



**HAL**  
open science

## Characterisation and identification of non-native French accents

Bianca Vieru, Philippe Boula de Mareüil, Martine Adda-Decker

► **To cite this version:**

Bianca Vieru, Philippe Boula de Mareüil, Martine Adda-Decker. Characterisation and identification of non-native French accents. *Speech Communication*, 2011, 53 (3), pp.292-310. halshs-00668927

**HAL Id: halshs-00668927**

**<https://shs.hal.science/halshs-00668927>**

Submitted on 10 Feb 2012

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Characterisation and identification of non-native French accents

Bianca Vieru<sup>1</sup>, Philippe Boula de Mareüil, Martine Adda-Decker

*LIMSI-CNRS, BP 133, 91403 Orsay CEDEX, France*

---

## Abstract

This paper focuses on foreign accent characterisation and identification in French. How many accents may a native French speaker recognise and which cues does (s)he use? Our interest concentrates on French productions stemming from speakers of six different mother tongues: Arabic, English, German, Italian, Portuguese and Spanish, also compared with native French speakers. Using automatic speech processing, our objective is to identify the most reliable acoustic cues distinguishing these accents, and to link these cues with human perception. We measured acoustic parameters such as duration and voicing for consonants, the first two formant values for vowels, word-final schwa-related prosodic features and the percentages of confusions obtained using automatic alignment including non-standard pronunciation variants. Machine learning techniques were used to select the most discriminant cues distinguishing different accents and to classify speakers according to their accents. The results obtained in automatic identification of the different linguistic origins under investigation compare favourably to perceptual data. Major identified accent-specific cues include the devoicing of voiced stop con-

---

<sup>1</sup>The author is now with Vecsys Research, Parc Orsay Université, 91400 Orsay, France

sonants, /b/~/v/ and /s/~/z/ confusions, the “rolled *r*” and schwa fronting or raising. These cues can contribute to improve pronunciation modeling in automatic speech recognition of accented speech.

*Key words:* Foreign accents; Non-native French; Perceptual experiments; Automatic speech alignment; Pronunciation variants; Data mining techniques; Automatic classification.

---

## 1 **1. Introduction**

2     Are we able to successfully recognise the accent of a speaker from a speech  
3 sample? Recent perceptual experiments on regional accents in English (Clop-  
4 per and Pisoni, 2004) and French (Woehrling and Boula de Mareüil, 2006)  
5 suggest a positive answer, as long as the regional identity is not too fine-  
6 grained. What is less known is the extent to which naive listeners are able  
7 to identify foreign accents in French, and what are the pronunciation traits  
8 that allow them to quickly identify someone’s mother tongue. A first goal of  
9 this article is to fill this gap by addressing non-native French from speakers  
10 of six different mother tongues. A second aim is to determine phonetic cues  
11 which may best characterise the corresponding foreign accents in French. A  
12 third objective is to build up an economical set of measurable and linguisti-  
13 cally meaningful features, in order to automatically identify foreign accents  
14 in French. To this purpose, we resort to automatic selection and classification  
15 techniques, which might compare to human perception in a similar task.

16     In the field of foreign-accented speech, a large body of studies focuses on  
17 segments to clarify the relative interplay between the two competing phone-  
18 mic systems of the mother tongue (L1) and a second language (L2). The

19 influence of the L1 system on the perception and the production of L2 is ad-  
20 dressed in a number of psycholinguistic studies on non-native speech. When  
21 comparing the L1 and L2 phonemic systems for a language pair, part of  
22 the phonemes may be shared between the L1 and the L2, whereas other  
23 phonemes may be specific to only one of the inventories. For instance, the  
24 French inventory includes a /y/, whereas in English this unit has no func-  
25 tional role (even though a similar sound can be heard in a word like *due*).  
26 Such L2-specific units may give rise to the perception of a foreign accent.  
27 When phonemes are shared (e.g. /t/), their acoustic realisations may also  
28 differ from one language to another (e.g. [t] with a varying burst energy),  
29 and these differences, if perceived, may be indicative of a foreign accent. Most  
30 studies focus on vowels (Flege et al., 2003), but consonant-related features are  
31 also addressed, for instance through voiced stops, which may index Arabic-  
32 accented (Flege and Port, 1981) or French-accented (Flege, 1984) English.  
33 Japanese speakers' confusions between /l/ and /r/ are also well-documented  
34 phenomena (Yamada et al., 1994). Among the clues that contribute to an  
35 impression of accentedness, a rather rich literature on Spanish-accented En-  
36 glish mentions factors affecting syllable structure, vowel quality, consonants  
37 (especially /s/~/z/ and /b/~/v/), as well as stress (Magen, 1998; Flege and  
38 Hammond, 1982).

39 Dimensions such as rhythm and intonation may also contribute to reveal  
40 a mother tongue that is different from the spoken L2. On the supraseg-  
41 mental level, researchers like Freland-Ricard (1996) showed that the mother  
42 tongue prosody tends to persist in non-native French speakers, unless some  
43 prosody-specific training is carried out. Also, in German-accented English

44 and English-accented German (Jilka, 2000), as well as in Spanish-accented  
45 Italian and Italian-accented Spanish (Boula de Mareüil and Vieru-Dimulescu,  
46 2006), prosody was found to play an important role. Concerning rhythm,  
47 a series of questions may arise with respect to foreign accent. Do rhythmic  
48 classes (stress-timed vs. syllable-timed) hypothesised for the languages them-  
49 selves remain valid for non-native speech? Will native Portuguese speakers  
50 – whose L1 is traditionally classified as stress-timed (Frota et al., 2007) –  
51 adopt a rhythm in French similar to the rhythm of their cousins of Romance  
52 syllable-timed language, Italian and Spanish? What will be the behaviour of  
53 Maghrebian speakers whose dialect may be considered as stress-timed and  
54 whose standard language is syllable-timed (Ghazali et al., 2002)? Parame-  
55 ters have been proposed to validate or contradict the existence of rhythmic  
56 classes (Arai and Greenberg, 1997; Ramus, 1999; Grabe and Low, 2002).  
57 These measurements, carried out on rather small, manually segmented and  
58 labelled corpora, have met with a certain success (Romano, 2010).

59 More recently, the foreign accent issue was addressed in the field of au-  
60 tomatic speech recognition (ASR), with the aim of reducing the impact of  
61 non-native speech on word error rates. Different directions have been ex-  
62 plored to deal with non-native speech for ASR: training strategies to build  
63 accent-specific acoustic models, which require amounts of generally scarce L2-  
64 L1 specific accented speech; adaptation strategies to include accent-specific  
65 variants within pronunciation dictionaries (Livescu and Glass, 2000; Silke  
66 et al., 2004; Cincarek et al., 2004; Bouselmi et al., 2006). The latter ap-  
67 proach requires linguistic knowledge on foreign accents, which may represent  
68 a bottleneck and ask for more in-depth accent-specific studies. Fewer studies

69 tackle automatic accent identification. Let us mention research on Mandarin-  
70 , German- and Turkish-accented English (Arslan and Hansen, 1997), Arabic-  
71 and Vietnamese-accented English (Kumpf and King, 1997; Berkling, 2001),  
72 Chinese-, Thai- and Turkish-accented English (Angkititrakul and Hansen,  
73 2003), African- and Brazilian-accented Portuguese (Rouas et al., 2008). The  
74 latest NIST LRE (language recognition evaluation) campaigns included au-  
75 tomatic dialect and accent verification, by relying on large amounts of ac-  
76 cented speech (Martin and Le, 2008). More linguistics-driven studies on  
77 foreign-accented speech exist (ten Bosch and Cremelie, 2002; Schaden, 2003;  
78 Raux, 2004; Bartkova and Juvet, 2004). Based on pronunciation alignment,  
79 like Goronzy (2004), these studies quantify ASR accuracy improvements, but  
80 they do not easily compare between each other, nor do they explicitly state  
81 how to identify the origin of a given foreign accent. Sangwan and Hansen  
82 (2009) do exploit phonological features; yet, it is in a perspective of accent  
83 analysis (of Chinese speakers of English) rather than accent identification.

84 This paper addresses foreign accents in French from a threefold perspec-  
85 tive: (i) to what extent are native listeners able to identify foreign accents  
86 in French? (ii) what acoustic evidence may contribute to corroborate a for-  
87 eign accent hypothesis, and finally (iii) what performance can be achieved  
88 by an automatic accent classification system based on perceptually salient  
89 features? A further issue is to broaden the scope of foreign accent studies  
90 by investigating a relatively large number of accents. More specifically, the  
91 presented work focuses on Arabic, English, German, Italian, Portuguese and  
92 Spanish accents which, according to statistics on immigration and tourism  
93 in France, should be most familiar to French listeners. The work described

94 hereafter combines perceptual experiments, linguistic knowledge on acoustic  
95 cues to accent differentiation, as well as automatic speech processing and  
96 data mining techniques to sort out which cues contribute to identifying the  
97 mother tongue of a non-native French speaker.

98 The corpus used in this work is described in Section 2. It comprises 84  
99 speakers (72 non-native and 12 native French speakers). The data were col-  
100 lected as two subsets of 42 speakers (balanced for accent and gender) during  
101 two separate recording phases. The earlier subset is used in a perceptual  
102 experiment and in subsequent acoustic analyses, which allow us to hypoth-  
103 esise features characterising the different accents; the later subset provides  
104 material for a second perceptual experiment and is kept aside to test our  
105 hypotheses via an automatic classification task.

106 Section 3 presents the perceptual tests, the experimental setup and pro-  
107 tocol, the tasks and corresponding results. Beyond accent identification, the  
108 speakers' degree of accentedness was judged by the native French listeners.

109 In Section 4, we examine some phonetic features concerning vowel quality,  
110 consonant articulation and prosody, including cues which were pointed out by  
111 our subjects during the perceptual experiments. The various acoustic analy-  
112 ses rely on automatic phonemic alignments by the LIMSI ASR system (Gau-  
113 vain et al., 2005), where acoustic models and pronunciation dictionaries can  
114 be manipulated as in Adda-Decker and Lamel (1999). The interest of the au-  
115 tomatic alignment method for linguistic studies has been shown in a number  
116 of studies (Gendrot and Adda-Decker, 2005; Woehrling and Boula de Mareüil,  
117 2006; Adda-Decker and Hallé, 2007; Woehrling et al., 2009). In a first step,  
118 acoustic feature measurements make use of automatic phonemic alignments

119 with standard French acoustic models and pronunciation dictionary. Next,  
120 foreign accent-related variants are added to the pronunciation dictionary,  
121 the most appropriate variants being selected during realignment. Finally,  
122 the standard French acoustic model set is extended with L1 acoustic models,  
123 so as to check whether non-native speakers remain closer to L1 productions  
124 rather than producing L2-like sounds.

125 Section 5 examines the relevance of all these features in an automatic  
126 classification task into 6 foreign accents and native French. Experiments are  
127 carried out on a held-out subset of the corpus, and the relative contributions  
128 of various linguistic feature sets (including vowel formants, consonant du-  
129 ration and voicing, prosodic cues, pronunciation variants derived from non-  
130 standard alignments with French and foreign acoustic units) are assessed.  
131 Classification results obtained with the best feature set are finally reported  
132 and compared to human perception.

## 133 **2. Corpus**

134 For this study, a corpus of more than 15 hours of speech was collected,  
135 including read and spontaneous speech of native and non-native speakers in  
136 similar recording and production conditions. As mentioned above, the corpus  
137 includes 6 non-native accents: Arabic, English, German, Italian, Portuguese  
138 and Spanish. Twelve speakers were recorded for each accent, in addition to  
139 12 native French speakers, who could be considered as control material. All  
140 the speakers (12 speakers per accent) were European or came from Arabic-  
141 speaking countries. A previous study showed the difficulty in discriminating  
142 the possible Algerian, Moroccan or Tunisian origins of speakers speaking



143 French (Boula de Mareüil et al., 2004). The Spanish speakers were neither  
144 Catalan nor Latin-American. As for the native French speakers, they were  
145 students who were born and grown up in the Paris region. Each speaker  
146 produced about 6 minutes of read speech and 5 minutes of spontaneous  
147 speech. The whole set of read speech recordings was checked for reading  
148 errors and fine-grained speech transcripts were produced. Only a small subset  
149 of the spontaneous speech was manually transcribed (totalling 6 minutes) in  
150 order to control for the content of the corresponding perceptual test.

151 *Material.* The read material stems from two short texts: a 400-word text  
152 from the *Phonology of Contemporary French* (PFC) project (Durand et al.,  
153 2003), and the International Phonetic Alphabet (IPA) text, *The North Wind*  
154 *and the Sun*, with 125 words in the French translation. Each reading of the  
155 two texts corresponds to an average of 5 minutes and 1 minute of speech  
156 respectively. Concerning the spontaneous speech part, the speakers talked  
157 freely for about 5 minutes in a face-to-face situation with the experimenter.

158 As mentioned earlier, the speakers were actually collected during two  
159 different recording phases resulting in two distinct subsets (termed A and B  
160 sets), each subset including 42 speakers (6 per L1), as depicted in Figure 1.  
161 The earlier A-set and the later B-set speakers were involved in perceptual  
162 tests, with spontaneous speech for the former and read speech for the latter  
163 (see Section 3). The A-set speakers were used for acoustic analyses (see  
164 Section 4) and for training the automatic classification system (see Section 5).  
165 The later B-set speakers were held out for the automatic accent classification  
166 task (see Section 5).

167 On average, the non-native A-set speakers (as many males as females,

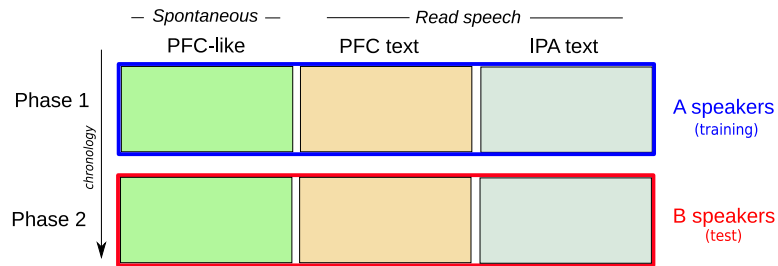


Figure 1: Overview of the composition of the foreign accent corpus, recorded during two phases. The earlier A set is used for acoustic analyses and automatic classifier training, the later B set for classifier testing. Each set includes 42 speakers (6 speakers of 6 foreign accents and of native French). Both sets are partially involved in perceptual tests.

168 all students) were 25 years old, had lived in France (in the Paris region)  
 169 for 15 months and had started to study French at the age of 15. The B-  
 170 set non-native speakers were 27 year-old students who had lived in France  
 171 for 21 months and had started to learn French at the age of 15. Overall,  
 172 the two sets were comparable in age and exposure to French. Speaker age  
 173 ranged from 24 to 27 years in the A-set, from 24 to 34 years in the B-set;  
 174 their duration of residence in the Paris region ranged from 6 to 37 months  
 175 in the A-set, from 13 to 37 months in the B-set; and the age of acquisition  
 176 of French as an L2 ranged from 10 to 24 years in the A-set, from 10 to 19  
 177 years in the B-set. Accent ratings should thus be comparable, even though  
 178 a weaker degree of accentedness could be hypothesised for the B-set where  
 179 one can notice a slightly longer immersive stay in France.

### 180 **3. Perceptual experiments**

#### 181 *3.1. Task, protocol, experiments and listeners*

182 *Task.* Perceptual experiments were conducted to determine to what extent  
183 French listeners are capable of identifying the investigated accents. This L1  
184 identification task was coupled with a minor task, the aim of which consisted  
185 in rating the speakers' degree of accentedness.

186 *Protocol.* First, the subjects were asked for their familiarity with the different  
187 accents and languages: they had to indicate whether yes/no they felt able to  
188 recognise this or that accent in French, and to rate their own proficiency in  
189 this or that language as (almost) nil, average or good. After a familiarisation  
190 phase, during which the subjects were presented a small set of typical excerpts  
191 of the 6 foreign accents together with their identity, the perception task  
192 proper was to identify the speakers' mother tongues and to evaluate the  
193 degree of accentedness on a 0-5 scale. The proposed degrees were paraphrased  
194 as follows: (0) no accent, (1) mild accent, (2) moderate accent, (3) rather  
195 strong accent, (4) strong accent, (5) very strong accent. The speech samples  
196 of the familiarisation phase were kept different, both in content and speakers,  
197 from those of the perceptual test material proper.

198 The perceptual experiments were run through a user-friendly interface to  
199 read the instructions, listen to the stimuli and capture the responses auto-  
200 matically. The stimuli were presented in a random order which changed for  
201 each listener. Each stimulus could be listened to as many times as judged  
202 useful by the subjects. In particular, the interface enabled a partial replay  
203 for any portion of interest, thus avoiding the full repetition of long stimuli.

204 Once a sample was processed (i.e. the subject made his/her choices among  
205 the possible L1s and degrees of accentedness with optional comments), it was  
206 no longer possible to go back to earlier samples.

207 *Experiments.* A first perceptual experiment (referred to as the 6-L1 test)  
208 consisted in a forced choice between Arabic, English, German, Italian, Por-  
209 tuguese and Spanish. For this experiment, a set of spontaneous speech ex-  
210 cerpts of about 10 seconds per speaker were selected from the non-native  
211 A-set speakers (see Section 2) according to the following criteria: absence of  
212 cultural references or morphosyntactic errors that could be typical of a given  
213 L1, few hesitations and coherence of the statement. The 6-L1 test was thus  
214 composed of 36 stimuli. The subjects, equipped with microphones, were in-  
215 vited to verbally react towards each excerpt (by imitating or caricaturing it)  
216 or to enter their comments into a text window. The experiment was run in a  
217 soundproof booth. The data were delivered stimulus by stimulus, and listen-  
218 ers were invited to specify the most relevant non-native features perceived in  
219 the speaker’s pronunciation and intonation.

220 A second perceptual experiment (referred to as the 7-L1 test), based on  
221 the reading of the IPA text (about 1-minute long), involved a forced choice  
222 between 7 possibilities: French, in addition to the six origins above. The  
223 rationale behind this experiment was to test which foreign accents could be  
224 most easily confused with native French and to examine these new results in  
225 relation with the degree of accentedness. For this 7-L1 test, subjects could  
226 also specify which cues they perceived as most salient, but only by adding  
227 final written comments.

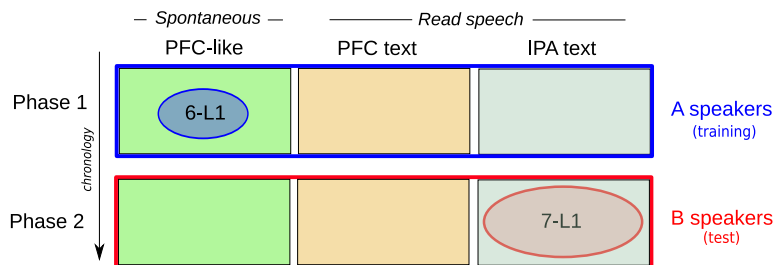


Figure 2: Parts of the corpus involved in the 6-L1 and 7-L1 perceptual experiments.

228 *Listeners.* Each experiment (which lasted 30-45 minutes per participant) in-  
 229 volved 25 untrained French listeners with normal hearing capacities, living  
 230 in the Paris region. All the 50 subjects had French as their mother tongue.  
 231 They were not paid for their participation.

### 232 3.2. Results

233 We first present a summary of the listeners' self-evaluations and the mea-  
 234 sured degrees of accentedness, before reporting on the perceptual identifi-  
 235 cation results proper. Perceptual distances between the 6 investigated for-  
 236 eign accents are then captured via multidimensional scaling and clustering  
 237 techniques, and synthetically displayed in a graphical representation. The  
 238 perceptual cues reported by the participants will follow.

239 *Degrees of familiarity and accentedness.* The majority of subjects reported  
 240 that they were capable of recognising Arabic, English and German accents  
 241 but fewer subjects felt able to recognise Italian, Portuguese and Spanish  
 242 accents in French. These trends do not match the listeners' proficiency in  
 243 the corresponding languages: as an example, almost all subjects self-reported  
 244 no or little knowledge in Arabic, whereas almost all of them were confident to

<b>Data set</b>	<b>Task</b>	<b>Ar</b>	<b>En</b>	<b>Ge</b>	<b>It</b>	<b>Po</b>	<b>Sp</b>	<b>Fr</b>
A-set, spontaneous	6-L1	2.4	3.0	2.2	3.1	2.4	2.9	–
B-set, read	7-L1	1.5	3.1	2.9	2.4	3.0	3.0	0.6

Table 1: Average degree of accentedness per speaker origin (on a 0-5 scale).

245 recognise an Arabic accent in French. For both experiments, the non-native  
 246 speakers’ degrees of accentedness achieve an average value of 2.7 on a 0-5  
 247 scale (see Table 1). The degrees of accentedness were comparable between the  
 248 different linguistic groups of non-native speakers, except for Arabic speakers  
 249 – a milder accent of 1.5 was measured in the 7-L1 test, while the Arabic  
 250 speakers of the 6-L1 test were rated as having an average degree of 2.4. The  
 251 latter difference cannot be easily explained by typical accent-related factors  
 252 such as age of acquisition and duration of residence: on average, the Arabic  
 253 speakers were older in the B-set (31 years) than in the A-set (27 years), but  
 254 they had spent a shorter time in France (27 months vs 37 months). Both  
 255 groups started to learn French at the age of 10, on average. Other factors like  
 256 speaking style might have played a role: more normative reading for Arabic  
 257 speakers vs more accented spontaneous speech.

258 *Accent identification results.* The results of the 6-L1 and 7-L1 accent identi-  
 259 fication tests are shown in Tables 2 and 3 respectively. For both experiments,  
 260 the average identification rates are above 50%, even though there is consid-  
 261 erable cross-accent variation. The overall identification rates are 52% in the  
 262 6-L1 experiment, whereas the 7-L1 experiment achieves 60% correct iden-  
 263 tification. The better 7-L1 task results are mainly due to the near-perfect

	(Acc)	Ar	En	Ge	It	Po	Sp
Ar	(2.4)	<b>77</b>	1	6	5	8	2
En	(3.0)	9	<b>49</b>	28	3	3	9
Ge	(2.2)	6	15	<b>63</b>	5	8	3
It	(3.1)	7	3	5	<b>40</b>	10	34
Po	(2.4)	17	8	17	12	<b>25</b>	21
Sp	(2.9)	5	3	3	19	11	<b>59</b>

Table 2: Confusion matrix of the 6-L1 perceptual identification test using spontaneous speech (%). Rows correspond to the reference while columns give the subjects' answers. Degrees of accentedness (Acc) are recalled within parentheses.

	(Acc)	Ar	En	Ge	It	Po	Sp	Fr
Ar	(1.5)	<b>36</b>	10	14	15	7	9	10
En	(3.1)	3	<b>73</b>	15	2	3	3	0
Ge	(2.9)	3	15	<b>65</b>	5	9	2	1
It	(2.4)	3	0	3	<b>46</b>	23	22	3
Po	(3.0)	11	5	11	19	<b>34</b>	19	1
Sp	(3.0)	2	0	1	15	15	<b>67</b>	0
Fr	(0.6)	1	0	2	1	0	0	<b>96</b>

Table 3: Confusion matrix of the 7-L1 perceptual identification test on the IPA text reading (%). Degrees of accentedness (Acc) are recalled for each L1.

264 identification of native French speakers. Excluding French natives, results  
265 drop to 54% (i.e. very close to the 6-L1 task results). It is interesting to  
266 note that despite the different experimental setups in the two experiments  
267 (10 seconds, spontaneous *vs* 1 minute, read), with different speakers (A-set  
268 *vs* B-set), very similar results are obtained.

269 For each L1,  $\chi^2$  tests show that correct identification scores are signif-  
270 icantly above chance level. For each linguistic origin, the most frequent  
271 answer is the right one, and this holds for most speakers (25 out of 36 in the  
272 first experiment, 28 non-native speakers and the 6 native French speakers in  
273 the second experiment). In both experiments, most frequent confusion pairs  
274 include Spanish/Italian and English/German accents. The lowest identifica-  
275 tion rates are observed for the Portuguese accent, which tends to be mistaken  
276 as any other accent but English. The hushing stereotype that tends to be  
277 (wrongly) associated to the Portuguese accent in French may contribute to  
278 the low identification rate of this accent. Among the best recognised accents  
279 appear Arabic, German and Spanish in the 6-L1 test, and English, German  
280 and Spanish in the 7-L1 test. In the latter test, a relatively high confusion  
281 rate with native French speakers can be observed for Arabic speakers (10%)  
282 – these rates do not exceed 3% for the other non-native speakers. This high  
283 rate is in line with the earlier mentioned low degree of accentedness of the  
284 Arabic B-set speakers. However, the link between degree of accentedness and  
285 identification rates does not appear to be straightforward. This issue will be  
286 addressed hereafter via statistical analyses.

287 *Graphical accent-distance representation.* The accent identification results  
288 can be represented graphically via multidimensional scaling and clustering



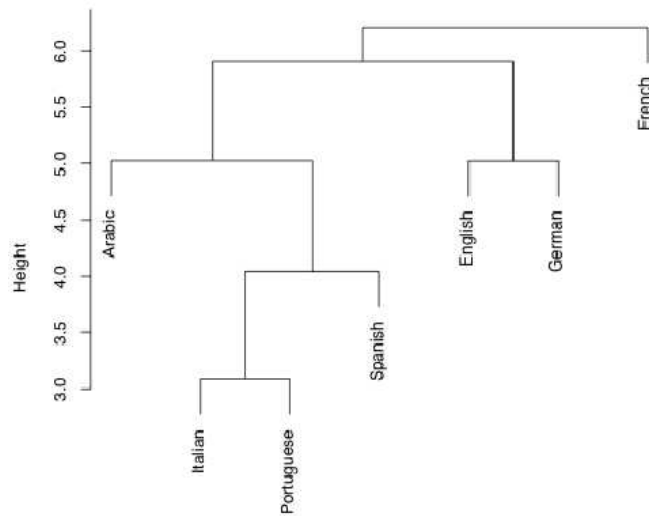


Figure 3: Dendrogram of the identification results of the 7-L1 perceptual experiment.

289 techniques, carried out using the R software (Ihaka and Gentleman, 1996).  
 290 For multidimensional scaling, similarity matrices are derived from the confu-  
 291 sion matrices and the features correspond to distances between the line pairs  
 292 of the confusion matrices (Woehrling and Boula de Mareüil, 2006). In the re-  
 293 sulting dendrograms, subtrees gather perceptually similar accents. Figure 3  
 294 shows the dendrogram of the 7-L1 perceptual experiment as produced by a  
 295 hierarchical agglomerative algorithm and a Euclidean distance. At level 0  
 296 of the tree, native French speakers are separated from non-native speakers.  
 297 At level 1, the speakers of Germanic languages are gathered in a subtree,  
 298 whereas at level 2 of the dendrogram, Arabic speakers are set apart from  
 299 the speakers of Romance languages. At least for non-native speakers, this  
 300 graphical representation yields subtrees in line with intuition and linguistic  
 301 knowledge on language typology.

302 *Statistical analyses.* Analyses of variance (ANOVAs) were conducted on the  
303 responses counted as right (1) or wrong (0) with the random factor *Subject*  
304 and the two within-subject factors *Familiarity* (with the accent) and *De-*  
305 *gree* of accentedness. Whether the listeners felt able to recognise the origin  
306 of the accent in the majority of cases (as in the case of Arabic, German  
307 and English) or not (as in the case of Italian, Portuguese and Spanish), two  
308 *Familiarity* levels were distinguished. As far as the *Degree* of accentedness  
309 is concerned, the speakers were split into three balanced groups, averaging  
310 listeners' evaluations. The ANOVAs show a major effect of listeners' *Famil-*  
311 *ilarity* [ $F(1, 24) = 56.5, p < 0.01$  in the first experiment;  $F(1, 24) = 25.3,$   
312  $p < 0.001$  in the second experiment] and speakers' *Degree* of accentedness  
313 [ $F(2, 48) = 21.4, p < 0.01$  in the first experiment;  $F(2, 48) = 40.8, p < 0.001$   
314 in the second experiment], with a marginal interaction between the two. De-  
315 spite an overall effect of the *Degree* of accentedness, it may be underlined  
316 that the difference of accent degree between Arabic and Portuguese speakers  
317 (the groups respectively identified best and worst in the first experiment) is  
318 not significant according to a *t*-test.

319 *Reported cues.* When looking at the listeners' comments collected during the  
320 first experiment, we can notice segmental features as well as some supraseg-  
321 mental cues. Segmental features most importantly include the *r* pronounci-  
322 ation, whether "rolled" (reminding of a Southern country) or uttered "in an  
323 English manner" (93 times); *yé* ([je]) instead of *je* (/ʒə/, English 'I'), [v]  
324 instead of /b/ and [s] instead of /z/ for Spanish speakers (38 times); [i]  
325 instead of /e/ in the case of Arabic speakers (31 times) and [z] instead of /s/  
326 for Germans (24 times); [u] instead of /y/ or vice versa and a bad realisation

327 of nasals (37 times) signaling a foreign accent rather than a particular origin.  
328 Features related to suprasegmentals remain very impressionistic in nature,  
329 and include “sing-song” sentences indicative of an Italian accent or a “rush”  
330 on certain words. Some of these cues were also pointed out by listeners of  
331 the second perceptual experiment, but they were not quantified.

### 332 *3.3. Conclusion*

333 According to the foregoing perceptual tests, the degree of accent of the  
334 non-native speakers was judged as moderate to rather strong (with an av-  
335 erage rate of 2.7 on a 0–5 scale). Via two experiments involving different  
336 speaker sets, native French listeners succeeded in identifying the foreign ac-  
337 cents in over 50% of cases. However, confusion rates were high for languages  
338 within a given linguistic family (e.g. Romance languages, Germanic lan-  
339 guages). Participants had the most difficulties with the Portuguese accent,  
340 which resulted in particularly high confusion rates. Language proficiency did  
341 not necessarily entail better identification scores. On the other hand, the 7-  
342 L1 perceptual test, which included French native speakers, showed that the  
343 subjects almost perfectly separated French natives from non-native speakers.

344 Salient features reported by the listeners included various *r* pronuncia-  
345 tions, [v] instead of /b/ and [s] instead of /z/ for Spanish speakers, [i]  
346 instead of /e/ for Arabic speakers, [z] instead of /s/ for German speakers  
347 and prosodic cues for speakers of different origins. In the following, acous-  
348 tic analyses are undertaken on both segmental and suprasegmental levels, to  
349 check whether accent-specific features can be measured objectively and, if  
350 so, whether they corroborate perceptual cues.

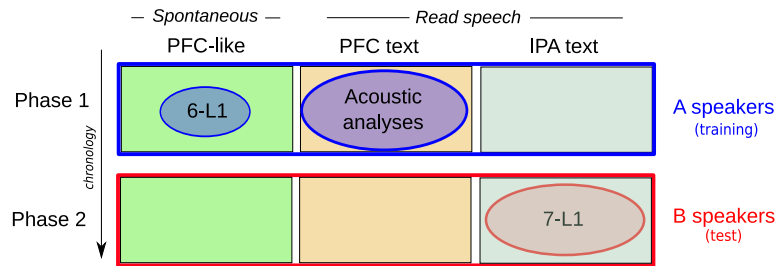


Figure 4: Sub-corpus of the acoustic analyses: PFC read speech from A-set speakers.

#### 351 4. Acoustic analyses using automatic alignments

352 For the acoustic analyses presented in this section, we used the PFC  
 353 text read by the A-set speakers (the 36 speakers used in the 6-L1 perceptual  
 354 experiment) and 6 native French speakers (3 males, 3 females, not used in the  
 355 7-L1 experiment). As the same linguistic content is produced by all speakers,  
 356 this material particularly lends itself to inter-speaker comparisons.

357 The corpus was automatically segmented and phone-labelled using the  
 358 LIMSI ASR system for French (Gauvain et al., 2005). The segmentation  
 359 was carried out using context-independent acoustic models, considered to  
 360 produce more reliable segment boundaries than context-dependent models:  
 361 previous work showed the reliability of the approach on different languages  
 362 and accents (Adda-Decker and Lamel, 1999; Gendrot and Adda-Decker, 2005;  
 363 Woehrling and Boula de Mareüil, 2006).

364 In the following, acoustic measurements proper (which include vowel for-  
 365 mants, consonant duration and voicing rates as well as prosodic cues) were  
 366 derived from a standard phonemic alignment: the pronunciation dictionaries  
 367 included standard French variants with optional schwas and liaisons, but no  
 368 accent-specific pronunciation variants. In addition to this first set of acoustic

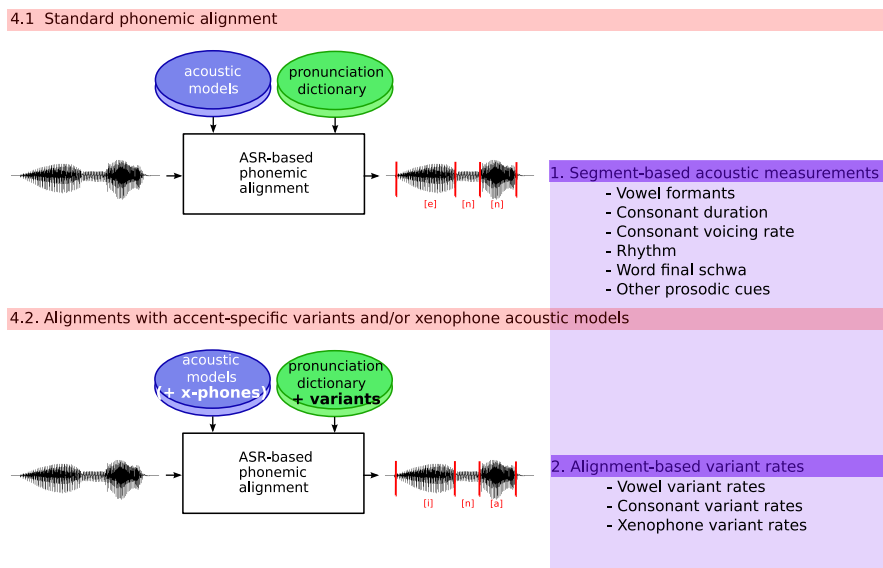


Figure 5: Synthetic overview of the experimental workflow of section 4 involving automatic alignments: standard phonemic alignment in subsection 4.1, accent-specific variants without/with xenophone acoustic models in subsection 4.2 to produce accent-related features (listed in the shaded box). See text for details.

369 measurements, we introduced *pronunciation variant rates* as in [Adda-Decker](#)  
 370 and [Lamel \(1999\)](#) to contribute to the foreign accent characterisation. The  
 371 idea was to introduce systematic options (e.g. /e/ pronounced as [e] or  
 372 [i]) within the pronunciation dictionary, since different foreign accents may  
 373 privilege different sets of variants. [Figure 5](#) summarises the major processing  
 374 steps to extract accent-specific features.

#### 375 4.1. Measurements based on standard phonemic alignment

##### 376 4.1.1. Vowel formants

377 Formant frequencies were measured on oral vowels (more than 500 vow-  
 378 els per speaker) using the Praat software ([Boersma, 2001](#)). The first two

379 formants (F1 and F2) as well as fundamental frequency ( $f_0$ ) values were  
380 measured every 10 ms using the standard settings of Praat. In order to get  
381 rid of aberrant formant values, the measurements were filtered using vowel-  
382 and gender-specific thresholds with respect to reference values in an average  
383 range of  $\pm 500$  Hz (Calliope, 1989; Gendrot and Adda-Decker, 2005). Fur-  
384 thermore, we only considered vowels that were voiced (i.e. the detected  $f_0$   
385 values were higher than 75 Hz) on more than half their durations. Each  
386 selected segment was then assigned formant (respectively  $f_0$ ) values by av-  
387 eraging the elementary measurements. The applied criteria resulted in an  
388 average 5.5% rejection rate. Formant values were then normalised using  
389 Nearey’s log-mean procedure (Nearey, 1989; Disner, 1980; Adank, 2003) to  
390 minimise differences due to speakers’ physiological characteristics. Without  
391 such a procedure, vowel triangles produced by vocal tracts of different sizes  
392 (between males and females especially) are not easily comparable: longer  
393 vocal tracts correlate with smaller vocalic triangles and vice-versa. The vo-  
394 calic triangles corresponding to the different accents (or linguistic origins)  
395 are displayed in Figure 6. For the sake of readability, the triangles of the six  
396 accents are separated into two subsets, the first one for Romance languages  
397 (Italian, Portuguese, Spanish) in the upper part, the second one for the re-  
398 maining accents (Arabic, English, German) in the lower part. The native  
399 French triangle is displayed in both parts as a reference.

400 A first observation concerns a difference in size for the French triangle,  
401 which tends to be smaller than the ones corresponding to foreign accents.  
402 Given that vocalic triangles tend to reduce their shapes with smaller segment  
403 durations (Gendrot and Adda-Decker, 2005), this is most likely due to the

404 fact that natives tend to speak faster than L2 speakers (see segment durations  
405 in Table 4). Also, English (and German, to a lesser extent) speakers' triangles  
406 are smaller than those of other non-native speakers: this may be related to  
407 vowel reduction in their mother tongues.

408 We may try to relate the average vowel locations in the F1/F2 space  
409 to what is known of the different languages' vowel characteristics and our  
410 listeners' comments. The /u/ fronting, well documented in English (Delattre,  
411 1965; Harrington et al., 2000), is here noticeable in English-accented French,  
412 as is the /y/ backing in Spanish and Italian speakers. Concerning /e/, the  
413 closest one to /i/ comes from Arabic speakers. The /e/~/i/ merger is rather  
414 common among Arabic speakers of French: it can be attributed to the fact  
415 that this distinction is not functional at least in the 3-vowel phonological  
416 system of standard Arabic. Observable differences concerning the /a/ vowel  
417 are less easily explainable. As far as the schwa is concerned, it is most closed  
418 for Portuguese speakers (tending towards the high central [i]) of their mother  
419 tongue (Veloso, 2007)), and fronted among Spanish and Italian speakers.

420 The /y/ realisations particularly differ between Spanish speakers on the  
421 one hand (where the /y/ is closer to [u]) and Arabic speakers on the other  
422 hand (whose /y/ goes closer to [i]). An interpretation is that the former  
423 speakers favour the [+round] feature, the latter speakers favour the [+front]  
424 feature. These rather well-known phenomena are often caricatured as in *tou*  
425 *m'as toué* for *tu m'as tué* ('you killed me') in Spanish-accented French and  
426 *Itats-Inis* for *États-Unis* ('United States') in Arabic-accented French. This  
427 accent-specific /y/ shift may also be evidenced by automatic scaling and  
428 clustering techniques: using for each speaker his/her average /y/ coordinates

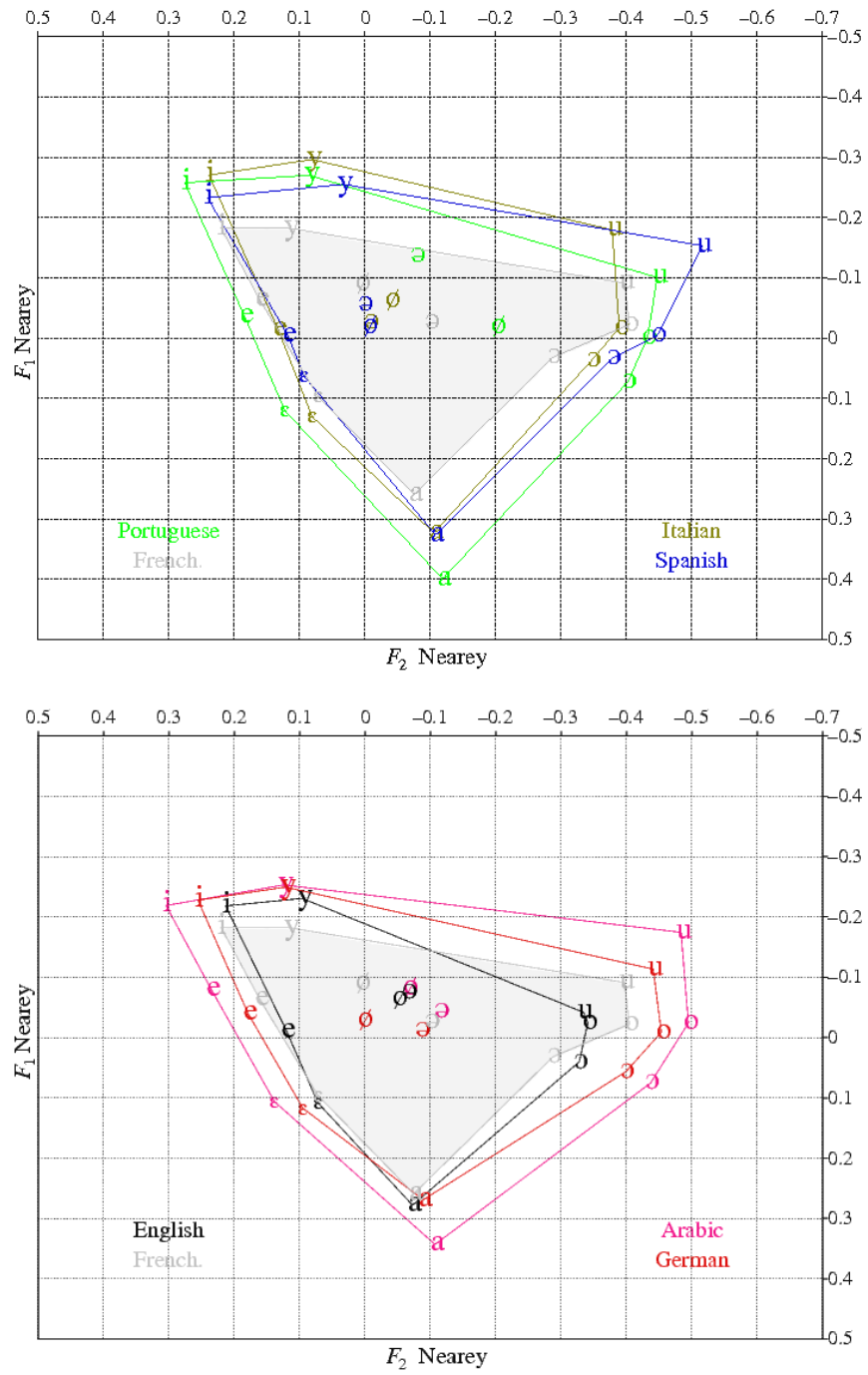


Figure 6: Vowalic triangles of native or foreign-accented French, for the PFC text.



	<b>Ar</b>	<b>En</b>	<b>Ge</b>	<b>It</b>	<b>Po</b>	<b>Sp</b>	<b>Fr</b>	<b>Mean</b>
/p/	88	84	89	79	81	80	67	81
/t/	82	92	89	81	84	84	75	84
/k/	95	95	90	86	83	82	82	88
/b/	83	64	79	92	89	63	74	78
/d/	78	68	65	78	79	76	60	72
/g/	79	69	65	74	74	64	62	70
/v/	79	67	69	77	91	62	61	72
/ʁ/	84	68	72	56	80	82	72	73
/φ/	91	90	89	91	94	89	73	88

Table 4: Duration of some consonants for the PFC text (in ms). The last column provides the average consonant duration over the 7 L1s. Mean values in the bottom line correspond to the overall segment duration including all vowels and consonants.

429 in the F1/F2 space, a cluster analysis results in a dendrogram with Arabic  
430 speakers on the one hand, Spanish and Italian speakers on the other hand.  
431 The same tendency could be found when using the spontaneous speech data.

#### 432 4.1.2. Consonant duration and voicing rates

433 Using the standard phonemic alignment, average consonant durations and  
434 voicing rates were measured. As a methodological caveat, it is important to  
435 notice that automatic segment boundaries may differ from manually assigned  
436 borders. The advantage of automatic processing, however, lies in consistently  
437 processed data by a repeatable procedure.

438 As shown in Table 4 (bottom line), average phoneme durations and cor-  
439 responding speech rates (measured as the average number of phone seg-

440 ments per second) are comparable across non-native speakers. For non-native  
441 speakers the average phone duration is close to 90 ms which yields a rate of  
442 11 segments per second, whereas native speakers have average segment du-  
443 rations close to 70 ms which results in 14 segments per second. Although  
444 observed duration differences may be due to many variation factors, it is in-  
445 teresting to note that (see Table 4) speakers of Arabic, English and German  
446 tend to have the longest voiceless stop segments. In these languages, voice-  
447 less stops are often aspirated while in French the voice onset time is most  
448 often low (Delattre, 1965; Abdelli-Beruh, 2004).

449 In the following, the voicing ratio is defined as the number of voiced  
450 measurements divided by the total number of measurements (every 10 ms).  
451 Table 5 shows the average voicing rates of some relevant consonants. The  
452 low voicing rates of voiced stops (/b/, /d/, /g/) in English-accented and  
453 German-accented French is noteworthy, reflecting a certain devoicing of these  
454 consonants in speakers of Germanic languages.

455 Tables 4 and 5 show that Spanish speakers yield a low voicing rate for the  
456 /z/ fricative (similar to /s/) and the shortest /b/ and /v/ durations among  
457 non-native speakers. As a matter of fact, there are no phonological voiced  
458 fricatives in Spanish, hence no /z/ and no /b/~/v/ distinction. Also, the /β/  
459 of Italian speakers is shorter and more voiced than for the other speakers.  
460 We will return to this in subsection 4.2.

#### 461 4.1.3. *Rhythm, word-final schwa and related prosodic cues*

462 Some of the subjects' comments during the perceptual experiments re-  
463 lated to rhythmic aspects. As mentioned in the introduction, parameters  
464 involving segment durations have been proposed to characterise rhythmic

$\varphi$	<b>Ar</b>	<b>En</b>	<b>Ge</b>	<b>It</b>	<b>Po</b>	<b>Sp</b>	<b>Fr</b>	<b>Mean</b>
/p/	18	32	33	37	32	28	21	29
/t/	17	28	32	39	33	31	18	28
/k/	16	25	28	34	27	28	20	25
/s/	22	23	34	40	39	36	20	31
/ʃ/	30	22	35	39	23	35	36	31
/b/	91	57	76	82	91	81	94	82
/d/	82	60	77	77	85	73	86	77
/g/	86	61	76	87	88	73	92	80
/v/	97	86	93	88	94	91	94	92
/β/	56	59	57	68	58	60	59	60
/z/	89	79	85	80	93	53	91	81
/ʒ/	83	71	82	84	83	77	78	80

Table 5: Voicing rates of consonants (% of defined  $f_0$  values) for the PFC text.

465 classes. [Ramus \(1999\)](#) considers the proportion of vocalic intervals (%V) and  
466 the duration variation of consonantal intervals in terms of standard deviation  
467 ( $\Delta C$ ). A consonantal interval is composed of one or more consecutive conso-  
468 nantal segments delimited by vowels or pauses. [Grabe and Low \(2002\)](#) pro-  
469 pose a slightly more complex approach: they measure the variability between  
470 successive vocalic and intervocalic intervals via so-called “Pairwise Variabil-  
471 ity Indices” (PVI<sub>s</sub>), possibly normalised to account for speech rate variation.  
472 These parameters do not explicitly take into account the notion of stress,  
473 but rely on the link between stress-timing, complexity of consonant clus-  
474 ters and vowel reduction. More recent work not investigated here aims at  
475 adapting these measures for quantifying second language proficiency ([Diez](#)  
476 [et al., 2008](#)) and to study the influence of speech rate on acoustic correlates  
477 of rhythm ([Dellwo, 2010](#)).

478 In the following, we measured Ramus and Grabe’s parameters on for-  
479 eign accents. The same complexity of consonant clusters is imposed to all  
480 speakers by the French language. Hence, the measured duration variability  
481 should be only poorly related to phonotactic differences. Concerning com-  
482 plex consonant clusters, one could imagine that speakers who are not used  
483 to such productions would tend to articulate the latter carefully (unless they  
484 resort to elisions), resulting in a total duration close to the sum of individual  
485 consonant durations. This could then lead to high  $\Delta C$  measurements in op-  
486 position to observations made in their mother tongues. Conversely, speakers  
487 who are used to complex consonant clusters may produce them with relatively  
488 shorter durations, thereby making %C decrease and %V increase. Resulting  
489 measurements would then contradict the tendency of stress-timed languages

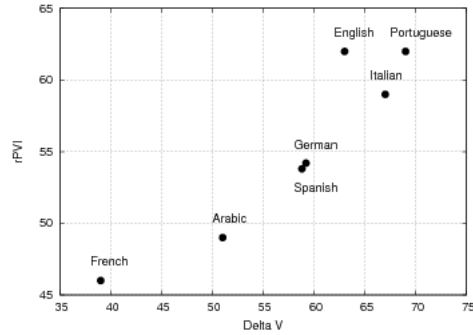


Figure 7: Rhythm characterisation of native and foreign-accented French combining Ramus’s  $\Delta V$  and Grabe’s PVI on vowels. L1s in increasing  $\Delta V$  order: French, Arabic, Spanish, German, English, Italian and Portuguese. Durations are in ms.

490 to reduce vowels. Actually, the duration of consonant clusters seems to be  
 491 speaker-specific. In the following, we only kept  $\Delta V$  (the standard deviation  
 492 of vocalic interval durations) and PVI measures on vowels. PVIs are not  
 493 normalised because speech rates are comparable across non-native speakers  
 494 (from 10.7 to 11.3 phonemes/second).

495 Figure 7 shows the results for the different L1s. As expected, French  
 496 appears in the left bottom corner. One might have further expected a sys-  
 497 tematic grouping of syllable-timed vs stress-timed L1s. However, no clear  
 498 rhythmic classes emerge. Results merely show an important difference (ac-  
 499 cording to  $\Delta V$  and PVI) between Arabic (closest to French) and e.g. Ital-  
 500 ian or Portuguese speakers. According to these measurements, Maghrebien  
 501 speakers do not tend to reproduce the stress-timed rhythm of their native  
 502 dialects.

503 In the remainder of this subsection, we focus on potential word-final  
 504 schwas of polysyllabic words, by measuring their realisation rates, the length-

	<b>Ar</b>	<b>En</b>	<b>Ge</b>	<b>It</b>	<b>Po</b>	<b>Sp</b>	<b>Fr</b>	<b>Mean</b>
$\%V_{schwa}$	10	15	14	23	15	11	11	14
$V/V_{schwa}$ duration ratio	2.0	1.6	1.9	2.4	2.2	2.1	1.9	2.0
$\Delta f_0(V, V_{schwa})$	1.3	0.9	-0.1	2.2	0.7	1.0	1.1	1.0

Table 6: Word-final schwa and related prosodic cues: realised word-final schwa rate (%), duration ratio (respectively  $f_0$  difference in semitones) between the vowel preceding a word-final schwa and the final schwa.

505 ening of the vowels preceding them and the corresponding  $f_0$  contours. There  
506 are 123 polysyllabic word-final schwa sites in the PFC text, potentially re-  
507 alised by the speakers and detected by the alignment system, as word-final  
508 schwas were set optional in the standard pronunciation dictionary used for  
509 alignment. Table 6 shows the results obtained. Italian speakers of French  
510 yield by far the highest word-final schwa realisation rate (with 23%), all the  
511 other groups of speakers keeping rates below 15%. These high numbers for  
512 Italian may be at least partially explained by the fact that in their mother  
513 tongue, content words ending with a consonant are extremely rare: speak-  
514 ers thus tend to add a word-final schwa in this situation. Table 6 further  
515 gives duration ratios (respectively  $\Delta f_0$  as difference in semitones) between  
516 the vowel preceding the word-final schwa and the pronounced final schwa.  
517 One may notice that Italians have both the highest duration ratio for the  
518 supposedly stressed syllable (whose nucleus is the vowel preceding the final  
519 schwa) and the highest  $\Delta f_0$  corresponding to a pitch lowering on the final  
520 schwa. By contrast, native Germans speaking French display a slightly ris-  
521 ing  $f_0$  contour on the final syllable (the only negative  $\Delta f_0$  in Table 6). Both

522 patterns seem to be perceptually salient and typical of these accents.

523 Since the figures were computed on a relatively small number of occur-  
524 rences we felt it necessary to perform ANOVAs. This was done with the  
525 measurements ( $\%V_{schwa}$ , duration ratio and  $\Delta f_0$ ) averaged by speaker as the  
526 dependent variables and the L1s as the independent variables. The difference  
527 did not reach significance level for the duration ratio, but the L1 effect was  
528 found to be significant for the schwa realisation rate [ $F(6, 35) = 3.86, p <$   
529  $0.01$ ] and  $\Delta f_0$  [ $F(6, 35) = 3.04, p < 0.05$ ].

#### 530 4.2. Measurements based on non-standard alignments

531 So far for the alignment, a standard French pronunciation dictionary was  
532 used, in which each entry was assigned generally one and sometimes several  
533 standard French pronunciation(s) due to optional schwas and liaisons. How-  
534 ever, perceptual results and acoustic measurements suggest that non-natives  
535 may produce variants that deviate markedly from the standard form, and  
536 that such deviations may be shared by speakers of a given mother tongue.

537 In the following, accent-specific variants are introduced, allowing a given  
538 (standard) phoneme to be replaced by one or several variants within the pro-  
539 nunciation dictionary. For each rule, the pronunciation dictionary is updated  
540 accordingly before alignment, in order to account for specific vowel (respec-  
541 tively consonant) substitutions between vowel (respectively consonant) pairs.  
542 The relevance of these variants can be measured after alignment via *variant*  
543 *rates*: the ratio of segments aligned with the non-standard symbols over  
544 the total number of segments. For example, given a rule stating that an  
545 /e/ can be produced either as an [e] or as an [i] (/e/→[e|i]), the cor-  
546 responding *variant rate* measures the proportion of segments aligned using

547 the non-standard symbols (e.g. /e/→[i], the proportion of /e/ aligned as  
548 [i]). In the experiments of subsection 4.2.1, the phonemic inventory and  
549 the corresponding acoustic model set remain unchanged. In subsection 4.2.2,  
550 the standard French phonemic inventory (respectively the acoustic model  
551 set) is augmented with xenophones (respectively phone models from other  
552 languages) for some phonemes whose foreign-accented realisations may be  
553 particularly different from the French ones. How this is done will be ex-  
554 plained in the next subsections.

#### 555 4.2.1. *Variants using the standard French acoustic model set*

556 Based on linguistic knowledge about the different L1s, comments from  
557 the perceptual tests and results of acoustic analyses, we defined a set of  
558 about 20 non-native French variant rules accounting for common variation  
559 processes such as voicing/devoicing, spirantisation and affrication for conso-  
560 nants, opening/closing, fronting/backing and denasalisation for vowels. Au-  
561 tomatic alignment using the corresponding variants simulates a binary, cate-  
562 gorical approach to foreign-accented speech, supplementing gradient analyses  
563 based on formant,  $f_0$  and duration measurements. A list of rules involving  
564 20 tested variants can be found in Table 7. Generally, they consist in a  
565 paradigmatic choice between phone pairs (/d/→[d|t]). In some cases, the  
566 rule may become more complex, with additional segment insertions and con-  
567 textual constraints (e.g. /ã/→[ã|ã̃n|an] or [ã|ã̃m|am] in a right [b|p]  
568 context). For any rule, the left phoneme can be aligned as any of the options  
569 listed on the right, conventionally noted between square brackets since they  
570 refer to phonetic realisations. For each rule, a specific pronunciation dictio-  
571 nary was generated, a distinct alignment was produced accordingly and a



572 corresponding variant rate was computed. Most rules propose a single al-  
573 ternative. In case of multiple choices (see last rule concerning French nasal  
574 vowels in Table 7), all non-standard alternatives are cumulated to compute  
575 the corresponding variant rate. Tested rules are listed in Table 7, with the  
576 percentages of non-standard variants aligned by the system (e.g. [v] in  
577 the case of /b/→[b|v]). Vowel diphthongisation rules were also designed  
578 for English-accented French, but the corresponding alignments yielded few  
579 diphthongised variants: the results achieved are not presented here.

580 As a preamble, we can observe that some non-standard variants are fre-  
581 quently selected by the alignment system, even for French natives (e.g. de-  
582 voiced /g/ and /z/, unrounded /y/, open underlying mid-closed vowels). As  
583 far as foreign accents are concerned, the displayed results tend to be in line  
584 with acoustic measurements of subsection 4.1 and prior knowledge. Some of  
585 the most interesting results are boldfaced in Table 7. As can be seen, Span-  
586 ish and English speakers produce the highest rates of non-standard variants,  
587 whereas results for Arabic and Portuguese speakers often remain close to  
588 the mean values or the figures measured for native French speakers. For  
589 example, 60% of the /b/ plosives are aligned as [v] fricatives for the Span-  
590 ish accent. Recall that the Spanish language does not have /b/ and /v/  
591 as distinct phonemes (Delattre, 1965): a [b] is realised after a pause or a  
592 nasal consonant; a [β] appears elsewhere (Quilis, 1993). This may favour  
593 the spirantised variant in many contexts, with an acoustic realisation closer  
594 to [v] than to [b], even when French is spoken. Continuing with Spanish,  
595 [s] is preferred to [z] (in 79% of cases), [j] to [ʒ] and [tʃ] to [ʃ]. Such  
596 pronunciations are rather well known in Spanish-accented French and En-

597 glish (Magen, 1998). For English (and more generally Germanic) speakers,  
598 voiced stops tend to be aligned with their unvoiced counterpart, reflecting  
599 tendencies in their own L1s. The Italian speakers' /y/ aligned as [u] (as in  
600 Spanish speakers) and /ʁ/ aligned as a sonorant liquid are also consistent  
601 with the results of subsection 4.1. For nasal vowels, all non-native French  
602 speakers yield high variant rates, with Spanish and Italian speakers reaching  
603 almost ten times as many nasal appendices as native French speakers do.  
604 This is well audible when listening to Spanish and Italian speakers. Other  
605 results are less conclusive: for instance, /e/~y/~i/ mergers, which tended  
606 to appear in the Arabic speakers' vocalic triangles, are not captured here.

#### 607 4.2.2. Variants using the xenophone-augmented acoustic model set

608 The previous variants were designed to evaluate potential confusions be-  
609 tween French phonemes made by non-native speakers. In this subsection,  
610 the standard French acoustic model set is supplemented with foreign acous-  
611 tic models, in order to account for non-native productions which may be “far”  
612 from the target units or “intermediate” between two French phonemes. Here-  
613 after, we address the cases of the French /b/, /v/, /ʒ/, /s/, /l/, /ʁ/ and /u/  
614 phonemes, mapped with L1 phonemes or allophones borrowed from different  
615 L1s: their realisations, motivated by linguistic mechanisms, may be partic-  
616 ular to speakers of certain origins (Delattre, 1965). For technical reasons,  
617 the L1s are limited to Spanish and English, for which extensively trained  
618 acoustic models are available within the corresponding LIMSI speech-to-text  
619 systems (Lamel et al., 2007). The models we added are [r], [ʝ], [j] and  
620 [β] from Spanish, [ɹ], [ɹ̥] and [ʊ] from English, as can be seen in Table 8  
621 with alignment results.

Rules	Variant rates							
	Ar	En	Ge	It	Po	Sp	Fr	Mean
/standard/ → [standard   variants]								
/b/ → [b   v]	8	30	32	8	22	<b>60</b>	3	23
/b/ → [b   p]	8	<b>55</b>	<b>42</b>	3	6	31	6	21
/d/ → [d   t]	9	<b>59</b>	<b>30</b>	6	9	30	12	22
/g/ → [g   k]	36	<b>67</b>	<b>59</b>	13	30	43	20	38
/s/ → [s   z]	1	3	4	<b>12</b>	7	4	1	5
/ʃ/ → [ʃ   tʃ]	5	7	2	6	3	<b>15</b>	3	6
/v/ → [v   b]	2	17	14	<b>28</b>	2	<b>23</b>	5	13
/v/ → [v   f]	9	<b>28</b>	8	5	12	15	12	13
/z/ → [z   s]	26	<b>47</b>	32	31	19	<b>79</b>	24	37
/ʒ/ → [ʒ   ʃ]	11	<b>26</b>	14	5	6	<b>25</b>	12	14
/ʒ/ → [ʒ   j]	7	11	7	1	7	<b>29</b>	4	9
/l/ → [l   w]	1	8	2	3	5	1	3	3
/ʁ/ → [ʁ   l]	4	<b>32</b>	7	<b>46</b>	6	7	6	15
/ʁ/ → [ʁ   w]	2	<b>12</b>	4	<b>14</b>	3	5	2	6
/e/ → [e   ε]	17	<b>50</b>	15	39	26	<b>47</b>	19	30
/e/ → [e   i]	15	15	18	8	7	11	9	12
/y/ → [y   u]	8	21	5	<b>35</b>	21	<b>32</b>	3	18
/y/ → [y   i]	32	34	34	26	30	36	26	31
/o/ → [o   ɔ]	18	<b>56</b>	16	32	38	<b>70</b>	45	39
/Ṽ/ → [Ṽ   ṼN   VN]	22	41	28	<b>63</b>	46	<b>69</b>	7	39

Table 7: Pronunciation variant rules for plosives, fricatives, liquids and vowels with corresponding non-standard variant rates (%) aligned using French acoustic models. In the bottom line, [Ṽ] stands for any of the three nasal vowels [ã], [ɛ̃] or [ɔ̃]; [V] for their oral counterparts [ɑ], [ɛ] or [ɔ]; [N] stands for [n] or [m].

622 In 4.1, we saw that English speakers tend to pronounce a fronted /u/.  
623 This is confirmed if we let the system choose an English lax [ʊ] for the  
624 French /u/: it appears that this centralised [ʊ] is aligned in over 50% of  
625 cases for English speakers.

626 The /l/ has a dark allophone in English and Portuguese, contrary to  
627 French (Delattre, 1965). Table 8 witnesses that the variant stemming from  
628 the English [ɫ] model, is more often aligned for English (and Portuguese)  
629 speakers of French than for other speakers – as was the [w] variant in Table 7.

630 For native English speakers, the /ʁ/ alignments paradoxically produce  
631 higher rates for the Spanish [r] than for the English [ɹ], which was checked  
632 perceptually: some speakers do pronounce rolled ‘r’s. These non-standard  
633 realisations of the French /ʁ/ are most often produced by Italian speakers.  
634 Results support the Italians’ tendency to pronounce rolled ‘r’s, as suggested  
635 in the previous subsections: the [r] xenophone shows up in more than 60% of  
636 cases (see Table 8). It is of note that this variant was aligned in less than 10%  
637 of cases for Spanish speakers, which rules out a possible bias stemming from  
638 the acoustic model origin. Instead, Spanish speakers tend to approximate  
639 the French /ʁ/ by a posterior [ɣ]-like sound.

640 Also, in Spanish speakers of French, the high rate of the [s̺] variant  
641 (56%) reflects the tendency in Spanish to realise an apical allophone for the  
642 /s/ phoneme (Alba, 2001). Finally, the palatal fricative [j] (in a majority  
643 of cases) and the [β] (aligned with either /b/ or /v/ in 43% of cases) are  
644 often preferred to the acoustic units corresponding to French phonemes. The  
645 previous alignments with only French acoustic units could not easily account  
646 for this phenomenon.

Rules	Variant rates							
	Ar	En	Ge	It	Po	Sp	Fr	Mean
/std/ → [std xeno-var] (xeno-L1)								
/u/ → [u ʊ] (En)	16	<b>56</b>	12	15	26	38	12	25
/l/ → [l ɫ] (En)	2	<b>10</b>	3	7	8	7	3	6
/ʁ/ → [ʁ ɹ] (En)	3	<b>21</b>	6	<b>22</b>	4	4	2	9
/ʁ/ → [ʁ r] (Sp)	7	33	14	<b>62</b>	12	9	8	21
/b/ → [b β] (Sp)	5	26	16	9	23	<b>43</b>	9	19
/v/ → [v β] (Sp)	8	19	26	36	5	<b>43</b>	5	20
/ʒ/ → [ʒ j] (Sp)	41	35	34	40	45	<b>55</b>	23	39
/s/ → [s ɰ] (Sp)	29	31	30	43	36	<b>56</b>	10	34

Table 8: Pronunciation variant rules involving xenophones for vowels, liquids, plosives and fricatives, with corresponding non-standard (xenophone) variant rates (%) measured from alignments using French acoustic models supplemented with xenophones from English or Spanish (xeno-L1).

### 647 4.3. Conclusion

648 Acoustic measurements including formants, fundamental frequency, seg-  
649 ment durations and voicing rates have been presented for the PFC read  
650 speech using automatically aligned data. Vocalic triangles were plotted, al-  
651 lowing interesting cross-accent comparisons. In particular, they provided  
652 evidence for /u/-fronting in English-accented French and /y/-backing in  
653 Spanish- and Italian-accented French. Also, low voicing rates of voiced stops  
654 (/b/, /d/, /g/) were measured for English and German speakers of French.

655 A series of automatic alignment-based experiments were then carried out  
656 with accent-related variants. Twenty pronunciation variant rules allowed  
657 typical vowel or consonant confusions to be accounted for (e.g. /y/~/u/,  
658 /b/~/p/). For example, voiced consonant devoicing measured in English  
659 and German speakers was highlighted by automatic alignment, through high  
660 non-standard variant rates. Other experiments including xenophone acoustic  
661 models also corroborated some of the results obtained in previous perceptual  
662 cues. The proposed method can thus be used in further experiments for  
663 investigating pronunciation specificities and identifying foreign accents.

## 664 5. Accent identification based on data mining techniques

665 This section investigates whether the cues measured in Section 4 are effec-  
666 tive for automatic accent identification, and which cues are most influential  
667 for a 7-L1 classification task. The experiments proposed in subsection 5.2 aim  
668 at disentangling the relative importance of vowels, consonants and prosody.  
669 Experiments of subsection 5.3 were designed to check which attributes are

670 selected by machine learning techniques. Results of automatic accent identifi-  
671 cation are presented. Comparisons with human perception are also proposed.

### 672 5.1. *Experimental setup*

673 Our experiments rely on the WEKA data mining software ([Witten and](#)  
674 [Frank, 2005](#)), which includes a set of 20 classification algorithms suited for our  
675 type of data, among which Bayesian Networks, Logistic Regression Models,  
676 Multilayer Perceptrons, Support Vector Machines, C4.5 decision trees and  
677 Random Forests.

678 Three experimental configurations were defined. In the first configuration  
679 (PFC-PFC), train and test speakers were different (A-set speakers for train-  
680 ing and B-set speakers for test), but the read material used was the same  
681 (PFC text). In a second configuration (PFC-IPA), not only were speakers  
682 different, the content of the read material also changed (A-set speakers and  
683 PFC text for training as in the former configuration, B-set speakers with IPA  
684 text for test). This contrast will indicate the dependence with respect to the  
685 content, or reversely the genericity of the features for foreign accent classifi-  
686 cation. In a third configuration ( $All_{lv1out}$ ) a leave-one-out method was used  
687 to maximise the available training data (84-1 training speakers reading both  
688 the PFC and the IPA texts). This cross-validation method enabled speaker-  
689 independent tests using one speaker at a time for testing and the rest for  
690 training – with a specific training session for each speaker. Since the content  
691 was shared between training and test, a comparison between PFC-PFC and  
692  $All_{lv1out}$  conditions will indicate the relative need for more data.

693 As a general rule, A-speakers were used for training, B-speakers for test  
694 (there was no development data set). The leave-one-out condition made use

695 of all the read material from all minus one speaker for training and the held-  
696 out speaker for test. All these data had thus to be aligned and processed to  
697 build the corresponding attribute vectors. Since classification performance  
698 may differ to a large extent according to the techniques and the subsets  
699 of cues, classification results were averaged across algorithms. This way,  
700 it is interesting to compare average results over 20 classifiers and perceptual  
701 results averaged over 25 subjects. In addition, automatic classification results  
702 were computed using a majority vote scheme: for a given speaker, all outputs  
703 from the different classifiers were pooled. The majority vote was then carried  
704 out on this answer set to determine the most frequently assigned L1 label for  
705 this speaker.

## 706 5.2. Classification based on linguistic sets of cues

707 The acoustic analyses described in Section 4 resulted in building up a  
708 feature set of 87 cues (termed *Full*), which may be summarised and decom-  
709 posed into feature subsets referred to as follows: *Formants* encompassing  
710 the first two formants of oral vowels ( $2 \times 10$ ), *Consonants* with duration and  
711 voicing measurements ( $2 \times 17$ ), *Prosody* with the two rhythm-related  $\Delta V$  and  
712 PVI measurements as well as the three schwa-related features (5), *French*  
713 *variants* (20) and *Xenophones* (8). In this subsection, we examine the contri-  
714 bution to accent classification of these subsets, keeping in mind that better  
715 classification results may be achieved with more features in a given subset.

716 For the *Full* feature set and each feature subset, Table 9 shows results of a  
717 7-L1 classification task in terms of correct identification averaged across the  
718 20 algorithms (in the left part) and issued by a majority vote (in the right  
719 part) for all experiments (PFC-PFC, PFC-IPA, All<sub>lv1out</sub>). The lower  $N$ -best



Attributes (#)	Average			Majority vote		
	PFC-PFC	PFC-IPA	All <sub>lv1out</sub>	PFC-PFC	PFC-IPA	All <sub>lv1out</sub>
<i>Full</i> (87)	47	35	64	69	50	74
<i>Formants</i> (20)	36	27	45	36	38	59
<i>Consonants</i> (34)	39	19	46	43	33	55
<i>Prosody</i> (5)	26	16	18	31	26	32
<i>French var.</i> (20)	36	32	60	60	38	68
<i>Xenophones</i> (8)	37	30	44	33	31	57
10-best (10)	45	36	56	55	45	60
12-best (12)	48	37	61	62	43	70
15-best (15)	47	35	63	62	45	69

Table 9: Percent correct identification obtained in a 7-L1 classification task by averaging the results of 20 algorithms (left) and by applying a majority vote (right), based on the *Full* set of attributes, linguistic subsets as well as  $N$ -best subsets.

720 rows will be discussed in 5.3.

721 As can be seen, the majority vote provides much better results than the  
722 average-based scheme for almost all experimental conditions. In particular,  
723 when the read text is shared between training and test speakers (PFC-PFC  
724 experiment), the *Full* set results increase from 47% to 69% with the ma-  
725 jority vote. In the All<sub>lv1out</sub> experiment, however, the shared text condition  
726 with more training data and speakers mainly benefits to the averaged re-  
727 sults (+17%), the majority vote results only increasing by 5% absolute to  
728 74%. The latter result corresponds to the best automatic classification rate.  
729 We notice a considerable performance decrease in the more realistic condition

730 (PFC-IPA experiment), where test speakers produce a relatively short speech  
731 sample (1 minute) differing in content from the training data. Nevertheless,  
732 the majority vote achieves 50% correct classification, which is relatively close  
733 to the perceptual results (see 3.2).

734 Concerning the different linguistic subsets of features in the average re-  
735 sults column, *Formants*, *Consonants* and *Prosody* feature sets seem to be  
736 quite sensitive to text and material duration changes. In the PFC-IPA ex-  
737 periment, *Consonants* and *Prosody* results are almost at chance level. For  
738 *Prosody*, the  $All_{lv1out}$  results are even worse than those of the PFC-PFC ex-  
739 periment, whereas for all other cues the  $All_{lv1out}$  results are the best. *French*  
740 *variants* and *Xenophones* prove to be more robust to corpus change: the ab-  
741 solute performance loss is only 4% and 7% respectively, between PFC-PFC  
742 and PFC-IPA experiments. Moreover, these feature sets achieve rather high  
743 results for relatively few features. In particular, the subset of *Xenophones*,  
744 with 8 attributes, achieves relatively stable and good results, as compared to  
745 the *Prosody* subset, with 5 attributes.

746 Considering the majority vote results, in the three experiments, the sub-  
747 set of *French variants* happens to perform best, while the *Prosody* subset  
748 performs worst. Even if the majority vote results are better than the average  
749 results, similar tendencies are observed for both, with maybe the exception of  
750 *Formants*: the latter achieve somewhat better results in the PFC-IPA exper-  
751 iment than in the PFC-PFC experiment, with the majority vote. Some vowel  
752 formants and other features may be more or less relevant for the classification  
753 task, as we are going to see.

754 *5.3. Automatic attribute selection and accent classification*

755 In this subsection, attribute selection is investigated to identify which  
 756 features are most relevant for accent classification. This selection also aims  
 757 at deleting unsuitable attributes, to potentially improve the performance of  
 758 learning algorithms (Guyon and Elisseeff, 2003).

We carried out experiments with 7 attribute selection algorithms im-  
 plemented in the WEKA toolkit, such as Information Gain and Principal  
 Components Analysis methods. As previously, we wanted to smooth the se-  
 lection results by averaging the outputs of the different algorithms. To this  
 aim, we introduced a rank score ( $rsc(j)$ ) for each attribute  $j$ , according to  
 the following formula:

$$rsc(j) = \frac{m(j)}{M} \sum_{i=1}^M \left(1 - \frac{r_i(j)}{J_{max}}\right) \quad (1)$$

759 where  $r_i(j)$  is the rank of attribute  $j$  by the  $i$ -th algorithm,  $m(j)$  is the  
 760 number of algorithms which selected attribute  $j$ ,  $M$  is the total number of  
 761 algorithms (here 7), and  $J_{max}$  corresponds to the total number of attributes.  
 762 The  $\frac{m(j)}{M}$  ratio gives more weight to attributes selected by more algorithms.  
 763 The  $N$ -best attributes then correspond to the attributes with the  $N$  highest  
 764  $rsc$  scores.

765 According to this rank scoring, the  $N$ -best attributes with  $N = 12$  are  
 766 the first two formants of /ə/, the second formants of /e/ and /a/, the rates  
 767 of nasal appendices as well as /z/→[s], /b/→[v], /b/→[p], /d/→[t],  
 768 /e/→[ɛ], /ʁ/→[l] and /ʁ/→[r] variant rates. Extending  $N$  to 15, ad-  
 769 ditional selected attributes are the PVI on vowels, the /ʁ/ length and the  
 770 /v/→[β] variant rate.

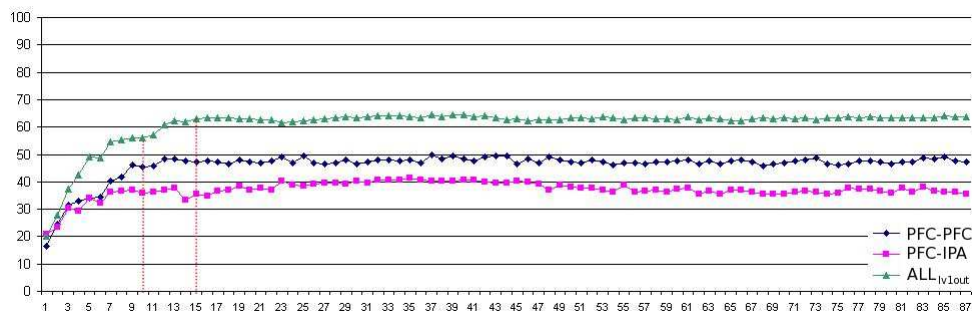


Figure 8: Average performance (%) of the PFC\_PFC, PFC\_IPA and ALL\_lv1out automatic classifications as a function of the number of attributes. Dotted lines indicate 10 and 15 attributes.

771 Results obtained with the 10, 12 and 15-best attributes are given in the  
 772 bottom lines of Table 9. They show how effective attribute selection is, as  
 773 classification results (especially average results), with few features, are quite  
 774 similar to those of the *Full* feature set. Concerning the majority vote, the  
 775 *Full* feature set consistently yields better results than the *N*-best feature sets.  
 776 However, the latter perform better than the linguistic subsets, especially in  
 777 the PFC-IPA experiment, which further demonstrates the effectiveness of  
 778 attribute selection. The *N*-best selected features keep making sense with  
 779 regard to linguistic knowledge and prove to be robust to corpus change.

780 Average results with progressively increasing *N* are shown in Figure 8.  
 781 They turn out to be quite stable above *N*=12 attributes. For example, the  
 782 average correct identification rate remains around 60% in the All\_lv1out exper-  
 783 iment. With a majority vote, results increase above 70%. In the following,  
 784 we will further develop comparisons between automatic and perceptual clas-  
 785 sifications.

hyp\ref	Ar	En	Ge	It	Po	Sp	Fr
Ar	27	<b>28</b>	16	2	2	3	23
En	6	<b>36</b>	16	16	4	7	16
Ge	20	17	<b>23</b>	11	7	8	15
It	11	10	6	<b>51</b>	2	4	17
Po	9	<b>31</b>	11	9	18	12	10
Sp	6	23	3	8	3	<b>50</b>	7
Fr	17	25	0	0	4	1	<b>53</b>

Table 10: Confusion matrix obtained for the PFC-IPA experiment by averaging the confusion matrices of 20 algorithms using the 12-best attributes (%).

786 *5.4. Comparisons with human perception*

787 Even though the results of perceptual experiments (reported in Tables 2  
788 and 3) and those of automatic classification are not directly comparable,  
789 some similarities and differences between them are worth noticing. We al-  
790 ready outlined the similarities and found that the automatically selected at-  
791 tributes were also cited by the perceptual test subjects as being most salient.  
792 Tables 10 and 11 correspond to the confusion matrices averaged over 20 clas-  
793 sifiers using the 12-best attributes for the PFC-IPA and All<sub>lv1out</sub> experiments  
794 respectively. We are here in conditions which are very similar to those of the  
795 7-L1 perceptual experiment. In the PFC-IPA experiment especially, the cor-  
796 rect identification rate is lower (37%) than in the 7-L1 perceptual experiment  
797 (60%), the All<sub>lv1out</sub> experiment achieving 61% correct identification.

798 Both tables consistently reveal that Italian speakers are better identi-  
799 fied by automatic classification than by human subjects' perception: with

hyp\ref	Ar	En	Ge	It	Po	Sp	Fr
Ar	<b>38</b>	8	18	3	10	2	21
En	9	<b>50</b>	20	4	7	5	7
Ge	14	15	<b>48</b>	5	8	6	4
It	4	7	4	<b>64</b>	11	2	8
Po	14	9	10	11	<b>46</b>	7	3
Sp	3	7	7	1	4	<b>77</b>	1
Fr	24	5	3	2	2	1	<b>63</b>

Table 11: Confusion matrix obtained for the All<sub>lv1out</sub> experiment by averaging the confusion matrices of 20 algorithms with the 12-best attributes (%).

800 at least 50% correct identification, Italian and Spanish speakers are well dis-  
801 criminated here, whereas perceptual results showed relatively high confusion  
802 rates between them (see Table 2). It is also interesting to note that auto-  
803 matic classification and human subjects produce many confusions between  
804 English and German speakers. However, these speakers are well identified  
805 in a relative majority of cases. Also, the most frequently assigned accent is  
806 the right one (on the diagonals of Tables 10 and 11) for almost each linguis-  
807 tic origin. The only exceptions are Arabic- and Portuguese-accented French  
808 in the PFC-IPA experiment (Table 10). Indeed, the Arabic speakers kept  
809 for the test have a mild accent (1.5 out of 5, see Table 3) and as already  
810 mentioned the Portuguese accent is difficult to capture.

811 The best score is obtained by native French speakers in the PFC-IPA  
812 experiment and by Spanish speakers in the All<sub>lv1out</sub> experiment. With 77%  
813 correct identification, the gain between these two experiments is notable for

814 Spanish speakers, who are characterised by quite a few robust features.

815 It may be instructive to follow the strategies of one particular classifica-  
816 tion algorithm for comparison with human judgements. The C4.5 algorithm  
817 (implemented in WEKA under the name J48) performs reasonably well and  
818 its decisions are directly interpretable. Figure 9 displays the C4.5 decision  
819 tree using the 12-best attributes (50% correct identification in the PFC-PFC  
820 experiment and 33% in the PFC-IPA experiment). As can be seen, the  
821 identification of Portuguese-accented French is based only on the first two  
822 (normalised) formants of /ə/. When the same algorithm is run using the  
823 15-best or all attributes, the same cues are used for isolating the Portuguese  
824 accent in French. As a matter of fact, only few features characterise this  
825 accent, which was also often mistaken in the perceptual experiments of Sec-  
826 tion 3. The schwa fronting then allows the algorithm to isolate Spanish and  
827 Italian speakers, who are separated by the /b/→[v] variant rate (Spanish  
828 speakers tending to have more /b/s aligned as [v]s than Italians). The  
829 /e/ raising/fronting brings together Arabic and German speakers, who are  
830 discriminated by the /d/→[t] variant rate.

831 The L1 groups gathered by the decision tree may be compared with the  
832 clustering resulting from listeners' answers in the perception test (Figure 3).  
833 Native French speakers were much better distinguished in the perception test.  
834 Since in the extraction of accent-characteristic patterns we were more inter-  
835 ested in foreign accents, we excluded speech rates from the possibly relevant  
836 features. Keeping them would certainly have improved the identification of  
837 native French speakers. The automatic classification of Germans is also dis-  
838 appointing with respect to human perception. An explanation may be that

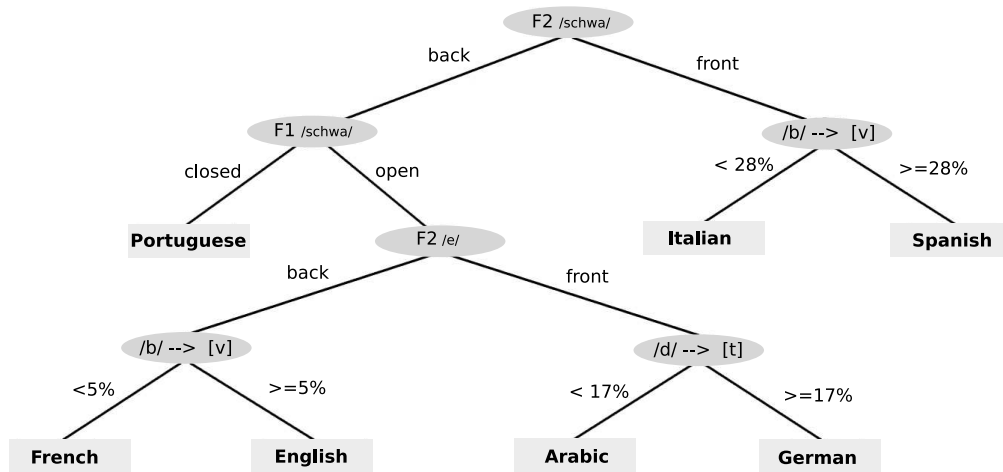


Figure 9: Decision tree resulting from the C4.5 learner implemented in WEKA (J48) using the 12-best attributes.

839 German speakers of the A-set (used to train the C4.5 learner) were judged  
 840 as having the weakest accent in the first perceptual experiment (2.2 out of  
 841 5, see Table 2).

### 842 5.5. Conclusion

843 In this section, classification algorithms of the WEKA data mining soft-  
 844 ware were used. They were either trained with a fixed training corpus (A-set  
 845 speakers) and tested with the remaining B-set speakers or trained and tested  
 846 using a leave-one-out scheme, maximising the training data volume. In the  
 847 former case, tests were carried out either on the text used for training (exper-  
 848 iment PFC-PFC) or in a more realistic setting where the linguistic content of  
 849 the test material was not observed during training (experiment PFC-IPA).  
 850 In the latter case (experiment  $All_{lv1out}$ ), the same read texts were included  
 851 in both the training and the test material, and cross-validation was applied.



852 (There was no development set.) Different feature sets were used, and classifi-  
853 cation results in a 7-L1 classification task were computed either by averaging  
854 them or by applying a majority vote over 20 algorithms. The majority vote  
855 achieved 74% correct identification in the most favourable All<sub>lv1out</sub> condition  
856 (shared text, maximum training data, full 87-feature set) corresponding to  
857 the best results overall. This rate drops to 50% in the more realistic PFC-  
858 IPA condition, where test speakers produce a relatively short speech sample  
859 (1 minute) differing in content from the training data. Classification results  
860 with linguistically-motivated subsets of features (formants, consonant dura-  
861 tion and voicing, prosody) showed that the role of prosody remained modest  
862 in comparison to segments. Globally the best results were achieved using the  
863 formant subset, even in the most difficult not-shared text condition. For ac-  
864 cent identification, our feeling is that improvements could be achieved by the  
865 adding MFCC (mel frequency cepstral coefficient) based features (Huckvale,  
866 2004; , 2007; Ferragne and Pellegrino, 2007). MFCCs are commonly used  
867 in automatic speech recognition systems. However, since they are hardly  
868 interpretable by humans, we did not use them in this study which aimed at  
869 gaining linguistic knowlegde rather than at targeting highest identification  
870 scores.

871 Data mining techniques were also used to rank the most discriminant  
872 cues (among vowel formants, consonant duration and voicing, prosodic cues,  
873 French variants and xenophones), with the objective of determining a concise  
874 set of accent-characteristic features. Restricted sets of 12 and 15 best fea-  
875 tures yielded classification results (69 and 70% by applying a majority vote)  
876 similar to those obtained with the full set of 87 attributes (74%). Formant

877 measurements and aligned pronunciation variant rates which make sense with  
878 respect to linguistic knowledge appear to be the most effective features. Some  
879 of the best-ranked attributes are the first two formants of /ə/, the second  
880 formants of /e/, as well as the percentages of /z/→[s], /b/→[v], /b/→[p],  
881 /d/→[t], and /ʁ/→[r] alignment rates.

## 882 6. Summary and future work

883 This paper described a study of foreign accents in French including Ara-  
884 bic, English, German, Italian, Portuguese and Spanish accents, stemming  
885 from the major foreign languages spoken in France according to statistics  
886 on immigration and tourism. A specially designed corpus of more than 15  
887 hours of read and spontaneous speech from 84 native and non-native speak-  
888 ers was recorded. The read speech part comprises the IPA fable *The North*  
889 *Wind and the Sun* and the PFC text, both widely used in phonetic stud-  
890 ies. Spontaneous speech corresponds to face-to-face conversations in French.  
891 Moreover, free translations of the IPA fable in the speakers' mother tongues  
892 allow L1/L2 comparisons for the same speakers (Vieru-Dimulescu, 2008).

893 The proposed study aimed at investigating foreign accents from percep-  
894 tion, speech processing and data mining perspectives. In perception exper-  
895 iments, an accent degree was assigned to each speaker and his/her L1 was  
896 identified by French subjects on the basis of speech samples. Speech pro-  
897 cessing allowed us to produce objective acoustic measurements and to auto-  
898 matically identify foreign accents by using these measurements together with  
899 classification techniques. The challenge was to check whether accent-specific  
900 features could be highlighted with different approaches, in particular with

901 features estimated from automatically segmented and labelled read speech  
902 data. The main findings with respect to perception, acoustic measurements  
903 and automatic classification are briefly summed up below, according to the  
904 three questions listed in the introduction.

905     Concerning the first question (i) *are native listeners able to identify for-*  
906 *eign accents in French?* perceptual results showed that naive and native  
907 listeners presented with foreign accents which they judged as average suc-  
908 ceeded in identifying the speaker's L1s in slightly above 50%, in an experi-  
909 mental setup with 6 foreign accents. Also, subjects were able to provide a list  
910 of segmental and prosodic cues which characterise a given accent (or more  
911 generally a foreign accent).

912     Regarding the next question (ii) *what acoustic evidence may contribute to*  
913 *corroborate a foreign accent hypothesis?* most of our measurements were car-  
914 ried out on a segmental level, including vowel formants, consonant duration  
915 and voicing rates. Most interesting prosodic patterns were related to rhythm  
916 and words ending with a pronounced schwa. On the other hand, classification  
917 results with linguistically motivated subsets of features (vowels, consonants,  
918 prosody) showed that the relative weight of prosody remained modest in  
919 comparison to segments for recognising foreign accents in French. Major  
920 identified accent-specific cues included schwa fronting or raising, devoicing of  
921 voiced stop consonants as well as /b/~/v/ and /s/~/z/ confusions.

922     As for the third question (iii) *what performance can be achieved by an*  
923 *automatic accent classification system based on perceptually salient features?*  
924 our results showed that rates of 50% correct identification could be achieved  
925 in the most realistic test condition (that of unseen data). Nonetheless, the

926 increase of correct identification scores obtained by applying a leave-one-  
927 out cross-validation method suggests that significant improvements could be  
928 achieved if more training data were available.

929 To the best of our knowledge, this is the first study comparing six non-  
930 native French accents to native French within the same experimental frame-  
931 work, combining perceptual identification, acoustic analyses and automatic  
932 classification. The same methodology can be extended to other accents and  
933 other languages. However, identifying a foreign accent remains a difficult  
934 task for both humans and automatic classification devices. Hints of foreign  
935 accents may be more or less frequent, which we did not take into account  
936 in our experiments based on automatic speech processing. Indeed, they may  
937 be grasped, not continuously throughout a speaker's speech flow, but only  
938 on some episodic events. In the future, we would like to address the issue  
939 as to whether foreign-accented speech could be more accurately identified  
940 in specific experimental setups, including more informed test material and  
941 more informed native subjects. A system capable of yielding a measure of  
942 accentedness like the one developed by [Sangwan and Hansen \(2009\)](#) would  
943 be another possible application.

944 Even though our brain does not work like a machine, an important  
945 byproduct of this work was the demonstration that perceptually relevant  
946 features could be used relatively successfully to identify foreign accents in  
947 French. The automatic alignment-based approach is thus particularly promis-  
948 ing since, unlike perceptual tests, it enables a number of pronunciation-  
949 related hypotheses to be tested quite rapidly. We hope that the large panel  
950 of results which emerged from different automatic alignments will contribute

951 to inspire upcoming research directions both within phonetic sciences and in  
952 the speech processing domain.

953 Future work is scheduled to measure subsegmental features, namely the  
954 voice onset time of stop consonants, expecting that this piece of information  
955 will be useful for automatic identification. Additionally, the most relevant  
956 pronunciation variants aligned can be used so as to make other measure-  
957 ments. Concerning further research in automatic accent identification, the  
958 proposed method deserves to be compared and combined with more standard  
959 approaches (such as cepstral features along with Gaussian Mixture Models  
960 and Support Vector Machines). For automatic speech recognition research,  
961 specific acoustic models and pronunciation dictionaries can be designed in  
962 order to reduce error rates on foreign-accented speech. Finally, benefit can  
963 be derived from speech synthesis: better than human imitators who are prone  
964 to reinforce and caricature certain features, speech synthesis may be a good  
965 simulation tool. Last but not least, we hope the outcomes of this work could  
966 be useful for learning and teaching French as a foreign language.

## 967 **References**

- 968 Abdelli-Beruh, N., 2004. The stop voicing contrast in french sentences. *Pho-*  
969 *netica* 61, 201–219.
- 970 Adank, P., 2003. Vowel normalisation: a perceptual acoustic study of Dutch  
971 vowels. Ph.D. thesis, Radboud University Nijmegen.
- 972 Adda-Decker, M., Hallé, P., 2007. Bayesian framework for voicing alternation  
973 and assimilation studies on large corpora in French. In: Proc. of the Inter-

- 974 national Conference of Phonetic Sciences (ICPhS). Saarbrücken, Germany,  
975 pp. 613–616.
- 976 Adda-Decker, M., Lamel, L., 1999. Pronunciation variants across system con-  
977 figuration, language and speaking style. *Speech Communication* 29, 83–98.
- 978 Alba, O., 2001. *Manual de fonética hispánica*. Editorial Plaza Mayor, San  
979 Juan.
- 980 Angkititrakul, P., Hansen, J., 2003. Use of trajectory models for auto-  
981 matic accent identification. In: *Proc. Interspeech*. Geneva, Switzerland,  
982 pp. 1353–1356.
- 983 Arai, T., Greenberg, S., 1997. The temporal properties of spoken Japanese  
984 are similar to those of English. In: *Proc. of the European Conference on*  
985 *Speech Communication and Technology (Eurospeech)*. Rhodes, Greece, pp.  
986 1011–1014.
- 987 Arslan, L., Hansen, J., 1997. A study of temporal features and frequency the  
988 *Acoustical Society of America* 102 (1), 28–40.
- 989 Bartkova, K., Jouvét, D., 2004. Foreign accent processing in automatic speech  
990 recognition. In: *Proc. SPECOM - International Conference on Speech and*  
991 *Computer*. Saint-Petersburg, Russia, pp. 22–28.
- 992 Berkling, K., August 2001. Scope, syllable core and periphery evaluation:  
993 Automatic syllabification and foreign accent identification. *Speech Com-*  
994 *munication* 35 (1-2), 125–138.

- 995 Boersma, P., 2001. Praat, a system for doing phonetics by computer. *Glott*  
996 *International* 5 (9/10), 341–345.
- 997 Bouselmi, G., Fohr, D., Illina, I., Haton, J.-P., 2006. Multilingual non-native  
998 speech recognition using phonetic confusion-based acoustic model modifi-  
999 cation and graphemic constraints. In: *Proc. Interspeech*. Pittsburgh, USA,  
1000 pp. 109–112.
- 1001 Calliope, 1989. *La parole et son traitement automatique*. Masson, Paris.
- 1002 Cincarek, T., Gruhn, R., Nakamura, S., 2004. Speech recognition for multiple  
1003 non-native accent groups with speaker-group-dependent acoustic models.  
1004 In: *Proc. Interspeech*. Jeju, Korea, pp. 1509–1512.
- 1005 Clopper, C., Pisoni, D., 2004. Some acoustic cues for the perceptual cate-  
1006 gorization of american english regional dialects. *Journal of Phonetics* 32,  
1007 111–140.
- 1008 Boula de Mareüil, P., Brahimi, B., Gendrot, C., 2004. Role of segmental and  
1009 suprasegmental cues in the perception of Maghrebian-accented French. In:  
1010 *Proc. Interspeech*. Jeju, Korea, pp. 341–344.
- 1011 Boula de Mareüil, P., Vieru-Dimulescu, B., 2006. The contribution of prosody  
1012 to the perception of foreign accent. *Phonetica* 63, 247–267.
- 1013 Delattre, P., 1965. *Comparing the phonetic features of English, French, Ger-*  
1014 *man and Spanish*. Julius Groos Verlag, Heidelberg.
- 1015 Dellwo, V., 2010. Influences of speech rate on the acoustic correlates of speech

- 1016 rhythm: An experimental phonetic study based on acoustic and perceptual  
1017 evidence. Ph.D. thesis, University of Bonn.
- 1018 Diez, F. G., Dellwo, V., Gavaldà, N., Rosen, S., 2008. The development of  
1019 measurable speech rhythm during second language acquisition. *Journal of*  
1020 *the Acoustical Society of America* 123 (5), 3886–3886.
- 1021 Disner, S., 1980. Evaluation of vowel normalization procedures. *Journal of*  
1022 *the Acoustical Society of America* 67, 253–261.
- 1023 Durand, J., Laks, B., Lyche, C., 2003. Le projet phonologie du français  
1024 contemporain. *La Tribune Internationale des Langues Vivantes* 33, 3–9.
- 1025 Ferragne, E., Pellegrino, F., 2007. Automatic dialect identification: A study  
1026 of british english. In: Müller, C., Schötz, S. (Eds.), *Speaker Classification*  
1027 *II/2*, Springer Verlag, Berlin, 243–257.
- 1028 Flege, J., 1984. The detection of french accent by american listeners. *Journal*  
1029 *of the Acoustical Society of America* 76 (3), 692–707.
- 1030 Flege, J., Hammond, R., 1982. Mimicry of non-distinctive phonetic differ-  
1031 ences between language varieties. *Studies in Second Language Acquisition*  
1032 5, 1–17.
- 1033 Flege, J., Port, R., 1981. Cross-language phonetic interference: Arabic to  
1034 english. *Language and Speech* 24 (2), 125–146.
- 1035 Flege, J., Schirru, C., MacKay, I., 2003. Interaction between the native and  
1036 second language phonetic subsystems. *Speech Communication* 40, 467–491.



- 1037 Fouché, P., 1959. *Traité de prononciation française*. Éditions Klincksieck,  
1038 Paris.
- 1039 Freland-Ricard, M., 1996. Organisation temporelle et rythmique chez les ap-  
1040 prenants étrangers. *etude multilingue. Revue de phonétique appliquée* 118–  
1041 119, 61–91.
- 1042 Frota, S., D’Imperio, M., Elordieta, G., Prieto, P., Vigario, M., 2007. The  
1043 phonetics and phonology of intonational phrasing in romance. In: Prieto,  
1044 P., Mascaró, J. (Eds.), *Segmental and Prosodic Issues in Romance Phonol-*  
1045 *ogy*. John Benjamins, Amsterdam/Philadelphia, 131–153.
- 1046 Gauvain, J.-L., Adda, G., Adda-Decker, M., Allauzen, A., Gendner, V.,  
1047 Lamel, L., Schwenk, H., 2005. Where are we in transcribing French broad-  
1048 cast news? In: *Proc. Interspeech*. Lisbon, Portugal, pp. 1665–1668.
- 1049 Gendrot, C., Adda-Decker, M., 2005. Impact of duration on F1/F2 formant  
1050 values of oral vowels: an automatic analysis of large broadcast news cor-  
1051 pora in French and German. In: *Proc. Interspeech*. Lisbon, Portugal, pp.  
1052 2453–2456.
- 1053 Ghazali, S., Hamdi, R., Barkat, M., 2002. Speech rhythm variation in Arabic  
1054 dialects. In: *Proc. Speech Prosody*. Aix-en-Provence, France, pp. 331–334.
- 1055 Goronzy, S., 2004. Generating non-native pronunciation variants for lexical  
1056 adaptation. *Speech Communication* 42, 109–123.
- 1057 Grabe, E., Low, F., 2002. Durational variability in speech and the rhythm

- 1058 class hypothesis. In: Gussenhoven C. / Warner N. (ed.), Papers in Labo-  
1059 ratory Phonology VII. The Hague: Mouton de Gruyter, 515–546.
- 1060 Guyon, I., Elisseeff, A., 2003. An introduction to variable and feature selec-  
1061 tion. *Journal of Machine Learning Research* 3, 1265–1287.
- 1062 Harrington, J., Palethorpe, S., C. Watson, C., 2000. Does the queen speak  
1063 the queen’s english? *Nature* 408, 927–928.
- 1064 Huckvale, M., 2004. ACCDIST: a Metric for Comparing Speakers’ Accents.  
1065 In: Proc. of the International Conference on Spoken Language Processing  
1066 (ICSLP). Jeju, Korea, pp. 29–32.
- 1067 Huckvale, M., 2007. Hierarchical clustering of speakers into accents with the  
1068 ACCDIST metric. In: Proc. of the International Conference of Phonetic  
1069 Sciences (ICPhS). Saarbrücken, Germany, pp. 1821–1824.
- 1070 Ihaka, R., Gentleman, R., 1996. R: A language for data analysis and graphics.  
1071 *Journal of Computational and Graphical Statistics* 5 (3), 299–314.
- 1072 Jilka, M., 2000. The contribution of intonation to the perception of foreign  
1073 accent. Ph.D. thesis, University of Stuttgart, Stuttgart, Germany.
- 1074 Kumpf, K., King, R., 1997. Foreign speaker accent classification using  
1075 phoneme-dependent accent discrimination models and comparisons with  
1076 human perception benchmarks. In: Proc. of the European Conference on  
1077 Speech Communication and Technology (Eurospeech). Rhodes, Greece, pp.  
1078 2323–2326.

- 1079 Lamel, L., Gauvain, J.-L., Adda, G., Barras, C., Bilinski, E., Galibert, O.,  
1080 Pujol, A., Schwenk, H., Zhu, X., 2007. The LIMSI 2006 TC-STAR EPPS  
1081 Transcription Systems. In: Proc. of the International Conference on Acous-  
1082 tics, Speech, and Signal Processing (ICASSP). Hawaii, USA, pp. 997–1000.
- 1083 Livescu, K., Glass, J., 2000. Lexical modeling of non-native speech for au-  
1084 tomatic speech recognition. In: Proc. of the International Conference on  
1085 Acoustics, Speech, and Signal Processing (ICASSP). Istanbul, Turkey, pp.  
1086 1683–1686.
- 1087 Magen, H., 1998. The perception of foreign-accented speech. *Journal of Pho-*  
1088 *netics* 26, 381–400.
- 1089 Martin, A., Le, A., 2008. NIST 2007 Language Recognition Evaluation. In:  
1090 *Odyssey - The Speaker and Language Recognition Workshop*. Stellenbosch,  
1091 South Africa, paper 016.
- 1092 Nearey, T. M., 1989. Static, dynamic, and relational properties in vowel  
1093 perception. *Journal of the Acoustical Society of America* 85, 2088–2113.
- 1094 Quilis, A., 1993. *Tratado de fonología y fonética españolas*. Biblioteca  
1095 románica hispánica, Madrid.
- 1096 Ramus, F., 1999. *Rythme des langues et acquisition du langage*. Ph.D. thesis,  
1097 EHESS, Paris, France.
- 1098 Raux, A., 2004. Automated lexical adaptation and speaker clustering based  
1099 on pronunciation habits for non-native speech recognition. In: Proc. Inter-  
1100 speech. Jeju, Korea, pp. 613–616.

- 1101 Romano, A., 2010. Speech rhythm and timing: structural properties and  
1102 acoustic correlates. *La dimensione temporale del parlato*, (S. Schmid, M.  
1103 Schwarzenbach, D. Studer, editors), Torriana: EDK Editore, 45–75.
- 1104 Rouas, J.-L., Trancoso, I., Viana, C., Abreu, M., 2008. Language and vari-  
1105 ety verification on broadcast news for portuguese. *Speech Communication*  
1106 50 (11-12), 965–979.
- 1107 Sangwan, A., Hansen, J., 2009. On the use of phonological features for auto-  
1108 matic accent analysis. In: *Proc. Interspeech*. Brighton, UK, pp. 172–175.
- 1109 Schaden, S., 2003. Crosstowns: Automatically generated phonetic lexicons  
1110 of cross-lingual pronunciation variants of European city names. In: *Proc.*  
1111 *of the International Conference on Language Resources and Evaluation*  
1112 (LREC). Lisbon, Portugal, pp. 1395–1398.
- 1113 Silke, G., Stefan, R., Ralf, K., 2004. Generating non-native pronunciation  
1114 variants for lexicon adaptation. *Speech Communication* 42 (1), 109–123.
- 1115 ten Bosch, L., Cremelie, N., 2002. Pronunciation modelling and lexical adap-  
1116 tation using small training sets. In: *ISCA Workshop on Pronunciation*  
1117 *Modeling and Lexicon Adaptation*. Aspen Lodge, USA, pp. 111–116.
- 1118 Veloso, J., 2007. Schwa in European Portuguese: The phonological status of  
1119 ə. In: *Actes des 5es Journées d'Études Linguistiques*. Nantes, pp. 55–60.
- 1120 Vieru-Dimulescu, B., 2008. Caractérisation et identification d'accents  
1121 étrangers en français. Ph.D. thesis, University Paris Sud, Orsay, France.

- 1122 Witten, I., Frank, E., 2005. Data Mining: Practical Machine Learning Tools  
1123 and Techniques. Morgan Kaufmann.
- 1124 Woehrling, C., Boula de Mareüil, P., 2006. Identification d'accents régionaux  
1125 en français : perception et analyse. *Revue Parole* 37, 25–65.
- 1126 Woehrling, C., Boula de Mareüil, P., Adda-Decker, M., 2009. Linguistically-  
1127 motivated automatic classification of regional French varieties. In: Proc.  
1128 Interspeech. Brighton, UK, pp. 2183–2186.
- 1129 Yamada, R., Strange, W., Magnuson, J., Pruitt, J., Clarke, W., 1994. The  
1130 intelligibility of Japanese speakers's productions of American English /r/,  
1131 /l/ and /w/, as evaluated by native speakers of American English. In: Proc.  
1132 of the International Conference on Spoken Language Processing (ICSLP).  
1133 Yokohama, Japan, pp. 2023–2026.