

Comparing Praat and Snack formant measurements on two large corpora of northern and southern French

Cécile Woehrling & Philippe Boula de Mareüil

LIMSI-CNRS, Orsay, France

{woehrlin;mareuil}@limsi.fr

Abstract

We compare formant frequency measurements between two authoritative tools (Praat and Snack), two large corpora (of face-to-face and telephone speech) and two French varieties (northern and southern). There are both an evaluation of formant tracking (as well as related filtering techniques) and an application to find out salient pronunciation traits. Despite differences between Praat and Snack with regard to telephone speech (Praat yielding greater F1 values), results seem to converge to suggest that northern and southern French varieties mainly differ in the second formant of the open /O/. /O/ fronting in northern French (with F2 values greater than 1100 Hz for males and 1200 Hz for females) is by far the most discriminating feature provided by decision trees applied to oral vowel formants.

Index Terms: Praat and Snack formant tracking evaluation, northern and southern French accent identification, phonetics

1. Introduction

Freely-available signal processing tools performing formant tracking are now widely used. The Praat software (www.fon.hum.uva.nl/praat) and the Snack Sound Toolkit (<http://www.speech.kth.se/snack/>) appear as the most widespread ones. They have both been developed to enable the creation of multi-platform applications with scripting languages. However, to our knowledge, they lack comparative evaluation on large databases [7]. Since the latter do not allow a manual labelling, the accuracy of this or that technique may depend on what is intended. This study focuses on the characterisation of two varieties of the same language: northern and southern French. Two large corpora both including northern (i.e. standard) and southern French speakers were analysed to assess the impact of the recording conditions. The first corpus comes from the PFC project (“Phonology of Contemporary French”) [3]: following a Labovian protocol [9], it comprises reading and spontaneous speech. The present study is based upon a dozen investigation points from this corpus: over 100 speakers who were firmly rooted in their places of residence. The second corpus is less controlled, but contains conversational telephone speech (CTS) of over 500 speakers from 7 broad French regions. CTS may cause specific problems for formant tracking, and there may be a need for parameter tuning. The two corpora, which have been segmented into phonemes and phonetically transcribed by automatic alignment, are introduced in further detail in the next section.

Various formant normalisation procedures have been proposed so far, to rule out speaker-specific physiological characteristics, but they have drawbacks [1]. We believe our corpora are large enough to handle formant measures in Hertz (Hz) without resorting to intrinsic or extrinsic normalisation. Other questions are raised in section 3: how many measurements per vowel are needed? How to cope

with measurement errors?

Results are reported in section 4, in terms of correlations and distances between formant values of males and females. The comparison between Praat and Snack on the two corpora is completed by a comparison between northern and southern French vocalic triangles. To try and discriminate these two varieties, data mining techniques are then used (section 5), and a linguistic discussion on a possible sound change in progress (namely /ɔ/ fronting in northern French) concludes this paper.

2. Corpora

2.1. PFC corpus

Following [12; 14], the PFC project [3] has undertaken to collect recordings covering a wide French-speaking territory, with about ten speakers per investigation point. We analysed 12 investigation points: 6 in the North of France (Brécey, Brunoy, Dijon, Lyon-Villeurbanne, Roanne, Treize-Vents), 1 in Romand Switzerland (Canton de Vaud) and 5 in the South of France (Biarritz, Douzens, Lacaune, Marseilles, Rodez). The material is made up of over 100 informants: as many males as females of balanced age categories, from varying educational and professional backgrounds, who were born and have spent most of their lives in the same place. Totalling tens of hours of speech, the data represent 12,000 different word tokens. For each speaker, we have at our disposal the reading of a 100 word list and a 20 sentence text, as well as 10 minutes of directed interviews and free conversations.

2.2. CTS corpus

The CTS corpus is composed of 367 telephone conversations between speakers calling each other from one of 7 broad regions of France: North, West, East, Centre, Paris, South-East and South-West. There are around 70 speakers per region. About one third of them are males, and the average duration of a conversation is 14 minutes. The total amount of data represents about 85 hours of spontaneous speech and 110,000 different word tokens.

2.3. Automatic alignment & phonetic transcription

The two corpora have been segmented into phonemes by automatic alignment, using different context-independent acoustic models with Gaussian mixtures for face-to-face and telephone speech (as in [2; 6]). From a speech signal and its orthographic transcription, given acoustic models as well as a pronunciation dictionary with variants, the shortest path in the pronunciation graph provides the most likely sequence of phonemes. Standard French (spoken in Paris) may be considered as possessing 10 phonological oral vowels: /a ε e i ɔ o u œ ø y/ — when realised, the schwa is encoded /œ/, because this pronunciation gains ground [12]. The system

may be reduced to 7 in southern French, due to the merger of mid vowels [4]. It may also undergo the influence of the norm. Here, it has been automatically transcribed with the allophones of northern French.

3. Formant tracking: method

Formant measurements are automated via scripts designed for either Praat or Snack. They may be taken at as many points as desired. Since the alignment precision is not greater than 10 ms, we decided to base our comparisons on formant measurements every 10 ms.

We chose to extract only the first 3 formants (F1, F2, F3) because it was enough for our analysis and because it allowed us to process telephone and non-telephone speech in the same way (the maximum frequency in the signal is 3300 Hz in telephone speech). Accordingly, the settings were adapted to look for 3 formants (under 3000 Hz for males and 3300 Hz for females with Praat). For both Praat and Snack, we set the frame width to 50 ms. The other parameters are the default ones (e.g. autocorrelation method for Snack).

Filters were designed to discard aberrant values with respect to reference values in an average range of ± 500 Hz [6]. Adapted to each vowel, distinguishing males and females, they were applied before and after averaging the measures over each vowel.

4. Results

To analyse the results (distinguishing males and females), we computed the F1 and F2 mean values for each of the 10 oral vowels (1) per speaker, (2) averaged over the PFC and CTS corpora. Correlation coefficients (1) and distances (2) were calculated on separate vowels and all vowels aggregated.

Correlation coefficients calculated on all vowels merged are limited to F1 values, because pooling F2 values clearly results in a distribution that is not normal. One F1 mean value per vowel and per speaker (processed as equilikely) is retained: the degree of freedom is then $df = 10n - 2$, where n is the number of speakers considered. For correlation coefficients merely computed on isolated vowels, $df = n - 2$.

Several distances between sound systems have been proposed. They are suited for phoneme clustering tasks. For comparing two sets of values, e.g. the Mahalanobis distance assumes these sets have the same covariance matrix. Instead, we employed simple distances along the F1 and F2 axes separately. The distance along F1 (resp. F2) between two systems X and Y of 10 vowels (/a e i o u œ ø y/) was also averaged as expressed in equation 1 (resp. 2).

$$\delta_1(X, Y) = \frac{1}{10} \sum_{i=1}^{10} |F1(X_i) - F1(Y_i)| \quad (1)$$

$$\delta_2(X, Y) = \frac{1}{10} \sum_{i=1}^{10} |F2(X_i) - F2(Y_i)| \quad (2)$$

4.1. Influence of the formant filter

First, we measured the values of the first two formants yielded by Praat and Snack with filters applied before or after averaging. The rates of rejected vowels are reported in Table 1. They are of course higher (7% vs. 4% on average) if filtering is applied afterwards than if all the measures need be wrong for a vowel to be rejected. Although less severe criteria could be applied, the difference is striking in PFC male speakers whose formants were measured with Snack.

Table 1. Percentage of rejected vowels with filters applied before or after averaging formant measures.

%reject		PFC		CTS	
		males	females	males	females
Praat	before	2.6	2.7	1.8	2.2
	after	6.0	6.6	4.5	5.2
Snack	before	8.4	5.2	2.4	3.8
	after	14.6	9.0	4.4	6.3

Overall correlations exceed 0.85, and average distances are below 30 Hz between formant measures filtered before or after averaging. Except maybe for closed vowels such as /u/, the differences are negligible. Filters are used before averaging in the following.

4.2. Influence of the tool: Praat vs. Snack

Praat–Snack distances and correlations in terms of formant measures are indicated in Table 2. At the bottom of this table, overall correlations between Praat and Snack F1 measures are all greater than 0.85. Distances (summed up in absolute values as in equations 1 and 2) are small for F2 (< 50 Hz). They are more important (> 50 Hz) for F1 on telephone speech.

Table 2. Correlations and distances (in Hz) between Praat and Snack measures.

Praat – Snack		PFC				CTS			
		males		females		males		females	
		corr	dist	corr	dist	corr	dist	corr	dist
a	F1	.91	14	.91	42	.94	53	.91	93
	F2	.80	-77	.93	-48	.75	-14	.21	-2
e	F1	.89	17	.87	41	.90	54	.87	79
	F2	.92	-38	.93	-23	.46	-21	-.03	-18
e	F1	.86	19	.80	41	.90	40	.80	53
	F2	.86	-50	.88	-47	.07	-25	-.18	-37
i	F1	.82	28	.75	54	.81	51	.64	69
	F2	.88	-58	.86	-113	.23	-53	.07	-78
o	F1	.88	25	.93	43	.89	61	.89	86
	F2	.86	-36	.97	-26	.95	40	.86	31
o	F1	.90	34	.90	44	.85	60	.83	69
	F2	.78	1	.89	18	.89	65	.86	58
u	F1	.82	45	.78	63	.81	69	.78	76
	F2	.83	4	.89	32	.81	91	.80	83
œ	F1	.89	16	.84	35	.84	64	.80	88
	F2	.86	-56	.89	-51	.80	27	.50	39
ø	F1	.87	14	.74	24	.86	46	.82	59
	F2	.68	-53	.82	-54	.89	17	.66	10
y	F1	.91	22	.82	38	.89	58	.89	75
	F2	.88	-35	.81	-45	.68	5	.35	-18
all	F1	.96	23	.95	42	.92	56	.90	75
	F2	–	41	–	46	–	36	–	37

The correlations between the values measured for each vowel by Praat and Snack are on the whole very good: only 16% of

them are under 0.7. On the PFC corpus, the lowest value is 0.68, and all the measures seem to correlate even if formant measures differ (distances up to 63 Hz for F1 and 113 Hz for F2 in absolute value). We observe more variation on the CTS corpus: for a given vowel, the correlation may be good and the distance important between Praat and Snack measures (e.g. F2 of females' /a/) or inversely (e.g. F2 of females' /ɛ/). Some F2 measures show very weak correlations (e.g. /a/, /œ/, /y/ for females; /i/, /e/, /ɛ/ for males and females). This is manifested by vocalic triangles of different shapes. On both corpora, distances computed on the first formant are always positive: that is, Praat's F1 values are always greater than Snack's ones, resulting in a vertical shift of vocalic triangles.

Praat and Snack exhibit substantial differences, especially on F1 and certain vowels. Therefore comparisons among vowel spaces stemming from different signal processing tools must be taken with caution. Nevertheless, to bring to light cross-dialect differences in the vowel system of a given language, Praat and Snack may behave in a similar fashion. This issue is investigated in the remainder of this article.

4.3. Regional differences

Matrices of cross-regional and cross-corpus distances were established along the F1 and F2 dimensions. Based on female speakers analysed with Praat, Table 3 presents North/South differences on both corpora as well as PFC-CTS distances for northern and southern speakers. Matrices obtained for males and/or Snack cannot be reproduced due to lack of space, but they are very similar. They reveal more cross-corpus than cross-regional differences, probably due to the peculiarities of telephone speech. A more in-depth analysis is thus necessary.

Table 3. Distance matrix (along F1 / along F2) obtained for females analysed with Praat: cross-corpus distances on the diagonal, cross-regional distances on the antidiagonal.

females	North_PFC	South_CTS
North_CTS	81 / 46	8 / 17
South_PFC	10 / 76	82 / 83

Vocalic triangles were plotted for northern and southern French. The ones obtained with Praat for the female speakers of the PFC corpus are shown in Figure 1.

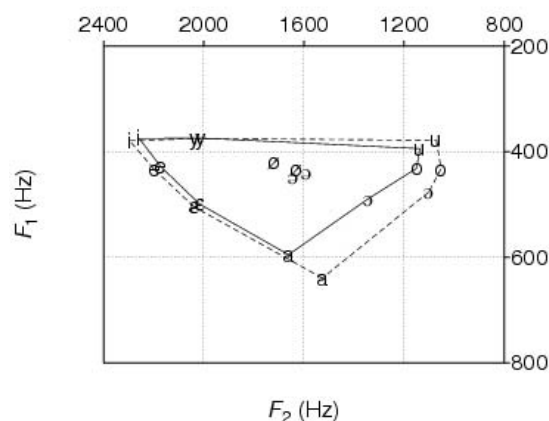


Figure 1: Praat vocalic triangles of northern females (in full lines) and southern females (in dotted lines) of the PFC corpus.

Irrespective of the corpus and the tool used, females as well as males display noticeable North/South differences as far as back vowels are concerned. Northern and southern speakers had a comparable speech rate (e.g. 12.1–12.2 phonemes/second in the PFC corpus) which does not suffice to explain these differences. The /a/ vowel is higher and more fronted in northern speakers, but the sharpest contrast that shows up is related to the realisation of the open /ɔ/ which tends to be centralised by northern speakers. As shown in Table 4, F2 differences for the /ɔ/ vowel are more marked on the PFC corpus (over 200 Hz) than on the CTS corpus (under 100 Hz). This seems to be due to the fact that the PFC speakers are supposed to be more representative of their locations. The reported values may be a useful and valuable by-product of this benchmark study for future work.

Table 4. Formant frequencies of the /ɔ/ vowel in northern (N) and southern (S) speakers. Values are rounded to 50 Hz.

/ɔ/		PFC				CTS			
		males		females		males		females	
		N	S	N	S	N	S	N	S
Praat	F1	450	400	500	500	500	500	600	550
	F2	1200	1000	1350	1100	1200	1100	1350	1300
Snack	F1	400	400	450	450	450	450	500	500
	F2	1250	1050	1400	1150	1150	1050	1300	1250

5. Discriminating northern and southern French varieties: /ɔ/ fronting

We wondered if the /ɔ/ vowel is the most discriminating one between northern and southern French, whatever the corpus and the tool used. We applied decision tree techniques, each speaker (either northern or southern) being represented by the average coordinates in the F1/F2 space of his/her vowels. There are more than 60 (resp. 370) northern speakers and 50 (resp. 150) southern speakers in the PFC (resp. CTS) corpus. We used the rpart library of the R software (<http://www.r-project.org>), whose output is a North/South bipartition as shown in Figure 2 for male speakers of the CTS corpus analysed with Snack.

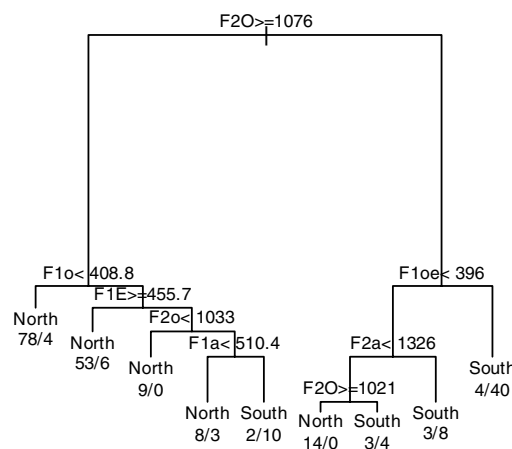


Figure 2: Decision tree discriminating northern and southern male speakers of the CTS corpus analysed with Snack. O = /ɔ/, E = /ɛ/.

Whether males or females are considered, on the PFC corpus and/or by using Praat, the discriminating nature of the second formant of /ɔ/ turned out to be obvious. This feature was always ranked first, even though the threshold changed (around 1100 Hz for males and 1200 Hz for females). This finding suggests the robustness of the /ɔ/ fronting parameter if one is interested in discriminating northern and southern French speakers: it results in 73–97% correct identification (the extremes being reached by Snacks’ formant values of females respectively from the CTS and the PFC corpus).

/ɔ/ is the default pronunciation of the graphic ‘o’, with a number of positional and morphological exceptions. Its fronting or centralisation has long been underlined in non-southern French [12]. Martinet [13] accounted for this sound change in terms of functional rendering (relatively weak and with little incidence on understanding for the /ɔ/~/œ/ opposition). More recently, this fronting phenomenon has been observed in studies concerned with vowel harmony [5; 10; 11]. Today, it would be a sign of affectedness, while from the XVIIth to the XXth century scholars attested fronted pronunciations which were then connoted as a working-class feature [8]. To our knowledge, this movement has not been studied in a systematic manner due to practical difficulties in carrying out phonetic fieldwork. Our measures allow us to revisit it and get insight into the diversity of French usage.

/ɔ/ fronting seems to be more achieved in spontaneous speech than in reading on the PFC data, where the two “styles” are represented. All other things being equal, we do not notice major differences between males and females on the one hand, or between speakers under 30 and over 60 on the other hand. Various influences may account for the /ɔ/→[œ] shift. The most frequent words with ‘o’ are not necessarily more fronted than the other ones, and the final/non-final position seems to have only little effect. However, front consonants in left and right contexts favour the /ɔ/ fronting. Examples are: *personne* (“nobody”); *notre*, *votre* (“our”, “your”); *jeux olympiques* (“Olympic games”), *visites officielles* (“official visits”); *socialisme* (and cognate words); *connais* (“know”). Even if the phonetically motivated change (originating in favourable contexts) is far from being consummated, there is evidence of an overall rapprochement of /ɔ/ and /œ/ in northern French.

6. Conclusions

The goal of this study was twofold: a comparison of two publicly available signal processing packages and a comparison between regional varieties of French. We tested two formant tracking tools on two large corpora comprising male and female speakers of standard and non-standard French varieties. The importance of filtering out erroneous measures was examined. The way of applying it happens not to be crucial, but this processing might become more influential if fewer measures per phoneme were extracted. It should also be adjusted if another scale (e.g. bark or mel) were used. This is in need of further investigation.

One may wonder: should Praat or Snack be preferred to analyse formant values of large corpora? Results prove very close on our corpus of face-to-face speech, but differ to some extent on telephone speech, Praat providing 50–100 Hz greater F1 values than Snack. Nonetheless, both Praat and Snack permit a distinction between northern and southern French speakers on the basis of formant frequencies (with

the help of data analysis techniques): F2 values for the /ɔ/ vowel greater than 1100 Hz for males and 1200 Hz for females are quite reliable features to identify northern speakers of French. Decision trees should also allow a comparison with better-known phenomena such as schwa and nasal vowel pronunciation [12; 14].

All these tools enabled us to outline a spreading linguistic change: /ɔ/ fronting in northern French. Successively indexical of working class and judged as a prestige variant, this feature would now index a geographical origin. The social dimension has been somewhat neglected here. Future work is planned to study whether *banlieue* Paris suburban youth adhere to this movement.

7. Acknowledgements

Part of this research was funded within the framework of the ANR PFC-COR project. We are grateful to Jacques Durand, Bernard Laks and Chantal Lyche for making available the PFC corpus. Our thanks also go to Lori Lamel and Martine Adda-Decker for her help in conducting this study.

8. References

- [1] Adank, P.M. *Vowel normalization: a perceptual-acoustic study of Dutch vowels*. PhD thesis, Radboud University Nijmegen, 2003.
- [2] Adda-Decker, M. & Lamel, L. “Pronunciation variants across system configuration, language and speaking style”. *Speech Communication* 29, 83-98, 1999.
- [3] Durand, J., Laks, B., Lyche, C. “Le projet ‘Phonologie du Français Contemporain’ (PFC)”. *La Tribune Internationale des Langues Vivantes* 33, 3–9, 2003.
- [4] Durand, J. & Lyche, C. “Structure et variation dans quelques systèmes vocaliques du français : l’enquête *Phonologie du français contemporain (PFC)*”. In: Coveney, A. & Sanders, C. (eds), *Variation et francophonie*, Paris: L’Harmattan, 217–240, 2004.
- [5] Fagyal, Z., Nguyen, N., Boula de Mareuil, P. “From dilation to coarticulation: is there vowel harmony in French?”. *Studies in the Linguistic Sciences* 32, 1–21, 2002.
- [6] 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”. *Proc. Interspeech* Lisbon, 2453–2456, 2005.
- [7] Harrison, P. *Variability of formant measurements*. MA thesis, University of York, 2004.
- [8] Hintze, M.-A., Pooley, T., Judge, A. (eds). *French accents: phonological and sociolinguistic perspectives*. London: AFLS/CiLT, 2000.
- [9] Labov, W. *Sociolinguistic patterns*. Philadelphia: University of Pennsylvania Press, 1972.
- [10] Landick M. “The Mid-Vowels in Figures: Hard Facts”. *French Review* 68(1), 88–102, 1995.
- [11] Malderez, I. *Contribution à la synchronie dynamique du français contemporain : le cas des voyelles orales arrondies*. PhD thesis, University of Paris 7, 1995.
- [12] Martinet, A. *La prononciation du français contemporain*, Paris: Droz, 1945.
- [13] Martinet, A. “C’est jeuli, le Mareuc !”. *Le français sans fard*, Paris: PUF, 345–355, 1969.
- [14] Walter, H. *Enquête phonologique et variétés régionales du français*, Paris: PUF, 1982.