors or these that came to their minds first. The author observed the tendency that multiple instances of the same error within one phrase (e.g. all /o/s were pronounced too open) were marked only once. The few comments about intonation, speaking rate and hesitations will be ignored, because the phrases appeared out of context and some unnatural features were evident in the speech of the controls, too. Altogether, more vowels than consonants were perceived as accented. In sound sequences consisting of an accented consonant and an accented vowel the latter was marked more often, e.g. 5 listeners underlined the /b/ in *Buch* 'book' (P2), whereas 7 listeners marked the /u/. Many noted errors were subphonemic in nature. Not all of them can be explained by L2 acquisition models.
Short open vowels are perceived too long/tense/closed. This goes with the prediction that Polish speakers assimilate non-existing open /t/ and /u/ to their nearest categories, their closed counterparts. For German listeners this, of course, leads to a category violation and is rated as error. The fact that the sounds are not only pronounced too closed, but in addition too tense and too long, probably enhanced listeners' reactions. Problematic for the L2 acquisition models is the fact that not only /t/ and /u/ were perceived as too closed, but also /e/. This finding is quite unexpected. /e/ exists in Polish nearly in the same place as in German. Speakers should assimilate it. As for /t/ and /u/ the German listeners notice the difference if the /e/ is too closed because it penetrates their higher category /e:/.

Not only short vowels were perceived as too long/closed, but also closed vowels as too open/short. This was mainly the case for various realizations of /o:/ and one word with /u:/. As already mentioned in the description of the languages, German /o:/ falls between the Polish categories /u/ and /o/. Obviously, L2 learners choose the lower category for assimilation. For German listeners this results in a change of the sound identity. They react to it harshly. The prediction by L2 acquisition models that sounds falling between native categories should be learned well failed. Interestingly, no similar case could be found for /e:/ for which the same assimilation problem exists. Only 4 listeners noted that speaker PI realized her /e:/ deviant from their native norm. In contrast to /o:/ it was not too open but almost a diphthong [e:]. Comparing the two pronunciation strategies, assimilation to the lower category (which matches the spelling) seems to be noticed easier than diphthongization. One could argue that during the gliding from the too open category /e/ to almost [i] speakers pass the right category. Therefore the deviation is not so clearly perceptible as in the realization [a]. In addition, gliding could be interpreted as a dialectal feature. The question remains if the pronunciation [a] for /o:/ would have gone unnoticed.

Although the focus of the analysis is on segmental accent, two error types that are related to rhythmic characteristics of speech have to be mentioned: vowel reduction and r-vocalization. The L2 acquisition models do not make any predictions about them. Both phenomena do not exist in Polish, but were not learned correctly. One possible influence on the learners is again orthography: We write <Spiegel> 'mirror' but pronounce [pi:gil]. Nevertheless, in this case the listeners could have been fooled, too. They had to mark phrases written in standard German orthography. Even more peculiar than the negative reaction to the missing vowel reduction is the criticism of missing r-vocalization. In most parts of Austria as well as in Austrian standard postvocalic /u/ gets vocalized to /e/, but

---

5 Most listeners named only one or two characteristics, but the point of reference will be the sum of answers.
6 Maybe the Polish /E/ is a bit more open, but no acoustic measures have been carried out to verify this assumption.
there are German varieties in which this is not the norm. Thus, the listeners assign foreign accent to speech lacking a process which is only optional. An interesting comparison would be the reaction to a realization of /b/ as [r]. Again, in German both variants exist. Unfortunately, all speakers pronounced their /b/ as velar.

Like in other studies on foreign accent (e.g. Flege 1984; Major 1987; Riney/Takagi 1999) wrong VOT played a crucial role in accent assignment. It was the most frequently named deviation in consonants. To remind of the relative positions of the plosives in the two languages, both Austrian variants are voiceless unaspirated with only a lenis-fortis distinction. They match the Polish voiceless category. To keep the contrast in German Polish learners are expected to use their truly voiced category for <b>. Speaker P1 pronounced it as voiceless unaspirated, but so lax that the listeners put it into their lenis category. This suggests that keeping the contrast by producing <b> more voiced than necessary does not guarantee a good differentiation. As the Polish <p> includes both Austrian phonemes, learners find it difficult to choose the right part of their category.

Even more subtle is the difference between /ʃ/ and /l/ in the two languages. For both sounds L2 acquisition models assume complete assimilation. Whereas the problem of /ʃ/ has been discussed above, the status of /l/ is not completely clear. The listeners who marked it as accented noted that the Polish variant is somewhat "darker" than the German one. The Polish /ʃ/ was correctly labeled "too back". Correct predictions were also made for the "new" categories of front rounded vowels. Hardly anyone perceived mispronunciations in these sounds. The relation of perceived wrong segments to global accent ratings will be discussed after the analysis of Experiment 3.

**Experiment 3: Accentedness of Single Sounds**

Experiment 3 was designed to check if listeners are able to judge the accent of single sounds independently of global accent. The inspiration for this part of the study was taken from Cunningham-Andersson (1997). Generally, the task resembled Experiment 2, with the difference that this time two sounds (all letters representing these sounds) in the written phrases were underlined already. Below each phrase two 4-point scales were provided for accent/ goodnes ratings of the sounds. In addition the listeners were asked to indicate if they thought the correct pronunciations of one of these sounds would improve global ac-

---

7 The letters <b> and <p> are taken as representatives for all plosives. They were the most frequent ones in the study and were mispronounced in Puppe "doll" (P1), Bett "bed" and Buch "book" (P2).
cent ratings. The sounds were chosen on the basis of the results of Experiment 2: Some had been marked as accented by many listeners, some by few and some served as correct controls. In the phrases spoken by the Austrians those sounds were assigned for accent ratings that seemed to show slight regional features. Although the rating task was rather complex, the listeners declared that they had no problem doing it.

RESULTS AND DISCUSSION (3)

The results of experiment 3 show that, indeed, listeners are able to distinguish between global accent and single sounds contributing to it. In addition, it was possible to differentiate sounds according to their divergence from the norm although they were presented in whole phrase contexts. Nevertheless, this is true with one restriction: As in Experiment 1 the best grade was reserved for native speakers. All of their sounds were rated 1, even if they actually did not match the norm. The best sounds in globally accented speech were given a 2. Thus, context does play a role to some extent, but the assignment of the other three points was quite promising.

To mention some examples, in Experiment 2 nobody marked the /y/ in Bücher 'books' (P1). /y/ – or front rounded vowels in general are stereotypically difficult to pronounce. Therefore it could have been the case that listeners perceive it as deviant although it is not. The correct decision of not marking the sound was confirmed in Experiment 3 – the /y/ got the score 2. The fact that the listeners did not react to regional variation is also evident in the judgments of the non-native speakers. According to the author P1 voices the /s/ in Vase 'vase' as it is the norm in many German regions. It got the score 2. On the other hand, sounds that were marked as accented in Experiment 2 by many listeners constantly received a 4 in Experiment 3. This was the case for /ɛ/ in Bett 'bed' and Heft 'booklet', the /ʊ/ in Puppe 'doll' (P1), <z> /ts/ in Zimmer 'room' (P3) as well as the devoiced /s/ in Wolke 'cloud' and too open /o/s in P5's first phrase. Another hint that the listeners distinguish single sounds from global accent is the fact that many of them judged the pronunciation of <o> in Unordnung 'mess' (P3) as quite good although it was erroneously stressed making the whole word sound accented.

Concerning the question of possible accent improvement due to a better pronunciation of certain segments, the listeners remained skeptical. They believe that the correction of single sounds has little or no influence on the global accent.

GENERAL DISCUSSION

Three experiments investigated the perception of Polish accent in Austrian German. The question was whether the listeners are able to differentiate good L2 learners from native controls, even if the latter differ from the language norm in slight regional speech characteristics.
Experiment 2 and 3 showed the ability to locate the perceived foreign accent on single sounds. The listeners proved to be sensitive to very subtle details in the signal. They named not only accented sounds, but also judged some possible regional features as foreign accent. Almost no features were marked in the sentences produced by the native controls. L2 acquisition models explain most but not all of the mispronunciations perceived by the native listeners.

In this chapter some interesting individual speech characteristics will be discussed and compared to other studies in the foreign accent literature. P4's global accent ratings almost reached a native-like level. In Experiment 2 she got even more "native-like" ratings than in Experiment 1. This finding is in contrast to the study by Flege/Fletcher (1992) where listeners gave worse ratings after a task that focused their attention on fine speech characteristics.

Comparing the best speaker P4 with the worst speaker P5 (best and worst with respect to their global accent ratings in Experiment 1) no difference in the absolute number of perceived errors in Experiment 2 could be established. This lack of difference holds, if the number of listeners that marked the single accented segments is taken into account. The results of Anderson-Hsieh et al. (1992) could not be replicated. They found a correlation between the number of errors and global accent ratings in a group of speakers with different L1s. Unfortunately, in my study the quality of errors cannot be compared as all speakers produced different phrases and therefore made different mistakes. Maybe also non-linguistic factors played a role in the assignment of accent ratings. P4 spoke slightly faster and more self-confident than P5 and the others. This might be an artifact resulting from the recording situation. Some of the speakers felt insecure speaking into a microphone (although they wore a headset).

The listeners' explanations why they accepted P4 as native speaker reveal different, partly complementary motives. In phrase 1 the realization of /a/ in Lampe 'lamp' was perceived as [ɔ] which reminded the judges of Austrian dialectal shift of /a/ to [ɔ] (cf. Moosmüller 1991: 141 ff). Given the fact that in Experiment 2 P4 was judged a native speaker because (intentionally or not) she used regional speech characteristics, the clear differentiation between natives and non-natives in Experiments 1 and 3 could be reinterpreted. The regional features in the speech of the Austrian controls were not a distraction but a help to identify them as natives. On the other hand, missing the application of optional processes like r-vocalization, which is also counted as a regional variable, did not show the same effect. To get a clearer picture of the interaction between the allowed (unnoticed) native and non-native variation, probably a list of features must be collected to establish the variables in the listeners' own varieties. After all, two listeners did not accept the Austrian speaker from Tyrol as native. Unknown speech characteristics seem to be judged as foreign accent.

But still, there is another possible interpretation. It results from the comments on P4's third phrase. Again, many listeners assigned her native-like ratings. This time, however, they guessed that she is German. The reason is the erroneously stressed word Telefon 'tele-
phone'. In Austria this word is correctly stressed on the last syllable. In Germany stress falls
on the first vowel. P4 chose the latter variant. So, would an intentional inclusion of certain
mistakes lead to better ratings? Many of the listeners reported that they would assign worse
ratings, if a foreign accent was combined with dialectal features. In that case probably so­
cial factors cannot be excluded any more.

CONCLUSION

This study can be taken as a first step into research on the distribution of accepted vs. not
accepted variation in German (i.e. accepted as native). Obviously, the distinction will never
be complete. Not every native speaker of German is familiar with all the different regional
variants. Therefore it is likely that some characteristics of foreign accent will be assigned a
dialectal status and vice versa. This was the case in this study. An investigation of the most
disturbing accent features in given regions might allow for some interesting conclusions in
L2 acquisition theory. Thinking of the great variety of German regional characteristics, the
argument that dialectal features are predictable, whereas accent is not, does not seem con­
vincing. Foreign accent is not only the way learners produce the L2, but also what the na­
tive speakers of the target language perceive as such. If language learners assimilate foreign
sounds to their native inventory, why don't the listeners assimilate them back?

From the methodological point of view a greater variety of native and non-native speakers
should be included in the design. How do listeners really react to speech that combines for­
eign accent with regional features? If German speakers from other regions were included as
additional controls, would P4 still have passed as native in some conditions? Given the as­
sumption that controls will be recognized as such, the possibility to make direct comparisons
to different native speakers should have consequences on the perception of the learners.

Foreign accent literature provides another method to investigate the differentiation of
native-speakers vs. L2 learners. Munro (1998) shows that good L2 speakers are perceived
much worse in a noisy background. Although the ratings without noise could not predict the
performance in the test condition, a comparison of accent vs. dialect speakers in bad listening
conditions could reveal some crucial differences in the two types of language variation.

A problem excluded so far from the discussion is the nature of a real communication
situation in comparison to experimental tasks. Studies working on the phonetic / phono­
logical level of language have the disadvantage that their features of investigation are rela­
tively unimportant in a natural communication as long as the speaker is comprehensible.
The essential factor is the content of the conversation. Besides, the best L2 pronunciation
cannot deceive native speakers if certain grammatical errors are present in the language. Of
course there is regional variation in syntax and lexical items, too, but some factors like the
wrong use of determiners are exclusive to non-natives.
Placing the study in the context of other studies on foreign accent it clearly adds some pieces to the puzzle about the nature of the phenomenon. Nevertheless, much work has to be done in order to put the pieces together. The great variety of methods, languages and subject populations under investigation not only enrich the whole enterprise, but unfortunately complicate the comparison of the results. Adding the component of regional variation, even more factors will have to be controlled in future. Nevertheless, an investigation of different factors of language variation is of interest in many other areas. The findings should be taken into account in models of speech perception; they can be applied in language teaching or automatic speech recognition and may serve as background for social studies.

REFERENCES


Flege, J.E./Yeni-Komshian, G.H./Liu, S. 1999 Age constraints on second-language acquisition, Journal of Memory and Language 41: 78–104.


What we (don't) perceive as foreign accent


**APPENDIX**

<table>
<thead>
<tr>
<th>SPEAKER (GENDER)</th>
<th>GLOBAL AC. MEDIAN</th>
<th>PHRASES INCLUDING ALL MARKED SOUNDS</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1 (f)</td>
<td>4</td>
<td>Ich sehe ein Zimmer und in dem Zimmer befinden sich verschiedene Gegenstände</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>Es gibt einen Schrank mit Bücher und auf dem Schrank steht eine Vase mit einer Blume und einer Kerze</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>Auf dem Bett sehe sich ein Heft eine Puppe und ein Teddybär</td>
</tr>
<tr>
<td>P2 (m)</td>
<td>3</td>
<td>Auf dem Boden liegen Gegenstände und Kleiderstücke verstreut</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>Auf den Wänden hängen Bilder und ein Kalender und dazu hängt noch eine Uhr neben dem Fenster</td>
</tr>
<tr>
<td>P3 (f)</td>
<td>5</td>
<td>Ein kleines Zimmer und hier befinden sich einige Gegenstände</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>Und auch ein Spiegel vielleicht - etwas Rundes das ist vielleicht ein Spiegel</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>Es ist Unordnung in diesem Zimmer</td>
</tr>
<tr>
<td>P4 (f)</td>
<td>2</td>
<td>Also ich sehe eine Lampe und eine Spinne da unten</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>Fußball etwas zum Spielen Schuhe Stiefel</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>Am Tisch - ah - ist das Telefon Schere ein Blatt Papier etwas zum Schreiben</td>
</tr>
<tr>
<td>P5 (m)</td>
<td>6</td>
<td>Im Fenster gibt's zwei Vögel und man kann - kann - auch eine Wolke da hinten sehen</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>Was haben wir noch - auf dem Boden sehe ich ein T-Shirt oder so was</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>Auf dem Regal befinden sich viele Bücher</td>
</tr>
<tr>
<td>A1 (m)</td>
<td>1</td>
<td>An der hinteren Wand ist ein Fenster ein - ah - mit zwei Vorhängen</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>An der linken Wand hängt außerdem noch ein Bild mit einer Sonne drauf und einem Fluss über den eine Brücke führt</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>Links vorne sieht man ein Bett mit einem Polster oder einem Kissen</td>
</tr>
<tr>
<td>A2 (f)</td>
<td>1</td>
<td>Und - ahm - ein Aquarium mit einem Fisch</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>Ein Rucksack liegt am Boden Hemden Schuhe</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>Auf dem Bild an der Wand - ah - ist abgebildet ein Prinz und eine Prinzessin</td>
</tr>
</tbody>
</table>

Table 4: Summary of global accent ratings per phrase in Experiment 1 and marked sounds in Experiment 2.

Eva Reinisch
Wien
eva.reinisch@gmx.net