CHARACTERIZING NON-NATIVE FRENCH ACCENTS USING AUTOMATIC ALIGNMENT

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

LIMSI-CNRS, Orsay, France
{bianca;mareuil;madda}@limsi.fr

ABSTRACT

The goal of this study is to investigate the most relevant cues that differentiate 6 foreign accents in French (Arabic, English, German, Italian, Spanish and Portuguese). We took advantage of automatic alignment into phonemes of non-native French recordings. Starting from standard acoustic models, we introduced pronunciation variants which were reminiscent of foreign-accented speech: first allowing alternations between French phonemes (e.g. [s]~[z]), then combining them with foreign acoustic units (e.g. a rolled *r*). Results reveal discriminating accent-specific pronunciations which, to a large extent, confirm both linguistic predictions and human listeners' judgments.

Keywords: foreign accent, non-native French, speech processing.

1. INTRODUCTION

We are often able to successfully recognize the mother tongue of somebody speaking a foreign language. For instance in [1], in the majority of cases, listeners accurately identified the mother tongue (L1) of Arabic, English, German, Italian, Portuguese and Spanish people speaking French (L2). This study addresses the phonetic cues which may best characterize foreign accents.

Among the clues that contribute to an impression of accentedness, a rich literature on Spanish-accented English mentions factors affecting syllable structure, vowel quality, consonants (especially [s]~[z] and [b]~[v]), as well as stress [2, 3]. In the automatic speech recognition (ASR) domain, studies on foreign-accented English aim at generating pronunciation variants which are specific to non-native speech [4, 5, 6]. More linguistics-driven studies on foreign-accented speech exist [7, 8]. These studies quantify ASR accuracy improvement but do not try and explicit how to identify the origin of a foreign accent.

The work described here combines linguistic knowledge with speech processing techniques (in particular automatic alignment) in order to sort out which cues contribute to identifying the origin of a given foreign accent. The conducted work is based upon the recordings of 6 native and 36 non-native speakers of French who were used in previous perceptual experiments. The speakers' degree of accentedness was evaluated by 25 native French listeners who were also asked to identify the origin of the accents and to indicate the cues which helped them make their decisions. The 'r' pronunciation was often put forth, together with vowel quality (e.g. [u] instead of /y/ or vice versa). Some comments were also concerned with prosody.

The corpus described in Section 2 was segmented into phonemes by automatic alignment, in order to measure discriminating phonemic and prosodic features. The technique relies on acoustic models and a pronunciation dictionary which we can manipulate, as in [9]. Previous work based on the LIMSI ASR system (e.g. [9]) showed the validity of the method. In section 3, standard French acoustic models and pronunciation dictionary were used. We measured vowel formants, consonant duration and voicing, as well as prosodic cues related to the pronunciation of the final schwa.

In Section 4, we introduced variants related to foreign accents into the pronunciation dictionary, still using French acoustic models. Introducing variants inspired by listeners' comments, the system outputs the pronunciation which best matches what was said.

In Section 5, we examined the extent to which L2 learners produce some L2 sounds just as if they were the corresponding L1 sound, that is, without any modification, through the mechanism of category assimilation. French phonemes are unable to account for all non-native productions, which may be "far" from the target units or "intermediate" between two phonemes of the L2. To cope with this, we merged French and foreign acoustic units that non-native speakers may borrow from their mother tongue.

2. CORPUS

We analysed 36 speakers of Arabic, English, German, Italian, Portuguese or Spanish mother tongue as well as 6 native French speakers. Among other things, they read a 400 word text in French (about 5 minutes). There were as many males as females. All the speakers were European except native Arabic speakers who came from the Maghreb. The French speakers were students with no marked accent who were born and lived in the Paris region. On average, the non-native speakers (who were all students) were 24 years old, had lived in France (in the Paris region) for 15 months and had started to study French at the age of 17. In a perceptual test [1], their origin was correctly identified in 52%, which was much greater than chance. For each linguistic origin, the most frequent answer was the right one. The speakers' degree of accentedness was comparable across the different foreign origins. It was rated 2.7 on average on a 0–5 scale.

3. STANDARD ALIGNMENT ANALYSIS

3.1 Vowel formants

Acoustic analyses were carried out on the basis of the corpus segmented into phonemes with the help of automatic alignment derived from the LIMSI ASR system for standard French. Formant frequencies were measured on vowels (over 500 vowels per speaker) using a script written for the PRAAT software (www.fon.hum.uva.nl/praat/). Formants as well as fundamental frequency (F_0) were measured every 10 ms using the standard options of PRAAT. Since the method is automatic, filters were considered, adapted to each vowel (males and females distinguished) to discard aberrant formant values with respect to reference values in an average range of ±500 Hz [10]. We also retained the sole vowels that were voiced on more than half their duration before averaging the values. Only 5.5% of vowels were rejected. The formant values were then normalized with the help of Nearey's log-mean procedure [11]. The fronting of /u/ in English is noticeable (see [12]), as is the backing of /y/ in Spanish and Italian speakers. Among /e/s, the closest one to /i/ is that of Arabs. The Arabic /e/~/i/ merger is notorious and it is stigmatized. As for schwa, it is most closed among Portuguese speakers, and fronted among Spanish and Italian speakers. Due to lack of space, only a few F1 and F2 values are reported in Table 1. As in the following, extreme values are emboldened.

Table 1: First and second formant values, normalized using log-mean method (for Arabic, English, Italian, Portuguese, Spanish and French speakers).

F1	Ar	En	Ge	It	Po	Sp	Fr	Mean
/u/	340	**390**	340	340	360	340	350	350
/y/	310	**320**	300	300	300	310	300	310
/i/	320	**330**	310	310	310	310	310	310
/e/	**370**	400	370	400	380	400	370	380
/ə/	390	380	380	390	**350**	380	370	380

F2	Ar	En	Ge	It	Po	Sp	Fr	Mean
/u/	1050	**1210**	1140	1160	1100	1020	1140	1120
/y/	1900	1850	1980	1820	1880	**1760**	1920	1870
/i/	**2280**	2080	2260	2130	2250	2160	2160	2190
/e/	**2120**	1900	2100	1920	2050	1910	2020	2000
/ə/	1510	1570	1620	1670	1580	**1710**	1570	1610

3.2 Consonant duration and voicing

Consonant duration and voicing were measured. Interestingly (see Table 2), speakers of Arabic, English and German have the longest /p/, /t/ and /k/. In these languages, voiceless stops are often aspirated while in French the voice onset time (VOT) is most often short [12, 13]. In the Arabic language, there is no phonological /p/. Our measurements suggest that Arabic speakers exhibit hypercorrection by producing aspirated variants when speaking French. As shown in Table 4, speech rates in terms of overall phoneme duration are comparable across non-native speakers.

The voicing ratio is defined for each phoneme as the number of voiced measures divided by the total number of measures (every 10 ms). In Table 3, the devoicing of voiced stops (/b/, /d/, /g/) in English-accented and German-accented French is noteworthy. A partial devoicing of /v/ and /ʒ/ is also measured for English speakers.

Table 2: Duration of consonants (in ms).

	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
/v/	79	67	69	77	91	**62**	61	72
/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
/ʁ/	84	68	72	**56**	80	82	72	73

In Spanish, there are no phonological voiced fricatives, hence no /b/~/v/ distinction. We observe that Spanish speakers of French very much devoice the /z/ fricative (which they seem to equate to /s/) and produce the shortest non-natives /b/s and /v/s. Also, the 'r's produced by Italian speakers are shorter and more voiced than they are for other speakers. We return to it in Section 4.

Table 3: Voicing of consonants (% defined F_0 values).

	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
/v/	97	**86**	93	88	94	91	94	92
/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
/ʁ/	56	59	57	**68**	58	60	59	60
/s/	22	23	34	40	39	36	20	31
/z/	89	79	85	80	93	**53**	91	81
/ʃ/	30	**22**	35	39	23	35	36	31
/ʒ/	83	**71**	82	84	83	77	78	80

3.3 Final schwa and related prosodic cues

We here focus on polysyllabic word final schwas.

They represent 123 occurrences in our text, possibly maintained or deleted. They were optional in the standard pronunciation dictionary we used.

It is obvious from Table 4 that Italian speakers of French realize more final schwas (with 23%) than all other speakers. This may be explained by the fact that, in Italian, ending a content word with a consonant is extremely rare.

We also measured the lengthening of the vowel preceding a pronounced schwa together with the ΔF_0 (in semitones) between the average pitch of the penultimate vowel and that of the final schwa. We notice that Italians lengthen the supposedly stressed syllable and lower the pitch on the final schwa, whereas native Germans speaking French display a rising F_0 contour on the final syllable (see the only negative ΔF_0 in Table 4). Both patterns are perceptually salient and typical of these accents.

Table 4: Final schwa and related prosodic cues; Dur.$\mathcal{R}$ (respectively ΔF_0) stands for the duration ratio (resp. F_0 difference) between the penultimate vowel and the final schwa. Ph.dur represents the overall phoneme duration.

	Ar	En	Ge	It	Po	Sp	Fr	Mean
%schwa	10	15	14	**23**	15	11	11	14
Dur. $\mathcal{R}$	2.0	1.6	1.9	**2.4**	2.2	2.1	1.9	2.0
ΔF_0(ST)	1.3	0.9	**-0.1**	2.2	0.7	1.0	1.1	1.0
Ph.dur (ms)	91	90	89	91	94	89	73	88

4. PAIRED VARIANTS & NASAL VOWELS

For the alignment described in the previous section, we used a pronunciation dictionary in which each entry was assigned one or several standard French pronunciation(s). We have seen that non-native speakers frequently produce variants that deviate markedly from the canonical form(s), and our acoustic measures suggest that such deviations may be common to speakers of a given mother tongue. On this basis and as mentioned above, we defined a set of 20 non-native French variants. Building such a kind of mapping table enables a binary, categorical approach to foreign-accented speech completing formant, F_0 and duration-based analyses. In some cases, it yields a more thorough analysis: for example, in addition to the paired variants listed in Table 5, we allowed French nasal vowels (/ɑ̃/, /ɛ̃/ and /ɔ̃/) to be denasalized and followed by nasal appendices ([n] or [m] before *p*/*b*). In any pair (e.g. /b/~/v/) the first phoneme could be identified as the second one and inversely. For each type of variant a specific pronunciation dictionary was generated and a distinct alignment was produced accordingly.

Most of the results reported in Table 5 are unsurprising in view of section 3 and prior knowledge. Recall that the Spanish language does not have /v/ as a distinct phoneme [12]: a [b] is realized after a pause or a nasal consonant; a [β] appears elsewhere. As a consequence, we expect Spanish speakers of French to prefer the French fricative /v/ instead of the French plosive /b/ in many contexts. Indeed, we measure that Spaniards tend to pronounce [v] for /b/, as well as [s] for /z/ (in 79% of cases), [j] for /ʒ/ and [tʃ] for /ʃ/. The English (and more generally Germanic) speakers' tendency to devoice voiced stops is well reflected. The Italian speakers' /y/ aligned as [u] and /ʁ/ aligned as a liquid are also consistent with section 2. After nasal vowels, native speakers of Romance languages speaking French show almost ten times as many nasal appendices as native French speakers, which is well audible for Spanish and Italian speakers. The other resultats are less conclusive.

Table 5: % variants aligned using French acoustic units. The bottom line shows the percentage of nasal appendices ([n]/[m]) aligned after (possible denasalized) nasal vowels.

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

5. ADDING FOREIGN ACOUSTIC UNITS

The previous variants were designed to evaluate confusions between French phonemes made by non-native speakers. In this section, we examine phonemes or allophones as [r], [ɹ], [ɫ], [ʊ], [s̺], [ʝ] and [β], which may be particular to speakers of certain origins [12].

As previous sections suggest, Italian speakers of French produce more rolled 'r's than other speakers. We therefore introduced an [r] xenophone into the French acoustic model set. It was borrowed from Spanish for which we had extensively trained models at our disposal. Our measures suggest that Italians prefer [r] to the

French variant in more than 60% of cases (see Table 6). This variant was aligned in less than 10% of cases for Spanish speakers, which is relevant and rules out a possible bias since Spaniards approximate the French /ʁ/ by a [ɣ] or [x]-like sound. Native English speakers paradoxically realize more [r]s, which we checked perceptually: some of them do pronounce rolled 'r's. If in the same way we introduce the English approximant [ɹ] option for the French /ʁ/, this variant is most aligned for Italian and English native speakers. We would have expected a higher rate of [ɹ] for English speakers. Nevertheless, this result shows how the pronunciations of Italian and English speakers are remote from the French /ʁ/.

The /l/ has a dark allophone in English and Portuguese, contrary to French [12]. Table 6 witnesses that the variant stemming from English models is more often aligned for English (and Portuguese) speakers of French than for other speakers — as was the /l/→[w] variant in Table 5.

In Section 3, 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 centralized [ʊ] is aligned in over 50% of cases for English speakers.

The final remarks concern Spanish acoustic units. The Spaniards' tendency to realize an apical [s̺] shows up in the high percentage of this variant (56%) when French is spoken. Also, 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 French phonemes. The previous alignments with only French acoustic units could not easily account for this phenomenon.

Table 6: %variants aligned using foreign acoustic units.

	Ar	En	Ge	It	Po	Sp	Fr	Mean
/ʁ/→[r]	7	33	14	**62**	12	9	8	21
/ʁ/→[ɹ]	3	21	6	**22**	4	4	2	9
/l/→[ɫ]	2	**10**	3	7	8	7	3	6
/u→[ʊ]	16	**56**	12	15	26	38	12	25
/s/→[s̺]	29	31	30	43	36	**56**	10	34
/ʒ/→[ʝ]	41	35	34	40	45	**55**	23	39
/b/→[β]	5	26	16	9	23	**43**	9	19
/v/→[β]	8	19	26	36	5	**43**	5	20

6. DISCUSSION

We measured cues which may contribute to identify the origin of foreign accents in French. The aim of this study consisted in shedding light on non-native pronunciations, motivated by linguistic mechanisms that can be implemented in a speech recognition system. We automatically aligned data from Arabic, English, German, Italian, Portuguese and Spanish speakers of French using a standard French system in order to measure vowel formants, consonant duration and voicing. We then introduced variants into the pronunciation dictionary and foreign phonemes into the acoustic models to better capture accent-specific features that are influenced by the speakers' mother tongue. Relevant cues seem to be the Arabic /e/~/i/ merger, the English and German aspirated or devoiced stops, the Italian rolled [r] and prosody, the Portuguese schwa pronunciation, the Spanish [s̺] (possibly substituting the French /z/), and /b/~/v/ confusions. These variants can be reintroduced in the alignment process to make new acoustic measurements. Currently, we are working on using data mining techniques to automatically rank the most discriminating cues to recognize the origin of a foreign accent in French. A study on VOT and a more general L1/L2 comparison are also planned. We hope this work could be useful for learning and teaching French as a foreign language.

7. REFERENCES

[1] Vieru-Dimulescu, B., Boula de Mareüil, P. 2006. Perceptual identification and phonetic analysis of 6 foreign accents in French. *Proc. Interspeech*. Pittsburgh. 411-414.

[2] Magen, H.S. 1998. The perception of foreign-accented speech. *Journal of Phonetics*. 26, 381-400.

[3] Flege, J.E., Hammond, R. 1982. Mimicry of Non-distinctive phonetic differences between language varieties. *Studies in Second Language Acquisition*. 5, 1-17.

[4] Bouselmi, G., Fohr, D., Illina, I., Haton, J.-P. 2006. Multilingual Non-Native Speech Recognition using Phonetic Confusion-Based Acoustic Model Modification and Graphemic Constraints. *Proc. Interspeech*. Pittsburgh. 109-112.

[5] Livescu, K., Glass, J. 2000. Lexical modeling of non-native speech for automatic speech recognition. *Proc. ICASSP*. Istanbul. 1683-1686.

[6] Silke, G., Stefan, R., Ralf, K. 2004. Generating non-native pronunciation variants for lexicon adaptation. *Speech Communication*. 42(1), 109-123.

[7] Bartkova, K., Jouvet, D. 2004. Foreign accent processing in automatic speech recognition. *Proc. SPECOM*. Saint-Petersburg. 22-28.

[8] Schaden, S. 2003. CrossTowns: Automatically generated phonetic lexicons of cross-lingual pronunciation variants of European city names. *Proc. LREC*. Lisbon. 1395-1398.

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

[10] Gendrot, C., Adda-Decker, M. 2005. Impact of duration on F1/F2 formant values of oral vowels: an automatic analysis of large broadcast news corpora in French and German. *Proc. Interspeech*. Lisbon. 2453-2456.

[11] Disner, S.F. 1980. Evaluation of Vowel Normalization Procedures. *JASA*. 67, 253-261.

[12] Delattre, P. 1965. *Comparing the phonetic features of English, French, German and Spanish*. Heidelberg: Julius Groos Verlag.

[13] Abdelli-Beruh, N.B. 2004. The Stop Voicing Contrast in French Sentences. *Phonetica*. 61, 201-219.