

# (How) can listeners identify the L1 in foreign accented L2 speech?

Marie-José KOLLY and Volker DELLWO

Phonetisches Laboratorium der Universität Zürich

Par son accent étranger, un locuteur/une locutrice révèle son origine, sa langue maternelle. Ainsi, la majorité des Suisses alémaniques seront reconnus comme tels, en parlant une langue seconde. A partir de cet accent 'suisse allemand', est-ce qu'on pourra aussi deviner la région dialectale d'où provient le locuteur/la locutrice? Et, si la perception humaine permet l'identification de ces subtilités, quels en sont les indices pertinents dans le signal linguistique? Les expériences de perception conduites avec nos sujets suisses alémaniques démontrent, dans un premier temps, que des différences d'accent dues à un dialecte particulier peuvent être perçues non seulement dans du matériel linguistique allemand standard, mais aussi quand lesdits locuteurs parlent français. Une expérience ultérieure explore l'importance des données temporelles, ou rythmiques, pour l'identification d'un accent étranger.

## 1. Introduction

(How) does the foreign accent of a speaker allow us to take guesses about his/her origin? The present article addresses these two questions by means of two perception experiments: first, we investigate whether the linguistic origin of L2 speakers can be identified perceptually at all. If so, we further explore how listeners identify the origin of foreign accents, i.e. based on which acoustic cues.

In certain situations it is typically easy for listeners to identify the L1 of an L2 speaker. If we think of L2 German, for example, the use of uvular trills for German /r/ and nasal vowels often leads listeners to identify a French accent. Likewise, the use of alveolar approximants for German /r/ and the unrounding of front rounded vowels will often point towards an English accent. Typically, such cues for accent identification result from the transfer of certain phonetic characteristics of the speaker's L1 to his L2 speech. Listeners can thus typically discriminate English-accented German speech from French-accented German speech (cf. Boula de Mareuil *et al.*, 2008, for comparable language constellations). In other situations however, it might be more difficult to identify or discriminate foreign accents. First, French and English in the above examples are two distinct languages that differ in many aspects. Could listeners also recognise foreign accents that stem from more closely related varieties, like dialects? Second, there are different kinds and combinations of cues that create the impression of particular accents (cf. Cunningham-Andersson & Engstrand, 1987). The above examples present segmental cues for accent identification. Do other

cues, for example prosodic cues, also allow listeners to identify foreign accents?

In our first experiment we examine whether foreign accents stemming from closely related varieties can be identified perceptually. Closely related varieties are found e.g. in German-speaking Switzerland, a region well-known for having a diverse dialectal landscape in a relatively small space. We investigate whether the dialectal origin of Swiss German speakers (from Bern and St. Gallen) is perceivable in L2 speech of these speakers, e.g. L2 Standard German or L2 French.

The Standard German variety spoken in Switzerland (*Schweizerhochdeutsch*) is not a real L2 for Swiss German dialect speakers; it is better described as an "extended" version of their L1 and its acquisition is considered to be an "erweiterter Erstspracherwerb mit einigen Zügen von Zweitspracherwerb" (Häcki Buhofer & Burger, 1998: 137). French, however, is taught as a first L2 in the majority of Swiss German primary schools. Swiss German accented French is commonly referred to as *français fédéral* (cf. Kolly, 2010). Swiss German as well as Swiss French listeners were tested in this experiment: Swiss German listeners heard Standard German and French stimuli and had to indicate the dialectal origin of the speakers (open response). Swiss French listeners heard French stimuli and had to indicate whether the speaker was from Bern or from St. Gallen (cf. 2.1.3)<sup>1</sup>.

We expect that Swiss German listeners will more easily recognise dialectal accents than Swiss French listeners. Swiss German listeners are used to hearing people with different dialectal backgrounds talk Standard German or French (e.g. at school). It has been shown that they can recognise Swiss German dialects as well as dialectal accents in Standard German speech above chance when confronted with a multiple choice task (cf. Guntern, 2011, including 8 Swiss dialect regions). Swiss French listeners naturally have less contact with Swiss German-accented speech and are not expected to have much knowledge about the Swiss German dialectal landscape. Even in Fribourg/Freiburg, the bilingual town where our listeners come from, Swiss French speakers' contact with Swiss German dialects is rather small (cf. Muller, 1998, for the situation in Biel/Bienne, a comparable bilingual Swiss town with a proportionally larger Swiss German population). Given the difficulties, the lack of interest and the negative attitudes that go with the acquisition of German for many Swiss French people (cf. Muller, 1998; Fuchs & Werlen, 1999; Kolly, 2011), the hypothesis that this group of listeners could recognize particular Swiss German

---

<sup>1</sup> Results from this experiment are also presented in Kolly (2010) and Kolly (2013) in German, where more weight is given to the particularities of each speaker. The present article provides a more general approach to the question whether dialectal accents can be recognised perceptually.

dialectal accents is ambitious. The question *if* listeners can identify particular foreign accents, investigated in this first experiment, leads to our second question: *how* can listeners identify particular foreign accents?

In our second experiment we examine some of the acoustic correlates of perceptual accent recognition. L2 speech contains a large amount of acoustic cues which are perceptually salient to the listener, and particular accents are often characterised by a certain amount and combination of those cues. Although acoustic cues for this kind of perceptual task were traditionally assumed to rely predominantly on a segmental level (cf. Boula de Mareuil *et al.*, 2008), research more and more considers prosodic aspects of L2-speech (cf. Flege, 1992; Tajima *et al.*, 1997; Jilka & Möhler, 1998; Hirschfeld & Trouvain, 2007; Atterer & Ladd, 2004; Missaglia, 2007). Within this research on the prosody of foreign accents, people have been trying to capture speaker origin in different ways and have shown that foreign accents can be recognized in several types of degraded speech signals (cf. Van Els & De Bot, 1987; Boula de Mareuil & Vieru-Dimulescu, 2006, for foreign accent recognition; Bush, 1967; Leemann & Siebenhaar, 2008; White *et al.*, 2012, for dialect recognition). These studies have focused on the frequency-domain of speech, while there has been relatively little research on the time-domain (cf. White *et al.*, 2012, for the perceptual discrimination of dialects based on the durational variability of vocalic and consonantal intervals). To our knowledge the time-domain has not been tapped into in terms of the perceptual recognition of foreign accents. However, the idea that temporal prosodic characteristics (for example speech rhythm) play a role in non-native speech has a long tradition; Lloyd James (1929), for example, discussed the transfer of L1 rhythmic cues to English by French speakers, which has an effect on their intelligibility.

Research about so-called "rhythm classes" (cf. Dellwo, 2006, for an overview) has shown that there is a perceivable difference in timing patterns between allegedly stress-timed languages like English, Dutch and German and allegedly syllable-timed languages like French, Italian and Spanish. Even if those rhythm classes are discussed controversially nowadays (cf. e.g. Arvaniti, 2012), they seem to have some perceptual relevance. If we assume that speakers transfer prosodic characteristics and, in particular, timing patterns from their native language to a second language, we expect perceivably different timing patterns in L2 German spoken by a native speaker of another stress-timed language, e.g. English, and in L2 German spoken by a native speaker of a syllable-timed language, e.g. French. If this is true, we expect French-accented German to sound rhythmically different from and, in particular, more marked than English-accented German. Our second experiment thus investigates whether

French-accented German can be distinguished from English-accented German in degraded speech signals that contain primarily temporal cues.

Temporal cues are found on many levels of the speech signal. We explore different types of speech durational characteristics in order to test whether they contain cues relevant to the perceptual impression of French and English accented German. We conducted perception experiments with Swiss German listeners and used three different types of signal degraded speech to draw subjects' attention to different temporal and rhythmic prosodic aspects: (a) amplitude envelope timing characteristics (in noise vocoded speech, cf. Shannon *et al.*, 1995), (b) segment durations (in 1-bit requantised speech), and (c) the durational variability of voiced and unvoiced intervals (in *sasasa*-speech, cf. Ramus & Mehler, 1999; Fourcin & Dellwo, 2009). Each of three listener groups are presented with one of the signal conditions and had to decide, for each stimulus, whether they heard a French or an English accent. The three signal conditions are severely degraded in the frequency-domain and each preserve primarily one particular type of durational characteristic. We thus examine whether listeners can distinguish foreign accents if their perception is restricted to temporal characteristics of speech. We expect that some types of temporal cues will lead to higher recognition rates than others.

## 2. Experiment 1: Can listeners recognise the dialectal origin of a foreign accent?

If listeners are familiar with the native language of a speaker, they are often able to identify this native language only by hearing the speaker's L2 speech (cf. Boula de Mareuil *et al.*, 2008). Can listeners also differentiate between dialectal foreign accents, e.g. between Bern dialect-accented Standard German/French and St. Gallen-accented Standard German/French? Our Bern speakers come from the city of Bern and our St. Gallen speakers from the city of St. Gallen. The varieties spoken in the cities of Bern and St. Gallen stand for a western and an eastern Swiss German dialect. The two varieties differ in a number of linguistic and, in particular, phonetic variables (cf. Kolly, 2010, 2013).

### 2.1 Method

#### 2.1.1 Subjects

Our within-subject design involved a group of 60 native Swiss German and 20 native Swiss French listeners. 46 of the Swiss German subjects were students from Bern University, 14 from Zurich University. This listener group was assumed to have a comparable knowledge of Standard German and French due to school education in Switzerland: In Swiss German

primary schools, Standard German is introduced at the beginning of primary school and French is learned as a first L2. Therefore, Swiss German listeners were also assumed to have a comparable experience with Swiss German-accented Standard German and French. The Swiss French listeners were students from the School of Business Administration<sup>2</sup> in Fribourg. This listener group was assumed to have similar knowledge of Standard German due to school education in Switzerland, where Standard German is learnt as a first L2 in Swiss French primary schools. Furthermore, the Swiss French listeners were assumed to have a comparable experience with Swiss German-accented French due to their similar education, which implies similar chances for contact with Swiss German native speakers. Subjects' age ranged from 18 to 31 years. None of the listeners reported any significant problems with hearing or sight.

### 2.1.2 Material

Speech was collected from eight speakers: four native speakers of Bern Swiss German and four native speakers of St. Gallen Swiss German (two males and two females each). All speakers, as well as their parents, grew up and lived in their respective cities and all of them had higher education, i.e., comparable proficiency in Standard German and French. Speakers' age ranged between 21 and 28. During the perception experiment, speakers were rated for accent degree on a five-point scale for a related investigation (cf. Kolly, 2010, 2011).<sup>3</sup> Accent degree ranged between 2.77 and 3.92 in Standard German speech and between 2.44 and 4.06 in French speech.

Speakers read a short text in Standard German and French (the fable *The Northwind and the Sun*, cf. The International Phonetic Association, 1999/2003: 81, 89). They also spoke spontaneously about their morning routine. Before the recordings, speakers did not familiarise themselves with the material. Recordings took place in a quiet room in their respective home or office, with a *Fostex FR-2LE* solid-state recorder (sampling rate of 48kHz, 16-bit quantisation) and a *Sennheiser* clip-on *MKE 2p-c* microphone. We thus collected 4 samples per speaker: read and spontaneous speech in Standard German and in French (mean duration: 40 s for read German samples, 37 s for spontaneous German samples, 53 s for read French samples, 51 s for spontaneous French samples). All of those samples were used as stimuli to construct a Standard German and a French perception experiment, containing 16 stimuli each (8 speakers \* 2 speaking styles).

---

<sup>2</sup> Tertiary institution.

<sup>3</sup> 1 = no accent; 2 = rather no accent; 3 = slight accent; 4 = clearly perceivable accent; 5 = strong accent.

### 2.1.3 Procedure

Swiss German listeners were presented with the stimuli of the Standard German as well as the French perception experiment (i.e., 32 stimuli) in a classroom at Bern resp. Zurich University. Swiss French listeners were presented with the French perception experiment only (i.e., 16 stimuli), in a classroom at the School of Business Administration in Fribourg. Stimuli were presented over loudspeakers. They were presented in two blocks, where the Standard German experiment was conducted before the French experiment. Within the blocks, stimuli were presented in a randomised order.

Listeners submitted their responses in a paper and pencil setting. For each stimulus, listeners had to guess the dialectal origin of the speaker. Concerning this task, two options were considered: an alternative forced choice task, where listeners have to attribute each stimulus to either a Bern or a St. Gallen dialectal accent; and an open response task, where listeners have no knowledge about which or how many different Swiss dialectal accents are represented in the stimuli and take their guesses in a completely open manner. Therefore, a pilot experiment was conducted with Swiss German as well as Swiss French listeners. Some listeners were presented the stimuli with an alternative forced choice task, others with an open question. The pilot showed a ceiling effect for Swiss German listeners in the alternative forced choice task. The open response task was thus chosen for this group of listeners. The Swiss French listeners, however, were not able to perform the open response task in the pilot experiment: in most cases, all fields were left blank. Therefore, the alternative forced choice task was chosen for Swiss French listeners. This difference in tasks depending on listener group entails the need for two different analysis methods.

The forced alternative choice task used for the Swiss French listeners represents the typical Bernoulli trial with two possible outcomes: success (correct identification) and failure (false response). We thus applied one-tailed binomial tests with an alpha-level of 0.05. For between-condition comparisons we used paired Wilcoxon signed rank tests, since the small sample does not allow the assumption of normally distributed data. The open question used for the Swiss German listeners, however, yields a variety of different responses that cannot be analysed as easily. Results are thus presented in a descriptive framework. Responses were categorised according to a careful analysis of the Swiss phonetic landscape (for details cf. Kolly 2010, 2013). We have created three decreasingly "strict" categories to be used for descriptive results, as presented in Table 1 and illustrated in Figure 1. A distinct categorisation of continuous dialectal spaces, as used here, is obviously an artificial

construct (cf. Christen, 2010: 269–272). Nevertheless, based on the work presented in Kolly (2010, 2013), this categorisation can approximate the reality of continuous dialectal spaces in a way that is systematic and based on extensive research about the Swiss dialectal landscape – including the Atlas of German-speaking Switzerland (cf. *Sprachatlas der Deutschen Schweiz* 1962–2003; Bachmann, 1908; Hotzenköcherle, 1961, 1984; Haas, 1985, 2000; Siebenhaar, 1994; Leemann & Siebenhaar, 2008).

	Accepted responses for Bern dialectal accent	Accepted responses for St. Gallen dialectal accent
"narrow" category	BE	SG
"middle" category	BE SO, FR	SG TG, SH, AP, (north-)eastern Switzerland
"broad" category	BE SO, FR BA, LU, AG, VS, UR, western Switzerland, western midland	SG TG, SH, AP, (north-)eastern Switzerland, ZH, GR, GL, SZ eastern Switzerland, eastern midland

Table 1: Categorization of responses using the abbreviations for names of Swiss cantons (an administrative entity usually associated with broad dialect groups, cf. Christen, 2010: 273–281; Kolly, 2010)

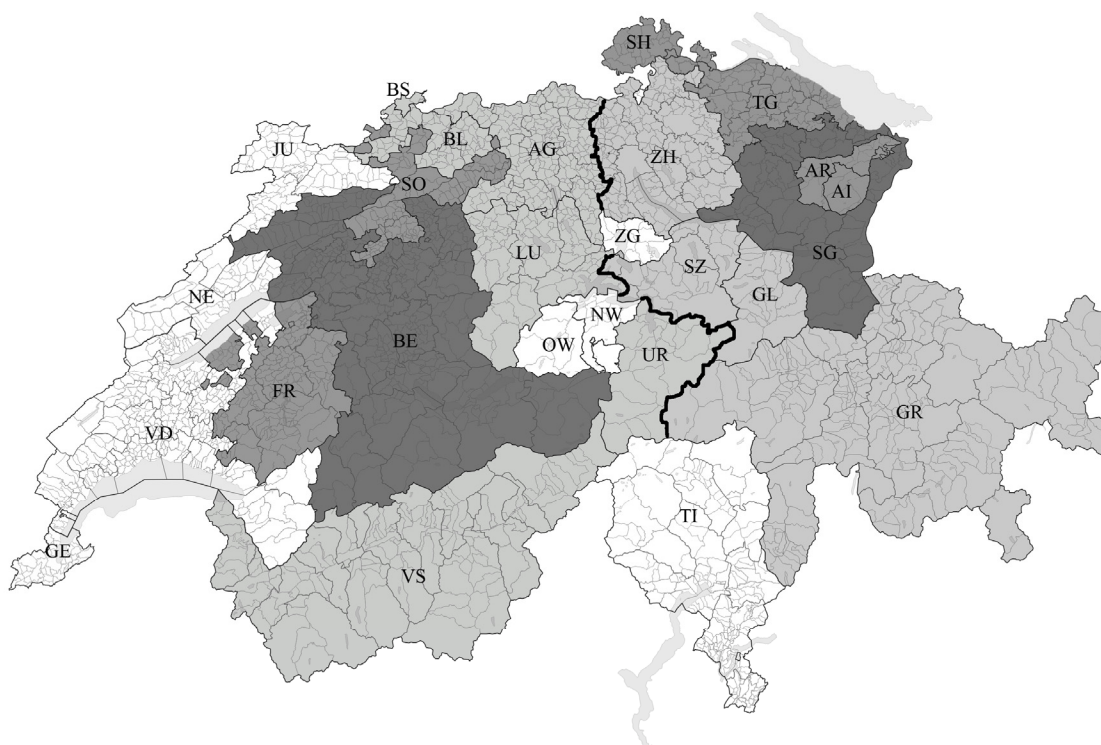


Figure 1: Categories "narrow", "middle" and "broad" for the west/east contrast (thick line), coded by increasingly light shades of grey

## 2.2 Results

Descriptive results from the perception experiment with Swiss German listeners are presented in figures 2–4. The *id*-sectors in the pie charts are based on the "middle" category described above. Added to the sectors *grey area*, they represent the "broad" category, i.e. listeners' ability to identify a western/eastern dialectal accent in the speech material (see section 2.1.3). Figure 2 shows the identification rates for all speakers; Figures 3 and 4 present the same for Bern and St. Gallen speakers separately.

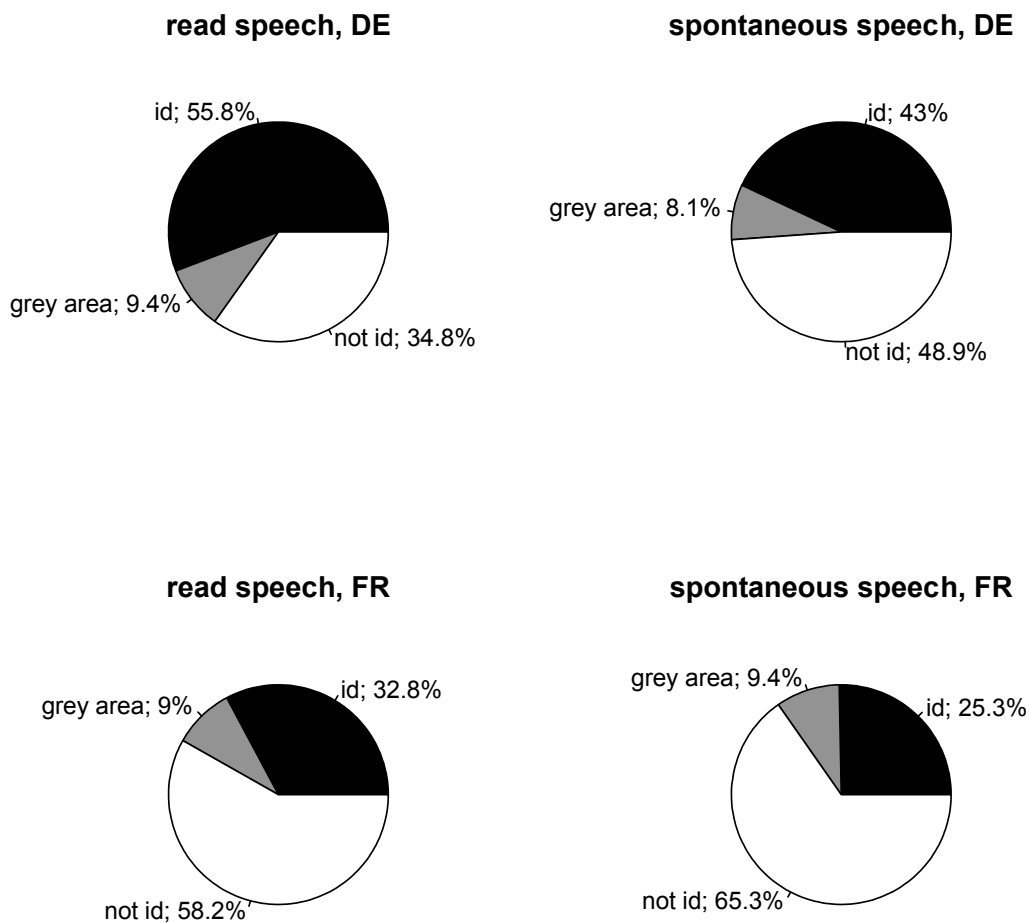


Figure 2: Average recognition rates over all stimuli per language and speaking style; *id* = correctly identified as being from BE/FR/SO (Bern speakers) or north-eastern Switzerland (St. Gallen speakers); *id* + *grey area* = correctly identified as being from western (Bern speakers) or eastern (St. Gallen speakers) Switzerland; *not id* = not correctly identified

We observe an average dialect recognition rate over both speaking styles of 49.4% in Standard German speech and of 29.1% in French speech (cf. Figure 2, sectors *id*). If we consider listeners' general ability to identify



western vs. eastern dialectal accents, we obtain identification rates of 58.2% in Standard German and 38.8% in French speech (cf. Figure 2, sectors *id* + *grey-area*).

The average recognition scores over both speaking styles in Standard German speech are similar for Bern (50.25%) and St. Gallen (48.6%) accents; however, the accent recognition scores differ for French speech (Bern: 32.6%, St. Gallen: 25.5%; cf. Figures 3, 4, sectors *id*).

Further, we observe an average accent recognition rate over both L2s of 45.1% for read speech and 29.1% for spontaneous speech.

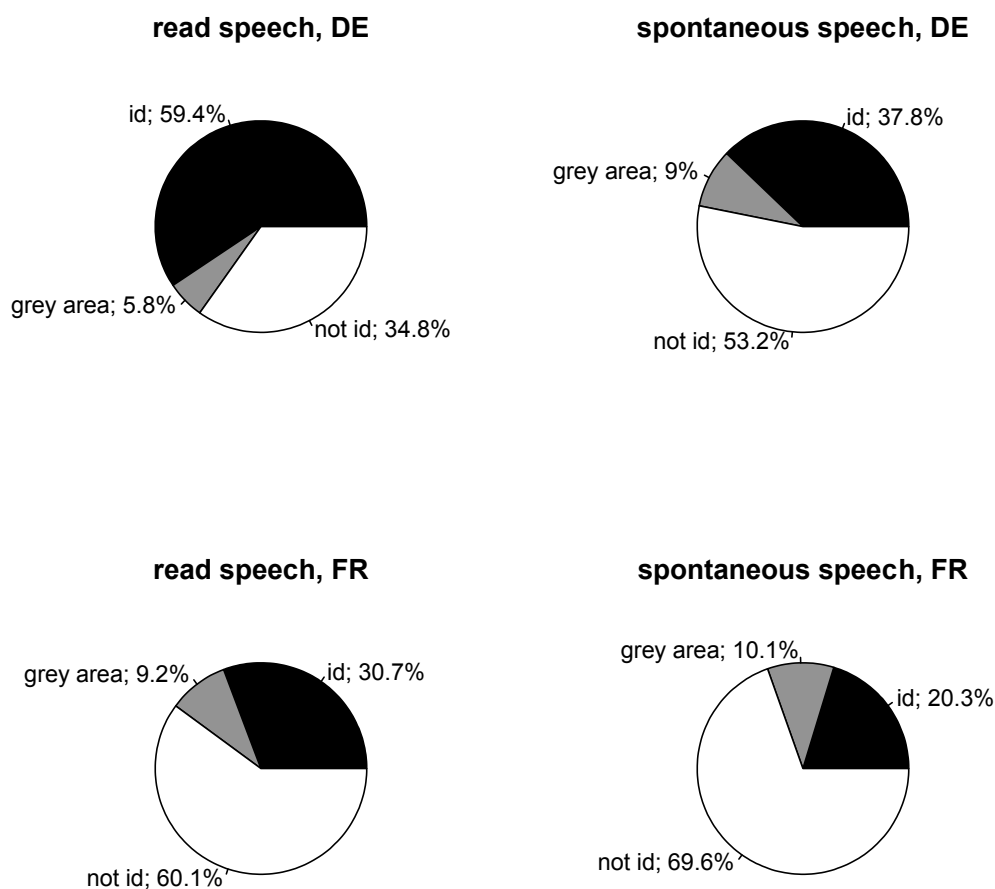


Figure 3: Average recognition rates over all St. Gallen-accented stimuli per language and speaking style; *id* = correctly identified as being from north-eastern Switzerland; *id* + *grey area* = correctly identified as being from eastern Switzerland; *not id* = not correctly identified

A one-tailed binomial test shows that Swiss French listeners can significantly discriminate the two dialectal accents ( $p < 0.01$ ). As for the Swiss German listeners, the mode is always the correct response.

A paired two-sample Wilcoxon test shows that there is no significant difference in the recognition of the Bern and the St. Gallen accent by Swiss French listeners.

Another paired two-sample Wilcoxon test shows that there is no significant difference between accent recognition in read and in spontaneous speech by Swiss French listeners.

In both the Swiss German and the Swiss French listener group there is considerable variability between the recognition scores of particular speakers. This variability seems to rely to some part on accent degree (cf. Kolly, 2011).<sup>4</sup>

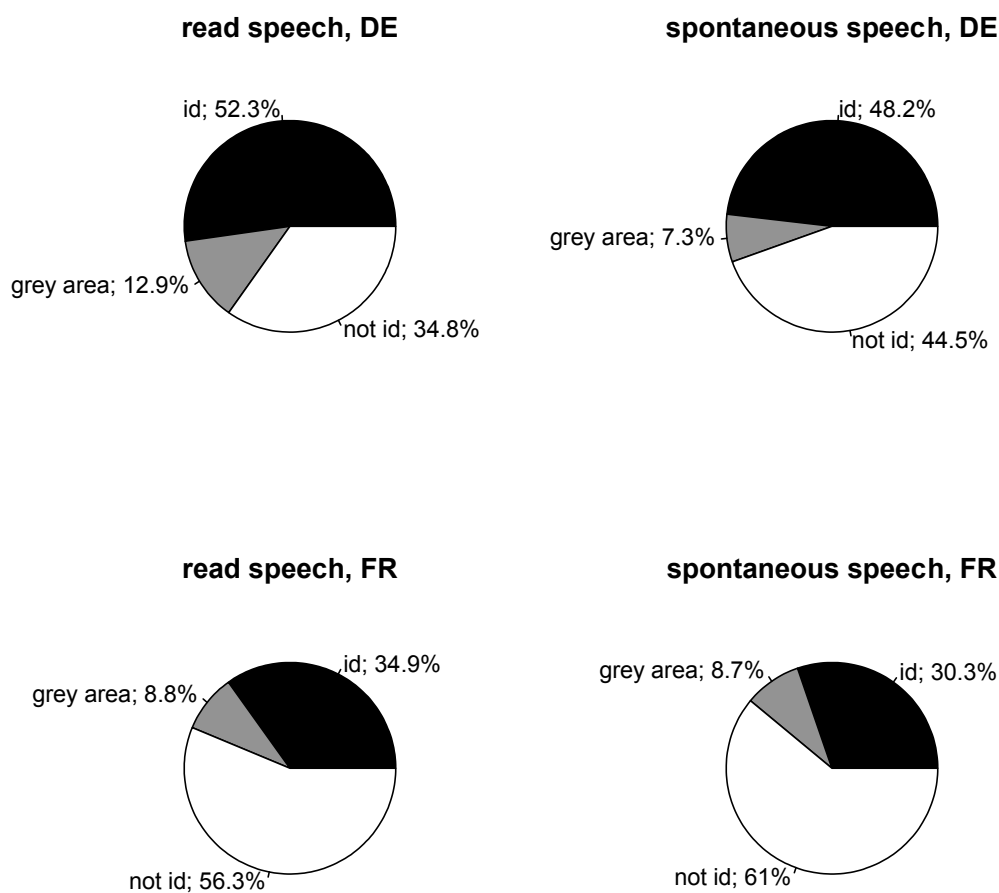


Figure 4: Average recognition rates over all Bern-accented stimuli per language and speaking style; *id* = correctly identified as being from BE/SO/FR; *id* + grey area = correctly identified as being from western Switzerland; not *id* = not correctly identified

<sup>4</sup> Speaker-specific results and a discussion of the reasons why particular speakers were easier or more difficult to identify are presented in Kolly (2010, 2013).

### 2.3 *Discussion*

We have tested whether Swiss German and Swiss French listeners are able to perceive the dialectal origin of Bern and St. Gallen speakers in accented Standard German and French L2 speech. Because of their different experience with Swiss German-(accented) speech, different response tasks were designed for the Swiss German and the Swiss French listeners (cf. 2.1.3). Therefore, two different methods of analysis had to be used and results have to be read with the experimental procedure in mind: Swiss German listeners had no knowledge of which and how many Swiss German dialectal accents were represented in the material (open response) – an alternative forced choice task would obviously have yielded different identification scores. Swiss French listeners had to respond in an alternative forced choice task. Results show that the dialectal accents are recognised not only in the Standard German, but also in the French speech material.

The hypothesis that dialectal accents of Swiss German speakers can be identified in Standard German speech is confirmed for the two dialects at hand. This result is in line with Guntern (2011). On average, dialectal accents in Standard German speech are correctly identified by about 50% of our Swiss German listeners when considering the above defined "middle" category (i.e., accepted responses: dialects of Bern, Solothurn, Fribourg for Bern accented stimuli; north-eastern Switzerland for St. Gallen accented stimuli). Dialectal accents in French stimuli were correctly identified by 30% of the Swiss German listeners; they were significantly discriminated by Swiss French listeners.

We thus note that Swiss German listeners reach higher recognition scores when hearing dialectal accents in Standard German speech than in French speech. This was to be expected, since Swiss German listeners have more experience Standard German spoken by Swiss German dialect speakers than with French spoken by Swiss German dialect speakers.

The result for the Swiss German listener group is remarkable, as very subtle accent distinctions could be recognised. However, it is in line with the fact that dialects occupy an important role for the identity of many Swiss German people and are a frequent topic of conversation among them. Swiss dialects (as opposed to Standard German) are the common variety used in conversational situations and, other than dialects in other linguistic regions, they are the prestige, not the stigma variety in the Swiss diglossic situation (cf. Hotzenköcherle, 1984; Werlen, 2000; Haas, 2004; Christen, 2010). However, given the literature about the contact situation between Swiss French and Swiss German people as well as the attitudes of many Swiss French people towards the acquisition of German (cf. 1, Muller, 1998; Fuchs

& Werlen, 1999), the result for the Swiss French listener group is even more remarkable.

The distinction of a more global western or eastern dialectal accent in Standard German speech is recognised by almost 60% of the Swiss German and 40% of the Swiss French listeners. This illustrates the prominence and the perceptual salience of an east/west divide in the Swiss dialectal and, in particular, phonetic landscape. Compared to dialect recognition rates of about 36% as described by Bauvois (1996: 300f.) for regional Belgian French accents, many of the above described rates are surprisingly high. Bauvois (1996) uses a similar method since her listening task involves an open question. A comparison with further accent perception studies like the ones described by Boula de Mareuil *et al.* (2008), Guntern (2011) and others is delicate because of their different experimental task (cf. 2.1.3).

Both accents reach similar recognition scores in Standard German speech; in French speech, however, the Bern dialectal accent is easier to identify than the St. Gallen accent, for Swiss German listeners. This is not the case for Swiss French listeners: no significant difference in recognition rates between the two accents is observed. The result observed in the Swiss German listener group may have to do with the fact that the St. Gallen dialect shares more phonetic features with French than the Bern dialect (typically vowel qualities and the /r/-sound, see Kolly, 2010, 2013). We hypothesise that a Bern accent in French speech sounds more salient than a St. Gallen accent and is thus easier to recognise. However, this result is bound to the "middle" category presented in the figures. If we take look at the "narrow" category that only accepts responses containing "Bern" or "St. Gallen", the Bern accent yielded higher recognition scores: Bern stimuli often triggered the precise response "Bern", whereas the St. Gallen stimuli mostly provoked responses like "Nordostschweiz" 'north-eastern Switzerland' or "Ostschweiz" 'eastern Switzerland'. Here we have to indicate the overrepresentation of listeners from Bern University which might entail a bias: listeners perceive more differences in varieties that are linguistically close to their own – "aus der Ferne dagegen mögen Sprecherinnen und Sprecher aus dem Schaffhausischen und dem Thurgau recht ähnlich in den Ohren klingen, was die Betroffenen selbst natürlich ganz anders sehen..." (Christen, 2005: 21). A further explanation is that the (dialectologically very diverse) canton of Bern seems to represent a single dialect in the mental representations of many Swiss Germans, whereas the (north-)eastern dialect varieties are perceived as a unity.<sup>5</sup>

Accents are easier to identify in read than in spontaneous speech, for Swiss German listeners. The result is in line with Kolly (2011), who found that read

---

<sup>5</sup> Cf. Christen (2010: 277–278) for the special status of the category 'eastern Switzerland' that is often used in a similar way as canton names for referring to dialects.

speech is perceived as being more accented than spontaneous speech. For the French listener group, however, there is no significant difference in recognition rates between the two speaking styles. The fact that read speech allowed higher accent identification rates for Swiss German listeners possibly has to do with the controlled speech material that allowed for better comparison between stimuli. Also, stimulus duration was usually higher in the read samples, giving time for potentially more acoustic accent cues to occur during stimulus presentation. Further, speakers might have been cognitively more involved when reading a text than when speaking spontaneously – what might have left less resources for the phonetic implementation of L2 speech.

The fact that we observe no difference between recognition rates for St. Gallen and Bern accents or in spontaneous and read speech in our Swiss French listener group might have to do with the different methods of analysis. Also, Swiss French listeners might perceive less detail information in Swiss German-accented speech, since they have less experience in the processing of Swiss German-accented speech than Swiss German listeners.

The accent recognition scores vary considerably between speakers. This is related to speakers' accent degree to some part (cf. Kolly, 2010, 2013). To another part, there must be certain acoustic accent cues (present in some speakers' L2 speech) that are more salient to listeners than other cues (present in other speakers' L2 speech). In fact, certain speakers' accent was well recognised by Swiss German as well as Swiss French listeners. Other speakers' accent was well recognised by Swiss German, but not by Swiss French listeners – and vice versa. Therefore, we hypothesise that French listeners focus on different acoustic cues than Swiss German listeners, when they perceive and categorise accented French speech. More research is needed to explore the types and combinations of acoustic cues that are salient indicators of a particular accent to a particular listener group. In the following experiment, we explore the perceptual importance of different durational cues for the recognition of English- and French-accented German by Swiss German listeners.

### **3. Experiment 2: how do listeners perceive L1 of an L2 speaker – do temporal characteristics play a role?**

As we have established that listeners can identify very subtle accent distinctions in Experiment 1, we now turn to the question on which basis listeners take such perceptual decisions. As explained in 1, we investigate the recognisability of French- and English-accented German in cases where listeners are restricted to primarily temporal cues. Three different types of signal degradations are used in order to explore the perceptual

salience of three types of temporal cues in relation with the foreign accents at hand.

In noise vocoded speech, spectral information is removed from the signal and replaced by band-limited noise. The resulting signal is strongly degraded in the frequency-domain and does not contain any vocal fold vibration;

- durational characteristics of voicing are absent from the signal
- segmental durations are not or hardly perceivable
- subjects' attention is drawn to amplitude envelope temporal characteristics like syllable beats: so-called low-frequency temporal cues.

In monotone 1-bit requantised speech, the amplitude of every sample is set to either 0 or -1. The resulting signal is strongly degraded in the frequency-domain and does not contain any intonational information;

- amplitude information is absent from the signal
- subjects' attention is drawn to segment durations and the durational variability of unvoiced and voiced intervals: so-called high-frequency temporal cues.

In monotone *sasasa*-speech based on voiced and unvoiced intervals, every unvoiced sound is replaced with a synthesised [s] and every voiced sound with a synthesised [a]. The resulting signal does not contain any original frequency-domain information;

- original amplitude information is absent from the signal
- segmental durations are not perceivable since voiced/unvoiced sounds have been merged to voiced/unvoiced intervals
- subjects' attention is drawn to cues about voice timing only.

The signal degraded sentences are unintelligible to the listeners. However, if presented with the corresponding lexical information, listeners learn to parse degraded speech (cf. Davis *et al.*, 2005).

### 3.1 Method

#### 3.1.1 Subjects

Our between-subject design involved three groups of ten listeners each, all of which were native speakers of Swiss German dialects. Most of them were students from Zurich University, some students from other Swiss Universities. The subjects were assumed to have similar knowledge of French and English due to school education in Switzerland: in Swiss German primary schools, French is learned as a first and English as a second L2. Due to their higher education, our listeners were also assumed to have a comparable experience with French- and English-accented Standard German. In a multilingual country like Switzerland and, in

particular, in Swiss Universities, opportunities for contact with L2 speakers of German are frequent. Subjects' age ranged from 19 to 32 years. None of the listeners reported any significant problems with hearing or sight.

### 3.1.2 Material

Speech was collected from twelve speakers: six native speakers of French and six native speakers of English (three males and three females each). All the French speakers grew up and live in the French speaking part of Switzerland (five in the canton of Fribourg, one in the canton of Vaud). The English speakers grew up in the US or in Canada, one female speakers in the UK; all of them were students or staff members of Zurich University. Speakers' age ranged between 23 and 56. Their self-estimated proficiency in German ranged from B1 to B2 for the French speakers and from A1 to B2 for the English speakers (cf. Council of Europe, 2011).

Speakers read a list of 19 German sentences (cf. Appendix). Sentences were taken from a list of Italian sentences used by Nazzi *et al.* (1998) and translated to German. Before the recording, they familiarised themselves with the material by reading the sentences aloud. Speakers were recorded in a quiet room at Zurich University or in their respective homes with a *Fostex FR-2LE* solid-state recorder (sampling rate of 48kHz, 16-bit quantisation) and a *Sennheiser* clip-on *MKE 2p-c* microphone. If filled pauses occurred during a sentence, speakers repeated the sentence spontaneously or, if not, they were asked to do so. nine sentences per speaker were chosen for each of three experimental conditions to contain 108 stimuli. We have used a different combination of sentences from each speaker such that each of the 18 used sentences appears six times in the experiment: three times spoken native speakers of French and three times by native speakers of English.

Manipulated stimuli were created using Praat signal processing software (Boersma & Weenink, 2012).<sup>6</sup>

- Noise vocoded speech was obtained by bandpass filtering each sentence between 50 Hz and 8000 Hz. The signal was then divided into 6 logarithmically spaced frequency bands by bandpass filtering with cutoff frequencies at 50 Hz, 116.5 Hz, 271.4 Hz, 632.5 Hz, 1473.6 Hz, 3433.5 Hz and 8000 Hz. These cutoff frequencies were subsequently used to filter white noise in order to obtain six noise bands. The amplitude envelope of each speech band was extracted by half-wave rectification and low-pass filtering at 10 Hz. Each

---

<sup>6</sup> Praat scripts for delexicalisation were written by the second author and are available at <http://www.pholab.uzh.ch/leute/dellwo/software.html>.

amplitude envelope was then multiplied with the corresponding noise band and, finally, the six noise signals were added together.

- Monotone 1-bit requantised speech was obtained by first creating a flat pitch line: every pitch point of a sentence was replaced by the mean pitch value of the sentence. The amplitude value of each sample was then set to 0 (for amplitude values  $> 0$ ) or to -1 (for amplitude values  $< 0$ ). The quantisation rate of the signal was thus converted to 1-bit.
- Monotone *sasasa*-speech was created with the Praat plug-in tool *Sasasa delexicaliser* (cf. Dellwo, accepted, for a more detailed description) and based on voiced and unvoiced intervals (cf. Fourcin & Dellwo, 2009) instead of the method used by Ramus & Mehler (1999). The latter constructed *sasasa*-speech by transforming every consonantal interval of the speech signal to [s] and every vocalic interval to [a]. We chose to use unvoiced and voiced instead of consonantal and vocalic intervals and thus preserve only voice timing characteristics from the original sound signal.

### 3.1.3 Procedure

Three groups of ten native Swiss German listeners were presented the 108 stimuli in a randomised order on a laptop computer. Listeners were tested in a quiet room at university or in their own homes. Groups of ten listeners each were presented either noise vocoded, 1-bit-requantised or *sasasa*-speech over high-quality earphones. For each stimulus, the corresponding sentence was presented visually on the computer screen about two seconds preceding the acoustic stimulus and until the acoustic presentation ended. Thus, subjects had access to lexical information while listening to the delexicalised stimulus and could concentrate on the accent cues relevant for the research question. For each stimulus listeners had to indicate whether they had heard German with a French or an English accent by clicking on the corresponding button, using a computer interface created in Praat.

As a sensitivity measure we have chosen  $d'$  from signal detection theory (cf. Swets & Green, 1966). The measure  $d'$  obtains the sensitivity of each listener, eliminating response bias, where perfect sensitivity (i.e., perfect discrimination of both types of signals) starts at a  $d'^2$ -value of 4 and a  $d'$ -value of 0 indicates sensitivity at chance level.

## 3.2 Results

Figure 5 as well as one-sample t-tests based on  $d'$  show that French and English accents could be identified above chance based on 6-band noise



vocoded ( $p < 0.001$ ) and 1-bit-requantised ( $p < 0.001$ ) stimuli, but not based on *sasasa*-speech.

We further computed a univariate ANOVA which shows a significant effect between conditions ( $F[2, 30] = 50.58$ ;  $p < 0.001$ ). Tukey post-hoc tests show that all group comparisons are highly significant ( $p < 0.001$ ). In particular, recognition scores were higher in 1-bit requantised than in 6-band noise vocoded speech.

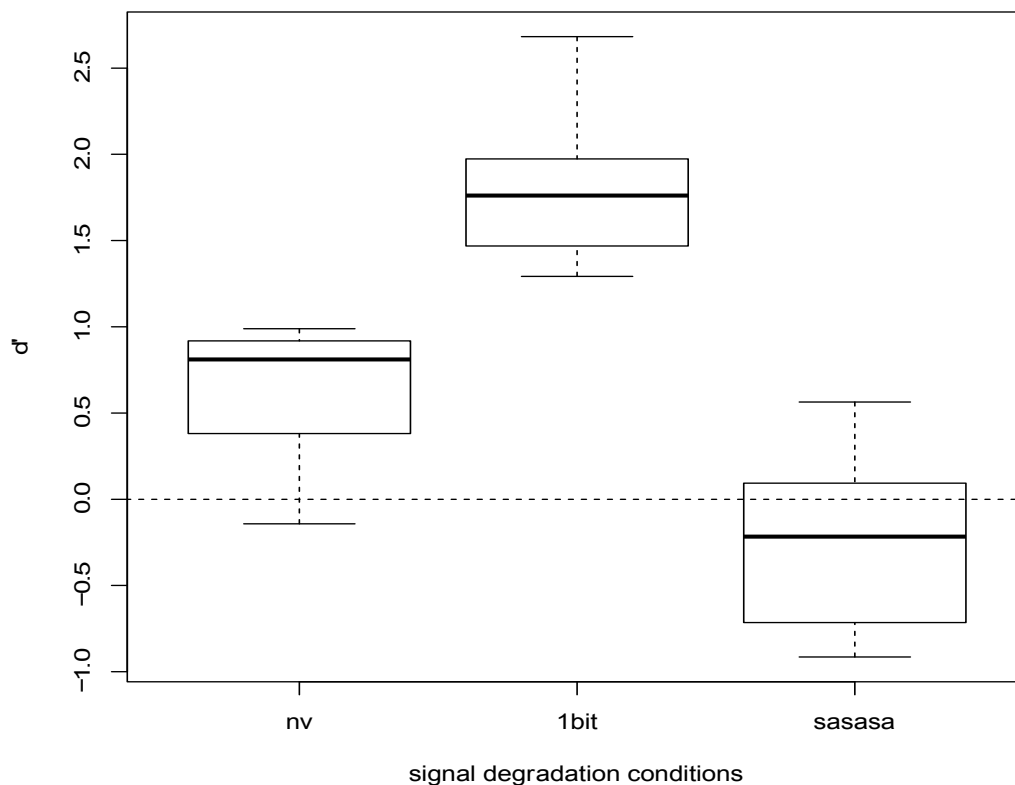


Figure 5: Perceptual identification of a French or English accent in delexicalised German speech. The dashed line indicates performance at chance.

### 3.3 Discussion

We have tested whether Swiss German listeners can distinguish French-accented German from English-accented German when presented with speech signals that are heavily degraded in the frequency-domain. An alternative forced choice task was conducted with three groups of listeners, each presented with a different type of delexicalised speech. Signal types each contained a different type of durational characteristic. Results show that listeners can discriminate English- from French-accented German based on primarily temporal cues. Further, listeners' performance depends essentially on the type of delexicalisation applied, that is, on the type of temporal characteristic conveyed by the particular signal.

The present experiment showed that French- and English-accented German speech can be identified above chance in 6-band noise vocoded as well as in 1-bit requantised speech. *Sasasa*-speech based on voiced and unvoiced intervals did not allow this distinction. Primarily temporal cues and, in particular, the absence of amplitude or even vocal fold vibration are sufficient to identify French and English accents in German L2-speech.

The type of temporal information contained in the delexicalised stimuli differs for each condition. In noise vocoded speech, subjects' attention is drawn to amplitude envelope temporal characteristics or syllable beats: so-called low-frequency temporal cues. 1-bit requantised speech, on the other hand, draws subjects' attention to high-frequency temporal cues like segment durations. *Sasasa*-speech based on voiced and unvoiced intervals renders cues about voice timing only. Since 1-bit requantised speech was better recognized in the experiment, listeners possibly rely more on segment durations than on lower frequency (and possibly rhythmic) cues for accent recognition when listening to French- or English-accented German. Cues about voice timing alone are not sufficient to solve this perceptual task.

#### 4. General discussion and Outlook

The experiments reported in the present article show that (a) listeners can discriminate very subtle accent distinctions and (b) the time-domain is relevant for the recognition of such foreign accents.

Acoustic correlates of (a) were assumed to lie in segmental as well as in prosodic aspects, since recent research stresses prosodic differences between Swiss German dialects (cf. e.g. Leemann & Siebenhaar, 2008; Leemann *et al.*, 2012; Leemann, 2012). As prosodic deviances seem to be a widely discussed feature of foreign accents, we investigated temporal and rhythmic cues for foreign accent identification and showed that speech can be strongly degraded in the spectral domain and still provide enough cues for listeners to identify a French or an English accent in German sentences. Furthermore, we have illustrated that the different types of durational cues contained in the different types of delexicalised speech yield different identification rates, with segment durations being the most effective cue to the accent recognition task conducted for the present paper.

Future research will further explore the amount of frequency-domain and time-domain information needed for listeners to recognise foreign accents. To this end, we use different types of signal manipulations such as noise vocoded speech with different numbers of bands. We expect these further conditions to tell us more about the perceptual processing of foreign accented speech and, in particular, about the possibility of identifying foreign accents based on temporal characteristics only.

We expect that our research will have implications on a variety of fields such as forensic phonetics: our aim is to better explain foreign accented speech, which may be applied in areas like the analysis of speaker origin (LADO) or speaker identity. The research might further contribute to the field of L2-acquisition where knowledge of prosody and rhythm in particular could help learners acquire a more native-like pronunciation.

## Acknowledgements

We would like to thank all our speakers and listeners for their contribution to this article. Further we thank Stephan Schmid and Adrian Leemann (Experiment 2) and Elke Hentschel (Experiment 1) for helpful comments on the experiment design as well as two anonymous reviewers for their valuable feedback on a first version of this manuscript. We also thank Stephan Schmid for the translation of Italian sentences. Part of this work (Experiment 2) was supported by the Swiss National Science Foundation (SNF; grant number: 100015\_135287).

## Bibliography

- Arvaniti, Amalia (2012): The usefulness of metrics in the quantification of speech rhythm. In *Journal of Phonetics*, 40, (2), 351–373.
- Atterer, M. & Ladd, R. (2004): On the Phonetics and Phonology of ‘Segmental Anchoring’ of F<sub>0</sub>. Evidence from German. In *Journal of Phonetics*, 32, (2), 177–197.
- Bachmann, A. (1908): [Schweizer-]Sprachen und Mundarten. In Knapp, C./Borel, M. (eds.), *Geographisches Lexikon der Schweiz*. Vol. 5. Neuchâtel (Attinger), 58–76.
- Bauvois, C. (1996): Parle-moi, et je te dirai peut-être d'où tu es. In *Revue de Phonétique Appliquée*, 121, 291–309.
- Boersma, P. & Weenink D. (2012): Praat: doing phonetics by computer. (<http://www.praat.org>).
- Boula de Mareuil, P./Vieru-Dimulescu, B. (2006): The contribution of prosody to the perception of foreign accent. *Phonetica*, 63, (4), 247–267.
- Boula de Mareuil, P., Vieru-Dimulescu, B., Woehrling, C. & Adda-Decker, M. (2008): Accents étrangers et régionaux en français. Caractérisation et identification. In *Traitement Automatique des Langues*, 49, (3), 135–163.
- Bush, C. N. (1967): Some Acoustic Parameters of Speech and Their Relationships to the Perception of Dialect Differences. *TESOL Quarterly*, 1, (3), 20–30.
- Christen, H. (2005): 'Tour de Suisse' der Deutschschweizer Dialekte. In *Forum Helveticum* (ed.), *Dialekt in der (Deutsch)Schweiz – Zwischen lokaler Identität und nationaler Kohäsion. Le dialecte en Suisse (alémanique) – entre identité locale et cohésion nationale*. Lenzburg (*Forum Helveticum*), 21–25.
- (2010): Was Dialektbezeichnungen und Dialektattributionen über alltagsweltliche Konzeptualisierungen sprachlicher Heterogenität verraten. In Anders, C. A., Hundt, M. & Lasch, A. (eds.), "Perceptual dialectology". *Neue Wege der Dialektologie*. Berlin/New York (de Gruyter), 269–290.

- Council of Europe (2011): Common European Framework of Reference for Languages: Learning, Teaching, Assessment. ([http://www.coe.int/t/dg4/linguistic/Source/Framework\\_EN.pdf](http://www.coe.int/t/dg4/linguistic/Source/Framework_EN.pdf)).
- Cunningham-Andersson, U. & Engstrand, O. (1987): Perceived strength and identity of foreign accent in Swedish. In *Phonetica*, 46, 138–154.
- Davis, M./Johnsrude, I./Hervais-Adelman, A./Taylor, K./McGettigan, C. (2005): Lexical Information Drives Perceptual Learning of Distorted Speech: Evidence From the Comprehension of Noise-Vocoded Sentences. *Journal of Experimental Psychology*, 134, (2), 222–241.
- Dellwo, V. (accepted): Acoustic correlates of speech rhythm: Are consonantal and vocalic intervals or syllables the more salient units?. To appear in: Van de Vijver, R. & Vogel, R. (eds.), *Rhythm in Phonetics, Grammar and Cognition* Berlin (De Gruyter).
- Flege, J. E. (1992): Speech Learning in a Second Language. In Ferguson, C. A., Menn, L. & Stoel-Gammon, C. (eds.), *Phonological Development. Models, Research, Implications*. Maryland (York Press), 565–604.
- Fourcin, A. & Dellwo, V. (2009): Rhythmic classification of languages based on voice timing. London (UCL Eprints) (<http://eprints.ucl.ac.uk/15122/>).
- Fuchs, G./Werlen, I. (1999): Zweisprachigkeit in Biel-Bienne. Untersuchung im Rahmen des Bieler-Bilinguismus-Barometers 1998. Im Auftrag des Forums für Zweisprachigkeit Biel-Bienne. Bericht erstellt auf Basis der Daten des GfS-Forschungsinstituts Bern. Biel (Stiftung Forum für die Zweisprachigkeit).
- Green, D. M. & Swets, J. A. (1966): *Signal detection theory and psychophysics*. New York (Wiley).
- Guntern, M. (2011): Erkennen von Dialekten anhand von gesprochenem Schweizerhochdeutsch. In *Zeitschrift für Dialektologie und Linguistik*, 78, 155–187.
- Haas, W. (1985): Sprachgeographie und Variationstheorie. In Werlen, I. (ed.), *Probleme der schweizerischen Dialektologie*. 2. Kolloquium der Schweiz. Geisteswissenschaftlichen Gesellschaft. *Problèmes de la dialectologie suisse*. 2e Colloque de la Société suisse des sciences humaines. Fribourg (Editions Universitaires Fribourg), 95–108.
- (2000): Die deutschsprachige Schweiz. In: Bickel, H. & Schläpfer, R. (Hrsg.): *Die viersprachige Schweiz*. Aarau/Frankfurt am Main/Salzburg (Sauerländer), 57–138.
- (2004): Die Sprachsituation der deutschen Schweiz und das Konzept der Diglossie. In: Christen, H. (ed.), *Dialekt, Regiolekt und Standardsprache im sozialen und zeitlichen Raum*. Beiträge zum 1. Kongress der Internationalen Gesellschaft für Dialektologie des Deutschen, Marburg/Lahn 5.–8. März 2003. Wien (Praesens), 81–110.
- Häcki Buhofer, A./Burger, H. (1998): Wie Deutschschweizer Kinder Hochdeutsch lernen. Der ungesteuerte Erwerb des gesprochenen Hochdeutschen durch Deutschschweizer Kinder zwischen sechs und acht Jahren. Stuttgart (Steiner). (= *Zeitschrift für Dialektologie und Linguistik*; Beihefte 98).
- Hirschfeld, U. & Trouvain, J. (2007): Teaching Prosody in German as a Foreign Language. In Trouvain, J. & Gut, Ulrike (eds.), *Non-Native Prosody. Phonetic Description and Teaching Practice*, Berlin/New York (de Gruyter), 171–187.
- Hotzenköcherle, R. (1961): Zur Raumstruktur des Schweizerdeutschen. In *Zeitschrift für Mundartforschung*, 28/3, 207–227.
- (1984): *Die Sprachlandschaften der deutschen Schweiz*. Ed. by Bigler, N. & Schläpfer, R. Aarau etc. (Sauerländer).
- Jilka, M. & Möhler, G. (1998): Intonational Foreign Accent. *Speech Technology and Foreign Language Teaching*. In *Proceedings of Speech Technology in Language Learning*, 25.–27.05.1998, Marholmen, Sweden, 113–116.
- Kolly, M.-J. (2010): Regionale Spuren in "Français Fédéral" und "Schweizerhochdeutsch". Eine Untersuchung im Schnittfeld von Akzent und Einstellung bei Schweizer Dialektsprechern. Master Thesis, University of Bern.

- (2011): Weshalb hat man (noch) einen Akzent? Eine Untersuchung im Schnittfeld von Akzent und Einstellung bei Schweizer Dialektsprechern. In *Linguistik online*, 50/6, 43–77.
- (2013): Akzent auf die Standardsprachen: Regionale Spuren in "Français Fédéral" und "Schweizerhochdeutsch". In *Linguistik online*, 58/1, 37–76.
- Leemann, A. (2012): *Swiss German Intonation Patterns*. Amsterdam/Philadelphia (Benjamins).
- Leemann, A. & Siebenhaar, B. (2008): Perception of Dialectal Prosody. In *Proceedings of Interspeech 2008, Brisbane 22.–26.9.2008*, 524–527.
- Leemann, A., Dellwo, V., Kolly, M.-J. & Schmid, S. (2012): Rhythmic variability in Swiss German dialects. In *Proceedings of Speech Prosody, 21.–25.5.2012, Shanghai, PRC*.
- Lloyd James, A. (1929): *Historical Introduction to French Phonetics*. London (ULP).
- Missaglia, F. (2007): Prosodic Training for Adult Italian Learners of German. The Contrastive Prosody Method. In Trouvain, J./Gut, U. (eds.), *Non- Native Prosody. Phonetic Description and Teaching Practice*, Berlin/New York (de Gruyter). 236– 258.
- Muller, N. (1998): "L'allemand, c'est pas du français!". *Enjeux et paradoxes de l'apprentissage de l'allemand*. Neuchâtel/Lausanne (Institut de Recherche et de Documentation Pédagogique/Loisirs et Pédagogie).
- Ramus, F. & Mehler, J. (1999): Language identification with suprasegmental cues: A study based on speech resynthesis. In *Journal of the Acoustical Society of America*, 105, (1), 512–521.
- Ramus, F./Nespor, M. & Mehler, J. (1999): Correlates of linguistic rhythm in the speech signal. In *Cognition*, 73, 265–292.
- Shannon, R. V., Zeng, F.-G., Kamath, V., Wygonski, J. & Ekelid, M. (1995): Speech recognition with primarily temporal cues. In *Science*, 270, 303–304.
- Siebenhaar, B. (1994): Regionale Varianten des Schweizerhochdeutschen. Zur Aussprache des Schweizerhochdeutschen in Bern, Zürich und St. Gallen. In *Zeitschrift für Dialektologie und Linguistik*, 61, 31–65.
- Sprachatlas der deutschen Schweiz (SDS)*. Ed. by Hotzenköcherle, R. et al. 1962–2003. Bern (I–VI), Basel: Francke (VII–VIII).
- Tajima, K./Port, R./Dalby, J. (1997): Effects of temporal correction on intelligibility of foreign-accented English. In *Journal of Phonetics*, 25, 1–24.
- The International Phonetic Association (1999/2003): *Handbook of the International Phonetic Association. A Guide to the Use of the International Phonetic Alphabet*. Cambridge (Cambridge University Press).
- Van Els, T./De Bot, K. (1987): The role of intonation in foreign accent. In *Modern Language Journal*, 71, (2), 147–155.
- Werlen, I. (ed.)(2000): *Der zweisprachige Kanton Bern*. Bern/Stuttgart/Wien (Haupt).
- White, L./Mattys, S. L./Wiget, L. (2012): Language categorization by adults is based on sensitivity to durational cues, not rhythm class. In *Journal of Memory and Language*, 66, (4), 665–679.
- Wiget, L., White, L., Schuppler, B., Grenon, I., Rauch, O., Mattys, S. L. (2010): How stable are acoustic metrics of contrastive speech rhythm? In *Journal of the Acoustical Society of America*, 127, 1559–1569.

## Appendix

- 01 Die Frau des Apothekers weiss immer, was sie will.
- 02 Das Theater hat viele neue Aufführungen geplant.
- 03 Er wollte sich seiner Schwächen einfach nicht bewusst werden.
- 04 Der öffentliche Verkehr lässt viel zu wünschen übrig.
- 05 Die schlechte Zahlungsbilanz lässt mich nicht zur Ruhe kommen.
- 06 Die Eltern geben ihm keine finanzielle Unterstützung.
- 07 Der starke Frühlingsregen hat grossen Schaden angerichtet.
- 08 Der schnellste Zug ist immer noch der ICE.
- 09 Der Wiederaufbau der Stadt wird sehr lange dauern.
- 10 Das Bildungsministerium hat den einfachsten Weg gewählt.
- 11 Diese Konditorei macht ausgezeichnete Kuchen.
- 12 Dieses Geschäft bietet sehr preisgünstige Ware an.
- 13 Sie haben die Wahrheit erst entdeckt, als er auspackte.
- 14 Für meine Mannschaft wird der Sieg ein Kinderspiel sein.
- 15 Die Meinungsumfragen sagen einen Sieg der Rechten voraus.
- 16 Die Strassen der Innenstadt wurden von der Polizei gesperrt.
- 17 Ein berühmtes Bild wurde aus dem Kunsthaus gestohlen.
- 18 Der Müssiggang ist bekanntlich aller Laster Anfang.
- 19 Frei schreiben zu können ist ein grosser Vorteil.