

# A diachronic study of prosody through French audio archives

*Philippe Boula de Mareüil, Albert Rilliard & Alexandre Allauzen*

LIMSI-CNRS, Orsay, France

{mareuil;rilliard;allauzen}@limsi.fr

## Abstract

This study makes use of advances in automatic speech processing to analyse French audiovisual archives. A 10-hour corpus covering five decades of broadcast news is investigated from the angle of prosody evolution. Two prosodic features are explored: word-initial stress (especially in the case of proper names) and penultimate vowel lengthening (especially penultimate nasal vowels before a pause), which may give an impression of emphatic style. Our measurements suggest that the following features as well as mean pitch have decreased since the forties: pitch rise and vowel duration associated with initial stress, and prepausal penultimate lengthening. In the meantime, the onsets of supposedly stressed initial syllables have become longer while speech rate has not changed. This puzzling outcome rises interesting questions for research on French prosody.

## 1. Introduction

It seems that we are able to recognise a recording made decades ago and distinguish it from a contemporary recording. Technical conditions (the distance to the microphone and its type) have evolved. They are at least partly responsible for a peculiar voice quality which is readily perceivable and can easily be caricatured. More punctual, maybe more subtle but salient linguistic features may also have changed. We here concentrate on French prosody. The overall goal of this study is to disentangle the parameters which enable us to characterise an announcer style that dates back to the forties, for instance, and contrast it with a current journalist style. We now have at our disposal archives spanning over fifty years of French broadcast news (BN) which make possible such an investigation.

Though French is traditionally said to possess a phrase-final stress, numerous theoretical, experimental and applied studies have highlighted the emergence of a word-initial stress in contemporary French [5] [8] [17] [7] [15]. This initial stress is complementary to the final one, which remains a major property of the French accentual system, but spreads particularly in journalist and didactic styles, in BN, public conferences and classrooms. This mutation, the origin of which is difficult to trace precisely, would date back to the late XIX<sup>th</sup> century and even earlier [6]. In line with it, recent accounts of French prosody have integrated the coexistence of primary and secondary stresses into phonological models [6] [18] [12]. They hypothesise a double marking of content words by an initial stress and a final stress, the former being essentially melodic and the latter characterised by lengthening [18] [2]. In some contexts, the underlying initial stress may be realised at the surface level as an emphatic stress (*accent d'insistance*), with more dynamic pitch patterns and additional lengthening associated with the prominent syllable onset. Whereas the emphatic stress has a proper pragmatic or paralinguistic function, the non-emphatic initial stress rather has a rhythmic function motivated by eurhythmic constraints

favouring well-balanced stress alternations [11]: this device avoids over-long stretches of unstressed syllables. However, phonetic descriptions rarely propose clear acoustic correlates which would allow word-initial stressed syllables to be distinguished satisfactorily from their unstressed counterparts without resorting to interpretation. Admittedly, it may be listener- as well as speaker-dependent [20] [9], and many influential factors call for further investigation.

Following criteria proposed by [19] especially, our own approach to initial stress is mainly concerned with clitic–nonclitic sequences. Such a sequence corresponds to the most frequent word chunk, and constitutes a good candidate for initial stress on the nonclitic word. We considered clitic–nonclitic sequences in an extensive way, in an attempt to find out prosodic differences over time. This empirical study is first intended to evaluate whether or not initial stress gains ground in journalists' style, through BN archives dating from 1940 to 1995. On the one hand, an emphatic way of speaking is typical of old recordings [14]; on the other hand, initial stress contributes to portraying today's journalists and political figures by contrast to casual speech. Our aim is twofold: to verify if a linguistic change is in progress (with an increasing or decreasing tendency to initial stress or other features) and to exhibit relevant pronunciation traits discriminating several speaking styles.

A second prosodic feature is analysed: the lengthening of the penultimate vowel preceding a pause. In that context, it may be particularly salient in some archaic ways of speaking [5] [13]. It is therefore interesting to see if we can measure an evolution in that respect.

After a description of the corpus (section 2), section 3 is devoted to word-initial stress through pitch rise and nucleus/onset lengthening. Section 4 deals with penultimate vowel lengthening. Finally, section 5 concludes this study.

## 2. Corpus and method

Our corpus consists of 10 hours of data which have been collected within the framework of the ECHO project [3]. This corpus contains two kinds of documents: war archives which are cinematographic newsreels and audiovisual BN gathered for the EURODELPHES project, dating from 1945 to 1995. The latter project aims to create multimedia pedagogical material to teach the history of Europe. The whole database is composed of 189 audiovisual documents, each one containing at least one speaker. The audio sequences range from 20 seconds to 20 minutes. Documents were orthographically transcribed by hand and segmented into phonemes by automatic alignment using extensively-trained context-independent acoustic models and a specifically-tuned pronunciation lexicon. The method described in [1] has been validated in several publications since then. Pitch values were then assigned to each phoneme by averaging fundamental frequency ( $F_0$ ) measurements taken every 10 ms by the PRAAT software (<http://www.praat.org>) with standard options — in which  $F_0$  values below 75 Hz appear as undefined. This

led to a rather raw but robust representation of prosody in which each phoneme is defined by its duration, mean pitch, the word it belongs to, and other information to facilitate further processing. In addition to these local tonal configurations, different melody stylisations are possible and wished for a closer look at the shape of the  $F_0$  curve [19] [16], but they were not applied here.

Along this article, four periods are considered so as to have balanced subsets of data: 1940–1959, 1960–1969, 1970–1979 and 1980–1995. Indeed, the sixties and seventies are more represented than the other decades in our data. Table 1 features the duration of each subset, the average phoneme duration which appears to be quite comparable between the retained periods, the percentages of vowels that are detected as unvoiced vowels, which are reassuringly low, as well as mean pitch. The latter is restricted to males due to the under-representation of females in the archives (<20% of the shows, only one show in the 40s and one show in the 50s).

Table 1: Duration of the data, phoneme duration, percentage of unvoiced vowels and mean pitch (for males).

	duration (s)	ph. dur. (ms)	unvoiced V (%)	mean pitch (Hz)
1940–1959	7434	79	4.8	175
1960–1969	11552	81	3.2	144
1970–1979	6053	83	4.7	141
1980–1995	9403	80	3.5	137

It is of note that males’ pitch was higher in the forties and the fifties and has regularly decreased since then. It also has the effect of making the 1940–1959 vocalic triangles larger than the other ones which are almost identical. (Formant estimation was performed by using PRAAT as in [9].) This higher pitch may be due to the recording conditions which compelled journalists to speak louder, standing in front of their microphones. It does not necessarily mean that males used to speak with a higher pitch. It is even not ascertained that journalists spoke with a higher pitch than today. The original recording support (record or tape) may impact parameters such as mean pitch. However, this is what we receive when listening to old archives. Additionally, more local parameters such as the following ones should not be affected.

### 3. Word-initial stress

#### 3.1. Pitch rise

A script was designed to compute pitch difference ( $\Delta F_0$  in semitones) between the last vowel of a clitic and the first vowel of the subsequent nonclitic word. A question arises as to the definition of what a clitic is. Based on prior knowledge, a first set of about 300 function words was established, including forms of auxiliary verbs such as *être* (“to be”) and *avoir* (“to have”). Based on the most frequent words of our corpus, we built up a second set of 30 function words: *le, la, les* (“the”), *un, une* (“a”), *du, des* (“some”), *de* (“of”), *à, pour* (“to”), *en, dans* (“in”), *et* (“and”), *que, qui* (“that”), *est* (“is”), *a* (“has”), *il* (“he”), *on, nous* (“we”), etc. The negations *pas* (“not”) and *plus* (“[no] more”) were excluded, because they can hardly be regarded as clitics. The resulting number of clitic–nonclitic sequences was 23,000 with the first set,

21,000 with the second set. We thus kept the latter, which is more controlled and displays a wide coverage. The drawback is that contiguous sequences such as *ne plus* (“ne plus”) then fall into clitic–nonclitic sequences. It was therefore particularly important to focus on clitic–polysyllabic word sequences, which additionally avoids the merger of initial and final stress in monosyllabic words. Table 2 reports the number of corresponding contexts, and the cumulative percentages of cases in which the pitch difference between the clitic vowel and the polysyllabic word-initial vowel is greater than 1, 2, 3 or 4 semitones (ST). The distribution with non-cumulated percentages is depicted in Figure 1. Of course, a higher difference in absolute value leads to a lower percentage. However for each epoch the ranking is the same. In Figure 1, it is noticeable that the more recent the document is, the lower the corresponding curve is on the right (positive) side and the higher it is on the left (negative) side.

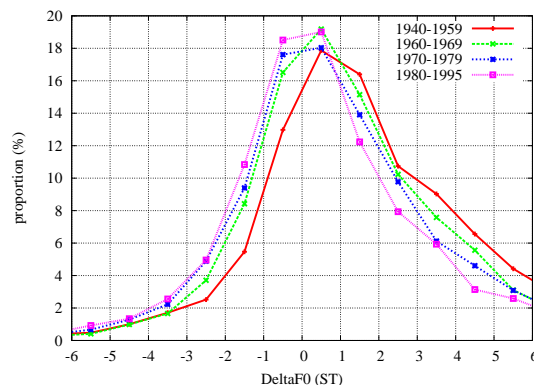


Figure 1:  $\Delta F_0$  between clitic vowels and polysyllabic word-initial vowels (in semitones).

In Table 2, the proportion of pitch differences that are greater than 1 ST, for example, is over 50% in the forties and the fifties, and decreases as time goes by. Following [19] and more recent studies on French [10], the 3 ST threshold is assumed to be a good estimate of the acoustic correlates of prosodic prominence. According to this interpretation, more than one clitic–polysyllabic word sequence out of four gives rise to a prosodic prominence in the 40s and the 50s, compared to less than one out of five in the 80s and the 90s. As expected, this proportion slightly increases with the size of polysyllabic words in the 40s and the 50s: 26% of disyllabic words, 31% of trisyllabic words and 32% of tetrasyllabic words exhibit pitch differences that are greater than 3 ST with the preceding clitics (see the first columns of Table 3). However this trend does not hold for the following decades.

Table 2: Number of clitic–polysyllabic contexts with the proportions of cases in which the pitch difference is greater than 1, 2, 3 or 4 semitones.

	#ctxts	%>1ST	%>2ST	%>3ST	%>4ST
1940–1959	2104	55	39	28	19
1960–1969	4154	47	32	21	14
1970–1979	2107	43	29	20	14
1980–1995	3793	38	26	18	12

Proper names are good candidates to receive initial stress. Their proportion among polysyllabic words in the studied contexts decreases as time goes by (see last column of Table 3). More interestingly, proper names seem to receive more prosodic prominence in the news clips of the 40s and the 50s than in the following decades: the remotest period is the only one in which the percentage of word-initially stressed proper names is greater than the percentage of proper names involved in clitic-polysyllabic word contexts (compare the last two columns of Table 3).

Table 3: Percentages of disyllabic, trisyllabic and tetrasyllabic words as well as proper names (PN) preceded by clitics with a pitch difference greater than 3 ST. The percentage of PNs involved in clitic-polysyllabic word sequences is also reported.

	%disyll. >3ST	%trisyll. >3ST	%tetrasyll. >3ST	%PN >3 ST	%PN
1940–1959	26	31	32	16	15
1960–1969	19	26	24	10	13
1970–1979	17	26	22	7	9
1980–1995	16	23	20	8	9

There is no clear tendency as to the lengthening of polysyllabic word-initial vowels with respect to the preceding clitic vowels. For instance, the percentages of duration differences that are greater than 20 ms are respectively 61%, 53%, 55% and 57% for the 1940–1959, 1960–1969, 1970–1979 and 1980–1995 periods. Nevertheless, other duration-related patterns may have evolved since World War II. This issue is addressed in the remainder of this paper.

### 3.2. Stressed syllable nucleus/onset lengthening

The duration of polysyllabic word initial onsets and vowels preceded by clitics was calculated in the same contexts as above. For example, in *de puissants*, (“of powerful”), the onset is [ptʃ] and the vowel nucleus is [i]. Table 4 reveals that onset duration increases over time. The 10 ms increase is even more regular when the analysis is restricted to simple onsets (composed of only one consonant). The forties and the fifties especially depart from the following decades, in which more than half of onsets exceed 70 ms. This proportion is below 50% in the case of the 40s and the 50s. Other results with respect to 60 ms and 80 ms thresholds are tabulated in Table 4 (thresholds are centred around mean values).

Table 4: Onset duration of polysyllabic words preceded by clitics (mean value and percentages of occurrences exceeding a given, variable threshold).

	All onsets			Simple onsets		
	mean dur.	%>70 ms	%>80 ms	mean dur.	%>60 ms	%>70 ms
1940–1959	70ms	49	37	63ms	58	42
1960–1969	77ms	60	47	71ms	69	54
1970–1979	80ms	62	49	72ms	70	56
1980–1995	79ms	63	50	73ms	71	58

This onset lengthening contradicts the tendency suggested in the previous paragraph. An opposite tendency should be

posited if onset lengthening were a correlate of initial stress (see [15] [11] [2]). An alternative interpretation is that the relative importance of initial stress correlates may have changed for half a century. In any case, a change is noticeable, while speech rates are almost equal in the period under study (see Table 1). This change is compared to vowel lengthening in the following.

As is apparent in Table 5, vowel duration in clitic-polysyllabic word contexts decreases from 1940 to 1995. The evolution is parallel if only vowels that are at least 3 ST higher than the preceding clitic vowels are considered. Mean duration and percentages are naturally greater than if all vowels are taken into account, and the fact that stressed vowels are a little shorter in the sixties than in the seventies is more marked.

Table 5: Vowel duration of polysyllabic words preceded by clitics (mean value and percentages of occurrences exceeding a given, variable threshold). The right part is restricted to vowels that are at least 3 ST higher than the preceding clitic vowels.

	All vowel nuclei			Vowels>3ST		
	mean dur.	%>70 ms	%>80 ms	mean dur.	%>70 ms	%>80 ms
1940–1959	79ms	61	48	85ms	66	55
1960–1969	73ms	51	40	75ms	56	42
1970–1979	73ms	52	38	79ms	58	47
1980–1995	71ms	47	34	78ms	54	39

All in all, the increase of onset duration and the small decrease of nucleus duration over time make the duration of supposedly stressed vowels stable. This leaves more time for pitch excursions in the 1940–1959 data. Even though other interpretations remain open, ours is that initial stress has decreased since the forties.

## 4. Penultimate vowel lengthening

A script was written to compare the duration of the last two vowels of polysyllabic words — the final schwa being excluded. At first glance, the distributions of penultimate-final vowel duration differences are very similar across the periods under investigation: percentages of positive differences range from 30% to 35%. Nonetheless, if the analysis is restricted to prepausal positions, the 1940–1959 patterns separate from the other ones. This leaves a large number of contexts, as shown in Table 6, and this position triggers a more salient perceptual effect. Table 6 presents the mean duration of penultimate vowels and the percentage of words in which the penultimate vowel is longer than the final one. The average percentage is strikingly stable in the most recent shows (18% from the sixties), but it goes up to 25% in the oldest recordings, speech rates being comparable (see Table 1). A normalisation by the mean duration of each vowel does not change these figures. However, it is well-known and we can measure on our data that French nasal vowels are intrinsically longer than are oral vowels (120 ms vs. 88 ms on average). We thus looked at penultimate nasal vowels in further detail. Of course, the right part of Table 6 shows fewer contexts and higher percentages than the left part. More importantly, the gap widens between the different periods. In the forties and the fifties, more than half of penultimate nasal

vowels are extra-long — longer than final vowels in spite of prepausal lengthening. The decrease of mean duration (from 110 ms to 91 ms) is also noticeable. Syllable-based patterns obtained by applying syllabification rules proposed by [1] are similar.

Table 6: Number of polysyllabic words preceding a pause, percentage of occurrences in which the penultimate vowel is longer than the final one and mean duration of penultimate vowels. The right part presents results for penultimate nasal vowels only.

	All vowels			Penultimate nasal V		
	#ctxt	%longer pen.	mean dur.	#ctxts	%longer pen.	mean dur.
1940–1959	1605	25	94ms	222	52	110ms
1960–1969	2781	18	86ms	312	40	101ms
1970–1979	1477	18	87ms	238	30	98ms
1980–1995	2492	18	82ms	337	33	91ms

## 5. Conclusion and future work

Speech processing (automatic alignment and pitch extraction) allowed a data-driven approach to linguistic changes, in particular prosodic changes in the French broadcast news style. From this well-defined context, it enabled the identification of epoch-specific patterns and a quantitative study that goes beyond usual impressionistic descriptions. Results showed a decrease of mean pitch, word-initial stress (at least as far as its melodic correlates are concerned) and prepausal penultimate lengthening which was more marked in the forties and the fifties, in particular for nasal vowels. These two decades are the most different from the other ones.

While speech rate has not evolved, the intriguing onset lengthening associated to what may be considered as initial stress rise interesting questions on French prosody. Despite the quantity of data manipulated, statistical analyses are probably necessary to understand the apparently paradoxical discrepancy between onset duration-related and pitch-related correlates of supposedly stressed syllables. Anyhow, average duration differences of 10 ms are not negligible, but they certainly result in fewer audible cases than 3 ST pitch differences. In comparison, the just noticeable difference is about 20 ms [4].

Speech processing enables us to select well-audible samples of these phenomena (e.g. words such as *nation* or *présent*, where the penultimate vowel is longer than the final nasal vowel). Informal listening tests using prosody modification/resynthesis may reveal how pitch and duration features contribute to the dated nature of some recordings. Further perceptual experiments are needed (and scheduled) on the role of prosody in the characterisation of old-fashioned speaking styles. We hope they will allow us to sort out the contribution of onset lengthening and pitch rise to the perception of initial stress in French.

## 6. Acknowledgements

We are grateful to L. Vinet and the French *Institut National de l'Audiovisuel* (INA) for the audiovisual corpus and its transcription which were made available for the EURODELPHES and ECHO projects. Our thanks also go to M. Adda-Decker and C. Woehrling for their help in this research.

## 7. References

- [1] Adda-Decker, M.; Boula de Mareüil, P.; Adda, G.; Lamel, L. 2005. Investigating syllabic structures and their variation in spontaneous French. *Speech Communication* 46(2), 119–139.
- [2] Astésano, C.; Bard, E.G.; Turk, A. 2002. Functions of the French Initial Accent: a Preliminary Study. *Speech Prosody*. Aix-en-Provence, 139–142.
- [3] Barras, C.; Allauzen, A.; Lamel, L.; Gauvain, J.-L. 2002. Transcribing audio-video archives. *ICASSP*. Orlando, 13–16.
- [4] Bartkova, K.; Sorin, C. 1987. A model of segmental duration for speech synthesis in French. *Speech Communication* 6(3), 245–260.
- [5] Carton, F.; Marchal, A.; Hirst, D.; Séguinot, A. 1977. *L'accent d'insistance (Emphatic stress)*. Montréal: Didier.
- [6] Di Cristo, A. 1999. Le cadre accentuel du français contemporain : essai de modélisation. *Langues* 2(3), 184–205.
- [7] Fant, G.; Kruckenberg, A.; Nord, L. 1991. Durational correlates of stress in Swedish, French and English. *Journal of Phonetics* 19, 351–365.
- [8] Fónagy, I.; Léon, P. 1980. *L'accent en français contemporain*. Montréal: Didier.
- [9] Gendrot, C. 2006. *Aspects perceptifs, physiologiques et acoustiques de différentes catégories prosodiques en français*. PhD thesis. Université de Paris III.
- [10] Goldman, J.-P.; Avanzi, M.; Simon, A.C.; Lacheret, A.; Auchlin, A. 2007. A methodology for the automatic detection of perceived prominent syllables in spoken French. *Interspeech*. Antwerp, 98–101.
- [11] Jankowski, L.; Astésano, C.; Di Cristo, A. 1999. The initial rhythmic accent in French: Acoustical and perceptual prosodic cues. *ICPhS*. San Francisco, 257–260.
- [12] Jun, S.-A.; Fougeron, C. 2000. A phonological model of French intonation. In *Intonation: Theories and Models*, A. Botinis (ed.). Dordrecht: Kluwer Academic Publishers, 209–242.
- [13] Léon, P. 1993. *Précis de phonostylistique. Parole et expressivité*. Paris: Fernand Nathan.
- [14] Méadel, C. 1994. *Histoire de la radio des années trente*. Paris: INA/ANTHROPOS.
- [15] Mertens, P. 1993. Accentuation, intonation et morpho-syntaxe. *Travaux de linguistique* 26, 21–69.
- [16] Mertens, P. 2004. The Prosogram: Semi-automatic transcription of prosody based on a tonal perception model. *Speech Prosody*. Nara, 23–26.
- [17] Padeloup, V. 1990. *Modèle de règles rythmiques du français appliqué à la synthèse de la parole*. PhD thesis. Université de Provence, Aix-Marseille I.
- [18] Rossi, M. 1999. *L'intonation, le système du français : description et modélisation*. Gap: Ophrys.
- [19] 't Hart, J.; Collier, R.; Cohen, A. 1991. *A perceptual study of intonation: an experimental-phonetic approach to speech melody*. Cambridge: CUP.
- [20] Vaissière, J. 1983. Language-independent prosodic features. In *Prosody: Models and measurements*, A. Cutler & D.R. Ladd (eds). Berlin: Springer-Verlag, 53–66.