Perceptual identification and phonetic analysis of 6 foreign accents in French

Bianca Vieru-Dimulescu & Philippe Boula de Mareüil

LIMSI-CNRS
BP 133, 91403 Orsay, France
{bianca;mareuil}@limsi.fr

ABSTRACT

A perceptual experiment was designed to determine to what extent naïve French listeners are able to identify foreign accents in French: Arabic, English, German, Italian, Portuguese and Spanish. They succeed in recognizing the speaker's mother tongue in more than 50% of cases (while rating their degree of accentedness as average). They perform best with Arabic speakers and worst with Portuguese speakers. The Spanish/Italian and English/German accents are the most mistaken ones. Phonetic analyses were conducted; clustering and scaling techniques were applied to the results, and were related to the listeners' reactions that were recorded during the test. They support the idea that differences in the vowel realization (especially concerning the phoneme /y/) seem to outweigh rhythmic cues.
Index Terms: perception, foreign accent, non-natif French.

1. INTRODUCTION

Can we recognize the origin of a speaker from a speech sample? This is possible in some cases: recent exper-iments confirmed it on regional accents in English [1] and French [2]. What is less known is the extent to which machine and naïve listeners are able to identify foreign accents, especially in French. The aim of this article is to fill this lack and to examine the contribution of different linguistic levels (vowel quality and rhythm) to the human perception of a foreign accent. Preliminary work suggests that vowel quality allow the identification of the origin of a foreign accent in French better than prosody does [3] [4]. Also, in the field of foreign-accented speech most studies concentrate on segmentals (the phoneme string, especially vowels [5]. Nevertheless other studies like [6] showed that for learners of French as a foreign language, the mother tongue prosody remains underlying in speech, in the absence of a specific training. This aspect is to be developed.

So as to conduct perceptual tests, our interest focused on accents with which French people would be most familiar: Arabic, English, German, Italian, Portuguese and Spanish. The choice of these languages was made by pooling statistics on immigration and tourism in France. We recorded spontaneous and read speech of about forty speakers of these different mother tongues.

The influence of the phonological system of the mother tongue (L1) on the perception and the production of a second language (L2) is addressed in a number of (psycho)linguistic studies on non-native speech: the fact that, for instance German possesses a /y/ in its phonemic inventory whereas in English (even though a close sound can be heard in a word like *due*) it has no functional role. If other dimensions such as rhythm are phonological categories, they could also betray a mother tongue that is different from the spoken language.

A series of questions therefore arises. Do rhythmic classes (stress-timed *vs* syllable-timed) established for the languages themselves remain valid for non-native speech? Will non-natives borrow the L2 rhythm or will they keep traces of the L1 rhythm? Will native Portuguese speakers (whose L1 is traditionally classified as stress-timed) adopt a rhythm in French similar to the rhythm of their cousins of Romance syllable-timed language, Italians and Spaniards? What will be the behavior of Maghrebian speakers whose dialect may be considered as stress-timed and whose standard language is syllable-timed [7].

Parameters have been proposed to validate or contradict the existence of rhythmic classes. Ramus [8] considers the proportion of vocalic intervals (%V) and the duration variation of consonantal intervals in terms of standard deviation (ΔC). Grabe [9] proposes a slightly more complex approach: she measures the variability between successive vocalic and intervocalic intervals via "Pairwise Variability Indices" (PVI), possibly normalized to account for speech rate variation. The %V/ΔC and PVI 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. Ramus and Grabe's studies were carried out on a very small corpus, which was segmented by hand. However the displayed duration differences are about 5 ms and the results happen to be corpus-dependent. Our corpus is larger and it will permit a comparison between measurements obtained for the same speakers on native and non-native speech.

The next section presents the protocol and the results of a perceptual identification test on the 6 accents under investigation. In Section 3, before concluding, we examine some phonetic features which subjects noted concerning vowel quality and rhythm. The acoustic analysis relies on the reading; elements of comparison with spontaneous speech are also provided.

2. PERCEPTUAL EXPERIMENT

2.1. Speakers and corpus

Among the speakers of Arabic, English, German, Italian, Portuguese or Spanish origin whom we recorded, 36 were used for the experiment itself (6 in each language). Six other ones (1 in each language) were used for a familiarization phase. In the test, we had as many females and males. All the speakers were European or came from various countries of the Maghreb as far as the Arabs are concerned — but a previous study showed the difficulty in discriminating the possible Algerian, Moroccan or Tunisian origins of speakers speaking French [4]. The Hispanic speakers were neither Catalan nor Latin-American. On average, the speakers (all students) was 24 years old, they had lived in France (in the Paris region) for about 15 months and had started to study French at the age of 17.

Among other things, the speakers read a 400-word text — that of the PFC project ("Phonology of Contemporary French") [10] — and they spoke freely for about 5 minutes in a face-to-face situation with the experimenter. A comparison was thus possible with data from native speakers, through the PFC project whose protocol was used and extended: the speakers were also recorded in their mother tongue.

For the perceptual test, a spontaneous speech excerpt of about 10 seconds was retained for all our speakers, according to the following criteria: absence of cultural references or morpho-syntactic errors that could be typical of a given origin, few hesitations and coherence of the statement. The selected stimuli were used in a perceptual experiment described below.

2.2. Protocol and subjects' task

The test took place in a soundproof booth. The subjects listened to the stimuli through loudspeakers with a comfortable sound level — equalized beforehand. The stimuli, in Wave format, were sampled at 22.05 kHz, 16 bits, mono.

The subjects were informed that they would be requested to make judgments on non-native French speech samples, through a user-friendly interface [11]. As a familiarization phase, they listened to spontaneous speech samples which illustrated Arabic, English, German, Italian, Portuguese and Spanish accents in French. These speakers whose origins were indicated were of course not used further in the test. In the next phase, the test strictly speaking, the subjects listened to the 36 stimuli presented in a random order which changed for each listener. As in the familiarization phase, each stimulus could be listened to as many times as necessary, stopped or replayed from a certain point. It was nonetheless impossible to go back to previous stimuli. The subjects had to evaluate the degree of accentedness of the stimulus they listened to on a continuous 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 subjects who were provided with a microphone were invited to react verbally towards the stimulus they listened to (by imitating it or caricaturing it) or to type their comments in a text window. The data were recorded stimulus by stimulus, and the instructions merely suggested to specify which non-native features in the speaker's pronunciation and intonation sounded most important. The listeners had to determine the mother tongue of the speaker who produced the stimulus before processing another excerpt. They had to make a forced choice between Arabic, English, German, Italian, Portuguese and Spanish. The experiment lasted about half an hour.

2.3. Listeners

The perceptual test was administered to 25 untrained native French listeners from the Paris region, with normal hearing. They were all members of a computer science laboratory (LIMSI) and they were not paid for this task.

The majority of subjects self-reported that they were capable of recognizing Arabic (20), English (20) and German (18) accents but fewer subjects thought they could recognize Italian (3), Portuguese (6) and Spanish (12) accents in French. These figures do not match with the listeners' proficiency in the corresponding language: as an example, 23 out of the 25 subjects self-reported no or little knowledge in Arabic, whereas 20 subjects self-reported a good knowledge of English.

3. RESULTS

From the listeners' judgments it turns out that the degree of accentedness is average (2.66 out of 5): 2.37 for Arabic, 2.18 for German, 3.00 for English, 3.07 for Italian, 2.39 for Portuguese and 2.94 for Spanish speakers. The results of the identification test show that our subjects succeed in distinguishing the presented foreign accents quite well. The global rate of correct identification (52.2%) is much greater than chance (16.7%). Chi-square tests reveal that for each language we are above chance level [$df = 3$; $p < 0.01$]. For each linguistic origin the most frequent answer is the good one — in bold in Table 1. The same happens for 25 out of the 36 speakers presented in the test, who are well recognized.

Table 1: Confusion matrix for 36 stimuli and 25 speakers (% with respect to 6×25 = 150 answers).

orig. \ answer	Arabic	English	German	Italian	Portuguese	Spanish
Arabic	**77.3**	1.3	6.0	5.3	8.0	2.0
English	8.7	**48.7**	28.0	2.7	3.3	8.7
German	6.0	14.7	**63.3**	4.7	8.0	3.3
Italian	7.3	3.3	5.3	**40.0**	10.0	34.0
Portuguese	16.7	8.0	17.3	12.0	**25.3**	20.7
Spanish	5.3	2.7	2.7	19.3	11.3	**58.7**

The speakers whose origin was identified best were the Arabs, main immigrant community living in France. The least recognized were the Portuguese speakers, whose phonemes are "close" to the French ones. The latter fact can explain their unmarked accent. Moreover, the hissing stereotype that is often associated to the Portuguese accent is far from reality. Interestingly, the degree of accentedness of Portuguese speakers is greater than that of Arabic speakers, even if this difference is not statistically significant according to a *t*-test.

Between these extreme cases, the most frequent confusions were Italian/Spanish (which corroborates previous studies [11]) and English/German. Multidimensional scaling and clustering techniques allow us to visualize this fact, yielding a kind of perceptual distance between the different accents. They gather Italians and Spanish speakers in one group, English and German speakers in another one, and Portuguese and Arabic speakers in two separate groups.

Our listeners claimed they based their judgements on segmental and suprasegmental features whose relevance can be checked by acoustic measurements. Among segmental features, the listeners noted: the "rolled" *r* which reminded them of a Southern country; [i] instead of /e/ in the case of Arabic speakers; *yé* instead of *je*, [v] instead of /b/ and [s] instead of /z/ for Spaniards; [z] instead of /s/ for Germans; [u] instead of /y/ or vice versa, and a bad realization of nasals revealing a foreign accent in general rather than a particular origin. Among suprasegmental features noted by the subjects we can cite: "singing" sentences which would be reminiscent of an Italian speaker of French or a "rush" on certain words. Hence our acoustic analyses, plotting the vocalic triangles of the speakers and a rhythm characterization. Studies on consonants and intonation are in progress.

4. PHONETIC ANALYSIS

4.1. Vowel spaces

The measurements presented in this section were made on the PFC text, read by the 36 speakers used in the perceptual test. This common material, which is less restricted than our spontaneous speech samples, facilitates the analysis and better lends itself to comparisons between speakers.

With the help of automatic alignment derived from the LIMSI speech recognition system [12], the PFC text was segmented into phonemes by using context-independent acoustic models for standard French. In the same way we segmented the PFC texts read by 6 speakers of standard French (3 males and 3 females), among the youngest speakers studied by [2]. The degree of accentedness of these speakers was estimated at 0.6 on a 0–5 scale by 25 French listeners following a similar protocol to ours. On the basis of this segmented corpus, acoustic studies were carried out. A script was written for PRAAT software (http://www.fon.hum.uva.nl/praat/) in order to measure formant frequencies on vowels — the PFC text contains over 500 vowels per speaker. Since the method is automatic, filters were foreseen, adapted to each vowel (males and females distinguished) to discard aberrant values with respect to reference values in an average range of ±500 Hz [13]. Only 4% of phonemes were rejected. The formant values were then normalized with the help of various procedures described by [14]. Using Nearey's normalization, the vocalic triangles corresponding to the different accents (or linguistic origins) are displayed in Figure 2, where formant values are averaged over the beginning, the mid-point and the end of vowels. We notice that the triangles of English and German speakers are more reduced than the other non-natif triangles. This can be related to vowel reduction observed in their mother tongues. The fronting of /u/ in English is also remarkable, as is the fact that among /e/s, the closest one to /i/ is that of Arabs.

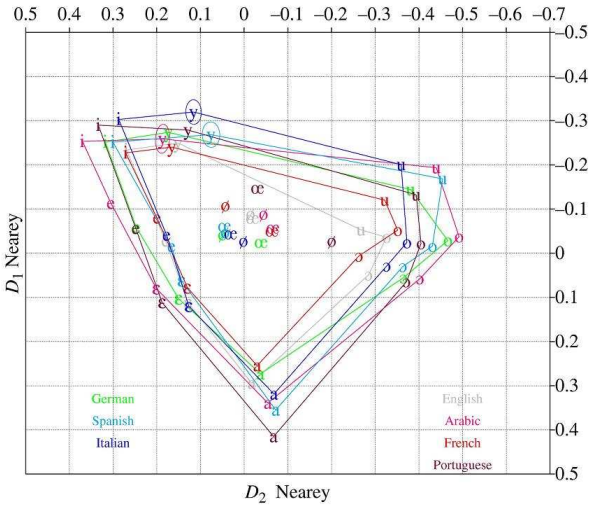

Figure 2 : vocalic triangles (after filtering out probable measurement errors and applying Nearey's normalization) corresponding to native or foreign-accented French, for the PFC text. From left to right the /y/s of Arabs, Italians and Spaniards speaking French are surrounded by ellipses.

The realization of French /y/ is very different between Spanish and Italian speakers on the one hand (where the /y/ is closer to [u]) and Arabic speakers on the other hand (from whom the /y/ is closer to [i]). The former speakers favor the [+round] feature, the latter speakers favor the [+front] feature. This phenomenon is rather well known and often caricatured, even though an explanation is not available. It is well evidenced by scaling or clustering techniques from a characterization of each origin by the average coordinates of the /y/ in the F1/F2 space. A chart in the form of a dendrogram is provided in Figure 3.

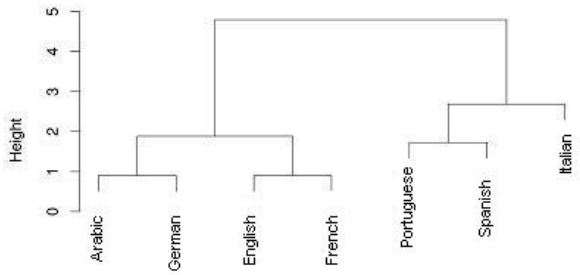

Figure 3 : Dendogram obtained from a divisive clustering algorithm using a Euclidean distance, from the characterization in the acoustic space of the /y/ pronunciation in standard French and in each accent.

We obtain very similar results introducing the third formant (F3), using another algorithm (agglomerative rather than divisive) and whatever the distance used (Euclidean or Manhattan). We find the same tendency in the spontaneous sentences presented to the listeners, which we transcribed, aligned and analyzed in the same way as the read text.

By representing averages by speaker rather than by language, with the same divisive algorithm using the Euclidean distance, we obtain a dendogram where Arabic speakers on the one hand and Italian and Spanish speakers on the other hand are fairy well grouped: 5 Arabic speakers out of 6 on the left side, 5 Italian and 4 Spanish speakers out of 12 on the right side of the tree. This /y/ could be an accurate discriminating clue.

4.2. Rhythmic parameters

Subjects in our experiment spoke about the /y/ but also (without resorting to the technical jargon of a specialist) about what can be perceived or understood as rhythmic aspects. Consequently, we carried on our data-driven approach by measuring Ramus and Grabe's parameters (see Introduction).

Since the complexity of consonant clusters is here imposed to all speakers by the French language, their duration variability cannot be interpreted easily. We could imagine that speakers who are not used to complex consonant clusters would pronounce the latter by adding up the duration of consecutive consonants. This could lead to a high ΔC in opposition to observations made on their L1. Actually in our measurements on non-native French, ΔC seems to correlate other parameters. Conversely, speakers who are used to complex consonant clusters could more easily shrink them, thereby making %C decrease and %V increase. This would be contrary to the tendency of stress-timed languages to reduce vowels. Therefore we only retained ΔV and PVI on vowels in Figure 4,

which shows that these parameters are closely related. PVIs are not normalized because speech rates are comparable across speakers (from 10.7 to 11.3 phonemes/second for non-natives). The results obtained with this type of approach show that there is a clear rhythm difference only between the Arabs (the closest speakers to the French rhythm) and Italians. Thus Maghrebian speakers do not reproduce the stress-timed rhythm of their dialect — whereas in French Portuguese speakers remain close to the stress-timed rhythm of English speakers. As for Italians, their high vowel duration variation is unsurprising since the ratio of stressed/unstressed vowels is particularly high in Italian. However Grabe does not give us comparative data since this language is unfortunately missing in her corpus of 18 languages.

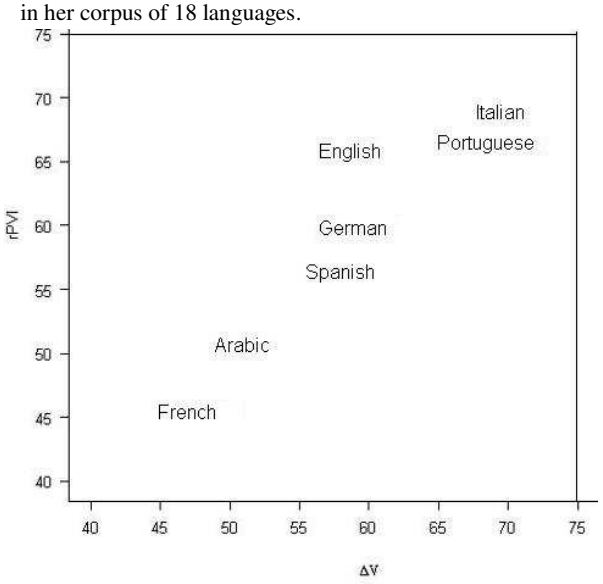

Figure 4 : Characterization of rhythm by combining Ramus and Grabe's parameters on vowels.

5. CONCLUSIONS

This paper demonstrated that it was possible to carry out a study based on perception, data analysis techniques and automatic speech processing on Arabic, English, German, Italian, Portuguese and Spanish accents in French. We recorded about forty talkers speaking French with an accent judged as average. Their origin was well identified by French subjects on the basis of speech samples — with 52%, *i.e.* 10% better than 6 regional accents of French studied in a similar task [2]. The listeners suitably pointed out acoustic features which sounded salient to them. It seems that the information carried by segments (at least vowel quality and in particular the /y/ pronunciation) enables the identification of the origin of an accent better than prosody (at least rhythm). The rhythmic parameters we measured did not exhibit clear-cut tendencies.

To refine the hierarchy, we are currently measuring the VOT (Voice Onset Time) of plosives and we intend to link rhythmic parameters with the duration ratio of stressed/unstressed syllables. The results should also be compared with the productions of our speakers in their L1.

Several prospects for automatic speech processing can be considered: addition of variants related to the different foreign accents (e.g. /y/ → [u] with French acoustic models); alignments with acoustic models from the languages of origin (e.g. English [ɹ]), targeting automatic accent identification and an improvement of non-native speech recognition scores. Benefit can also be derived from speech synthesis: better than human imitators who are prone to reinforce certain features, it is a good simulation tool. Finally, we hope this study could be useful for learning and teaching French as a foreign language.

6. ACKNOWLEDGEMENTS

We are indebted to Martine Adda-Decker, Cécile Woehrling and Cédric Gendrot for making available, launching, or adapting scripts in particular exploiting automatic alignment.

7. REFERENCES

[1] Clopper, C.G. & Pisoni, D.B., "Some acoustic cues for the perceptual categorization of American English regional dialects", *Journal of Phonetics*, **32**, pp. 111–140, 2004.

[2] Woehrling, C. & Boula de Mareüil, P., "Identification d'accents régionaux en français : perception et catégorisation", *Bulletin PFC*, **6**, pp. 89–103, 2006.

[3] Lauret, B., *Aspects de phonétique expérimentale contrastive : « l'accent» anglo-américain en français*, PhD thesis, Université Paris III, 1998.

[4] Boula de Mareüil, P., Brahimi,B. & Gendrot, C., "Role of segmental and suprasegmental cues in the perception of Magrebian-accented French", *Interspeech*, Jeju, 2004.

[5] Flege, J.E., Schirru, C., and MacKay, I.R.A., "Interaction between the native and second language phonetic subsystems", *Speech Communication*, **40**, pp. 467–491, 2003.

[6] Freland-Ricard, M., "Organisation temporelle et rythmique chez les apprenants étrangers. Étude multilingue", *Revue de phonétique appliquée*, **118–119**, pp. 61–91, 1996.

[7] Ghazali, S., Hamdi, R. & Barkat, M., "Speech Rhythm Variation in Arabic Dialects", *Speech Prosody*, Aix-en-Provence, pp. 331–334, 2002.

[8] Ramus, F., *Rythme des langues et acquisition du langage*, PhD thesis, EHESS, Paris 1999.

[9] Grabe, E. & Low, F.L., "Durational Variability in Speech and the Rhythm Class Hypothesis", in Gussenhoven, C. & Warner, N. (Eds), *Laboratory phonology VII*, Mouton de Gruyter, Berlin, pp. 515–546, 2002.

[10] Delais-Roussarie, É. & Durand, J., *Corpus et variation en phonologie du français. Méthodes et analyses*, Presses Universitaires du Mirail, Toulouse, 2003.

[11] Vieru-Dimulescu B. & Boula de Mareuil, P., "Contribution of prosody to the perception of a foreign accent", *ISCA Workshop on Plasticity in Speech Perception*, London, pp. 66–69, 2005.

[12] Adda-Decker, M., Boula de Mareüil, P., Adda, G. & Lamel, L., "Investigating syllabic structures and their variation in spontaneous French", *Speech Communication*, **46**, pp. 119–139, 2005.

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

[13] Adank, P.M., *Vowel normalization: a perceptual-acoustic study of Dutch vowels*, PhD thesis, Radboud University Nijmegen, Nijmegen, 2003.