Characterisation and identification of non-native French accents

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

Characterisation and identification of non-native French accents

Bianca Vieru[1], 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-

[1]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. Introduction

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

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

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

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

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

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

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

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

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

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

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

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

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

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

2. Corpus

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

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

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

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

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

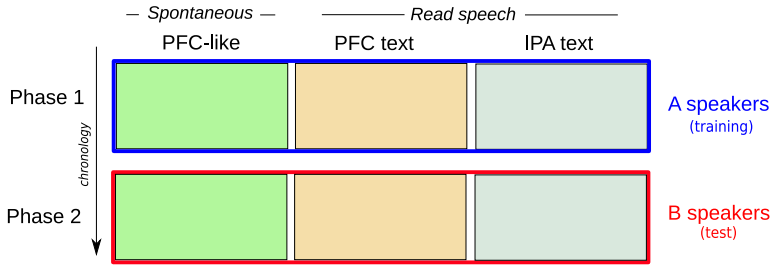

Figure 1: Overview of the composition of the foreign accent corpus, recorded during two phases. The earlier A set is used for acoustic analyses and automatic classifier training, the later B set for classifier testing. Each set includes 42 speakers (6 speakers of 6 foreign accents and of native French).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

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

3. Perceptual experiments

3.1. Task, protocol, experiments and listeners

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

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

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

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

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

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

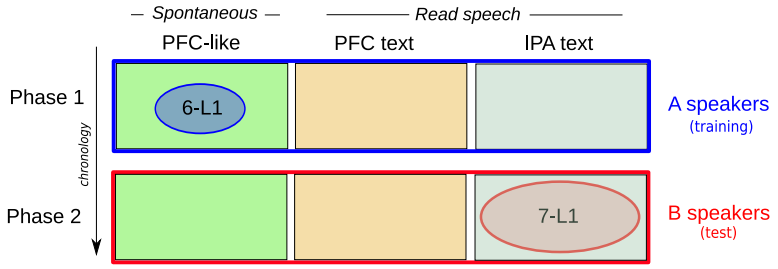

Figure 2: Parts of the corpus involved in the 6-L1 and 7-L1 perceptual experiments.

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

3.2. Results

We first present a summary of the listeners' self-evaluations and the measured degrees of accentedness, before reporting on the perceptual identification results proper. Perceptual distances between the 6 investigated foreign accents are then captured via multidimensional scaling and clustering techniques, and synthetically displayed in a graphical representation. The perceptual cues reported by the participants will follow.

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

Data set	Task	Ar	En	Ge	It	Po	Sp	Fr
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).

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

Accent identification results. The results of the 6-L1 and 7-L1 accent identification tests are shown in Tables 2 and 3 respectively. For both experiments, the average identification rates are above 50%, even though there is considerable cross-accent variation. The overall identification rates are 52% in the 6-L1 experiment, whereas the 7-L1 experiment achieves 60% correct identification. The better 7-L1 task results are mainly due to the near-perfect

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

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

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.

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

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

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

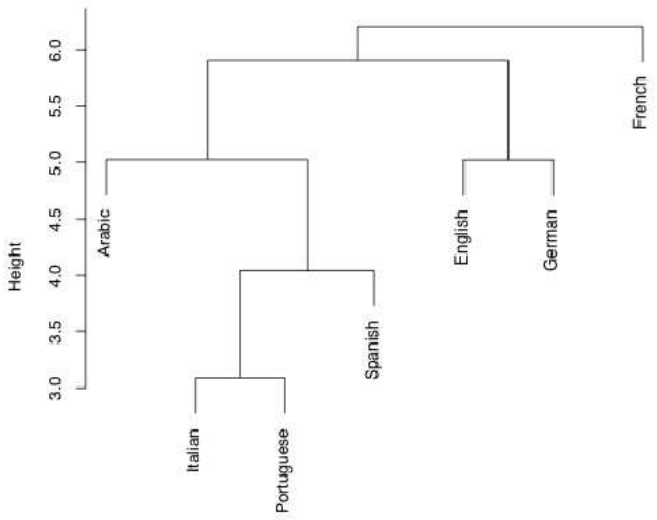

Figure 3: Dendrogram of the identification results of the 7-L1 perceptual experiment.

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

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

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

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

3.3. Conclusion

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

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

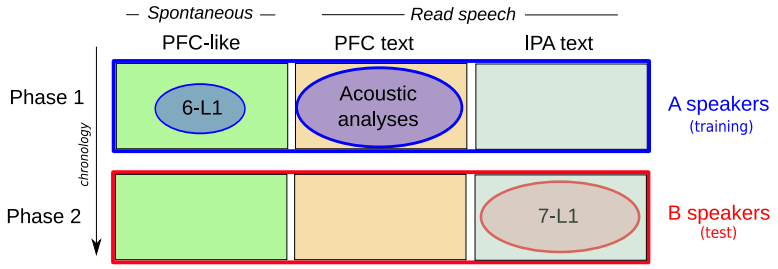

Figure 4: Sub-corpus of the acoustic analyses: PFC read speech from A-set speakers.

4. Acoustic analyses using automatic alignments

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

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

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

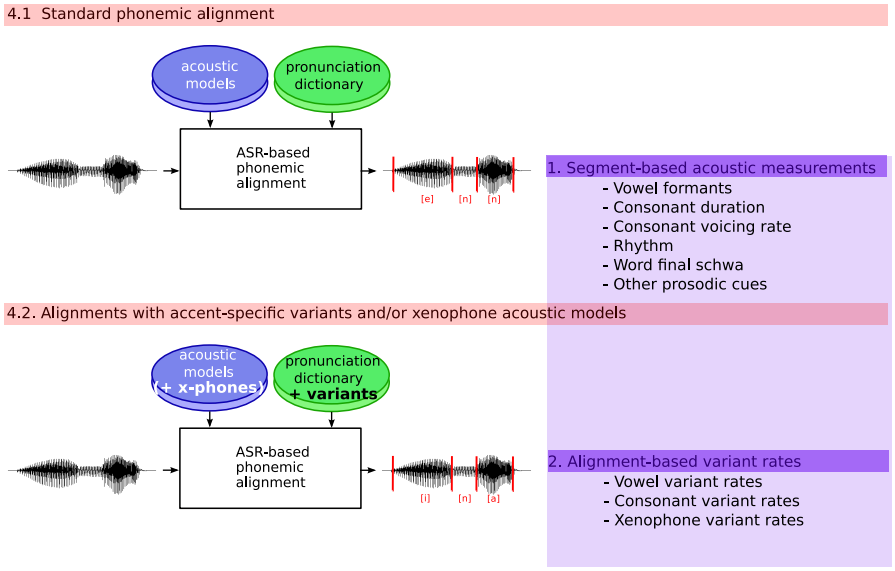

Figure 5: Synthetic overview of the experimental workflow of section 4 involving automatic alignments: standard phonemic alignment in subsection 4.1, accent-specific variants without/with xenophone acoustic models in subsection 4.2 to produce accent-related features (listed in the shaded box). See text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

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

4.1. Measurements based on standard phonemic alignment

4.1.1. Vowel formants

Formant frequencies were measured on oral vowels (more than 500 vowels per speaker) using the Praat software (Boersma, 2001). The first two

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

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

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

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

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

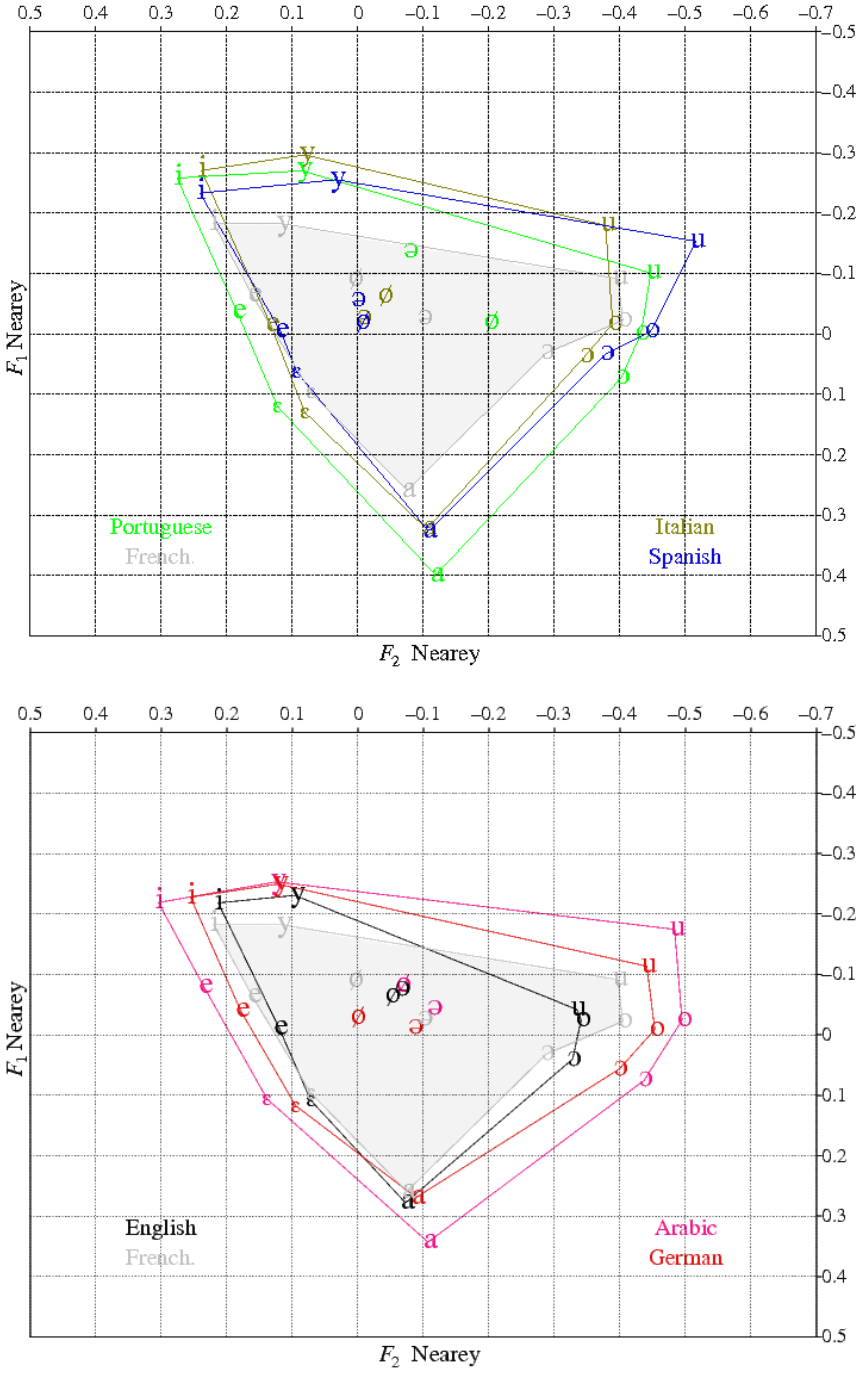

Figure 6: Vocalic triangles of native or foreign-accented French, for the PFC text.

	Ar	**En**	**Ge**	**It**	**Po**	**Sp**	**Fr**	**Mean**
/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.

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

4.1.2. Consonant duration and voicing rates

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

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

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

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

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

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

Some of the subjects' comments during the perceptual experiments related to rhythmic aspects. As mentioned in the introduction, parameters involving segment durations have been proposed to characterise rhythmic

φ	**Ar**	**En**	**Ge**	**It**	**Po**	**Sp**	**Fr**	**Mean**
/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
/ʁ/	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.

classes. Ramus (1999) considers the proportion of vocalic intervals (%V) and the duration variation of consonantal intervals in terms of standard deviation (ΔC). A consonantal interval is composed of one or more consecutive consonantal segments delimited by vowels or pauses. Grabe and Low (2002) propose a slightly more complex approach: they measure the variability between successive vocalic and intervocalic intervals via so-called "Pairwise Variability Indices" (PVIs), possibly normalised to account for speech rate variation. These parameters do not explicitly take into account the notion of stress, but rely on the link between stress-timing, complexity of consonant clusters and vowel reduction. More recent work not investigated here aims at adapting these measures for quantifying second language proficiency (Diez et al., 2008) and to study the influence of speech rate on acoustic correlates of rhythm (Dellwo, 2010).

In the following, we measured Ramus and Grabe's parameters on foreign accents. The same complexity of consonant clusters is imposed to all speakers by the French language. Hence, the measured duration variability should be only poorly related to phonotactic differences. Concerning complex consonant clusters, one could imagine that speakers who are not used to such productions would tend to articulate the latter carefully (unless they resort to elisions), resulting in a total duration close to the sum of individual consonant durations. This could then lead to high ΔC measurements in opposition to observations made in their mother tongues. Conversely, speakers who are used to complex consonant clusters may produce them with relatively shorter durations, thereby making %C decrease and %V increase. Resulting measurements would then contradict the tendency of stress-timed languages

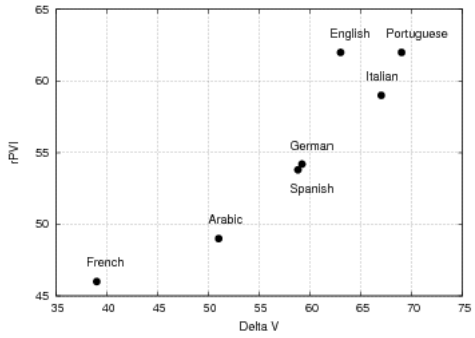

Figure 7: Rhythm characterisation of native and foreign-accented French combining Ramus's ΔV and Grabe's PVI on vowels. L1s in increasing ΔV order: French, Arabic, Spanish, German, English, Italian and Portuguese. Durations are in ms.

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

Figure 7 shows the results for the different L1s. As expected, French appears in the left bottom corner. One might have further expected a systematic grouping of syllable-timed vs stress-timed L1s. However, no clear rhythmic classes emerge. Results merely show an important difference (according to ΔV and PVI) between Arabic (closest to French) and e.g. Italian or Portuguese speakers. According to these measurements, Maghrebian speakers do not tend to reproduce the stress-timed rhythm of their native dialects.

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

	Ar	**En**	**Ge**	**It**	**Po**	**Sp**	**Fr**	**Mean**
$\%V_{schwa}$	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.

ening of the vowels preceding them and the corresponding f_0 contours. There are 123 polysyllabic word-final schwa sites in the PFC text, potentially realised by the speakers and detected by the alignment system, as word-final schwas were set optional in the standard pronunciation dictionary used for alignment. Table 6 shows the results obtained. Italian speakers of French yield by far the highest word-final schwa realisation rate (with 23%), all the other groups of speakers keeping rates below 15%. These high numbers for Italian may be at least partially explained by the fact that in their mother tongue, content words ending with a consonant are extremely rare: speakers thus tend to add a word-final schwa in this situation. Table 6 further gives duration ratios (respectively Δf_0 as difference in semitones) between the vowel preceding the word-final schwa and the pronounced final schwa. One may notice that Italians have both the highest duration ratio for the supposedly stressed syllable (whose nucleus is the vowel preceding the final schwa) and the highest Δf_0 corresponding to a pitch lowering on the final schwa. By contrast, native Germans speaking French display a slightly rising f_0 contour on the final syllable (the only negative Δf_0 in Table 6). Both

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

Since the figures were computed on a relatively small number of occurrences we felt it necessary to perform ANOVAs. This was done with the measurements ($\%\mathrm{V}_{schwa}$, duration ratio and Δf_0) averaged by speaker as the dependent variables and the L1s as the independent variables. The difference did not reach significance level for the duration ratio, but the L1 effect was found to be significant for the schwa realisation rate [$F(6,35) = 3.86, p < 0.01$] and Δf_0 [$F(6,35) = 3.04, p < 0.05$].

4.2. Measurements based on non-standard alignments

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

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

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

4.2.1. Variants using the standard French acoustic model set

Based on linguistic knowledge about the different L1s, comments from the perceptual tests and results of acoustic analyses, we defined a set of about 20 non-native French variant rules accounting for common variation processes such as voicing/devoicing, spirantisation and affrication for consonants, opening/closing, fronting/backing and denasalisation for vowels. Automatic alignment using the corresponding variants simulates a binary, categorical approach to foreign-accented speech, supplementing gradient analyses based on formant, f_0 and duration measurements. A list of rules involving 20 tested variants can be found in Table 7. Generally, they consist in a paradigmatic choice between phone pairs (/d/→[d|t]). In some cases, the rule may become more complex, with additional segment insertions and contextual constraints (e.g. /ɑ̃/→[ɑ̃|ɑ̃n|ɑn] or [ɑ̃|ɑ̃m|ɑm] in a right [b|p] context). For any rule, the left phoneme can be aligned as any of the options listed on the right, conventionally noted between square brackets since they refer to phonetic realisations. For each rule, a specific pronunciation dictionary was generated, a distinct alignment was produced accordingly and a

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

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

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

4.2.2. Variants using the xenophone-augmented acoustic model set

The previous variants were designed to evaluate potential confusions between French phonemes made by non-native speakers. In this subsection, the standard French acoustic model set is supplemented with foreign acoustic models, in order to account for non-native productions which may be "far" from the target units or "intermediate" between two French phonemes. Hereafter, we address the cases of the French /b/, /v/, /ʒ/, /s/, /l/, /ʁ/ and /u/ phonemes, mapped with L1 phonemes or allophones borrowed from different L1s: their realisations, motivated by linguistic mechanisms, may be particular to speakers of certain origins (Delattre, 1965). For technical reasons, the L1s are limited to Spanish and English, for which extensively trained acoustic models are available within the corresponding LIMSI speech-to-text systems (Lamel et al., 2007). The models we added are [r], [s̪], [j] and [β] from Spanish, [ɪ], [ɬ] and [ʊ] from English, as can be seen in Table 8 with alignment results.

Rules			Variant rates							
/standard/	→	[standard | variants]	**Ar**	**En**	**Ge**	**It**	**Po**	**Sp**	**Fr**	**Mean**
/b/	→	[b | v]	8	30	32	8	22	**60**	3	23
/b/	→	[b | p]	8	**55**	**42**	3	6	31	6	21
/d/	→	[d | t]	9	**59**	**30**	6	9	30	12	22
/g/	→	[g | k]	36	**67**	**59**	13	30	43	20	38
/s/	→	[s | z]	1	3	4	**12**	7	4	1	5
/ʃ/	→	[ʃ | tʃ]	5	7	2	6	3	**15**	3	6
/v/	→	[v | b]	2	17	14	**28**	2	**23**	5	13
/v/	→	[v | f]	9	**28**	8	5	12	15	12	13
/z/	→	[z | s]	26	**47**	32	31	19	**79**	24	37
/ʒ/	→	[ʒ | ʃ]	11	**26**	14	5	6	**25**	12	14
/ʒ/	→	[ʒ | j]	7	11	7	1	7	**29**	4	9
/l/	→	[l | w]	1	8	2	3	5	1	3	3
/ʁ/	→	[ʁ | l]	4	**32**	7	**46**	6	7	6	15
/ʁ/	→	[ʁ | w]	2	**12**	4	**14**	3	5	2	6
/e/	→	[e | ɛ]	17	**50**	15	39	26	**47**	19	30
/e/	→	[e | i]	15	15	18	8	7	11	9	12
/y/	→	[y | u]	8	21	5	**35**	21	**32**	3	18
/y/	→	[y | i]	32	34	34	26	30	36	26	31
/o/	→	[o | ɔ]	18	**56**	16	32	38	**70**	45	39
/$\tilde{V}$/	→	[$\tilde{V}$ | $\tilde{V}$N | VN]	22	41	28	**63**	46	**69**	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, [$\tilde{V}$] stands for any of the three nasal vowels [$\tilde{a}$], [$\tilde{\varepsilon}$] or [$\tilde{\jmath}$]; [V] for their oral counterparts [ɑ], [ɛ] or [ɔ]; [N] stands for [n] or [m].

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

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

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

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

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

4.3. Conclusion

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

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

5. Accent identification based on data mining techniques

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

selected by machine learning techniques. Results of automatic accent identification are presented. Comparisons with human perception are also proposed.

5.1. Experimental setup

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

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

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

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

5.2. Classification based on linguistic sets of cues

The acoustic analyses described in Section 4 resulted in building up a feature set of 87 cues (termed *Full*), which may be summarised and decomposed into feature subsets referred to as follows: *Formants* encompassing the first two formants of oral vowels (2×10), *Consonants* with duration and voicing measurements (2×17), *Prosody* with the two rhythm-related ΔV and PVI measurements as well as the three schwa-related features (5), *French variants* (20) and *Xenophones* (8). In this subsection, we examine the contribution to accent classification of these subsets, keeping in mind that better classification results may be achieved with more features in a given subset.

For the *Full* feature set and each feature subset, Table 9 shows results of a 7-L1 classification task in terms of correct identification averaged across the 20 algorithms (in the left part) and issued by a majority vote (in the right part) for all experiments (PFC-PFC, PFC-IPA, All_{lv1out}). The lower *N*-best

Attributes (#)	Average			Majority vote		
	PFC-PFC	PFC-IPA	All_{lv1out}	PFC-PFC	PFC-IPA	All_{lv1out}
Full (87)	47	35	64	69	50	74
Formants (20)	36	27	45	36	38	59
Consonants (34)	39	19	46	43	33	55
Prosody (5)	26	16	18	31	26	32
French var. (20)	36	32	60	60	38	68
Xenophones (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.

rows will be discussed in 5.3.

As can be seen, the majority vote provides much better results than the average-based scheme for almost all experimental conditions. In particular, when the read text is shared between training and test speakers (PFC-PFC experiment), the *Full* set results increase from 47% to 69% with the majority vote. In the All_{lv1out} experiment, however, the shared text condition with more training data and speakers mainly benefits to the averaged results (+17%), the majority vote results only increasing by 5% absolute to 74%. The latter result corresponds to the best automatic classification rate. We notice a considerable performance decrease in the more realistic condition

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

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

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

5.3. Automatic attribute selection and accent classification

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

We carried out experiments with 7 attribute selection algorithms implemented in the WEKA toolkit, such as Information Gain and Principal Components Analysis methods. As previously, we wanted to smooth the selection 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:

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

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

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

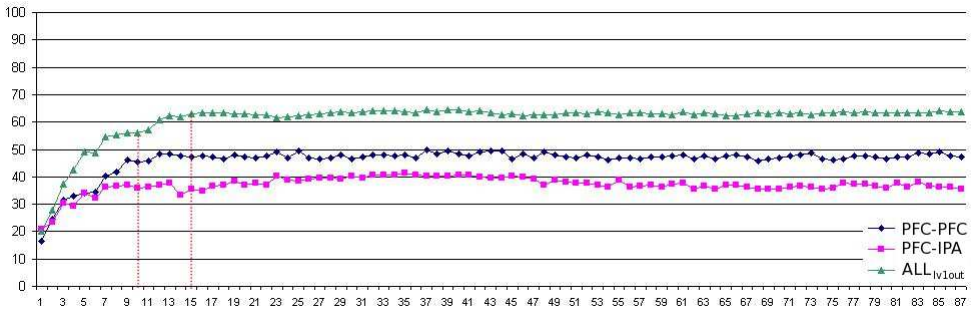

Figure 8: Average performance (%) of the PFC_PFC, PFC_IPA and ALL_lv1out automatic classifications as a function of the number of attributes. Dotted lines indicate 10 and 15 attributes.

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

Average results with progressively increasing N are shown in Figure 8. They turn out to be quite stable above N=12 attributes. For example, the average correct identification rate remains around 60% in the All_{lv1out} experiment. With a majority vote, results increase above 70%. In the following, we will further develop comparisons between automatic and perceptual classifications.

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

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

5.4. Comparisons with human perception

Even though the results of perceptual experiments (reported in Tables 2 and 3) and those of automatic classification are not directly comparable, some similarities and differences between them are worth noticing. We already outlined the similarities and found that the automatically selected attributes were also cited by the perceptual test subjects as being most salient. Tables 10 and 11 correspond to the confusion matrices averaged over 20 classifiers using the 12-best attributes for the PFC-IPA and All_{lv1out} experiments respectively. We are here in conditions which are very similar to those of the 7-L1 perceptual experiment. In the PFC-IPA experiment especially, the correct identification rate is lower (37%) than in the 7-L1 perceptual experiment (60%), the All_{lv1out} experiment achieving 61% correct identification.

Both tables consistently reveal that Italian speakers are better identified by automatic classification than by human subjects' perception: with

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

Table 11: Confusion matrix obtained for the All_{lv1out} experiment by averaging the confusion matrices of 20 algorithms with the 12-best attributes (%).

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

The best score is obtained by native French speakers in the PFC-IPA experiment and by Spanish speakers in the All_{lv1out} experiment. With 77% correct identification, the gain between these two experiments is notable for

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

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

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

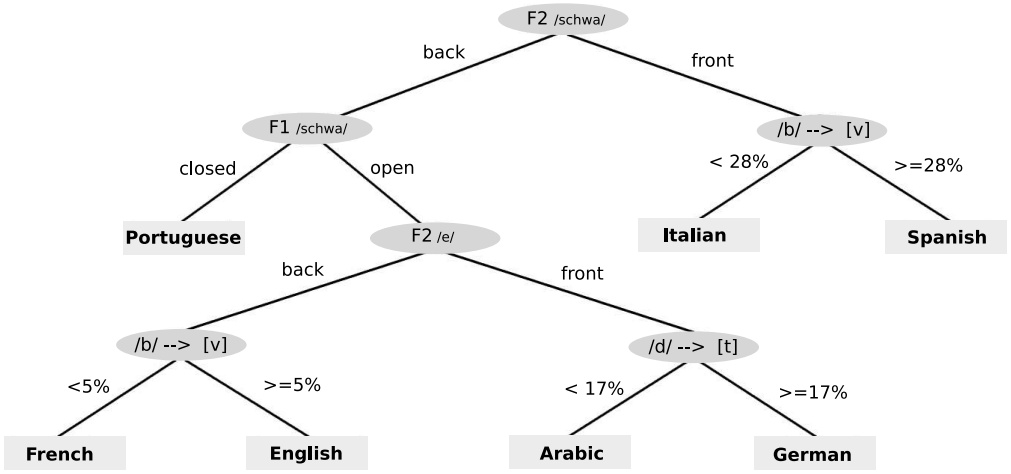

Figure 9: Decision tree resulting from the C4.5 learner implemented in WEKA (J48) using the 12-best attributes.

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

5.5. Conclusion

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

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

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

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

6. Summary and future work

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

The proposed study aimed at investigating foreign accents from perception, speech processing and data mining perspectives. In perception experiments, an accent degree was assigned to each speaker and his/her L1 was identified by French subjects on the basis of speech samples. Speech processing allowed us to produce objective acoustic measurements and to automatically identify foreign accents by using these measurements together with classification techniques. The challenge was to check whether accent-specific features could be highlighted with different approaches, in particular with

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

Concerning the first question (i) *are native listeners able to identify foreign accents in French?* perceptual results showed that naive and native listeners presented with foreign accents which they judged as average succeeded in identifying the speaker's L1s in slightly above 50%, in an experimental setup with 6 foreign accents. Also, subjects were able to provide a list of segmental and prosodic cues which characterise a given accent (or more generally a foreign accent).

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

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

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

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

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

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

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

References

Abdelli-Beruh, N., 2004. The stop voicing contrast in french sentences. Phonetica 61, 201–219.

Adank, P., 2003. Vowel normalisation: a perceptual acoustic study of Dutch vowels. Ph.D. thesis, Radboud University Nijmegen.

Adda-Decker, M., Hallé, P., 2007. Bayesian framework for voicing alternation and assimilation studies on large corpora in French. In: Proc. of the Inter-

national Conference of Phonetic Sciences (ICPhS). Saarbrücken, Germany, pp. 613–616.

Adda-Decker, M., Lamel, L., 1999. Pronunciation variants across system configuration, language and speaking style. Speech Communication 29, 83–98.

Alba, O., 2001. Manual de fonética hispánica. Editorial Plaza Mayor, San Juan.

Angkititrakul, P., Hansen, J., 2003. Use of trajectory models for automatic accent identification. In: Proc. Interspeech. Geneva, Switzerland, pp. 1353–1356.

Arai, T., Greenberg, S., 1997. The temporal properties of spoken Japanese are similar to those of English. In: Proc. of the European Conference on Speech Communication and Technology (Eurospeech). Rhodes, Greece, pp. 1011–1014.

Arslan, L., Hansen, J., 1997. A study of temporal features and frequency the Acoustical Society of America 102 (1), 28–40.

Bartkova, K., Jouvet, D., 2004. Foreign accent processing in automatic speech recognition. In: Proc. SPECOM - International Conference on Speech and Computer. Saint-Petersburg, Russia, pp. 22–28.

Berkling, K., August 2001. Scope, syllable core and periphery evaluation: Automatic syllabification and foreign accent identification. Speech Communication 35 (1-2), 125–138.

995 Boersma, P., 2001. Praat, a system for doing phonetics by computer. Glot
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, Corea, 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, Corea, 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

rhythm: An experimental phonetic study based on acoustic and perceptual evidence. Ph.D. thesis, University of Bonn.

Diez, F. G., Dellwo, V., Gavaldà, N., Rosen, S., 2008. The development of measurable speech rhythm during second language acquisition. Journal of the Acoustical Society of America 123 (5), 3886–3886.

Disner, S., 1980. Evaluation of vowel normalization procedures. Journal of the Acoustical Society of America 67, 253–261.

Durand, J., Laks, B., Lyche, C., 2003. Le projet phonologie du français contemporain. La Tribune Internationale des Langues Vivantes 33, 3–9.

Ferragne, E., Pellegrino, F., 2007. Automatic dialect identification: A study of british english. In: Müller, C., Schötz, S. (Eds.), Speaker Classification II/2, Springer Verlag, Berlin, 243–257.

Flege, J., 1984. The detection of french accent by american listeners. Journal of the Acoustical Society of America 76 (3), 692–707.

Flege, J., Hammond, R., 1982. Mimicry of non-distinctive phonetic differences between language varieties. Studies in Second Language Acquisition 5, 1–17.

Flege, J., Port, R., 1981. Cross-language phonetic interference: Arabic to english. Language and Speech 24 (2), 125–146.

Flege, J., Schirru, C., MacKay, I., 2003. Interaction between the native and 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., Mascaro, 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

class hypothesis. In: Gussenhoven C. / Warner N. (ed.), Papers in Laboratory Phonology VII. The Hague: Mouton de Gruyter, 515–546.

Guyon, I., Elisseeff, A., 2003. An introduction to variable and feature selection. Journal of Machine Learning Research 3, 1265–1287.

Harrington, J., Palethorpe, S., C. Watson, C., 2000. Does the queen speak the queen's english? Nature 408, 927–928.

Huckvale, M., 2004. ACCDIST: a Metric for Comparing Speakers' Accents. In: Proc. of the International Conference on Spoken Language Processing (ICSLP). Jeju, Korea, pp. 29–32.

Huckvale, M., 2007. Hierarchical clustering of speakers into accents with the ACCDIST metric. In: Proc. of the International Conference of Phonetic Sciences (ICPhS). Saarbrücken, Germany, pp. 1821–1824.

Ihaka, R., Gentleman, R., 1996. R: A language for data analysis and graphics. Journal of Computational and Graphical Statistics 5 (3), 299–314.

Jilka, M., 2000. The contribution of intonation to the perception of foreign accent. Ph.D. thesis, University of Stuttgart, Stuttgart, Germany.

Kumpf, K., King, R., 1997. Foreign speaker accent classification using phoneme-dependent accent discrimination models and comparisons with human perception benchmarks. In: Proc. of the European Conference on Speech Communication and Technology (Eurospeech). Rhodes, Greece, pp. 2323–2326.

Lamel, L., Gauvain, J.-L., Adda, G., Barras, C., Bilinski, E., Galibert, O., Pujol, A., Schwenk, H., Zhu, X., 2007. The LIMSI 2006 TC-STAR EPPS Transcription Systems. In: Proc. of the International Conference on Acoustics, Speech, and Signal Processing (ICASSP). Hawaii, USA, pp. 997–1000.

Livescu, K., Glass, J., 2000. Lexical modeling of non-native speech for automatic speech recognition. In: Proc. of the International Conference on Acoustics, Speech, and Signal Processing (ICASSP). Istanbul, Turkey, pp. 1683–1686.

Magen, H., 1998. The perception of foreign-accented speech. Journal of Phonetics 26, 381–400.

Martin, A., Le, A., 2008. NIST 2007 Language Recognition Evaluation. In: Odyssey - The Speaker and Language Recognition Workshop. Stellenbosch, South Africa, paper 016.

Nearey, T. M., 1989. Static, dynamic, and relational properties in vowel perception. Journal of the Acoustical Society of America 85, 2088–2113.

Quilis, A., 1993. Tratado de fonología y fonética españolas. Biblioteca románica hispánica, Madrid.

Ramus, F., 1999. Rythme des langues et acquisition du langage. Ph.D. thesis, EHESS, Paris, France.

Raux, A., 2004. Automated lexical adaptation and speaker clustering based on pronunciation habits for non-native speech recognition. In: Proc. Interspeech. Jeju, Corea, 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.

Witten, I., Frank, E., 2005. Data Mining: Practical Machine Learning Tools and Techniques. Morgan Kaufmann.

Woehrling, C., Boula de Mareüil, P., 2006. Identification d'accents régionaux en français : perception et analyse. Revue Parole 37, 25–65.

Woehrling, C., Boula de Mareüil, P., Adda-Decker, M., 2009. Linguistically-motivated automatic classification of regional French varieties. In: Proc. Interspeech. Brighton, UK, pp. 2183–2186.

Yamada, R., Strange, W., Magnuson, J., Pruitt, J., Clarke, W., 1994. The intelligibility of Japanese speakersŠ productions of American English /r/, /l/ and /w/, as evaluated by native speakers of American English. In: Proc. of the International Conference on Spoken Language Processing (ICSLP). Yokohama, Japan, pp. 2023–2026.