



ELSEVIER

Speech Communication 16 (1995) 293–311

**SPEECH**  
COMMUNICATION

## Durational variations in speech and didactic accent during reading <sup>☆</sup>

Claire Gerard <sup>\*</sup>, Delphine Dahan

*Laboratoire de Psychologie Expérimentale, Université René Descartes, CNRS, URA 316, 28 rue Serpente, 75006, Paris, France*

Received 19 May 1993; revised 19 May 1994

---

### Abstract

The aim of this research is to analyze the main durational changes occurring in the spoken string when a reader produces an emphatic accent (“didactic accent”) on target-words included in texts, and to verify whether these changes are organized into structured prosodic forms. Three experiments address the following questions: (1) Are the durational changes concentrated in the immediate vicinity of the target (last syllable and pause preceding the target, duration of enunciation of the target, pause following the target)? (2) Are there intercorrelations among the durational variations observed? (3) Does the typographical realization of the target produce different effects on time changes? (4) Does the semantic weight of the targets change the way in which speakers produce didactic accents? Two speakers had to read 16 different texts (each text once with and once without target-words). After a description of the amplitude and location of durational variations in their performance, analyses are performed in order to locate precisely where slowing down occurs. The conclusions are relativized when confronting the two speakers’ strategies. Pre-target variables are intercorrelated, but post-target pauses vary independently. Speakers are insensitive to typographical and semantic determinants. The results are compatible with the hypothesis that mental representations of prosodic forms govern the temporal structure of speech in loud reading, and they show the importance of cognitive determinants during continuous reading.

### Zusammenfassung

Es wird angenommen, dass die zeitliche Struktur der Sprache beim lauten Lesen von mentalen Repräsentationen prosodischer Formen gesteuert wird. Die vorliegende Arbeit beschäftigt sich mit den zeitlichen Änderungen, die während einer Sprachkette eintreten, wenn der Leser seine Aufmerksamkeit auf Ziel-Worte richtet die im Text integriert sind und analysiert ob diese in strukturierten prosodischen Formen organisiert sind. Drei Experimente versuchen die folgenden Fragen zu beantworten: (1) Sind die zeitlichen Änderungen auf die unmittelbare Umgebung (letzte Silbe und Pause die dem Ziel-Wort vorausgeht, Dauer des Aussprache des Ziel-Wortes, Pause die dem

---

<sup>☆</sup> This research was supported in part by a grant no. 89 1B 226 from the Centre National d’Etudes des Télécommunications. We gratefully acknowledge the contribution of Françoise Emerard to the ideas which led to this experiment. We are also grateful to Catherine Rigaut for her assistance in the analysis of the data, and to Maya Hickmann and Hélène Mateescu for their valuable comments on the English version of the paper.

<sup>\*</sup> Corresponding author.

Ziel-Wort folgt) der Ziel-Worte konzentriert? (2) Sind diese verschiedenen zeitlichen Änderungen korreliert? (3) Haben die unterschiedlichen typographischen Realisierungen des Ziel-Wortes verschiedene Wirkungen auf die zeitlichen Änderungen? (4) Variiert die Art und Weise der Betonung mit dem semantischen Gewicht des Ziel-Wortes? Zwei Leser haben 16 verschiedene Texte jeweils zweimal mit und ohne Ziel-Wort gelesen. Die Ziel-Worte waren durch eine Änderung der Typographie gekennzeichnet. Nach einer vorausgehenden Beschreibung des Ortes und der Amplitude der beobachteten zeitlichen Änderungen, erlauben die Analysen des akustischen Signals die Ursprünge der Sprachverlangsamung einzugrenzen, die jedoch je nach der vom Leser angewandten Strategie voneinander abweichen. Die Werte der gemessenen Variablen auf und vor dem Ziel-Wort sind interkorreliert, der Wert der Pause die dem Ziel-Wort folgt ist jedoch unabhängig. Beide Leser sind unempfindlich für typographische und semantische Determinanten. Die Ergebnisse weisen auf kognitive Beschränkungen hin, die die Entscheidungen des Lesers beim kontinuierlichen Lesen von Texten beeinflussen.

## Résumé

La structure temporelle de l'énoncé, dans la lecture à voix haute, est guidée par des représentations internes des formes prosodiques auxquelles un lecteur se doit d'obéir. Cette recherche porte sur les changements temporels survenant dans la chaîne parlée lorsqu'un lecteur réalise un accent d'insistance (accent didactique) sur des mots-cibles inclus dans le texte, et analyse les formes prosodiques structurées qui en résultent. Les trois expériences présentées tentent de répondre aux questions suivantes: (1) Les changements temporels sont-ils limités à l'entourage immédiat des cibles (dernière syllabe et pause précédant la cible, durée d'énonciation de la cible, pause suivant la cible)? (2) Ces divers changements temporels sont-ils corrélés? (3) La nature même des marques typographiques permettant de repérer les cibles produit-elle un effet sur les modifications temporelles réalisées? (4) L'accent didactique produit varie-t-il avec l'importance relative de la cible dans la signification du texte? Deux locuteurs ont lu 16 textes différents présentés chacun deux fois, avec et sans mots-cibles signalés par un changement de typographie. Après une description préliminaire de l'emplacement et de l'amplitude des changements temporels observés, les analyses du signal sonore ont permis de circonscrire les sources de ralentissement, lesquelles peuvent diverger selon la stratégie adoptée par le locuteur. Les valeurs des indices mesurés sur et avant la cible sont inter-corrélés, mais la valeur de la pause suivant la cible est indépendante. Les deux locuteurs sont insensibles aux déterminants typographiques et sémantiques. Les résultats témoignent des contraintes cognitives qui pèsent sur les décisions du locuteur lors de la lecture continue de textes.

*Keywords:* Prosody; Durational variations; Loud reading; Didactic accent

## 1. Introduction

Speech communication implies that speakers and listeners have a shared knowledge of the prosodic manifestations of various intentions: e.g., interrogation, assertion, insistence. Sorin (1989) and Calliope (1989) therefore assume that some abstract mental representations of time, of intensity and of melodic prosodic variations govern the production as well as the perception of speech. Temporal variations in prosody during the continuous loud reading of texts are not yet well known and text-to-speech synthesis reveals that relevant data are missing. The aim of this research is to shed some light on the prosodic *durational patterns* occurring when a French reader insists on

target-words included in texts, producing what we will later define as a didactic accent.

During the unfolding of spontaneous discourse, pauses (silent or not) play an important role in the cognitive decisions made by a speaker. For spontaneous speech, Goldman-Eisler (1972a, 1972b) showed that about 40 to 50 percent of the total time of speech is constituted by pauses and assumed that these are synchronized with the cognitive processes involved in the production of a spoken string. A series of experiments by Grosjean and Deschamps (1972, 1973, 1975), aiming at measuring various temporal parameters and at comparing these parameters in French and English, gave the following results: (1) The duration of pauses was significantly shorter for interviews

(0.52 seconds) than for descriptions of humorous drawings (1.32 sec) and the number of pauses was smaller in the first task. This result would be due to the cognitive constraints required by a descriptive task as compared to the relative freedom inherent in an interview. (2) The syntactic distribution of these pauses was similar for the two tasks: 62.43% of the pauses for descriptions and 60.46% for interviews were located at the ends of sentences. Therefore, a subject having to describe a given stimulus delays his vocalizations, losing some time to ensure a difficult coding: the articulation rate slackens, the vocal strings shorten and become more spaced. In reading, cognitive constraints are lighter, since the speaker does not have to create the concepts, but rather only to discover and translate them. Lucci (1983) for French, as well as Howell and Kadi-Hanifi (1991) for English compared read and spontaneous speech and noticed a general decrease in the number of pauses, empty or filled, in reading. In loud reading, the total duration of reading is highly correlated with the total duration of pauses, and the mean duration of pauses at the end of a sentence is twice as long as the duration of other pauses (2 sec versus 1.1 sec, Saint-Bonnet and Boë, 1977). The speakers' competence plays a role in the number and duration of pauses. Kowal and O'Connell (1980) studied pauses in various types of discourse: e.g., prose, poetry, public speeches. Comparing the readings of a biblical text by two groups of subjects, "naive" readers and dramatic art students, they observed that students made less silent pauses than naive subjects, but that the duration of their pauses was longer. The kind of text to be read is as important as the kind of task to be achieved: the percentage of pauses doubles in dictation and the frequency of pauses due to punctuation (compared with the total number of pauses), which is around 80% in loud reading, is only 50% for dictation (Le Nouveau, 1987). Thus, syntactic determinants of pauses, usually reflected in punctuation, is nearly – not totally, as shown by Grosjean (1980) – necessary and sufficient to generate pauses in usual reading, but not in dictation.

In musical as well as in many linguistic studies (Dell, 1984; Lucci, 1983; Martin, 1972, 1979;

Rossi, 1985, 1988; Gérard and Auxiette, 1992), *accent* is a generic term qualifying any event whose perceptual salience draws the listener's attention. The first function of an accent in a language is to divide the spoken string into significant subunits – words (as in English or Italian, Spanish or German), or syntagms (as in Arabic or French) (Rossi, 1988). In contrast to many other languages, French is not a stressed language, which means that no specific rule (included in the lexicon) specifies the existence of a "stress" or its location in a polysyllabic word. However, two kinds of accents (not to be confused with stress) are spontaneously produced by native French speakers: the *primary accent* (also called "intensity accent" or "normal", "etymological", "historical" accent), located on the last syllable of a word or of a group of words, and the *secondary accent* (also called "contrastive accent" (Bolinger, 1961), and for French, "accent of insistence" (Rossi, 1980; Garde, 1968), or "expressive", "emphatic", "emotional", "didactic", "barytonic" accent), which may be located at the beginning of a word or of a group of words, as well as in the middle, or at the end of an internal morpheme in a polymorphemic word. In the framework of this paper, we will adopt Lucci's concept of "didactic accent" (Lucci, 1983), which corresponds best to our reading task in comparison to the concepts of "expressive", "contrastive" or "emphatic accent", or of "secondary accent" (Pardeloup, 1988, 1989), or of "barytonic accent" (Fonagy, 1980). Similarly to the changes observed for lexical stress in stressed languages, changes in the rhythm, intensity and melodic contour of utterances result from the "didactic accent" due to the insistence on a word or group of words in a spoken string (Lieberman, 1960; Chafe, 1974; Séguinot, 1977; Cutler and Fodor, 1979; Bock and Mazella, 1983; Cooper et al., 1985; Rossi, 1985, 1987; Eady and Cooper, 1986; Needham, 1990; Pardeloup, 1990). If melodic variations due to this accent are understood with adequate accuracy, intensity variations are not well understood. As for time variations, there is no agreement among researchers concerning the dispersion around the target word. Dahan (1994), Dahan and Bernard (1994) and Gérard and Da-

Gérard and Dahan (1992) have all shown differences in this dispersion depending on the kind of reading task (isolated sentences or texts), on the one hand, and on the strategies developed by various speakers, on the other hand. We assume that in continuous text reading the didactic accent is prepared by the speaker in advance, and that it still weighs on time organization after the target word has been realized. But the relative importance of time variations compared to a neutral statement is still to be assessed, together with their position in the spoken string, and their sensitivity to such factors as target visual or syntactic and semantic determinants.

This research is therefore intended to study how speech changes *in time* when a didactic accent is placed on target words within an extended corpus in texts that are read out loud. The corpus under study corresponds to two French speakers reading 16 texts presented twice, once without required accentuation (the “monotonous” versions) and a second time with 14 target words (underscored by changing their typography) which have to be accentuated by the speaker (the “relief” versions).

Records were made of only two speakers. The resulting description is therefore not an average which is valid for any abstract speaker, but an analysis of the strategies of an individual subject, with the study of the other speaker only in order to relativize conclusions and to indicate alternative strategies. All of the following comparisons concern the differences between the “monotonous” and “relief” versions of the same texts read by the same speaker. In the first part (Experiment I), *global measurements* are performed (reading duration, speech rates, pauses translating the visual text organization into paragraphs and its syntactic organization by punctuation marks), as well as *local measurements* in the immediate vicinity of target words (duration of the syllable preceding the target and of the vowel contained in that syllable, pre- and post-target pauses, duration of the target). As the sum of local changes does not explain the total slowing down observed in accentuated texts, a follow-up study is carried out after this first part of the study in order to take into account longer speech

fragments around the targets (Experiment II). Finally, Experiment III (based on a more limited corpus extracted from the previous recordings) addresses more specific questions related to semantic determinants of the didactic accent.

## 2. Experiment I

### 2.1. Method

#### 2.1.1. Material

In a first phase of the experiment, 24 texts were selected from newspapers, or internal information in industry or administrations, their content corresponding to classified advertisements, job offers, meeting announcements, news bulletins or instructions for use, . . . , all of which are characterized by a clear structure and dense semantic information, but without attractive episodes evoking any emotion or personal implication on the part of the reader. These texts were read a first time by a speaker (presented below), and two judges (students in linguistics at the University Paris V) listened to the resulting recordings in order to select all the words which seemed accentuated. This task was repeated three times. Only 16 texts about which the judges achieved total agreement in their selection of accentuated words were retained as experimental texts. For the experiment, each text had to be presented in two forms: the *monotonous* form (*m*), corresponding to the text structured in separate paragraphs, but without any visual emphatic sign on particular words, and the *relief* form (*r*), corresponding to the same text divided into paragraphs, but including 14 target words, the print of which was contrasted with the rest of the text.

In a second phase, the experimental target-words had to be selected. This selection obeyed the following principles: no target word could be situated at a syntactic boundary marked by a punctuation sign; no target word could be adjacent to another target word, and at least six other words had to separate them; half of the target words had to be chosen from the previously spontaneously accentuated words, and the other half among non-accentuated words; no target word

could appear in the first two sentences; and finally (and more intuitively), the 14 target words selected in a text had to be distributed in such a way that loud reading did not become too hard a task. The typing fonts retained to mark the target words in the relief version were italics, capital letters, bold characters, underlining, mainly typographical elements drawing the reader's attention. A few parentheses and inverted commas were added, but it was found difficult to insert artificial additions into already drafted texts. Approximately 25% of the targets began with a vowel and 75% with a consonant, these percentages corresponding to the distribution of initial phonemes in French (Lucci, 1983). Appendix B presents two of the 16 texts as examples.

### 2.1.2. *Speakers*

On the basis of a pre-test (loud reading of a scientific article in French), two women speakers among four candidates were selected for the good quality of their articulation in continuous reading by the same two judges as previously. Speaker 1 (who read the first 24 texts) was a teacher at the University Paris V, born and living in Paris, without regional accent. Speaker 2 was a research engineer at the National Center for Scientific Research (CNRS), born in Alençon, who had been student in Tours and had worked in Paris since 1970. She had no regional accent.

### 2.1.3. *Recording procedure*

We chose to have four distinct recording sessions, half a day each, for each of the two speakers. One session corresponded to the recording of the two versions of four texts, i.e. a block of eight recordings. A text was thus read out loud in its two versions during the same session. The order of presentation of the two versions was balanced so that the monotonous and relief forms were presented at first an equal number of times. The influence of the repetition of the same semantic content was checked: a given text was presented in its second version only after three intermediate texts and a half-hour break. Seated in an anechoic room, the reader first analyzed the text to be read by working on its theme, the technical terms, the pronunciation difficulties. Recording

of a given text took place after an analysis of about 10 minutes. A recording session therefore lasted a little more than two hours. The speaker was informed that a listener outside the anechoic room listened to him in real time and took notes during the course of reading.

### 2.1.4. *Recording and signal analysis systems*

Recording was conducted on PRO-X/L500 videocassettes using an AKG D 190E microphone facing the speaker in the anechoic room. UNICE (VECSYS) sound signal editing software was used to display the sound records. This software allows measurement of the intensity, melody and duration variations of utterances. The temporal accuracy of our system was 12.8 msec (one frame).

## 2.2. *Results*

As indicated in the introduction, we will first present the global measurements, then circumscribe the temporal changes observed between monotonous and relief versions of the same texts. The very role of articulatory determinants of the didactic accent was not of interest for us (psychologists), and we will consider the intrinsic phonetic nature of targets as secondary. We will first insist on the way speakers globally master their speech rhythm.

### 2.2.1. *General organization of texts and pauses*

Each text appeared visually as a series of distinct paragraphs (code VO = vertical organization), each paragraph being divided in sentences and syntactic sub-groups marked by punctuation signs (code P). Between two punctuation signs, target words were included in groups of words which we will call (in the context of this experiment) "sense units" (code SU). The number of actual pauses (longer than 100 msec)<sup>1</sup>, their mean durations and standard deviations are presented in Table 1.

<sup>1</sup>In the general text organization, only those pauses lasting more than 100 msec were counted so as to exclude articulatory silences. In the later study of targets as such, all the pauses have been recorded.

Table 1

Mean number ( $n$ ), mean durations and standard deviations (sd) of pauses related to vertical organization (VO), punctuation (P) and sense units (SU) for the monotonous ( $m$ ) and relief ( $r$ ) versions

	VO			P			SU		
	$m$	$r$	$r - m$ %	$m$	$r$	$r - m$ %	$m$	$r$	$r - m$ %
Speaker 1									
mean	2155.6	1997.5	-8.26	857	849	-0.94	332.7	338.6	1.77
sd	551	505		369	348		204	206.5	
$n$	6.63	6.56		20.2	19.6		19.1	25.7	
Speaker 2									
mean	3652.1	3683.2	0.72	1516.2	1556.2	2.79	470.5	473.7	0.68
sd	907	829		905	958		302	350	
$n$	6.43	6.50		19.4	19.6		22.1	27.1	

The number of pauses hardly varied, irrespective of the version being considered, monotonous ( $m$ ) or relief ( $r$ ) for VO or P, but increased for sense units (SU). The appearance of target words within sense units therefore resulted in additional stops in the speech of the two speakers. Due to the variability of pause durations, as measured by their standard deviation, the mean duration of the observed pauses did not vary significantly when the text version changed, irrespective of the speaker concerned. The differences between speakers were much stronger than between text versions (less than 0.54% as an average between  $m$  and  $r$ ). The fact that the duration of the pauses relating to the general text structure or to punctuation hardly changed has to be considered more closely.

Punctuation indicates the syntactic organization of texts. Punctuation pauses reflect the integration of that organization by the speaker. Is this process modified when the texts are presented in their "relief" version? Table 2 presents the mean durations of pauses corresponding to full stops, colons, semi-colons and commas for each

monotonous ( $m$ ) or relief ( $r$ ) version of the texts and for each speaker.

For the two speakers S1 and S2, a distinction can be made among three categories: "long" pauses corresponding to full stops; "intermediate" pauses corresponding to colons, signs of enumeration or subsequent explanations; "short" pauses corresponding to commas and semi-colons. Speaker 2 marked punctuation by longer stops than speaker 1. However, for the two speakers, increases of the pause duration between  $m$  and  $r$  remained small (less than 9% of the total duration) and were not statistically significant in view of the large calculated standard deviations for each and all the texts. Therefore, accentuating the target-words does not seem to significantly affect the duration of the stops at syntactic boundaries.

### 2.2.2. Slowing down of speech

2.2.2.1. *Global slowing down.* Table 3 presents total reading durations for each version, i.e. for each text T1 to T16 and each speaker S1 and S2.

Table 2

Mean durations (msec) of punctuation pauses

	Point		Colon		Comma and semi-colon	
	$m$	$r$	$m$	$r$	$m$	$r$
Speaker 1	1236.2	1217.8	950.1	950.1	610.1	612.5
% $r - m$	-2.2		0		0.74	
Speaker 2	2610.0	2834.0	1583.3	1696.3	893.8	910.3
% $r - m$	8.73		8.99		3.73	

Table 3  
Speech rate slowing down<sup>a</sup>

Total number of			Total duration (sec)				Speech rate (syll/sec)			
			S1		S2		S1		S2	
words	syll.	<i>m</i>	<i>r</i>	<i>m</i>	<i>r</i>	<i>m</i>	<i>r</i>	<i>m</i>	<i>r</i>	
T1	306	583	173.16	190.55	159.88	221.91	3.36	3.06	3.64	2.63
T2	237	484	142.19	157.34	165.51	171.89	3.40	3.07	2.92	2.82
T3	231	413	121.00	137.87	149.39	159.40	3.41	3.00	2.76	2.59
T4	255	460	142.61	162.28	177.69	207.66	3.22	2.83	2.58	2.21
T5	235	501	160.90	158.21	178.21	192.53	3.11	3.16	2.81	2.60
T6	174	380	107.64	124.39	124.06	130.37	3.53	3.05	3.06	2.91
T7	299	427	155.27	162.28	167.49	192.52	2.75	2.63	2.55	2.22
T8	186	387	120.83	144.99	147.92	162.54	3.20	2.67	2.62	2.38
T9	191	338	96.22	106.93	114.94	128.23	3.51	3.16	2.94	2.64
T10	208	395	108.70	120.37	130.57	141.13	3.63	3.29	3.02	2.80
T11	193	418	131.63	142.50	150.10	162.45	3.17	2.93	2.78	2.50
T12	244	674	151.14	178.62	196.96	204.85	4.46	3.77	3.44	3.30
T13	236	517	156.85	172.60	175.65	228.90	3.30	2.99	2.94	2.26
T14	188	392	109.18	122.86	150.06	167.00	3.59	3.19	2.61	2.35
T15	263	514	153.64	164.38	205.15	230.42	3.34	3.13	2.50	2.23
T16	258	567	170.38	190.38	233.93	237.85	3.33	2.98	2.42	2.38
<b>M</b>	<b>231.5</b>		<b>137.58</b>	<b>152.28</b>	<b>164.22</b>	<b>183.73</b>	<b>3.39</b>	<b>3.06</b>	<b>2.85</b>	<b>2.55</b>
			Standard deviation:				0.35	0.26	0.33	0.30
			Slowing down:				9.73%		10.53%	

<sup>a</sup> T = texts; S1 = speaker 1; S2 = speaker 2. *m* = monotonous version; *r* = relief version.

As a function of the total number of syllables contained in each text, a speech rate (or number of syllables per second) can be measured. Measurement shows that both speakers slowed down their speech rates in the relief versions. Speaker 1 globally read more quickly than speaker 2, but the slowing down rate in the “relief” version was of the same order of magnitude for both speakers, about 10%. For a one-page text (231 words on the average) with 6% of words to be accentuated (14 targets/231 words), we found 10% total slowing down as compared with monotonous reading. The next problem was to locate slowing down accurately.

2.2.2.2. *Slowing down at target level.* Five parameters were measured (durations of pauses preceding and following the target<sup>2</sup>, durations of enun-

ciation of the *syllable* preceding the target, of the *vowel* in that syllable, and of the *target* itself), and submitted to analyses of variance (ANOVAs) by groups (speakers 1 and 2), texts (T1 to T16) and versions (*m* and *r*), the 14 targets being the random factor. The results are presented in Table 4.

Table 4  
Mean durations (msec) of the five parameters in monotonous (*m*) and relief (*r*) versions for the two speakers S1 and S2

	Version	Monotonous	Relief	Difference
Mean enunciation duration for target	S1	763.88	988.62	224.74
	S2	770.49	1066.16	295.67
	Mean	767.19	1027.39	
Mean duration of pre-target pauses	S1	86.16	237.32	151.16
	S2	104.29	210.18	105.89
	Mean	95.22	223.75	
Mean duration of post-target pauses	S1	225.67	539.96	314.29
	S2	419.60	665.47	245.85
	Mean	322.64	602.71	
Enunciation duration of pre-target syllable	S1	246.21	319.82	73.61
	S2	240.63	318.17	77.54
	Mean	243.42	319.00	
Enunciation duration of the included vowel	S1	119.46	165.89	46.43
	S2	122.99	156.21	33.22
	Mean	121.05	160.63	

<sup>2</sup> As indicated in footnote 1, we now measure all the silences detected by our analysis system. Consequently, the articulatory silences will be included. Our reasoning was the following: when speakers accentuate a word, they vary melodic and intensive parameters as well as durational ones. When intensity increases, articulatory silences also increase for voiceless plosives. If we want to find all sources of durational changes, we have to measure all pauses.

The mean *enunciation durations of the target* were significantly longer in the “relief” version ( $F(1-208) = 608.66, p < 0.00001$ ). The difference between speakers was significant in the accentuated version ( $F(1-208) = 25.20, p < 0.0001$ ): speaker 2 slowed down more than speaker 1 when pronouncing targets, as shown by the calculated differences  $r - m$  (295.6 msec and 224.7 msec). Target enunciation durations were text-dependent ( $F(15 - 208) = 3.06, p < 0.0005$ ). This effect can be explained by the intrinsic variations of target lengths among texts, which is confirmed by the fact that the text effect was significant irrespective of the version, whether “monotonous” ( $F(15-208) = 3.12, p < 0.0001$ ) or “relief” ( $F(15-208) = 2.94, p < 0.0005$ ).

The mean durations of the *pre-target pauses* were more than twice as long in the “relief” version than in the monotonous one ( $F(1-208) = 189.29, p < 0.00001$ ). Globally no difference was noted between the two speakers ( $F(1-208) = 0.28, p > 0.50$ ) but for speaker 1 the durational increase (difference  $d$ ) was greater than for speaker 2 and the interaction version  $\times$  speaker was significant ( $F(1-208) = 9.66, p < 0.005$ ). Didactic accents on targets were therefore always translated by significant durational increases, especially for speaker 1 who marked shorter pauses in the monotonous versions than speaker 2. No significant difference was induced by the nature of the text ( $F(15-208) = 1.77, p > 0.40$ ).

The durations of *post-target pauses* were almost twice as long in the  $r$  version than in the  $m$  version ( $F(1-208) = 239.78, p < 0.00001$ ). It

should be noted that duration is over half a second in version  $r$  (about 600 msec), which approximates the values observed for some punctuation pauses when reading the texts. The difference between speakers was significant ( $F(1-208) = 28.17, p < 0.0001$ ): speaker 1 always marked shorter pauses but increased the pause durations more than speaker 2 with version  $r$  in comparison to version  $m$  ( $d = 314$  and  $245$  msec, respectively). For this type of pauses, the text nature introduced significant differences ( $F(15-208) = 3.482, p < 0.0001$ ) and an interaction also appeared between texts and speakers ( $F(15-208) = 3.214, p < 0.0001$ ).

Global lengthening was 31.05% (75.58/243.42 msec) for the *syllable preceding the target*, and 32.85% (39.83/121.05) for the *vowel* in the syllable ( $F(1-208) = 385.64, p < 0.00001$  for the syllable globally and  $F(1-208) = 296.71, p < 0.00001$  for the vowel only). Half of the total syllabic lengthening was therefore found to occur on the vowel. No difference due to the text or the speaker was observed for the two types of segments under consideration.

**2.2.2.3. Total lengthening in the immediate vicinity of targets.** The durational increases due to target level when reading the texts are summarized in Table 5. This table shows the following information: the total text reading duration; for all the targets, the total durations of the pre-target pauses (code Pr) and of the post-target pauses (code Po), as well as the durations of enunciation of the last syllables preceding the targets (code

Table 5

Percentage of lengthening due to targets summing up pre-target pauses (Pr), post-target pauses (Po), duration of enunciation for the syllable preceding the targets (syll) and durations of enunciation for the targets (T) in monotonous ( $m$ ) and relief ( $r$ ) versions

	Total duration (seconds)		Pr		Po		Syll		T	
	$m$	$r$	$m$	$r$	$m$		$m$	$r$	$m$	$r$
Speaker 1	2201	2447	20	56	53	119	54	71	171	221
$r - m$	246		36		66		17		50	
	Syll + Pr + Po + T: 169									
	% lengthening due to targets: 68.7									
Speaker 2	2627	2944	24	47	98	155	53	70	173	239
$r - m$	317		23		57		17		66	
	Syll + Pr + Po + T: 163									
	% lengthening due to targets: 51.28									



SYLL) and of the target words (code *T*). In each column the durational increase between the monotonous and “relief” versions is calculated (difference  $r - m$ ); the total lengthening observed in the vicinity of targets (corresponding to the sequence SYLL + Pr + *T* + Po) is then related to the global reading lengthening. For speaker 1, 69% of speech slowing down were observed in the immediate vicinity of targets (which represented only 6% of the text words). Nevertheless, approximately 30% were thus located elsewhere. For speaker 2, barely more than 50% of the global lengthening were due to the immediate vicinity of the targets. Therefore, the two speakers do not distribute durational changes in the same way, and speaker 1 “concentrates” them more around the targets.

### 2.2.3. Correlations among temporal parameters

So far we have studied each durational parameter separately. However, it can be assumed that all durations are not handled by the speaker independently from one another in the spoken string and that slowing down has some inertia. As a result, correlations should appear if durational increases are handled in a consistent way by the speaker and if they correspond to the same function. Conversely, lack of correlation would imply either a solution in the reading strategy continuity or different functions for the different increases. We therefore examined the data to determine if such correlations could be found. The matrices (Table 6) present the values of the Bravais–Pearson correlation coefficient  $r$ .

The correlations were always significant between duration scales considered pairwise for “syllables”, “vowels” and “pre-target pauses”, i.e. any element preceding the target, for both versions and for both speakers. Out of the 24 coefficients, 6 were low (code a): each time the element concerned was the link of one of the pre-target elements with the post-target pause. Cochran and Hotelling tests (allowing for a comparison between the correlation coefficient values) were calculated to confront correlations between pre-target elements and correlations between post-target pause and pre-target pause, and between post-target pause and duration of

Table 6

Correlations among durational parameters surrounding the targets (in parentheses: “relief” versions, without parentheses: monotonous versions)

	Pr	Po	Syllable	vowel	
Speaker 1	Pr		0.20 <sup>c</sup>	0.423 <sup>d</sup>	0.258 <sup>d</sup>
	Po	(0.181) <sup>b</sup>		0.092 <sup>a</sup>	0.136 <sup>a</sup>
	Syllable	(0.373) <sup>d</sup>	(0.14) <sup>a</sup>		0.654 <sup>d</sup>
	Vowel	(0.40) <sup>d</sup>	(0.115) <sup>a</sup>	(0.483) <sup>d</sup>	
Speaker 2	Pr		0.192 <sup>c</sup>	0.468 <sup>d</sup>	0.636 <sup>d</sup>
	Po	(0.214) <sup>b</sup>		0.14 <sup>a</sup>	0.215 <sup>c</sup>
	Syllable	(0.458) <sup>d</sup>	(0.131) <sup>a</sup>		0.777 <sup>d</sup>
	Vowel	(0.539) <sup>d</sup>	(0.197) <sup>c</sup>	(0.669) <sup>d</sup>	

<sup>a</sup>  $p > 0.01$ ; <sup>b</sup>  $p < 0.01$ ; <sup>c</sup>  $p < 0.005$ ; <sup>d</sup>  $p < 0.0000001$ .

Pr = pre-target pause; Po = post-target pause.

the syllable. The results confirm that the duration of the post-target pauses does not seem to be linked with the overall speech slowing down (in that case the correlation would be positive and significant), nor is it used to compensate insufficient slowing down (in which case the correlation would be negative and significant). It cannot therefore be excluded that post-target pauses have a function which is different from other durational indices in the spoken communication process. We propose the hypothesis that the post-target pauses differ functionally from the preceding durational parameters. While all the material that precedes the target would give evidence for the processes developed by the speaker to prepare himself for accentuation, the post-target pauses would serve as a signal specifically intended for the listener to integrate the accentuated word.

### 2.2.4. Visual determinants: typographical characteristics

For the speaker to acknowledge the presence of targets prosodically, it is necessary that the latter be detected visually by an initial sensory and perceptual process. During this first stage, it is not excluded that visual characteristics of the stimulus such as the letter size or shape (capital letters, italics), the energy concentration in typographical characters (bold characters), or the addition of non-pronounceable signs (parentheses, inverted commas) represent different degrees of

detectability. Moreover, during a second more cognitive stage, it is possible that cultural tradition requires that different degrees of insistence be allocated to typographical signs: for example, capital letters may place more emphasis on a given word than parentheses which might on the contrary indicate a restriction. The part played by typography had therefore to be checked. Four parameters (pre- and post-target pauses and durations of enunciation of the last vowel and last syllable preceding the target) were submitted to analyses of variance by groups (speakers 1 and 2), version (*m* and *r*), and typographical characters (the 6 characters used). As the global effects of the text versions and the differences between speakers are now known, we will present only an analysis of possible differences due to typography for each variable. For this analysis two categories were considered. The first category included the main typographical characters for which many observations were made and which were expected to stand out clearly in the text (capital letters, italics, underlinings and bold characters). The second included the last two characters (parentheses and inverted commas), less frequent in the texts and visually different because they are located only *around* the target, rather than *on* the target, when the others change the visual nature of the target itself. The role of the text version was first examined for each typographical character. Between- and within-category contrasts were then calculated. The differences between categories resulted from the fact that the text version changed durational indices for the first category, but not for the second (Appendix A presents the results). None of the contrasts calculated within categories was significant. We can therefore consider that, for the two speakers and with our typographical layout, the differences between typographical characters of the first category apparently played no important role.

### 3. Experiment II

Didactic accentuation of targets was marked by pronounced speech slowing down for the two speakers, but the immediate vicinity of the targets

did not explain it totally. It is of interest to know “from when” and “until when” the speaker slowed down his speech rate. We therefore explored the spoken string over larger distances from the targets in order to try and explain this phenomenon partially. This analysis was carried out on a smaller corpus: 23 sentences extracted from the texts. We chose the 23 sentences in such a way that the target was preceded and followed by a rather extended spoken string, making sure that targets were not placed at the beginning or end of a line for fear that line changes would affect the reading rhythm. The fragments of the spoken string being considered represented (on the average) 15 syllables prior to the target, and 12.8 syllables after the target. These long fragments were then divided into two parts: the so-called “proximate” fragments, i.e., 4 syllables on either side of the target, and the so-called “remote” fragments, i.e., the first 11 syllables (15 – 4) in the long fragments preceding the target, and the last 8.8 syllables (12.8 – 4) in the long fragments following the target.

The *total duration* and *articulation rate* (in syllables per second) for each fragment were measured. Articulation rates were calculated by subtracting non-articulation pauses (more than 100 msec) from total durations. The data were submitted to analyses of variance by groups (speakers 1 and 2) and versions (monotonous and relief). These data are presented in Table 7.

#### 3.1.1. Durational variations before the target

For *proximate fragments* (4 syllables), the enunciation duration was higher in *r* than in *m* ( $F(1-22) = 26.18$ ,  $p < 0.00001$ ) and the articulation rate lower ( $F(1-22) = 38.55$ ,  $p < 0.00001$ ). In neither cases was there any significant difference between speakers. Recall that the mean durational increase recorded for the last syllable before target over all the texts was 75.58 msec (Table 4). The mean durational increase recorded for proximate fragments (315.49 msec) is therefore not limited to the last syllable, but it also concerns other syllables. For *remote fragments* (11 syllables), the duration of enunciation did not significantly increase between *m* and *r* ( $F(1-22) = 1.02$ ,  $p > 0.30$ ). The articulation rate did not

Table 7

Enunciation durations (msec) and articulation rates of proximate and remote fragments around the target for the two speakers S1 and S2

		Enunciation durations			Articulation rates		
		<i>m</i>	<i>r</i>	<i>r - m</i>	<i>m</i>	<i>r</i>	<i>r - m</i>
Proximate pre-target fragments	S1	972.23	1221.01	248.78	4.91	4.51	-0.4
	S2	951.10	1333.39	382.19	5.13	4.00	-1.13
	<i>M</i>	961.67	1277.20	315.49	5.02	4.22	
Remote pre-target fragments	S1	3030.39	3041.95	11.56	4.26	4.40	0.14
	S2	3259.55	3378.68	111.13	4.37	4.10	-0.27
	<i>M</i>	3144.97	3210.32	65.35	4.32	4.25	
Proximate post-target fragments	S1	868.17	956.10	87.93	4.75	4.42	-0.33
	S2	855.93	983.36	127.43	4.81	4.26	-0.55
	<i>M</i>	862.05	969.73	107.68	4.78	4.34	
Remote post-target fragments	S1	2296.28	2311.79	15.51	4.26	4.29	0.03
	S2	2531.06	2707.48	176.42	4.45	4.24	-0.21
	<i>M</i>	2413.67	2509.64	95.97	4.36	4.27	

*m* = monotonous version; *r* = relief version.

vary either from one version to the other ( $F(1-22) = 0.55$ ,  $p > 0.40$ ). Both speakers therefore began to slow down about 4 syllables prior to the target word.

### 3.1.2. Durational variations after the target

For *proximate fragments* (4 syllables), both the duration of enunciation and the articulation rate were markedly longer in *r* than in *m*, equally for the two speakers (for duration of enunciation,  $F(1-22) = 24.74$ ,  $p < 0.00005$ , for articulation rate,  $F(1-22) = 31.16$ ,  $p < 0.00005$ ). The speakers therefore did not recover their rhythm immediately after the post-target pause. For *remote fragments* (8.8 syllables), only speaker 2 had a longer duration of enunciation in *r* than in *m* ( $F(1-22) = 5.29$ ,  $p < 0.05$ ). This slowing down was not related to the articulation rate, which did not vary between the two versions, either for S1 or for S2. Speaker 2 therefore increased the number and/or duration of her pauses in those fragments.

Thus, the two speakers signal the presence of a target about 4 syllables before and after this target, and they do it in a similar way. Once the target word has been pronounced, and after about 4 syllables, speaker 1 recovers the same speech rate as in the monotonous version, whereas speaker 2 goes on lengthening her duration of

enunciation, not by modifying her articulation rate, but by means of more pauses. We find here the same results as above, i.e., a difference between the two speakers as regards the “concentration” of durational increases in the immediate vicinity of the target (the estimated percentage increase was 68.7% for speaker 1 and 51.28% for speaker 2 in Table 5). The location of durational variations resulting from the presence of a didactic accent is therefore not systematic, being more or less concentrated around the targets, depending on the strategy adopted by the speaker.

## 4. Experiment III

When a reader has to insist on a salient word and to transmit it to a listener, he undergoes articulatory and communicative constraints, as well as syntactico-semantic decisions. Therefore, the precise semantic content of the texts and the relative semantic weight of the targets within a text were supposed to have an effect on the way in which prosodic accentuation is performed by the reader. The following study tests the hypothesis that “secondary” details would benefit from more accentuated temporal variations than “basic” information, in order to attract the listener’s attention.

### 4.1. Method

*Phase A.* Only two of the 16 texts were selected. An examination of the targets intuitively led to the following distinction between two categories: “important concepts” (nouns, verbs) and “secondary modulators” (adverbs, adjectives). The distinction led to classifying the targets as follows: T1 = 3 concepts and 11 modulators, and T2 = 7 concepts and 7 modulators (hence there is a larger number of concepts for T2 than for T1).

*Phase B.* The mere *grammatical nature* of words (adjectives, nouns, adverbs) was perhaps not a precise enough translation of the *functions* of these words in the text. In order to find a psychologically more relevant typology, we changed the criterion and took as a basis the

Table 8

Mean duration (msec) of pre-target pauses (Pr), post-target pauses (Po), target enunciation (*T*) and standard deviation of fundamental frequencies (FO) in each phase of Experiment III

	Phase A				Phase B			
	T1		T2		T1		T2	
	<i>m</i>	<i>r</i>	<i>m</i>	<i>r</i>	<i>m</i>	<i>r</i>	<i>m</i>	<i>r</i>
Column no.	1	2	3	4	5	6	7	8
Pr (ms)	27.1	207.0	116.4	215.0	26.4	109.3	52.9	207.9
Po (ms)	204.3	557.9	127.2	478.6	105.7	456.4	5.7	445.0
<i>T</i> (ms)	637.9	883.6	669.9	932.9	646.4	839.7	695.0	972.9
FO (Hz)	27.0	547.8	521.4	952.6	626.4	251.0	723.31	52.83
Statistical values								
Comparisons between T1 and T2 (relief versions)								
	Phase A				Phase B			
Pr	$F(1-24) = 0.009, p > 0.90$				$F(1-26) = 2.20, p > 0.10$			
Po	$F(1-24) = 0.663, p > 0.40$				$F(1-26) = 0.192, p > 0.60$			
<i>T</i>	$F(1-24) = 0.243, p > 0.60$				$F(1-26) = 1.52, p > 0.20$			
FO	$F(1-24) = 0.727, p > 0.40$				$F(1-26) = 0.118, p > 0.70$			
Comparisons between the two relief versions of a text (phase A and B)								
	T1				T2			
Pr	$F(1-26) = 0.747, p > 0.30$				$F(1-26) = 0.470, p > 0.40$			
Po	$F(1-26) = 0.445, p > 0.50$				$F(1-26) = 3.37, p > 0.05$			
<i>T</i>	$F(1-26) = 0.083, p > 0.70$				$F(1-26) = 0.204, p > 0.60$			
FO	$F(1-26) = 0.100, p > 0.40$				$F(1-26) = 0.067, p > 0.70$			

judgments of “naive” listeners as to the relative importance of the words in the text. The two texts read out monotonously were presented to 92 listeners who were required to take notes. Processing these notes allowed ranking each word as a function of its relative recall frequency. Other targets were then redefined on the basis of these classifications with the aim of re-recording the same texts read out by the same speaker. For T1 the 14 chosen targets were the first elements recalled (and are therefore assumed to be important concepts for the listener), whereas for T2 the 14 chosen targets were the last elements recalled (and are therefore assumed to be secondary for the listener).

Two types of criteria<sup>3</sup> therefore governed tar-

<sup>3</sup> The “intuitive” and “empirical” criteria of phases A and B are not contradictory. A majority of the words recalled the best in phase B were nouns and verbs (or “concepts” in the sense of A), while the words that were less well recalled were adverbs and adjectives (or “modulators” in A). Reciprocally, if recall of phase B is used to test targets A, it is noted that for T2 most targets corresponded to the words that were best recalled, but not for T1.

get selection in phases A and B. Moreover, the classification of targets as “important” or “secondary” in the two texts was inverted between phases A and B. The same speaker (speaker 1) recorded one monotonous and two “relief” (A and B) versions of the same two texts. Texts and targets are presented in Appendix B.

#### 4.2. Results

For all recordings we measured the following four parameters: (1) the pre-target pauses (or, for the monotonous version, the pauses preceding those words intended to become targets in the relief version); (2) the post-target pauses; (3) the target enunciation duration; (4) in order to measure the changes in voice fundamental frequency, we assumed that a speaker would have a more contrasted pronunciation for targets in the relief version (with both lower-pitched and higher-pitched voiced sounds): therefore, the standard deviation of the F0 values of target voiced sounds was also measured as the fourth parameter. The mean values obtained are presented in Table 8.

As for all the other texts, the “relief” version resulted in a slowing down in enunciation and a lengthening of pauses, together with significant changes in melodic contrasts, in comparison to the monotonous version. However, the text type had no effect on the data, nor had the target type for the same text. A comparison was made for each parameter between the *relief* versions, both *within* each phase for the two different texts (comparisons between columns 2–4 and 6–8 of Table 8) and *between* the two phases for the same text (comparisons between columns 2–6 and 4–8). As shown at the bottom of Table 8, neither comparison was significant, either within phase A or B, where targets and texts T1 and T2 were differentiated by their semantic contents, nor between phases A and B, where the two “relief” versions of a text only differed by the relative “semantic importance” of the selected targets.

Neither the text content nor the relative semantic importance of targets modified the duration of pre- or post-target pauses, the duration of enunciation of targets, or the melodic contrasts within a target. It is therefore as though the prosodic changes made by the speaker were always of the same order whatever the intrinsic target significance or the target semantic value in a text.

### 4.3. *General discussion and conclusion*

Studying a large speech corpus has both advantages and drawbacks. One of the advantages is to provide a number of measurements, the mean values of which are reliable because they are not submitted to the strong variations which can be observed for a given speaker from time to time or from one text to another. Clearly strong variations also exist between speakers and only the analysis of a long spoken corpus can distinguish between the two sources of variability (within and between speakers). In this research we analyzed some duration-related prosodic variations, and only once another index, the standard deviation of the fundamental frequencies of voiced sounds (a melodic index, but undoubtedly varying to some degree with speech intensity). The differences in reading times by the two speakers for the large

corpus of 16 texts read out in two versions can in any case be considered as stable differences, which are specific to the concerned individuals, and which can supply information on distinct reading strategies. We will now review our findings.

Punctuation pauses represent a major part of the total pause duration in a spoken string, as recalled in the introduction. The absolute durations of paragraph change- or punctuation-related pauses strongly varied between speakers, but their relative durations were identically arranged, irrespective of the speaker. For paragraph changes speaker 1 stopped 2 seconds and speaker 2 almost twice as long: 3.6 seconds. For full stops, speaker 1 stopped 1.2 second in average, i.e. less than half the pauses of speaker 2 (2.7 seconds as an average). For colons, pauses lasted from 950 msec (speaker 1) to 1600 msec (speaker 2) and for semi-colons and commas from 600 msec (speaker 1) to 900 msec (speaker 2). Didactic accentuation on targets was translated into additional pauses, a durational increase in already existing pauses, and lengthening of the enunciation duration for the last syllables before the target and for the target itself. The two speakers differed as to the concentration of their durational increases around target words: almost 70% of global reading time lengthening was concentrated around the targets for speaker 1, only 50% for speaker 2. However, for both speakers, the mean duration of pre- and post-target pauses nearly doubled, the target enunciation duration increased by 34% (from 765 msec to 1027 msec as an average) and the enunciation duration for the pre-target syllable increased by 32% (from 243 msec to 319 msec).

When studying duration indices in the immediate vicinity of the target, we had noted that the text type did not influence the duration of the last pre-target syllable, but caused systematic variations in target enunciation duration and in post-target pauses. The part played by texts on target enunciation duration variations can perhaps be explained by the very length of the targets, which varied from one text to another, but the same explanation is not valid for the post-target pauses: these pauses were handled differ-

ently from the others. The study of the correlations between durational variations reinforced this idea. Thus, if significant links among syllable duration, vowel duration and pre-target pauses might result from the fact that these elements were uttered in the same articulatory movement, the absence of clear correlation between the post-target pause and the preceding elements might on the contrary imply that a different function is allocated to this particular pause. The hypothetical role of post-target pauses serving as “additional signals” for the listener would require further checking.

When looking for the sources of the slowing down in speech rate, we also showed that didactic accentuation changed the spoken string at least 4 syllables before and after the target. Differences between the speakers were again noted: speaker 2 continued to introduce pauses more than 4 syllables after the target, while speaker 1 returned more quickly to a “monotonous” rhythm. Studies by Folkins et al. (1975), Weismer and Ingrisano (1979) and Padeloup (1984) showed for short sentences that, when insistence occurred, there was an increase in the duration necessary to utter the other words in the sentence. Cooper et al. (1985) and Eady and Cooper (1986), using longer sentences, did not find any effect on neighbouring words. The authors assumed that in comparison to preceding research this difference was due to the sentence length: for shorter sentences (5 to 6 syllables) an effect would be evident on neighbouring words, but this effect would be inversely proportional to the distance with the accentuated word and would therefore not be found with longer spoken strings. Our analyses on target-proximate and target-remote fragments in long sentences confirm this hypothesis.

The two speakers under study seem to have been insensitive to factors such as the nature of the typographical characters of the targets or the target semantic weight. Generalizing our findings on semantic aspects to situations other than ours, i.e. continuous reading of many texts, can be difficult. The two speakers can be thought to have trained for efficient reading with a view to both avoiding articulation errors and obeying the

explicit instruction to transmit the targets to the listener. In a more natural situation and with speakers that are less trained, we have recently shown that the target semantic weight changes their melodic contour in the enunciation of isolated sentences (Dahan and Gérard, In preparation). Furthermore, as demonstrated by Howell and Kadi-Hanifi (1991), read material and spontaneous speech differ as to the location of stressed units, the distribution of primary stress, and the number and location of pauses. Would typographical aspects be in turn determining factors in other contexts? In silent reading, typographical characters have different impacts on eye movements. Similarly, a comprehensive view of a whole page is supposed to benefit from the visual structuration indices supplied by typographical changes. In loud reading, however, can the speaker master his vocalizations in such a sophisticated manner that graphemes can be translated into phonemes and that typographical differences can also be transmitted? Further studies involving more precise and natural situations are necessary to address these questions.

Semantic and pragmatic determinants for speakers’ prosodic variations have been tackled as research themes in the framework of the “given-new” distinction, but without any clear study of precise prosodic variations due to the speaker. However, potential dissociation between the speaker’s and listener’s behaviours facing semantic questions seems to be a theme for further research. In contrast to the numerous studies carried out on the effect of insistence on listeners’ perception (studies mainly initiated by Cutler and Darwin (1981) and Cutler (1987)), little is actually known about the effect of insistence on listeners’ behaviours. A study by Jakobi et al. (1988) analyzed the effect of emphasizing a word within a statement in an interview situation: the interviewer repeated a sentence uttered by the interviewee, putting emphasis on the subject, the verb or the complement, or putting no emphasis at all. Reiterated statements under such circumstances assumed the pragmatic status of requests for explanation. It was found that the interviewee’s argumentative structure was modified by the addition of emphasis and differed depending



For *pre-target pauses*, lengthening due to relief was significant for the first typographical category (the values increased from 75.9 msec to 205.4 msec as an average for the two speakers, and for the four typographical characters,  $F(1-197) = 188.96$ ,  $p < 0.00001$ ), whereas the mean values for parentheses and inverted commas (second category) were not significantly affected by the version. A number of contrasts was studied between the typographical characters of the first category, but none resulted in significant differences.

The behaviour of *post-target pauses* was the same as the preceding. The same contrast between the two typographical character groups was found: pauses significantly increased from versions *m* to *r* for the first category (from 247.4 to 535.2 msec on the average,  $F(1-197) = 314.49$ ,  $p < 0.00001$ ) and did not significantly increase for the second category (from 1134.4 to 1259.5 msec,  $F(1-19) = 2.03$ ,  $p > 0.10$ ). No other significant difference was found within the two typographical categories. In reality, before and after targets marked by inverted commas and parentheses, the speakers already produced important pauses in the monotonous version.

The duration of enunciation of the *pre-target syllable* increased for any and all of the typographical characters marking the target (the values of *F* were always significant,  $p < 0.0001$ ), except for the inverted commas for speaker 1 ( $F(1-8) = 0.04$ ,  $p > 0.40$ ). Identical results were found for the duration of the *pre-target vowel* with the same exception for inverted commas for speaker 1 ( $F(1-8) = 0.05$ ,  $p > 0.40$ ). No contrast was found to be significant within the first typographical category.

Thus, first category typographical characters (capital letters, italics, bold characters, underlinings) always induce durational increases in the four time parameters under consideration. Parentheses and inverted commas do not modify pause duration – they were already very long – but they lengthen the utterance of the last syllable. This phenomenon can be explained by the very choice of the particular targets, already submitted to accentuations: as underlined in the methods section, it was difficult to find words which lent

themselves to the addition of parentheses or inverted commas.

## Appendix B

The structure and difficulty of the two texts used in Experiment III are representative of all the other texts. They are presented here in their monotonous version and followed by the list of the 14 target words selected (phase A). The 14 other targets selected empirically (after the recall of 92 listeners) in phase B of the third experiment are then presented.

### *Text 1. Made in Zaïre du 14 au 22 novembre 1989*

La Chambre de Commerce et d'Industrie de Paris recevra du 14 au 22 novembre 1989 une importante délégation de l'Association Nationale des Entrepreneurs du Zaïre (ANEZA) et de la Chambre de Commerce et d'Industrie Franco-Zaïroise, délégation conduite par le commissaire d'Etat au Commerce Extérieur de la République du Zaïre.

Cette opération est organisée à l'occasion de l'exposition "Made in Zaïre", rassemblant une quinzaine d'entreprises zaïroises, parmi les plus compétitives et les plus importantes, à la Bourse du Commerce du 15 au 22 novembre 1989.

L'organisation de ces manifestations témoigne de la volonté des milieux d'affaires zaïrois de développer et d'élargir leurs relations économiques avec les pays occidentaux et plus particulièrement d'intensifier leurs échanges avec la France.

Le Zaïre, par ses potentialités, offre de très larges perspectives économiques. Il est le deuxième pays le plus peuplé d'Afrique Noire (40.000.000 habitants) et possède des ressources minières de toute première importance. Citons, à titre d'exemple, les productions de cuivre (500.000 tonnes par an), de cobalt (10.000 à 15.000 tonnes par an), de diamant, de pétrole brut, etc. . . .

Aussi, nous vous proposons de participer à une journée économique franco-zaïroise qui se tiendra au siège de la Chambre de Commerce et d'Industrie de Paris: 27, avenue de Friedland, 75008 Paris, mardi 14 novembre 1989 à 9h.



*Target words*

Phase A: conduite, exposition, quinzaine, manifestations, volonté, développer, économiques, occidentaux, échanges, perspectives, deuxième, minières, brut, journée.

Phase B: important, nationale, commissaire, extérieur, occasions, rassemblant, compétitives, témoignent, milieux, particulièrement, larges, toute première, proposons, siège.

*Text 2. Notice d'utilisation modèle XO 1*

L'étude technique très attentive de ce modèle lui confère des performances excellentes, mais son installation mérite un soin attentif afin de profiter pleinement de toutes ses qualités.

Installation: Les XO 1 seront installées contre un mur réfléchissant et sur un support très rigide d'environ 60 cms de haut, elles seront dégagées des angles de la pièce. Leur raccordement à l'électronique s'effectuera avec un câble de haute qualité (ne pas utiliser du fil de type scindex, désastreux sur le plan des résultats sonores). Les longueurs seront égales pour les deux canaux et les phases soigneusement repérées. L'écartement entre les deux enceintes sera d'environ 2m à 2m50. Elles feront face, si possible, à une surface absorbante (mur meublé, rideaux, tentures etc...). En aucun cas l'XO 1 ne sera utilisée à même le sol.

Fonctionnement: L'électronique sera utilisée dans les règles de l'art, c'est-à-dire sans utilisation des correcteurs à niveau d'écoute normal, ceux-ci ne seront actionnés que pour des écoutes à bas niveau (même chose pour le loudness s'il existe). Les commutations entre les différentes sources (par exemple, passage de phono à tuner) seront toujours effectuées le volume au minimum, il sera fait de même lors de leur branchement ou débranchement.

Protection: L'XO 1 est équipée d'un disjoncteur thermique réarmable manuellement et situé entre les bornes de raccordement, celui-ci la protège efficacement contre les surcharges de puissance. Il faut néanmoins savoir que, en cas d'écrêtage de l'électronique (fonctionnement au delà de sa puissance nominale, ce qui est en général obtenu bien avant que le bouton de volume soit au maximum) la protection est aléatoire;

il conviendra donc de limiter le niveau d'écoute dès l'apparition de toute forme de distorsion (saturation, talonnement du grave, sifflantes).

*Target words*

Phase A: pleinement, réfléchissant, rigide, haut, deux, soigneusement, normal, niveau, commutations, réarmable, surcharges, écrêtages, au-delà, avant.

Phase B: performances, mur, support, 60 cm, câble, longueurs, enceintes, faire face, surface, utilisation, phono, volume, thermique, protection.

**References**

- J.K. Bock and J.R. Mazella (1983), "Intonational marking of given and new information: Some consequences for comprehension", *Memory and Cognition*, Vol. 11, No. 1, pp. 64–76.
- D.L. Bolinger (1961), "Contrastive accent and contrastive Stress", *Language*, Vol. 37, No. 1, pp. 83–96.
- Calliope (1989), *La Parole et son Traitement Automatique* (Masson, Paris).
- W.L. Chafe (1974), "Language and consciousness", *Language*, Vol. 50, pp. 111–133.
- W.E. Cooper, S.J. Eady and P.R. Mueller (1985), "Acoustical aspects of contrastive stress in question-answer contexts", *J. Acoust. Soc. Amer.*, Vol. 77, pp. 2142–2156.
- A. Cozannet (1991), Message to speech, Note interne, Centre National d'Etudes des Télécommunications.
- A. Cutler (1987), "Speaking for listening", in *Language Perception and Production. Relationships between Listening, Speaking, Reading and Writing*, ed. by A. Allport, D.G. Mackay, W. Prinz and E. Echeerer (Academic Press, London), pp. 23–40.
- A. Cutler and C.J. Darwin (1981), "Phoneme-monitoring reaction time and preceding prosody: Effects of stop closure duration and of fundamental frequency", *Perception and Psychophysics*, Vol. 29, No. 3, pp. 217–224.
- A. Cutler and J.A. Fodor (1979), "Semantic focus and sentence comprehension", *Cognition*, Vol. 7, pp. 49–59.
- D. Dahan (1994), "Étude de l'influence des variations prosodiques dans les processus de traitement de la parole", *Proc. 20èmes Journées d'Etude sur la Parole, Trégastel, June 1994*.
- D. Dahan and J.M. Bernard (1994), "Production et perception des phénomènes prosodiques dans la parole: Comment identifie-t-on l'insistance malgré la variabilité interlocuteurs?", *J. Physique*, Vol. IV, Suppl. JP III, No. 5, pp. 501–504.
- D. Dahan and C. Gérard (In preparation), "The prosodic correlates of emphatic stress in French".

- F. Dell (1984), "L'accentuation dans les phrases en français", in *Forme Sonore du Langage. Structure des Représentations en Phonologie*, ed. by F. Dell, D. Hirst and J.R. Vergnaud (Hermann, Paris), pp. 65–122.
- S.J. Eady and W.E. Cooper (1986), "Speech intonation and focus location in matched statements and questions", *J. Acoust. Soc. Amer.*, Vol. 80, pp. 402–415.
- J.W. Folkins, C.J. Miller and F. Minifie (1975), "Rhythm and syllable timing in phrase level stress patterning", *J. Speech Hearing Res.*, Vol. 18, pp. 739–753.
- I. Fonagy (1980), "L'accent français: Accent probabilitaire", in *L'accent en Français Contemporain* (Studia Phonetica, Didier, Montréal), pp. 123–233.
- P. Garde (1968), *L'accent* (PUF, Paris).
- C. Gérard and C. Auxiette (1992), "The processing of musical prosody by musical and nonmusical children", *Music Perception*, Vol. 10, No. 1, pp. 93–126.
- C. Gérard and J. Clément (1994), "Patrons prosodiques et intentions des locuteurs: production et perception de formes expressives chez l'enfant et l'adulte", *Proc. 20èmes Journées d'Etude sur la Parole, Trégastel, June 1994*.
- C. Gérard and D. Dahan (1992), "Formes prosodiques et focalisation sémantique", *Proc. 19èmes Journées d'Etudes sur la Parole, Bruxelles*, pp. 507–511.
- C. Gérard and C. Rigaut (1994), "Patterns prosodiques et intentions des locuteurs: le rôle crucial des variables temporelles dans la parole", *J. Physique*, Vol. IV, Colloque C5, Suppl. JP III, No. 5, pp. 505–508.
- C. Gérard, D. Dahan and C. Rigaut (1991), *Etude des pauses: Relations entre la typographie d'un texte et sa structuration temporelle lors de la lecture*, CNET: Rapport de la convention de recherche no. 89 1B226.
- J. Goldman-Eisler (1972a), "La mesure des pauses: Un outil pour l'étude des processus cognitifs dans la production verbale", *Bulletin de Psychologie*, Vol. 26, pp. 383–390.
- F. Goldman-Eisler (1972b), "Pauses, clauses, sentences", *Language and Speech*, Vol. 15, pp. 103–147.
- F. Grosjean (1980), "Linguistic structures and performance structures: studies in pause distribution", in *Temporal Variables in Speech*, ed. by H.W. Dechert and M. Raupach (Mouton, The Hague), pp. 85–90.
- F. Grosjean and A. Deschamps (1972), "Analyse des variables temporelles du français spontané", *Phonetica*, Vol. 26, pp. 129–156.
- F. Grosjean and A. Deschamps (1973), "Analyse des variables temporelles du français spontané II. Comparaison du français oral dans la description avec l'anglais (description) et avec le français (interview radiophonique)", *Phonetica*, Vol. 28, pp. 191–226.
- F. Grosjean and A. Deschamps (1975), "Analyse contrastive des variables temporelles de l'anglais et du français: Vitesse de parole et variables composantes, phénomènes d'hésitation", *Phonetica*, Vol. 31, pp. 144–184.
- P. Howell and K. Kadi-Hanifi (1991), "Comparison of prosodic properties between read and spontaneous speech material", *Speech Communication*, Vol. 10, No. 2, pp. 163–169.
- J.M. Jakobi, A. Blanchet and M. Bromberg (1988), "Les effets du focus intonatif dans les répétitions référentielles en situation d'entretien de recherche", *Revue Internationale de Psychologie Sociale*, Vol. 1, Nos. 3–4, pp. 345–363.
- J. Jayez (1991), "Un aspect de l'analyse automatique des conversations: Croyances, mouvements conversationnels, mouvements discursifs", *Cahiers de Linguistique Française*, Vol. 10, pp. 147–170.
- S. Kowal and D.C. O'Connell (1980), "Pausological research at Saint Louis University", in *Temporal Variables in Speech*, ed. by H.W. Dechert and M. Raupach (Mouton, The Hague), pp. 61–66.
- N. Le Nouveau (1987), *Quelques variables temporelles de la lecture orale pour une application en synthèse vocale*, Mémoire de DEA de Psychologie sous la direction d'A. Lieury and G. Poulain, Université de Rennes.
- P. Lieberman (1960), "Some acoustics correlates of word stress in american english", *J. Acoust. Soc. Amer.*, Vol. 32, No. 4, pp. 451–454.
- V. Lucci (1983), *Etude phonétique du français contemporain à travers la variation situationnelle*, Publication de l'Université des Langues et Lettres de Grenoble.
- J.G. Martin (1972), "Rhythmic (hierarchical) versus serial structure in speech and other behavior", *Psychological Rev.*, Vol. 79, No. 6, pp. 487–509.
- J.G. Martin (1979), "Rhythmic and segmental perception are not independent", *J. Acoust. Soc. Amer.*, Vol. 65, No. 5, pp. 1286–1297.
- W.P. Needham (1990), "Semantic structure, information structure and intonation in discourse production", *J. Memory and Language*, Vol. 29, pp. 455–468.
- S.G. Nootboom, J.P.L. Brokx and J.J. de Rooij (1978), "Contributions of prosody to speech perception", in *Studies in the Perception of Language*, ed. by W.J.M. Levelt and G.B. Flores d'Arcais (Wiley, New York), pp. 75–107.
- V. Padeloup (1984), *Etude acoustique et perceptive de la mise en valeur dans la phrase assertive en français*, Mémoire de Maîtrise de Linguistique, option phonétique, sous la direction de R. Gsell, Université Paris III.
- V. Padeloup (1988), "Essai d'analyse du système accentuel du français: Distribution de l'accent secondaire", *Proc. 17èmes Journées d'Etude sur la Parole, Nancy, 1988*.
- V. Padeloup (1989), "Essai de modélisation du rythme du français", *Journal d'Acoustique*, Vol. 2, No. 1, pp. 47–51.
- V. Padeloup (1990), *Modèle de règles rythmiques du français appliqué à la synthèse de la parole*, Thèse de 3ème cycle sous la direction de M. Rossi, Université de Provence, Aix-Marseille I.
- J.P. Pierrel and N. Carbonel (1993), "La reconnaissance automatique de la parole", *Le Courrier du CNRS*, Vol. 79, p. 11.
- M. Rossi (1980), "Le Français, langue sans accent?", in *L'accent en Français Contemporain* (Studia Phonetica, Didier, Montréal), pp. 13–51.
- M. Rossi (1985), "L'intonation et l'organisation de l'énoncé", *Phonetica*, Vol. 42, pp. 135–153.
- M. Rossi (1987), "Peut-on prédire l'organisation prosodique

- du langage spontané?”, in *Aspects Prosodiques de la Communication, Etudes de Linguistique Appliquée*, Vol. 66, pp. 20–48.
- M. Rossi (1988), “Prosodie et technologies vocales”, in *Journées Nationales du GRECO-PRC. Communication Homme – Machine. Parole, Langage Naturel et Vision* (EC2, Paris), pp. 63–80.
- M. Saint-Bonnet and L.J. Boë (1977), “Les pauses et les groupes rythmiques: Leur durée et distribution en fonction de la vitesse d’élocution”, *Proc. 8èmes Journées d’Etudes sur la Parole, Aix-en-Provence*, pp. 337–343.
- A. Séguinot (1977), “L’accent d’insistance en français standard”, in *L’accent D’insistance. Emphatic Stress*, ed. by F. Carton, D. Hirst, A. Marchal and A. Séguinot (Didier, *Studia Phonetica* 12, Montréal), pp. 1–58.
- C. Sorin (1989), “Perception de la parole”, in *Psychoacoustique et Perception Auditive*, ed. by M.C. Botte, G. Canevet, L. Demany and C. Sorin (Inserm/SFA/CNET, Paris), pp. 123–139.
- G. Weismer and D. Ingrisano (1979), “Phrase-level timing patterns in English: Effects of emphatic stress location and speaking rate”, *J. Speech Hearing Res.*, Vol. 22, pp. 516–533.