The influence of speaking style on lexical f_0 profiles in French

Rena Nemoto[1]*, Martine Adda-Decker*[2]*, Jacques Durand*[3]

[1]LIMSI-CNRS & Université Paris-Sud 11, Orsay, France
[2]LPP CNRS/Université Paris 3, Paris, France
[3]CLLE-ERSS (UMR5263) CNRS & Université de Toulouse-Le Mirail, Toulouse, France
nemoto@limsi.fr, martine.adda-decker@univ-paris3.fr, jacques.durand@univ-tlse2.fr

Abstract

This study presents a comparison of French lexical fundamental frequency (f_0) profiles for different speaking styles using phonemic, syllabic and lexical transcriptions as well as part-of-speech annotations. Three speaking styles (broadcast news, broadcast conferences and conversations) with over 20 hours of speech were used. Syllabic word length and POS were considered as influential factors. Results confirm word final syllable accentuation as common tendency in French. The study highlights noun word-initial accentuation after `determiner` for BN style speech. Journalistic prepared speech features lexical words with more dynamic f_0 profiles on average versus more stable flat profiles for our spontaneous data. Future works include localization of named-entity and/or focus of speech within the framework of discriminative classifiers.

Index Terms: fundamental frequency, lexical f_0 profiles, French, word-final accentuation, POS annotation, corpus-based study, automatic processing.

1. Introduction

This work aims at improving our knowledge of potential prosody changes between speaking styles in French. We propose to make use of large corpora and automatic speech processing tools to investigates fundamental frequency (f_0) regularities of French words for various speaking styles. We consider speaking styles as used by the automatic speech recognition (ASR) community: from prepared to spontaneous conversational speech. From a linguistic point of view, it is agreed upon that speech production changes with style (cf. hyper- and hypospeech [1]). Taking a more technological ASR perspective, systems produce word error rates that are typically thrice as high for spontaneous speech than for carefully prepared speech. These observed differences suggest major changes in the acoustic realizations. The question of interest in this study is whether differences can be highlighted via lexical f_0 profiles? The used methodology was introduced in our previous studies [2, 3] to question prosodic regularities of French words via f_0 profiles. It combines time-aligned phonemic and lexical transcriptions, as well as automatic prosodic and POS annotations to compare average f_0 profiles according to word classes of given syllabic length, word final-schwa, duration and syntagms. For prepared journalistic speech the average lexical word contour showed a final rise concurrent with a minimum f_0 on the penultimate syllable, which tended to be reinforced with syllabic word length. Average f_0 profiles tended to be raised in presence of final schwa and for longer syllabic durations. We also observed weak first syllable accentuation of noun words after determiner. The present study then aims at checking whether speaking style plays a role in f_0 profile patterns and whether the previous findings hold across the newly added conditions .

In the following, section 2 presents the different style speech corpora and recalls the methodology to extract and organize the measurements. Section 3 compares and discusses f_0 profiles across conditions. Conclusions and future perspectives are given in section 4.

2. Corpora and Methodology

2.1. Corpora

The current study makes use of male speech from three manually transcribed corpora:
- **TECHNOLANGUE-ESTER** [4]: French broadcast news (BN), prepared speech (13 hours). The data mainly include public French radio news reports.
-**QUAERO**: scientific and journalistic public conferences corresponding to semi-prepared speech (0.5 hour). This data is somewhere between news reports and free conversations.
- **PFC** (Phonology of Contemporary French) [5]: speech of various speaking styles and of various French-speaking regions, with speakers who are firmly rooted geographically. Only the spontaneous PFC speech data including *guided* and *free* conversations from male speakers are considered here (6.8 hours).
Table 1 gives a word level description of the 3 corpora according to mono- and polysyllabic words. The QUAERO subset being the smallest in volume and corresponding to a loosely defined speaking style, the related upcoming results should be taken only as indicative. However, on a methodological level it is interesting to check whether the measured f_0 profiles from a small corpus are similar to (some of) those of the larger corpora.

2.2. Methodology

Concerning the prosodic level in French, many authors noticed the correlation between accentuation (final and initial), lengthening and word or syntagm boundaries [6, 7, 8, 9]. In the following, we propose contrastive measurements on subsets with increasing proportions of potential prosodic phrase boundaries. Acoustic correlates, namely f_0 is examined with respect to supposed influential factors: syllabic word length expressed in number of syllables, presence or absence of word-final schwa, part-of-speech (POS). Figure 1 gives a schematic overview of the processing steps on the investigated data.
f_0 measurements: Fundamental frequency (f_0) values were measured every 5 milliseconds (ms) using the standard settings of Praat [10] which results in at least six f_0 samples for each segments (a minimum phoneme duration is 30 ms).
Lexical and phonemic alignment: The audio corpora were automatically aligned by the LIMSI speech recognition sys-

Table 1: *Quantitative ESTER,* QUAERO *& PFC corpus description with regard to (w.r.t.) word tokens of syllabic length n from 0 to 4. Counts are separated for words with/without realized final schwa (top/bottom). Syll.class n_s states n: the number of full syllables; s: presence/absence of final schwa.*

n	*Syll. class* n_s	Occurrences *#Words* ESTER	PFC	QUAERO	*Examples*
0	0_0	12578	13921	414	d' /d/; de /d/
1	1_0	72249	65521	2915	vingt; reste
2	2_0	36027	20346	1212	beaucoup
3	3_0	15994	4959	497	notamment
4	4_0	6053	1408	176	présidentielle
n	*Syll. class*	Occurrences *#Words + /ə/*			*Examples*
0	0_1	12295	5056	413	de /də/; le /lə/;
1	1_1	3918	1642	112	reste; test
2	2_1	2087	716	36	ministre
3	3_1	698	208	30	véritable
4	4_1	174	49	3	nationalistes

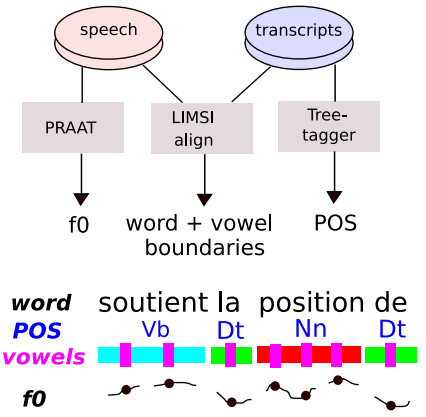

Figure 1: *Automatic processing steps and annotation levels: each vowel is tagged by an average f_0 value and its duration, by its rank within the word, by lexical and POS information.*

tem [11] producing word and phonemic segmentation. During the alignment, the pronunciation dictionary allows for optional word-final schwas, if the standard pronunciation ends with a consonant. For technical reasons, a phoneme segment is a minimum 30 ms duration and a boundary location precision of 10 ms.

Syllabic word length; syllabic length class: Each word token was annotated by its *syllabic word length*, corresponding to the number of full syllable in its aligned pronunciation. Word-final schwas did not count for the syllabic length, however, they were used to tag words into specific subsets. The word *test* ('test') with pronunciation [tɛst] was of syllabic length 1 with no word-final schwa, and was tagged as belonging to the *syllabic length class* 1_0. The same word pronounced [tɛstə] goes to the *syllabic length class* 1_1 (cf. *syll.class* in Table 1). Words of the same syllable class are merged to compute average f_0 profiles.

Part-of-speech tagging: To measure the influence of POS classes and sequences on f_0 realizations, the three corpora were POS-tagged. The annotation was carried out with WMATCH, the LIMSI regular expression general-purpose annotation engine [12] with a French version of TREETAGGER [13].

f_0 values, f_0 profiles: f_0 profiles were computed for each syllabic word class (*syll.class* n_s tags of Table 1). To compute these profiles only vowels with voicing ratios over 70% were used to minimize potential segmentation errors due to automatic alignment. This resulted in a rejection rate of about 10% for the prepared ESTER and semi-prepared QUAERO data and of 30% for the spontaneous PFC corpus. The discrepancy in rejection rates between prepared and spontaneous corpora suggest that there might be major changes in the acoustic realizations of these different speaking styles.

For each vowel a mean f_0 value was computed over all voiced frames of the vocalic segment. The values in Hz were converted to semitones (st), with 120 Hz as reference frequency (120 Hz is often considered as average male voice height) [14]. Perceptual studies [15] have shown that differences of 3 st play a role in the communicative situations even though weaker differences can already contribute to the perception of lexical demarcation for instance. Only words with all their vowels passing the voicing criterion were kept for further investigations. This selection aimed at reducing the impact of erroneous measurements, due to combined alignment and/or f_0 extraction errors. To each word from the prepared corpora including orthographic/phonemic transcribed pronunciation, a corresponding POS was associated. Each vowel was thus annotated with its mean f_0 in st and its syllable rank in the word. The *f_0 profile* of a word was then defined as a schematic f_0 contour connecting the f_0 values of the different vowels of increasing rank. Similarly, for a given syllabic word length class (*syll.class* in Table 1), the *f_0 profile* is defined as a schematic f_0 contour connecting the average f_0 values (computed over all the vowels of a given syllable rank and a given word subset) of the different syllables of a word. For example, given the `2_0` class of bisyllabic words without final schwa, the corresponding f_0 profile was computed as the contour connecting the average f_0 value of the rank 1 vowels (first syllable) of bisyllabic words to the average f_0 value of the rank 2 vowels (final syllable) of bisyllabic words. Word subsets can combine both syllabic word length and POS information.

3. f_0 profile results

In the following, f_0 profiles are compared across the three corpora. For the presentation given here, we only focus on words without final schwa (in Table 1). First, we present profiles for lexical words as opposed to grammatical words. The rationale is to empirically confirm whether grammatical words tend to remain unstressed which should then result in comparatively low f_0 profiles (subsection 3.1). Then we focus on nouns and noun phrases (`Determiner - Noun`) (subsection 3.2). As French tends to produce word-final accentuation, the graphical displays of the f_0 profiles of increasing syllabic word length were right-justified: the first syllable of monosyllabic words, the second syllable of bisyllabic words etc. are displayed at the final n-th position of the longest n-syllabic words.

3.1. Lexical vs. grammatical words

We present lexical and grammatical words f_0 profiles to check whether the known f_0 rise varies according to speaking style. Most occurrences of grammatical words in French are mono- or bisyllabic, whereas lexical words are frequently polysyllabic. Due to minimum word frequency criteria (#word tokens >100),

all profiles are limited to at most 4-syllabic lexical words. Table 2 shows the quantitative description of each corpus.

Table 2: *Quantitative ESTER,* QUAERO *& PFC corpus description with regard to (w.r.t.) word tokens of syllabic length* n *from* 1 *to* 4 *for* `lexical words` *(top) and from* 1 *to* 2 *for* `grammatical words` *(bottom)*

	n_s	ESTER	QUAERO	PFC
Lex.	1_0	30888	1272	29583
	2_0	33715	1125	18391
	3_0	15960	496	4854
	4_0	6036	176	1390
Gramm.	1_0	40921	1622	32382
	2_0	2237	83	1791

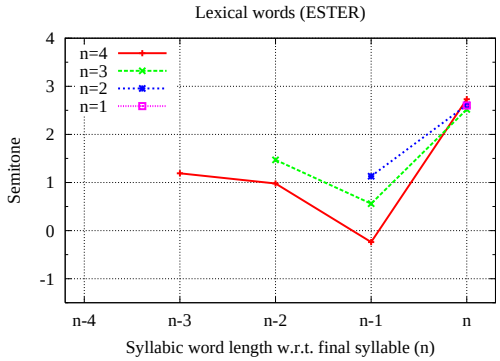

Figure 2: *ESTER corpus word* f_0 *profiles (lexical words)*

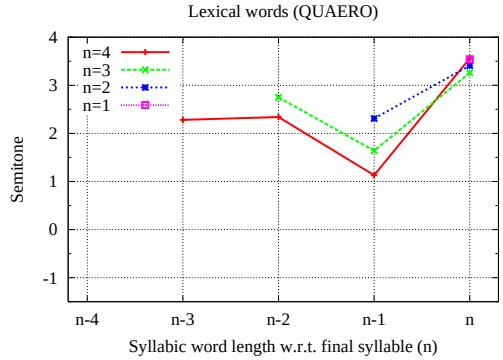

Figure 3: QUAERO *corpus word* f_0 *profiles (lexical words)*

Concerning the **lexical word** profiles (Figures 2,3,4), the 3 examined corpora (and speaking styles) share the following properties:

(i) Mean f_0 is higher for the final syllable n than for all preceding syllables.

(ii) For trisyllables or more, the f_0 difference between two consecutive vowels is maximal between penultimate and final vowels (Δ2–3 st for ESTER, Δ1–2 st for QUAERO, Δ0.8–1.5 st for PFC).

Contrary to ESTER and QUAERO corpora, which are (semi-) prepared speech, PFC corpus profiles show quite flat f_0 profiles except between final and penultimate syllables. It is not clear yet whether this result holds for any spontaneous speech corpus. QUAERO f_0 profiles have more movements than PFC f_0 profiles, but less than ESTER ones.

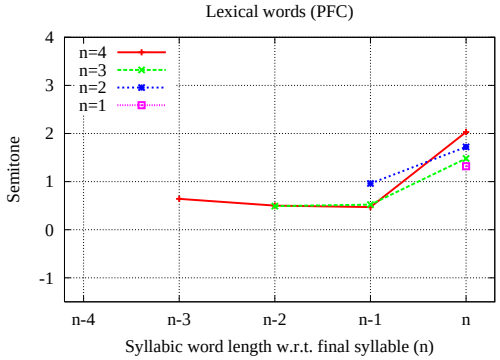

Figure 4: *PFC corpus word* f_0 *profiles (lexical words)*

For **grammatical words**, average f_0 contours of grammatical words feature flatter curves than the lexical word ones. It can be observed that final syllabic f_0 values of grammatical words are relatively lower than those of lexical words in the three corpora except for bisyllabic grammatical words in QUAERO. As outlined earlier, the latter results are only indicative, and ask for reexamination with a larger volume of data for this condition. From these results, we can observe similar, but different f_0 curves according to different speaking styles for lexical words. It is noteworthy, that the QUAERO f_0 profiles for the lexical data subset are very similar to (and in between the) two other speaking style corpora.

3.2. Noun vs. Noun phrase

Table 3: *Quantitative corpus description w.r.t.* `noun` *(top) and* `determiner # noun` *(bottom) sequence tokens of noun syllabic length* n *from* 1 *to* 4*.*

	n_s	ESTER	QUAERO	PFC
Noun	1_0	8222	325	5060
	2_0	11794	440	4990
	3_0	5120	188	1330
	4_0	2919	87	641
Det # Noun	det # 1_0	2243	88	969
	det # 2_0	2610	100	975
	det # 3_0	1403	65	267
	det # 4_0	862	18	141

In this subsection, mean f_0 profiles were calculated for noun phrases, limited to the `determiner noun`. In a previous study [3], we observed remarkable f_0 rise between determiner and first syllable of the noun. Here, we examined f_0 profiles according to different speaking styles to check if they exhibit similar distinct f_0 rise. Due to a small number of tokens in QUAERO corpus, we limited the noun and noun phrase comparison between ESTER (prepared speech) and PFC (spontaneous speech) corpora. Table 3 presents a quantitative description of the three corpora. Noun f_0 profiles are presented in Figures 5 and 6 (Left). These f_0 profiles are very similar to those of the lexical words in Figures 2 and 4.

From noun phrase figures (right Figures 5 and 6), we can observe that for ESTER, determiner f_0 values are located in under 0 st and difference between determiner and the first syllable of the noun are 1.2–3.5 st. For PFC corpus, f_0 profiles present Δ1.5 st for determiner–monosyllabic noun and Δ0.5 st for determiner–bisyllabic/4-syllabic nouns. These results point out less f_0 profile differences between determiner–first syllable of the noun for spontaneous speech. We can also observe an unexpected result for determiner–trisyllabic f_0 profile. f_0 value for determiner is slightly higher than to the first syllable of the noun. This may be again due to a small amount of sequence occurrences. The results from the different speaking style corpora suggest that f_0 values are low for determiner and f_0 profiles rise for the first syllable nouns. For spontaneous speech (PFC), the difference of f_0 values between determiner and noun are lower than for prepared speech (ESTER).

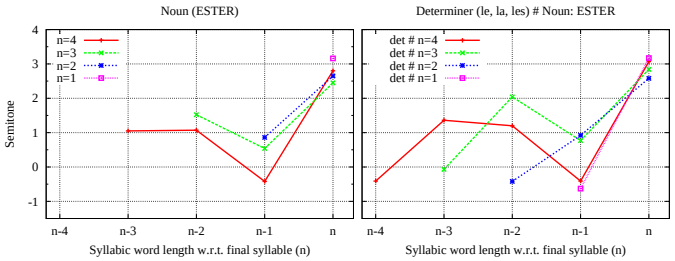

Figure 5: *ESTER corpus f_0 profiles* **Left***: Noun* **Right***: Noun phrase*

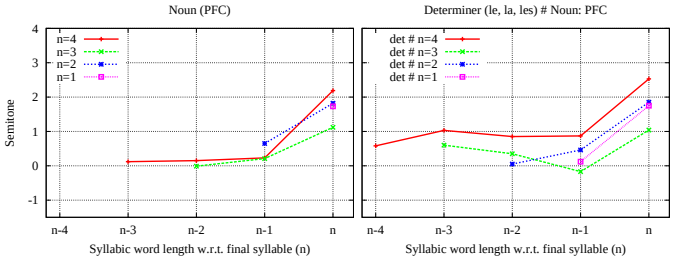

Figure 6: *PFC corpus f_0 profiles* **Left***: Noun* **Right***: Noun phrase*

4. Conclusions

This study presented a comparison of French lexical fundamental frequency (f_0) profiles for different speaking styles using phonemic, syllabic and lexical transcriptions as well as part-of-speech annotations. Three speaking styles (broadcast news, broadcast conferences and conversations) with over 20 hours of male speech were used. Syllabic word length and POS were considered as influential factors. The presented methodology combined automatically PRAAT-extracted and filtered f_0 contours with time-aligned phonemic and lexical transcriptions and POS annotations to compare average f_0 profiles according to word subsets of given word syllabic length and phrases. Results confirm word final syllable accentuation as common tendency in French across speaking styles. The study highlighted word-initial noun accentuation after `determiner` for journalistic (prepared) speech style. The latter features lexical words with more dynamic f_0 profiles on average versus more stable flatter profiles for our spontaneous data. The comparison between lexical/grammatical words showed interesting differences: marked variation of f_0 for lexical and low f_0 for grammatical words. This difference tended to decrease for the examined spontaneous PFC speech: our spontaneous speech f_0 profiles showed flatter average shapes, whereas in the prepared speech, more dynamic patterns were observed. Intermediate profiles characterize semi-prepared speech, for which more data will be added in the future. Future studies will also include in-depth focus on boundary specificities, such as localization of named-entity and/or focus of speech within the frame work of discriminative classifiers.

5. Acknowledgments

This work was partially funded by the DIGITEO research cluster - through *Région Ile-de-France* - doctoral grant to the first author, by OSEO under the *Quaero* program and by ANR PFC-Cor.

6. References

[1] Lindblom, B., "Explaining Phonetic Variation: A Sketch of the H&H Theory", in Hardcastle, W.J. and Marchal, A. [Eds], Speech Production and Speech Modeling, Kluwer Academic Publishers, Dordrecht, 403–439, 1990.

[2] Nemoto, R., Adda-Decker, M., and Durand, J., "Investigation of lexical f_0 and duration patterns in French using large broadcast news speech corpora", in Speech Prosody, 2010.

[3] Nemoto, R., Adda-Decker, M., and Durand, J., "Word boundaries in French: Evidence from large speech corpora", in LREC, 2010.

[4] Galliano, S. et al., "The ESTER Phase II Evaluation Campaign for the Rich Transcription of French Broadcast News", in Proceeding of Interspeech, Lisbonne, 2005.

[5] Durand, J. et al., "La phonologie du français contemporain: usages, variétés et structure", in C. Pusch & W. Raible [Eds] Romanistische Korpuslinguistik- Korpora und gesprochene Sprache/Romance Corpus Linguistics - Corpora and Spoken Language, Tübingen: Gunter Narr Verlag, 93–106, 2002.

[6] Vaissière, J., "Rhythm, accentuation and final lengthening in French", in Sundberg, J. et al. [Eds], Music, Language, Speech and Brain, 108–121, 1991.

[7] Hirst, D., Di Cristo, A., Intonation Systems : A Survey of 20 Languages, Cambridge University Press, Cambridge, 1998.

[8] Lacheret-Dujour, A. and Beaugendre, F., La Prosodie du Français, CNRS Éditions, Paris, 1999.

[9] Fougeron, C. and Jun, S. A., "Rate Effects on French Intonation: Prosodic Organization and Phonetic Realization", in Journal of Phonetics, 26:45–69, 1998.

[10] Boersma, P. and Weenink, D., "Praat: doing phonetics by computer [computer program]", Online: http://www.praat.org/, Tech. report, 2005.

[11] Gauvain, J.-L. et al., "Where Are We In Transcribing French Broadcast News?", in Proceedings Interspeech, Lisbonne, 2005.

[12] Galiber, O., Approches et méthodologies pour la réponse automatique à des questions adaptées à un cadre interactif en domaine ouvert. PhD. thesis, Universté Paris-Sud 11, 2009.

[13] Schmid, H., "Probabilistic Part-of-Speech Tagging Using Decision Trees", in Proceedings of International Conference on New Methods in Language Processing, Manchester, 1994.

[14] Léon, P. R., Phonétisme et prononciation du français, Armand Colin, Paris, 5e édition, 2007.

[15] 't Hart, J., "Differential sensitivity to pitch distance, particularly in speech", in Journal of Acoustical Society of America, 69(3):811–821, 1981.