

Identification of regional accents in French: perception and categorization

Cécile Woehrling & Philippe Boula de Mareuil

LIMSI-CNRS

BP 133 — F-91403 Orsay CEDEX

Cecile.Woehrling;Philippe.Boula.de.Mareuil@limsi.fr

ABSTRACT

This article is dedicated to the perceptual identification of French varieties by listeners from the surroundings of Paris and Marseilles. It is based on the geographical localization of about forty speakers from 6 Francophone regions: Normandy, Vendee, Romand Switzerland, Languedoc and the Basque Country. Contrary to the speech type (read or spontaneous speech) and the listeners' region of origin, the speakers' degree of accentedness has a major effect and interacts with the speakers' age. The origin of the oldest speakers (who have the strongest accent according to Paris listeners' judgments) is better recognized than the origin of the youngest speakers. However confusions are frequent among the Southern varieties. On the whole, three accents can be distinguished (by clustering and multidimensional scaling techniques): Northern French, Southern French and Romand Swiss.

Index Terms: perceptual dialectology, French regional accents

1. INTRODUCTION

In speech communication, considerable variability comes into play as a function of regional and social accents. Humans must cope with it, like machines for automatic processing. In order to account for various accents that may exist in French, a perceptual dialectology approach is developed, from speech corpora which computational techniques help analyzing.

We now have at our disposal a large number of recordings collected in different places of the French-speaking area, within the framework of the PFC project ("Phonology of Contemporary French") [1]. In the audio data, various accents are represented, as many deviations with respect to a norm. They may be detected by certain phonetic features which are salient enough to be recognized and characterized. Are naïve listeners capable of identifying these accents? How fine-grained is the distinction between, for example South-West and South-East accents? How many accents are experts or non-specialists able to discern? What is the impact of the requested subjects' geographical background? These questions are not novel in dialectometry [2], even if we here distinguish between accents and traditional dialects.

Clopper and Pisoni [3] recently demonstrated that, without prior training or feedback, American listeners, invited to listen to compatriots of various accents and to locate their geographical origin on a map of the USA, are able to distinguish three broad regions: New England, South, North/West. Studies were also devoted to British English, Dutch and Norwegian varieties [4, 5]. To our knowledge, results of a similar perceptual clustering task are not available

for the French language, even though a North/South distinction sounds obvious to any French speaker. To quantify to what extent listeners are able to determine the origin of a speaker, the key difficulty is that some parameters may be distinctive to certain listeners but not to others. A number of factors may affect the judgments and the awareness of identifiable differences. Based solely on pronunciation, locals are expected to perform with a finer-grained perception than non-locals. In sociolinguistics, studies focus on the representations of specific varieties, stored in long-term memory, more or less stereotyped, and possibly different from behavioral reactions to actual speech samples [6, 7]. However most studies are descriptive, and their findings do not enable us to reliably predict the most salient properties that are associated to a given accent.

The PFC project now allows systematic investigations. One of its objectives is to cover a vast territory throughout informants who are firmly rooted geographically — who were born and have spent most of their lives in the same place. Since several age categories and speech "styles" (reading and interview especially) are represented in the corpus, their influence on the listeners' performance also deserves being quantified. We examined this issue in two steps, with a pre-test whose goal was a mere accent rating and two experiments strictly speaking of accent origin identification, on the basis of about forty speakers from six Francophone regions. It is indeed interesting to carry out perceptual tests in different locations, to compare the results: the experiments we conducted involved listeners from the Paris region (pre-test and Experiment 1) and the Marseilles region (Experiment 2). The next section presents the corpus, the listeners' tasks and a phonetic analysis of the speech material. Results are reported in section 3 and further analyzed in Section 4, before concluding remarks.

2. EXPERIMENTS

2.1. Speakers and stimuli

As in [3] our experiment is based on six regions. In Southern France we retained Biarritz (Basque Country) in the South-West, Douzens (Languedoc) in the South, Marseilles (Provence) in the South-East. In the North, two points are located at the same latitude: Treize-vents (Vendee) in the West and the Canton of Vaud (Romand Switzerland) in the East. For the sake of symmetry wit was wished to choose a point in the Paris region. Nevertheless, since Paris represents the norm, asking subjects about "the Paris accent" would have been confusing. Facing this difficulty, we opted for a point situated on the same axis as



the Basque Country and Vendee: Brécey (Normandy). Another dialectal area is thus represented.

In each of the six regions (Normandy, Vendee, Romand Switzerland, Languedoc, Provence and the Basque Country), six speakers (3 males, 3 females) were selected, split into three age categories: 15–30 years-old, 30–60 years-old, over 60 years-old. For each speaker, two speech samples were chosen. The first one is a long read sentence (25 words), from the middle of the PFC text, which is identical for all speakers: “La côte escarpée du mont Saint-Pierre qui mène au village connaît des barrages chaque fois que les opposants de tous les bords manifestent leur colère”. The second one is an excerpt of spontaneous speech, from a guided interview. It was chosen according to the following criteria: assertive utterance whose length is equivalent to that of the read excerpt (9.4 seconds on average, compared to 8.3 for the read sentence), absence of reference to a location which would bias the identification, absence of intervention from the interlocutor and few hesitations from the speaker. With 33 words per excerpt on an average, the speech rate of spontaneous speech (10.6 phonemes/second) is comparable to that of reading (10.4 phonemes/second). In total, we have 3 age categories × 2 genders × 6 regions × 2 speech styles = 72 stimuli.

Previous studies often rely on the sole “read speech”, despite the problems it raises [3]. In our opinion, the use of both read and spontaneous speech offers several advantages. First, the display of an identical read sentence allows comparisons all other things being equal. Reading discards lexical and syntactic clues and makes sure that the differences between speakers are due to pronunciation. Spontaneous speech represents a register of speech which better reflects the true vernacular of the speakers — their natural way of speaking in everyday use. By using both types of speech, we are in a position to verify whether the one or the other favors the accent recognition, or if the two are equivalent — thereby allowing new generalizations. Finally, the display of two types of stimuli in a random order allows the listeners not to be tired: the subjects do not have to always listen to the same sentence, which contributes to keep their attention.

2.2. Listeners

The pre-test and the two experiments were administered to twenty-five listeners each. For the pre-test and Experiment 1, the listeners were residents of the Paris region. The group of listeners who participated in the actual experiment had not taken part in the pre-test. For Experiment 2 listeners were residents of the Marseilles region. They all had French as their mother tongue and no hearing impairment. On average, they were 32 years old, and had spent 20 years in their region of residence. According to self-reports, almost all of them were familiar with the Marseilles (Provence) and the Vaud (Swiss) accents, and unfamiliar with the other accents.

2.3. Protocol

The tests took place in a sound-isolated booth; the subjects were wearing closed headphones of the same model. The sound level, which had been equalized, was comfortable. The stimuli, in Wave format, were sampled at 22.05 kHz, 16 bits, mono. A user-friendly interface was used, among other things allowing the participants to input information about their familiarity with this or that accent, by clicking on buttons, and

to capture their responses. The listeners started by a brief familiarization with the same sentence read by a male or a female speaker from each of the six regions under consideration. These sentences were no longer used further in the test. During the following phase, the participants listened to 74 stimuli, the first two of which were not counted in the results (spontaneous utterances from a Northern male speaker and a Southern female speaker). The next 72 stimuli, read or spontaneous excerpts, were presented in a random order which was different for each listener.

- **Pre-test.** During the familiarization phase, a degree of accentedness was given as indicative for each displayed stimulus. During the test phase, the listeners had to rate a degree of accentedness to the excerpt they had just listened to. The proposed degrees, on a six-point scale, were paraphrased as follows: 0 (no accent), 1 (mild accent), 2 (moderate accent), 3 (rather strong accent), 4 (strong accent), 5 (very strong accent).

- **Experiments 1 and 2.** The speaker’s region of origin was indicated for each displayed stimulus during the familiarization phase. During the test, after each stimulus, the listeners had to specify the speaker’s origin among the six possibilities: Brécey (Normandy), Treizevents (Vendee), Canton of Vaud (Romand Switzerland), Biarritz (Basque Country), Douzens (Languedoc) and Marseilles (Provence). No feedback on the answer correctness was given.

The listeners were not urged to answer. Each stimulus could be listened to as often as needed, but it was impossible to come back to them after answering. Each of the three experiments lasted about twenty minutes.

2.4. Acoustic analysis of the corpus

The stimuli presented to the listeners as well as the whole PFC text were analyzed acoustically in terms of schwa and nasal appendix realization, and formant values. This analysis was facilitated by the segmentation into phonemes provided by automatic speech recognition [8]. The procedure is based on the automatic alignment of context-independent continuous-density HMMs. As in [9], formant frequencies were measured with the help of PRAAT software (<http://www.fon.hum.uva.nl/praat/>) and averaged over the beginning, the mid-point and the end of vowels.

We have been able to quantify the fact that Southern speakers of French produce twice as many schwas as Northern speakers, and nasal appendices after half of nasal vowels (especially /ã/ and /ɛ̃/) whereas Northern speakers do not. The tendency (in some cases) to pronounce [ɔ] in the South instead of /o/ was also observed: in a word such as *côte* (“slope”) on average, the F1 values we measured were 450 Hz in the South vs. 550 Hz in the North. Other differences concerning vowel quality are not so notorious. In particular, the fronting of the open /ɔ/ which is noticeable in Northern (actually standard) French does not apply to Southern French. The vocalic triangles of Figure 1 enable us to display this tendency, which is conspicuous for both the spontaneous stimuli and the whole text reading. The /R/ is also subject to variation: it is systematically apical in the oldest Languedoc speakers, close to the Spanish *jota* in Basque speakers and tends to be pharyngeal



([ʃ]) in Normandy. Pitch and duration analyses are in progress to verify if a peculiar prosodic pattern on word-penultimate syllables is reminiscent of the Swiss accent.

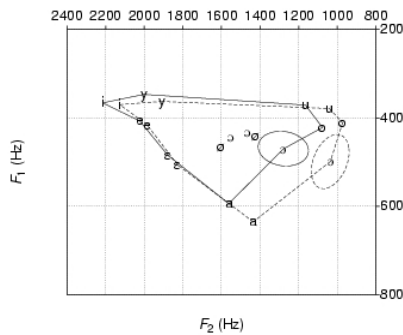


Figure 1: Vocalic triangles of speakers from the North (full lines) and from the South (dotted lines) computed on the stimuli listened to by the subjects.

3. PERCEPTUAL RESULTS

3.1. Pre-test

The pre-test enabled us to sort out the regions of our speakers by average degree of accentedness: from 0.8 for Normandy to 3.4 for Languedoc — the overall mean value is 2. The older the speakers, the stronger their accent. This result is unsurprising, even though, under some circumstances, nothing prevents young people from having a stronger accent than their elders, to assert their identity. Here, the average degrees for the three age categories are judged to be 1.4, 2.1 and 2.7. The difference is not so marked between reading, for which the average degree is 2.1 and spontaneous speech, assessed at 2.0. It is noteworthy that the only stimuli which were rated with a very strong accent (5/5) are those from speakers who “roll” /r/s.

3.2. Experiments 1 and 2

In the identification experiment, the listeners of the Paris region achieved 42.1% correct responses and those of Marseilles region 43.9%, which is significantly above chance according to chi-square tests [$df = 5; p < 0.01$]. A series of analyses of variance (ANOVAs) was also conducted on the responses counted as right (1) or wrong (0) with the random factor Subject: the between-subject factor Group (Paris or Marseilles region) and the within-subject factor Type of speech (read or spontaneous), speakers’ Age or Degree of accentedness. In each case the factor Group does not have a major effect, because the answer profiles are very similar. Contrary to the Type of speech, the speakers’ Degree of accentedness has a major effect [$F(5,245) = 18,1; p < .001$], and it interacts with the speakers’ Age [$F(8,392) = 8,11; p < .001$]. As a matter of fact, the origin of the oldest speakers (the ones who have the strongest accent) is better recognized than is the origin of the youngest speakers. For the oldest to the youngest speakers’ categories, we observe 47.3%, 43.9% and 37.8% correct answers. The difference is not so marked between reading, which results in a 42.3% score and spontaneous speech, which results in a 43.7% score. Among the 6 regions, Romand Switzerland is identified best. Despite its Franco-Provencal substratum, it is only marginally

mistaken with Southern varieties (in 2.5% of cases). Confusions are frequent among the Southern varieties, in such a way that their average identification rate is less than 35% (see Table 1). Symptomatically, the rather strong accent of the Languedoc speakers, which is the best identified as Southern (in 90% of cases) is more often associated to Marseilles than Marseilles itself. This result is in keeping with our intuition: the image we have of the Marseilles accent is actually a Southern stereotype which extends far beyond Marseilles.

Table 1: Confusion matrix for 50 listeners (%).

Response \ Origin	V	N	S	BC	L	P
Vendee	35.7	47.5	6.3	3.5	5.8	1.2
Normandy	37.3	48.5	2.3	4.5	6	1.3
Switzerland	11.7	13.8	72	1	1	0.5
Basque Country	8.3	5.3	1.8	38.7	35.2	10.7
Languedoc	5.7	3.7	0.2	26.3	31	33.2
Provence	11	10.3	1.3	16.5	28.7	32.2

4. CLUSTERING AND SCALING

Statistical analyses were also performed to provide graphic representations of the outcomes: in the form of dendrograms, by hierarchical clustering techniques and in the form of a two-dimension space, by multidimensional scaling [10]. The former technique results in trees where the horizontal distance between two leaves is a function of the distance which separates them in an observation matrix — in our case, the confusion matrix (see Figure 2). As far as the multidimensional scaling is concerned, it takes a dissimilarity matrix in input, which may be computed from the distances between the confusion matrix lines. This dissimilarity matrix is comparable to tables of distance between main towns, which can be found in pocket calendars. For this purpose, we used both the Euclidean distance and the Manhattan distance (also called *city blocks*), with several algorithms.

In almost all the configurations, the clustering yields a bipartition which opposes the North to the South. Switzerland is always well isolated too.

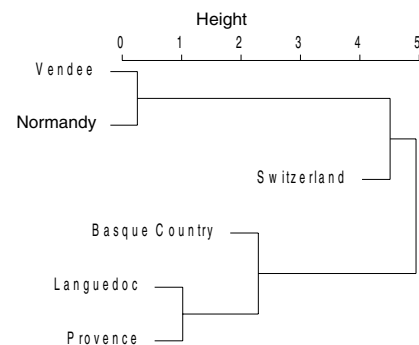


Figure 2: Dendrogram stemming from the hierarchical agglomerative clustering with a Euclidean distance.



From the multidimensional scaling visualization, three broad accents emerge: Northern French, Southern French and Romand Swiss (see Figure 3). Whatever the metrics used and the subset of responses considered, the plots are similar. Roughly speaking, the first dimension corresponds to the East-West axis: the second dimension corresponds to the North-South axis. These two dimensions account for a good proportion (90%–97%) of the variance.

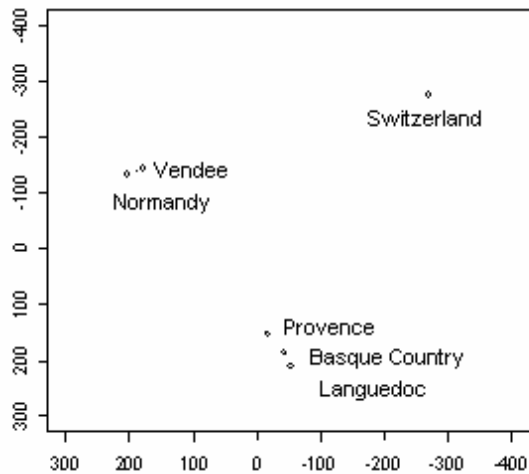


Figure 3: Result of the multidimensional scaling with the classical algorithm and the Euclidean distance.

5. CONCLUSIONS

Since accents can be identity markers, the question of their identification is particularly important. This preliminary work was based on perception (as [3]) and data analysis techniques (as [5]). From 36 speakers of 6 Francophone regions, it enabled us to draw 3 accents: North, South and East, without the results being affected by the speakers’ style and the listeners’ region of origin.

A couple of decades ago, [11] considered 15 accents in French, further subdivided by [2]. The distinction they made between Provence and Languedoc accents, for instance, seems to have vanished since then. Nonetheless, accents of Alsace and Corsica, among others, were missing in our study. Data fusions techniques will be considered, to add new points in the spaces obtained by multidimensional scaling.

The stimulus duration is comparable to previous studies [7]. By lengthening it, we could increase the accuracy and precision of responses. The results we obtained with 10 ms utterances challenge the myth according to which, for example, a Marseilles speaker can easily and readily be recognized. Yet, it is possible that social affiliation now outweighs geographic boundaries, contrary to what is traditionally claimed about the French language [2].

The subjects’ representativeness always raises sociolinguistic issues. Is there a stigmatized or idealized prototype for a given variety? If there is, is it reached? How to measure a distance with respect to this prototype, whether it is real or only imagined? The appreciation of an accent is dictated by the social status it is attached to, hence extremely complex

problems. In this context, speech processing may give precious indications, owing to its objectivity. Speech recognition, in particular, through its error rates and the proportion of pronunciation variants provided by automatic alignment, may be regarded as a tool to measure a distance with respect to a norm [12, 13]. We are currently working in this direction. In view of the quantity of speech data available, data-driven studies can be carried on a much larger scale than was previously possible, allowing perceptual and linguistic hypotheses to be validated or refuted.

6. ACKNOWLEDGEMENTS

This work was carried out within the framework of the CNRS VarCom project. We are grateful to Martine Adda-Decker and Noël Nguyen for making possible this comparative study.

7. REFERENCES

- [1] Delais-Roussarie, É. & Durand, J., *Corpus et variation en phonologie du français : méthodes et analyses*, Presses Universitaires du Mirail, Toulouse, 2003.
- [2] Walter, H., *Enquête phonologique et variétés régionales du français*, PUF, Paris, 1982.
- [3] Clopper, C.G., & Pisoni, D.B., “Some acoustic cues for the perceptual categorization of American English regional dialects”, *Journal of Phonetics*, **32**, 111–140, 2004.
- [4] van Bezooijen, R. & Gooskens, C., “Identification of Language Varieties. Contribution of Different Linguistic Levels”, *Journal of Language and Social Psychology*, **18**(1), 31–48, 1999.
- [5] Heeringa, W., *Measuring dialect pronunciation differences using Levenstein distance*, PhD thesis, Rijksuniversiteit, Groningen, 2004.
- [6] Preston, D.R., *Perceptual dialectology*, Foris Publications, Dordrecht, 1989.
- [7] Hintze, M.-A., Pooley, T., Judge, A. (eds), *French accents: phonological and sociolinguistic perspectives*, AFLS/CiLT, London, 2000.
- [8] Adda-Decker, M., Boula de Mareüil, P., Adda, G., Lamel, L., “Investigating syllabic structures and their variation in spontaneous French”, *Speech Communication* **46**(2), 119–139, 2005.
- [9] 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.
- [10] Manning, C.D. & Schütze, H., *Foundations of statistical natural language processing*, The MIT Press, Cambridge, 1999.
- [11] Carton, F. Rossi, M. Autesserre, D., Léon, P., *Les Accents des Français*, Hachette, Paris, 1983.
- [12] ten Bosch, L., “ASR, dialects and acoustic/phonological distances”, *ICSLP*, Beijing (pp. 1009–1012), 2000.
- [13] Hansen, J.H.L., Yapanel, U., Huang, R., Ikeno, A., “Dialect analysis and modelling for automatic classification”, *Interspeech*, Jeju (pp. 1569–1572), 2004.